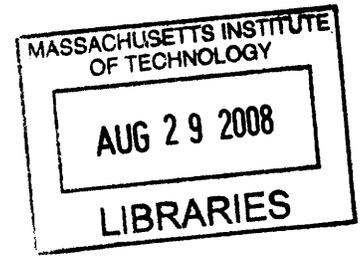


Quantifying Uncertainty in Computational Neuroscience with Bayesian Statistical Inference

By

Beau D. Cronin

B. A. Computer Science
University of California at Berkeley, 2000



Submitted to the Department of Brain and Cognitive Sciences in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Neuroscience

at the

Massachusetts Institute of Technology

September 2008

© 2008 Massachusetts Institute of Technology
All rights reserved

Signature of Author: _____
Department of Brain and Cognitive Sciences
August 12th, 2008

Certified By: _____
Mriganka Sur
Professor of Neuroscience
Thesis Supervisor

Accepted By: _____
Matthew Wilson
Professor of Neurobiology
Associate Department Head for Education

ARCHIVES

Quantifying Uncertainty in Computational Neuroscience with Bayesian Statistical Inference

By

Beau D. Cronin

Submitted to the Department of Brain and Cognitive Sciences in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Neuroscience

Abstract

Two key fields of computational neuroscience involve, respectively, the analysis of experimental recordings to understand the functional properties of neurons, and modeling how neurons and networks process sensory information in order to represent the environment. In both of these endeavors, it is crucial to understand and quantify uncertainty - when describing how the brain itself draws conclusions about the physical world, and when the experimenter interprets neuronal data. Bayesian modeling and inference methods provide many advantages for doing so.

Three projects are presented that illustrate the advantages of the Bayesian approach. In the first, Markov chain Monte Carlo (MCMC) sampling methods were used to answer a range of scientific questions that arise in the analysis of physiological data from tuning curve experiments; in addition, a software toolbox is described that makes these methods widely accessible. In the second project, the model developed in the first project was extended to describe the detailed dynamics of orientation tuning in neurons in cat primary visual cortex. Using more sophisticated sampling-based inference methods, this model was applied to answer specific scientific questions about the tuning properties of a recorded population. The final project uses a Bayesian model to provide a normative explanation of sensory adaptation phenomena. The model was able to explain a range of detailed physiological adaptation phenomena.

Thesis Supervisor: Mriganka Sur

Title: Professor of Neuroscience

Acknowledgements

My career as a graduate student has itself been something of a collaborative experiment - a test of the hypothesis that both computational and experimental neuroscience benefit when they are done together. I thank my thesis advisor Mriganka for supporting this idea from the very beginning, and for providing the wonderfully supportive environment that enabled such a positive result.

I also thank my co-advisor Konrad Kording, who has been supremely generous with his ideas, his time, and his joy in living a life dedicated to science.

I also warmly thank the other two members of my committee: Josh Tenenbaum, for providing the vision to situate the work in the context of Bayesian statistics in the brain and cognitive sciences, and Jim Dicarolo, for greatly improving the work by pushing me to articulate the advantages of these techniques over other methods.

The Sur Lab has been my intellectual home for the last five years, and I am so thankful for everyone's willingness accept - even welcome - a computationalist in their midst. In particular, I would like to thank those who entrusted me with their data: James Schummers, Jitendra Sharma, Caroline Runyan, and Rong Mao.

Vikash Mansinghka provided great insight on many matters, both philosophical and technical. It has been a pleasure to share my graduate career with Srinivas Turaga, who showed great patience in listening to many, many bad ideas.

Finally, I thank Naomi, who has always seen the best in me, and whose ferocious dedication to things that actually matter in life has kept me from floating off into the ether.

Contents

1	Introduction	11
1.1	Bayesian Formulations of Scientific Questions	11
1.2	The Road Ahead	14
2	Background	17
2.1	A Brief Introduction to Bayesian Statistics	17
2.2	Hierarchical Models and Graphical Models	18
2.3	Methods of Bayesian Statistical Inference	20
2.4	Markov chain Monte Carlo	22
2.4.1	Metropolis Sampling	23
2.4.2	Gibbs Sampling	25
2.5	Potential Pitfalls of Applying Bayesian Models	25
2.6	Conclusion	25
3	Analysis of Tuning Curve Data	29
3.1	Introduction	29
3.2	Methods	32
3.2.1	Hierarchical Model	33
3.2.2	MCMC Sampling	36
3.2.3	Bayesian Model Selection	38
3.3	Results	39
3.3.1	Estimating parameter values	39
3.3.2	Visualization of potential tuning curves	39
3.3.3	Assaying quantitative changes in response properties	41
3.3.4	Assessing selectivity to the stimulus feature using model selection	41
3.3.5	Comparing different selectivity models	43
3.3.6	Performance of the harmonic mean estimator	43
3.4	Comparison with the Bootstrap	45

3.4.1	The Bootstrap	45
3.4.2	Nonlinear least-squares optimization	46
3.4.3	Combining the bootstrap and nonlinear optimization	46
3.4.4	Comparing the methods	47
3.5	The BayesPhys Toolbox	51
3.6	Discussion	51
3.7	Next Steps	52
4	Analysis of Orientation Tuning Dynamics	55
4.1	Introduction	55
4.1.1	The experimental paradigm and standard analysis	57
4.1.2	Previous work	58
4.2	Methods	60
4.2.1	Hierarchical model description	60
4.2.2	Inference and sampling	65
4.3	Results	66
4.4	Comparison with the Bootstrap and Optimization	69
4.4.1	Bootstrap and curve fitting procedure	69
4.4.2	Comparison Results	70
4.5	Discussion	72
5	Neuronal Adaptation as Bayesian Inference	75
5.1	Introduction	75
5.2	Methods	77
5.2.1	Model of Gain Dynamics	77
5.2.2	Gain Estimation	79
5.2.3	Short-Term Synaptic Plasticity	79
5.2.4	Orientation Tuning Adaptation	79
5.3	Results	81
5.3.1	Short-Term Synaptic Plasticity	82
5.3.2	Orientation Tuning Adaptation	83
5.4	Discussion	86
5.4.1	Other Normative Models	86
5.4.2	Relationship to Mechanistic Models	87
5.4.3	Relationship to Kording et al., 2007	87
5.4.4	Limitations and Future Work	88

<i>CONTENTS</i>	5
6 Conclusion and Future Directions	89
6.1 Mainstreaming sampling approaches	89
6.2 Generalizing the model	90
6.3 Adaptive Experiments	92
6.4 Conclusion	94
A BayesPhys Usage	95
A.1 Preparing stimuli and data for analysis	95
A.2 Performing sampling	95
A.3 Interpreting sampling results	96
A.4 Performing model selection	98
B Gaussian Approximation of Synaptic Activity	99
Bibliography	101

List of Figures

2.1	DGM Example	20
2.2	Bayesian inference methods	21
2.3	Metropolis sampling	24
2.4	Gibbs sampling	26
3.1	Conceptual introduction to Bayesian modeling of tuning curve data	30
3.2	A graphical model of the tuning curve data analysis	34
3.3	Methods for sampling from tuning curve data	37
3.4	Parameter estimates from samples on tuning curve data	40
3.5	Model selection for tuning curve data	42
3.6	Performance of the harmonic mean estimator	44
3.7	Comparison 1 between sampling and the bootstrap	48
3.8	Comparison 2 between sampling and the bootstrap	49
3.9	Inclusion of prior knowledge to improve TC estimates	50
4.1	Schematic of the reverse correlation experimental setting.	57
4.2	Stimulus triggered average example	59
4.3	Schematic of the orientation dynamics analysis	61
4.4	Hierarchical model of orientation dynamics analysis	62
4.5	Discrete cosine transform proposals	67
4.6	Results from the orientation dynamics analysis	68
4.7	Simulated neuron with no tuning shift	71
4.8	Simulated neuron with change in preferred orientation	73
5.1	A DGM of the adaptation model	78
5.2	The network model of orientation adaptation	80
5.3	Comparison with Tsodyks and Markram	82
5.4	Repulsive orientation adaptation	84
5.5	Comparison with Dragoi et al.	85

6.1	A schematic of the linear-nonlinear-stochastic model	91
6.2	A simulated experiment using adaptive stimuli	93
A.1	BayesPhys results example	97

List of Tables

3.1	Example tuning curve functions	33
3.2	Probability models for different types of physiological data	35

Chapter 1

Introduction

Two key fields of computational neuroscience involve, respectively, the analysis of experimental recordings to understand the functional properties of neurons, and modeling how neurons and networks process sensory information in order to represent the environment. In both of these endeavors, it is crucial to understand and quantify uncertainty - when describing how the brain itself draws conclusions about the physical world, and when the experimenter interprets neuronal data. Bayesian statistics excels in both settings, because it provides a framework for consistently and correctly reasoning under conditions of uncertainty. Bayesian models, by their nature, require the explicit formulation of all assumptions that can affect the outcome of the statistical inference. This process of laying out all assumptions makes clear what we do not yet know, both in terms of the structure of the model, as well as the uncertainty which remains about the model's parameters after the data has been taken into account (i.e., error bars).

Rigorously specified probabilistic models can therefore aid our understanding of the inferential processes involved in sensory processing and improve our analyses of physiological data. In this thesis, I present three case studies which implement these principles and illustrate the advantages of Bayesian statistical inference in computational neuroscience.

1.1 Bayesian Formulations of Scientific Questions

The variables in a Bayesian model can represent any scientifically meaningful quantity. In the brain sciences, Bayesian methods have been most famously and extensively applied to the problem of perception. In this general formulation, the sensory system is conceived of as continually performing Bayesian statistical inference to comprehend its environment via the interpretation of sensory signals (see, for example, Knill and Richards, 1996; Yuille and Kersten, 2006).

Most generally, this computational-level theory of sensory perception as Bayesian inference can be stated using Bayes' rule as follows:

$$\Pr(\textit{world state}|\textit{sensory signal}, \textit{past signals}) \propto \frac{\Pr(\textit{sensory signal}|\textit{world state}, \textit{past signals}) \Pr(\textit{world state}, \textit{past signals})}{\Pr(\textit{sensory signal}|\textit{world state}, \textit{past signals})} \quad (1.1)$$

That is, the nervous system uses sensory signals to infer probable states of the outside world (the “posterior distribution”), given it’s prior knowledge about which world states are the most probable, and also its understanding of how these world states can induce particular patterns of sensory input. Specific models of perceptual processes will then introduce particular meanings for these variables, and impose appropriate probabilistic relations between them.

While this is a common framework in which to formulate behavioral models of perception, it has not been extensively applied to the analysis of individual neuronal responses. In Chapter 5 of this thesis, I present a general model of sensory adaptation that explains how an individual neuron can optimally adjust its own response properties in order to maintain its ability to successfully estimate the state of the world. This model may provide a bridge between the properties of individual neurons and the computational robustness displayed by the sensory system in performing inference about the nature of the environment.

Although the perceptual setting given in equation 1.1 is probably the most widely recognized application of Bayesian methods to the brain sciences, it is but one example. Another general application of Bayes rule, used throughout the sciences, might be stated as follows:

$$\Pr(\textit{model}|\textit{experimental data}, \textit{physical law}) \propto \frac{\Pr(\textit{experimental data}|\textit{model}, \textit{physical law}) \Pr(\textit{model}, \textit{physical law})}{\Pr(\textit{experimental data}|\textit{model}, \textit{physical law})} \quad (1.2)$$

This setting is applicable in any scenario in which noisy, error-prone, and incomplete experimental measurements are used to draw inferences about the “true” nature of the physical world. For example, cosmologists and astrophysicists have made extensive use of Bayesian methods in order to decide between models of different complexity that describe the basic properties of the universe (Kunz et al., 2006; Trotta, 2007). In this case, the model parameters must be estimated from noisy measurements of the cosmic microwave background radiation. In addition, the data are also used to determine which model is best supported by the set of observations.

Shifting focus back to the analysis of neuronal data, Bayesian methods have been applied, for example, to the difficult problem of spike sorting (Wood et al., 2006). In this application, the waveforms corresponding to action potentials from several different neurons have been

mixed into a single signal during recording, and must be divided into groups such that the resulting partitioning reflects as accurately as possible the neuronal sources of the spikes; this is clearly a statistical inference problem which fits into the general setting of equation 1.2. A primary advantage of Bayesian spike sorting methods over the various alternatives is their ability to quantify the degree of uncertainty which remains about the assignment of each spike to a particular neuron (Lewicki, 1998). This is crucial because spike sorting is not a goal in and of itself, but rather the first stage of many neuroscientific data analyses. Therefore, a quantification of uncertainty allows researchers to make informed decisions about which spike assignments to “count on” in these later stages.

This thesis provides a detailed treatment of another application of the Bayesian framework to the analysis of neuronal data: that of inferring the tuning properties of individual neurons in response to particular classes of sensory stimuli. In this setting, the relevant variables are the recorded neuronal responses, the stimuli, and the parameters that govern the “receptive field”:

$$\Pr(\text{tuning properties}|\text{neuronal recordings, stimuli}) \propto \Pr(\text{neuronal recordings}|\text{tuning properties, stimuli}) \Pr(\text{tuning properties}) \quad (1.3)$$

Note that this formulation is a direct specialization of equation 1.2. Just like measurements from other types of scientific instrumentation, all commonly used physiological methods are subject to noise, whether extracellular electrophysiology, intracellular recording using patch techniques, confocal or 2-photon microscopy, or recording of local field potentials. In addition, many neuronal responses display a high degree of intrinsic variability in the context of typical laboratory preparations; that is, the neuronal response to two repetitions of the same stimulus might vary quite widely, even once measurement noise has been taken into account. Making matters worse, the experimenter is often quite limited in the amount of time that is available to collect data from a given neuron. If stimuli could be presented and responses recorded forever, the noise and variability in neural systems could be overcome; in the limit of infinite data, simple averaging will suffice to answer virtually any scientific question. But the severe constraints on data collection imposed by the reality of typical neurophysiological experimental preparations require that sophisticated methods be employed to first extract as much information as possible from the available data, and also to quantify exactly how much uncertainty remains after the data has been taken into account.

A comparison of the two formulations in equations 1.1 and 1.3 reveals an underlying symmetry: just as neuroscientists investigate the properties of the brain by drawing inferences about neuronal behavior from physiological data, so do individual neurons (and the systems

which they comprise) investigate the properties of the environment. Furthermore, both types of investigation require assumptions: the neuroscientist employs knowledge about the likely response properties of neurons, and individual neurons maintain models of their past inputs for comparison to new signals.

While we believe that the Bayesian framework is underutilized in the analysis of neuronal data, there are a number of previous studies that have preceded (and inspired) the current work. Gillespie et al. (2001), for example, explicitly framed the analysis of tuning curve data as a Bayesian inference problem, an approach which we generalize and extend in Chapter 3. We believe that this is the only published example of a Bayesian approach being applied in such a “classical” neurophysiological setting. Sahani and Linden (2003) use a more sophisticated modeling approach to estimate the stimulus-response function of neurons in response to noise stimuli. They found that the incorporation of prior information, including assumptions about the smoothness and sparseness of the response properties, significantly improved the inferential power of their approach.

1.2 The Road Ahead

The next chapter presents the basic setting of Bayesian statistics, with an emphasis on hierarchical models, the directed graphical model formalism, and Markov chain Monte Carlo sampling techniques for performing approximate inference.

The bulk of the thesis comprises three case studies of the application of Bayesian statistics to computational neuroscience. Each study provides a fully specified probabilistic model, as well as a statistical inference scheme that allows for the principled quantification of uncertainty given the data.

Chapter 3 describes a model and inference method for the analysis of tuning curve data. This model is applicable whenever a neurophysiologist has collected neuronal response data in which one aspect of the stimulus - angular orientation, for example - has been manipulated. In these experiments, which have remained common over many decades, the experimenter typically wishes to answer questions such as: How selective is the neuron for the stimulus property being varied?, Does this selectivity change due to an experimental manipulation such as adaptation or learning?, and What is the best model with which to describe the neuron’s responses? In addition to providing a general modeling framework for formulating these and other scientific questions, I also provide a set of techniques and a software implementation that allow experimentalists to quickly and easily answer questions from their own data. This model presented in this chapter also provides a basis for the specification of more complex models.

One such complex model is described in Chapter 4. This model was developed to answer very specific scientific questions about the dynamics of neuronal responses: does the orientation tuning of neurons in primary visual cortex change over the time course of the response. For example, does tuning sharpen significantly between the early response (at, say, 60 ms latency) and the later response (after, say, 120 ms)? Because of the dynamic nature of the stimulus used in the experiment, as well as the additional complexity required to describe neuronal responses that change from one stimulus latency to the next, both the model and the scheme used to perform statistical inference on its parameter values are more sophisticated than the static tuning curve model developed in Chapter 3.

Chapter 5 demonstrates how the Bayesian framework can be applied to a different setting altogether. In this study, we provide a model of how individual neurons can optimally estimate the value of their drives in the presence of fluctuations on their synaptic inputs. In short, we formulate neuronal adaptation as a statistical inference problem that individual neurons must solve in order to achieve accurate and stable perceptual performance in the face of these fluctuations that exist over many time scales. This project demonstrates how simple, local adjustments can contribute to a high-level computational goal of the sensory nervous system.

Finally, Chapter 6 explains how these methods could be used in future studies, including adaptive experimental designs in which the optimal stimulus ensemble for each neuron is inferred “online” from the behavior of the cell.

Chapter 2

Background

2.1 A Brief Introduction to Bayesian Statistics

Probability theory is a calculus for reasoning under conditions of uncertainty. In fact, a result known as Cox's theorem shows that probability provides the unique means for doing so, given a few common-sense assumptions about the goals of such reasoning (Jaynes, 2003). Furthermore, to perform inference under uncertainty, one must make assumptions about the world; Bayesian probability theory provides the tools by which to make these assumptions explicit. In particular, a formula known as Bayes rule offers the beginnings of an algorithm by which to combine information from multiple sources and draw appropriate conclusions. It can be stated as follows:

$$\Pr(B|A) = \frac{\Pr(A|B) \Pr(B)}{\int_B \Pr(A|B) \Pr(B)} \quad (2.1)$$

In this presentation, A and B denote arbitrary sets of random variables. If we further suppose that the set A contains those random variables that are fixed by way of experimental observation, and B those whose values we wish to infer, then Bayes' rule offers a prescription for how to answer scientific questions from data.

I will therefore introduce a more specific formulation of Bayes' rule that incorporates these assumptions and is the starting point for all of the models described in this thesis. Note that the denominator in equation 2.1 is a number that does not depend on B (the variables of interest), and that is fixed for any particular setting of A (the data). We can therefore write the following form of Bayes' rule, introducing names for the random variables that match common usage in the applied statistics literature:

$$\underbrace{\Pr(\theta|\mathbf{x}, \phi)}_{\text{posterior}} \propto \underbrace{\Pr(\mathbf{x}|\theta, \phi)}_{\text{likelihood}} \underbrace{\Pr(\theta|\phi)}_{\text{prior}} \quad (2.2)$$

In equation 2.2, the commonly used name for each of the terms has been noted. The *likelihood* is the probability that the particular data we observe would be generated, given a particular setting of the parameter variables, θ , and the hyperparameters, ϕ . It encompasses all of the information that the data provides about the model parameters. The *prior*, on the other hand, represents our knowledge about which parameter values are likely and unlikely, before taking any information from the data into account. The *posterior* is a combination of these two terms, and is proportional to their product. All variables may be vector valued.

For our purposes, there are two primary advantages of the Bayesian framework. First, Bayesian statistics allow for the combination of multiple sources of knowledge into a single, consistent, probability-based model. As already outlined, these sources of knowledge can include experimental measurements and data, the prior knowledge of the experimenter about the system under study, and also partial inference results obtained from earlier analyses. Second, Bayesian statistics also offers a wide range of powerful inference methods that can be used to estimate the quantities of scientific interest, including the uncertainty that remains about those quantities after the data have been collected and taken into account. These inference methods can also be used to compare the evidence provided by the data in favor of different probability models, a process known as model selection. The following sections will explain in further detail how the Bayesian statistical framework delivers these advantages, and how we will make use of them in the applications to follow.

2.2 Hierarchical Models and Graphical Models

Bayes' rule allows our knowledge about the probability distribution on a given random variable (RV) to be specified in terms of our prior knowledge about one or more other random variables. This “generative” process, in which each RV is defined in terms of those already encountered, can be repeated through as many levels as are needed; this results in a so-called “hierarchical” model. In a hierarchical model, each RV's distribution is defined in terms of known constants, as well as previously defined random variables. For example, the following lines describe one of the simplest possible hierarchical models, a Gamma-Poisson model:

$$\begin{aligned}
\alpha &= c_1 \\
\beta &= c_2 \\
\lambda &\sim \text{Gamma}(\alpha, \beta) \\
x_i &\sim \text{Poisson}(\lambda)
\end{aligned}
\tag{2.3}$$

In the terminology of equation 2.2, $\phi = \{\alpha, \beta\}$ are the hyperparameters (which are fixed, by assumption, to the values c_1 and c_2), $\theta = \{\lambda\}$ is the single (unobserved) parameter, and the x_i are the data. Note that all of the variables in the model are either fixed and known (e.g., α and β), or specified as random variables in terms of previously-defined quantities (such as λ and the x_i). We assume that the data are observed, presumably as a result of some experimental process.

Hierarchical models such as the one in 2.3 can be thought of as defining a collection of conditional independence relations among the random variables in question. In other words, the model entails a particular factorization of the joint probability density of the random variables in the model. In many cases, this factorization significantly constrains the joint density; this is very important when it comes time to perform inference on the model. For example, the Gamma-Poisson model from above factorizes as follows:

$$\Pr(\mathbf{x}, \lambda, \alpha, \beta) = \Pr(\lambda | \alpha, \beta) \prod_i \Pr(x_i | \lambda)$$

This factorization of the joint density makes clear our assumption that the data x_i are independent and identically distributed (IID), given the Poisson rate parameter λ . The x_i are also independent of α and β , given λ . Relaxing these assumptions would require a different, and in general more complicated, hierarchical model.

Once the factorization of the joint density has been represented in this manner, it is simple to translate it into a directed acyclic graph; this is known as the directed graphical model (DGM) formalism. To construct a DGM from a particular factorization of the joint density, first create a graph node for each of the random variables in the model. Next, for each variable in the model, draw a directed edge to the corresponding variable node from each of the variables on which it directly depends. For example, the λ variable depends directly on both α and β , so there will be edges from both the α and β nodes to the λ node.

A DGM depicting the Gamma-Poisson model is shown in 2.1. By convention, those random variables which are observed (for example, those collected as experimental data) are shaded in the graph to indicate this status. In addition, plate notation (e.g., the square box

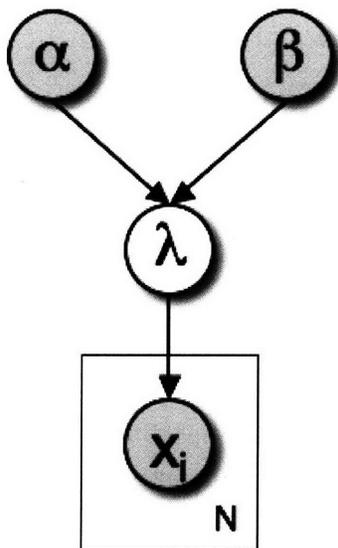


Figure 2.1: DGM Example

A Directed Graphical Model (DGM) representing the Gamma-Poisson conjugate model. Observed random variables are shaded. α and β are “hyperparameters” which define the Gamma prior distribution; λ is the rate parameter of the Poisson, drawn from the Gamma prior; and the x_i are the observed data points, drawn from the Poisson distribution.

around the x_i) is used to denote repeated structure.

Hierarchical and graphical models also make clear one of the primary advantages of Bayesian statistical inference: every quantity in the model is a random variable, and is therefore subject to uncertainty. While some variables are assumed to be fixed and known for the purposes of particular analyses, such assumptions can be relaxed when necessary. For example, Chapter 3 will provide an example of how a variable held fixed in one inferential setting (corresponding to the identity of the stimulus-response model) can instead be allowed to vary in order to perform inference on its value (i.e., which stimulus-response model is best supported by the data).

2.3 Methods of Bayesian Statistical Inference

One way to think about Bayesian statistical inference is as a process in which observations about the world (i.e., the data) are combined with a probabilistic model of how those data were produced in order to draw conclusions about the probability of different models and hypotheses. To the extent to which the model so inferred can be identified with particular conditions in the real world, this process of statistical inference can be used to reason and

	Analytical	Sampling
Exact	Conjugate Inference Junction-Tree	Rejection Sampling Coupling from the Past
Approximate	Variational Methods Belief Propagation MAP Estimation + Laplace's Method	Particle Filters Metropolis Sampling Gibbs Sampling Importance Sampling

Figure 2.2: Bayesian inference methods

draw conclusions about the nature of the environment. Inference methods vary widely, from closed-form analytical solutions to computationally-intensive, algorithmic approaches.

The (random) variables in a hierarchical model can be divided between those whose value are known, either through experimental observation or by assumption, and those whose values are unobserved. For example, in the Gamma-Poisson model of equation 2.3, both the data x_i and the hyperparameters α and β are observed. The unobserved variables can be partitioned into those whose values are currently of interest, and which therefore constitute the targets of statistical inference, and those considered “nuisance variables”. Again referring back to the Gamma-Poisson model, the single unobserved variable, θ , is almost surely a target of inference; there are no nuisance variables in this model. In order to successfully perform inference on the target variables, any nuisance variables must typically be “integrated out” of the joint density, through a process called marginalization. This may be accomplished either analytically or by approximation, with the most appropriate method determined by the structure of the model.

One common goal of inference in hierarchical models is to compute the posterior distribution of the target variables, given the assumed structure of the model and the observed data. There exists a wide range of inference techniques; a complete discussion of these methods is outside the scope of this thesis. We will note, however, two axes along which inference methods can be usefully distinguished (see Figure 2.2): whether the method is exact or approximate, and also whether it is analytical in nature or involves drawing samples from the target distribution. In some cases (referred to in Figure 2.2 as Exact-Analytical), inference can be performed without approximation, in closed form. This is the case, for example, when using conjugate models such as the Gamma-Poisson example above, as well as in a linear,

or Gaussian, setting such as the classical state-space model (Roweis and Ghahramani, 1999; Kalman, 1960). More often, however, the integration required to achieve analytical solutions is either practically impossible or computationally intractable, and it is necessary to use some form of approximate inference. Some applications in this thesis will therefore rely on a family of stochastic methods for approximate inference collectively known as Markov chain Monte Carlo (MCMC). In cases where conjugacy relations between different random variables can be exploited, it is possible to use a very efficient MCMC technique known as Gibbs sampling. In other cases, it will be necessary to fall back on a more general MCMC sampling technique known as the Metropolis sampling method (Metropolis et al., 1953). Because they are central to the arguments of this thesis, MCMC methods will be explained in some detail in the following section.

Aside from posterior parameter estimation, another common problem in statistical inference is model selection (Kass and Raftery, 1995; MacKay, 2003). In this inference setting, we compare the evidence provided by the data in favor of two or more competing models. That is, rather than asking which parameter values are most probable in light of the observations, we instead ask which model formulation is most consistent with the data. This type of inference is especially challenging, because it requires us to integrate over all possible parameter values. This integration is rarely possible to perform analytically, and it is often difficult to even obtain high-quality approximations.

Finally, we note that there are many inference methods, such as population-based Monte Carlo approaches (Smith et al., 2001), that we will not discuss here except to note that other researchers are currently working to apply them in related applications.

2.4 Markov chain Monte Carlo

We focus on two widely-used MCMC sampling methods, Metropolis and Gibbs sampling. All MCMC methods perform random walks through the probability landscape defined by the posterior distribution. The almost-magical property of valid MCMC methods is that, in the limit of infinite samples, the amount of time that the walk spends in any region of the probability landscape is proportional to the probability mass contained in that region. The collection of samples can therefore be used to approximate the full posterior distribution; in particular, any summary computed from the samples will converge to the value of that summary if it were computed on the true posterior. The method by which different MCMC methods achieve this goal are somewhat different, however.

Both Metropolis and Gibbs sampling are highly effective at characterizing which are the plausible values for a particular model parameter, given some set of experimental observa-

tions. In Chapters 3 and 4, these inference techniques will be used to determine which tuning parameter values, such as the preferred orientation and width of a circular Gaussian tuning curve, are most likely to be true, given a collection of recorded neuronal responses to a series of stimuli. Unlike some inference methods, which assume that there is a single “best” value for each parameter, sampling methods like Gibbs and Metropolis are able to entertain a full range of possibilities. For example these sampling methods would be able to discover if a particular data set provided strong support for two different hypotheses about the value of a parameter such as the preferred orientation.

2.4.1 Metropolis Sampling

Metropolis sampling proceeds as follows. We assume that the posterior density $\Pr(\theta|\mathbf{x}, \phi)$ can be evaluated for any valid setting of the (possibly vector-valued) parameter variable θ , which is the target of inference. The goal is to collect a set of M samples of θ , $\{\theta^{(1)} \dots \theta^{(M)}\}$. First, choose an initial sample value $\theta^{(1)}$. Next, given the current sample $\theta^{(t)}$, propose a new sample $\theta^* \sim q(\theta^{(t)})$, where $q(\theta)$ is the *proposal distribution*. The proposal distribution is chosen so that it is easy to sample from; in standard Metropolis sampling, it must be symmetric. If $\Pr(\theta^*|\mathbf{x}, \phi) > \Pr(\theta^{(t)}|\mathbf{x}, \phi)$, then let $\theta^{(t+1)} \leftarrow \theta^*$. If this condition is not met, then draw a random variable $u \sim U(0, 1)$. If $u < \frac{\Pr(\theta^*|\mathbf{x}, \phi)}{\Pr(\theta^{(t)}|\mathbf{x}, \phi)}$, then $\theta^{(t+1)} \leftarrow \theta^*$; otherwise, $\theta^{(t+1)} \leftarrow \theta^{(t)}$.

While this process may seem complicated, it has a simple intuition behind it. The proposal distribution is used to generate a tentative sample. If the posterior distribution, evaluated at this proposed sample, is greater in value than the posterior at the current sample, then the proposed sample is accepted. This is an “uphill” move in the posterior probability landscape. If the posterior at the proposed sample is lower in value than the current location, then the proposed sample is still sometimes accepted, with a probability equal to the ratio of the two posterior values. If the proposed sample is rejected, then the current sample value is retained as the new sample. This process is repeated until enough samples have been obtained. An example of Metropolis sampling on a simple landscape is shown in Figure 2.3 on page 24.

The Metropolis method is appropriate in a wide range of situations, because, unlike alternatives such as Gibbs sampling (described below), it is not necessary to derive or compute the conditional probability distribution of the variables of interest. This freedom comes at a cost, however: in many applications, it is crucial to use an efficient proposal distribution to suggest new sample values. In fact, much of the art in successfully applying Metropolis sampling lies in the invention of a proposal distribution that is appropriate to a given problem. Gibbs sampling avoids this difficulty, because the conditional distribution of the sampled variable serves as the proposal distribution. The particular proposal distributions used in each application will be presented in the relevant methods sections.

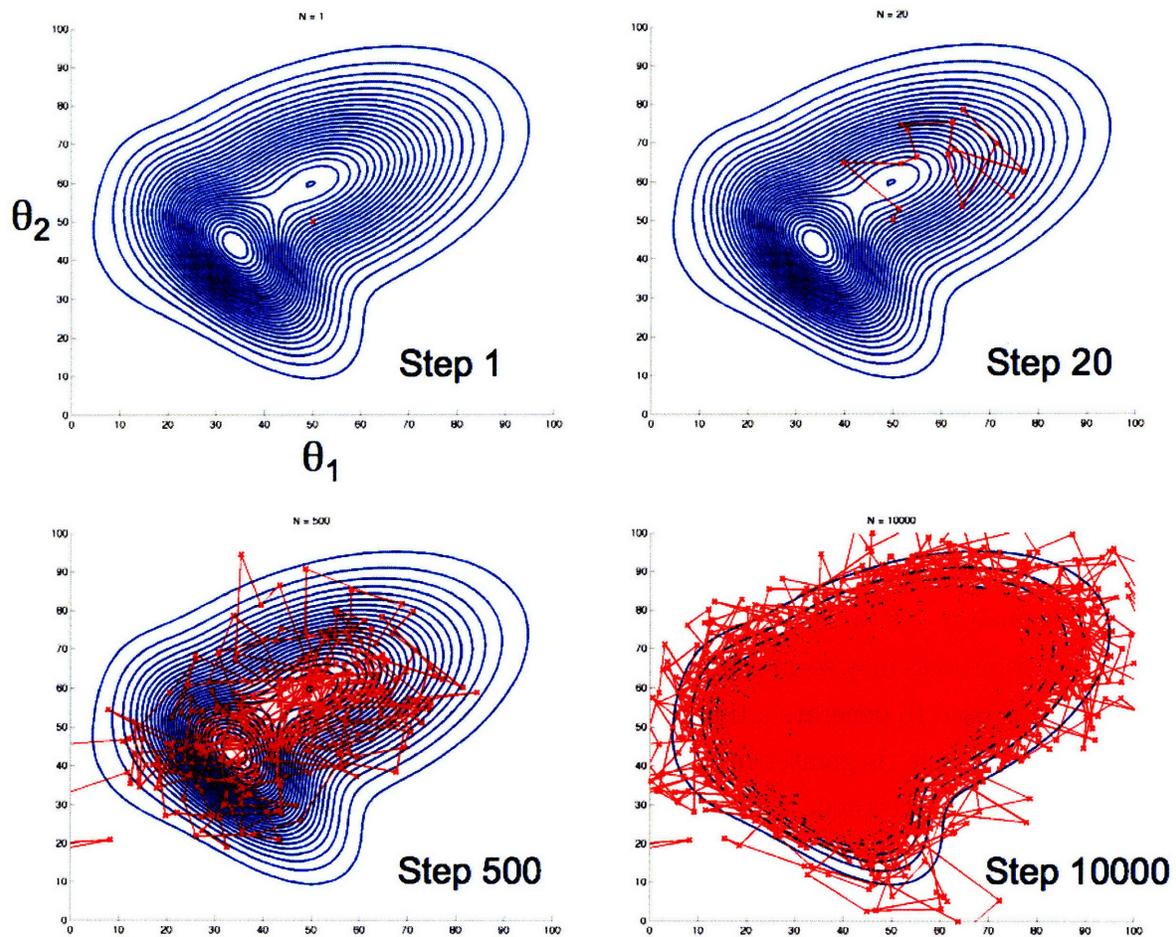


Figure 2.3: Metropolis sampling

An example of Metropolis sampling in a probability landscape of two dimension, θ_1 and θ_2 . Each vertex in the path (red x) represents an accepted sample, and successive samples in the chain are connected by red line segments. The equi-probability contours of the true posterior density landscape are shown in blue; in real sampling situations, the shape of this landscape is unknown, and it the goal of the sampling process to explore and characterize it. Snapshots are shown at the initial step, and also after 20, 500, and 10,000 samples.

2.4.2 Gibbs Sampling

Gibbs sampling proceeds as follows. We assume that the parameter variables are partitioned such that we can draw samples from the conditional distribution of one variable (or subset of variables), denoted θ_i , given all of the other variables, $\theta_{\setminus i}$; we denote this conditional distribution $\Pr(\theta_i|\theta_{\setminus i}, \mathbf{x}, \phi)$. We simply cycle through all of the unobserved variables, drawing samples from these conditional distributions in turn. Under very general conditions, this process is guaranteed to converge in distribution to the posterior.

Another way of viewing the Gibbs sampling process, depicted in Figure 2.4, is as taking successive “slices” through the posterior density landscape; these slices are the conditional distributions, and they are located at the current value of the variables $\theta_{\setminus i}$.

Because every sample is accepted, Gibbs sampling is usually more efficient than Metropolis sampling. Its use is limited by the necessity of deriving the conditional probability distribution of the target variables. While straightforward in many common hierarchical models that use exponential family distributions and conjugate priors, this derivation is difficult or impossible to perform analytically in other models.

2.5 Potential Pitfalls of Applying Bayesian Models

While Bayesian modeling and inference methods provide a very powerful set of tools by which to understand the working of the nervous system, they can also lead to erroneous conclusions if they are not applied with appropriate care. In particular, it is crucial to check the assumptions that underlie the models - while the Bayesian approach requires that these assumptions be stated, it is up to the statistician to ensure that they are reasonable. In general, there is no substitute for an intimate understanding of the data, the conditions under which they were collected, and the known properties of the neuronal systems in question. Model checking and model comparison, when feasible, can be employed to explicitly contrast the conclusions reached by different models. Even when comprehensive and rigorous checks are not possible - as may be the case for complex models - it is important to verify that the model provides a reasonable explanation of the data.

2.6 Conclusion

This chapter has outlined the framework of Bayesian statistical inference. I have introduced hierarchical modeling, a useful formalism for describing structured probability models that can combine information from experimental observations and prior knowledge. I have also in-

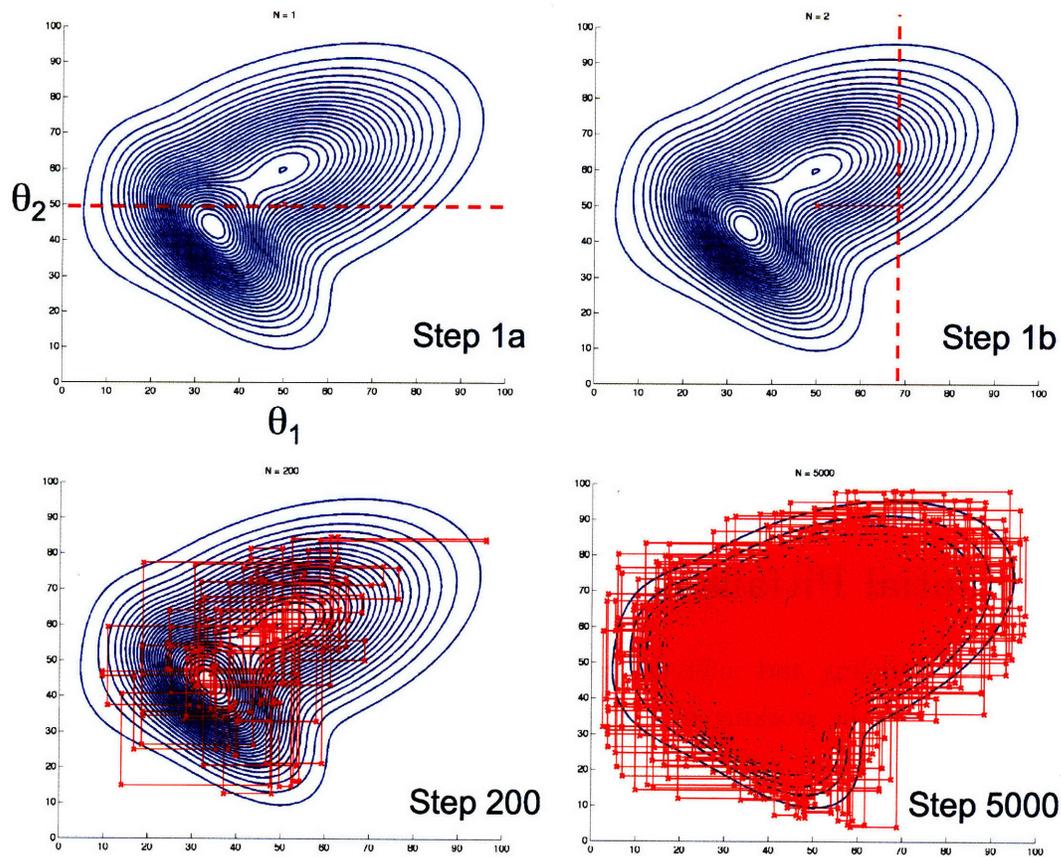


Figure 2.4: Gibbs sampling

An example of Gibbs sampling in a probability landscape of two dimension, θ_1 and θ_2 . The top two panels show the first and second samples, with the dotted red lines depicting the conditional probability distributions along which the Gibbs sampling was performed. For example, the horizontal dotted line in the first panel represents the conditional probability density $\Pr(\theta_1|\theta_2 = c)$. The bottom two panels show the samples obtained, after 200 and 5000 time points.

roduced both the Metropolis and Gibbs MCMC sampling methods, the inference techniques that are most relevant to the data analysis applications described in later chapters.

Each of the next three chapters builds on this background by describing a different application of Bayesian modeling and inference methods, and each chapter will provide further technical details about the specific methods used. Chapters 3 and 4 propose new hierarchical models, while Chapter 5 applies a standard model to a new domain. Chapter 4 also provides an innovative sampling methodology in order to make inference efficient on the high-dimensional model in question, while Chapters 3 and 5 rely on established inference schemes.

Chapter 3

Analysis of Tuning Curve Data

Tuned responses are essential for the coding of sensory information. Thus, a central theme of systems neuroscience is to characterize the selectivity and tuning of neural activity in response to sensory stimuli. Traditionally, tuning curves are determined in electrophysiological experiments that relate a neuron's activity to a range of stimulus parameters that vary along a given dimension. Central to the measurement's validity is to determine how precise this tuning is, and what the degree of uncertainty is in characterizing the tuning. Some of these questions are hard to answer in a principled way using traditional methods. Here we present a new sampling-based method that enables asking these questions in a coherent framework and that can easily be adapted to a wide range of statistical inquiries regarding tuning curves. This method also provides a useful way of visualizing which tuning curves are compatible with the data obtained from a single neuron, as exemplified by measurements of orientation and direction tuning in primary visual cortex.

3.1 Introduction

In many neuroscience experiments, a primary goal of analysis is to understand the relationship between the firing properties of a neuron and a single external variable. When the responses of sensory neurons are analyzed, this external variable represents a property of the stimulus, for example the orientation of a bar or a grating (Hubel and Wiesel, 1959; Hubel and Wiesel, 1962). When the responses of motor neurons are analyzed, this external variable typically refers to a property of the executed movement, for example the direction of movement (e.g. Georgopoulos et al. 1986).

Plotted in graphical form, these relationships are commonly called tuning curves (see

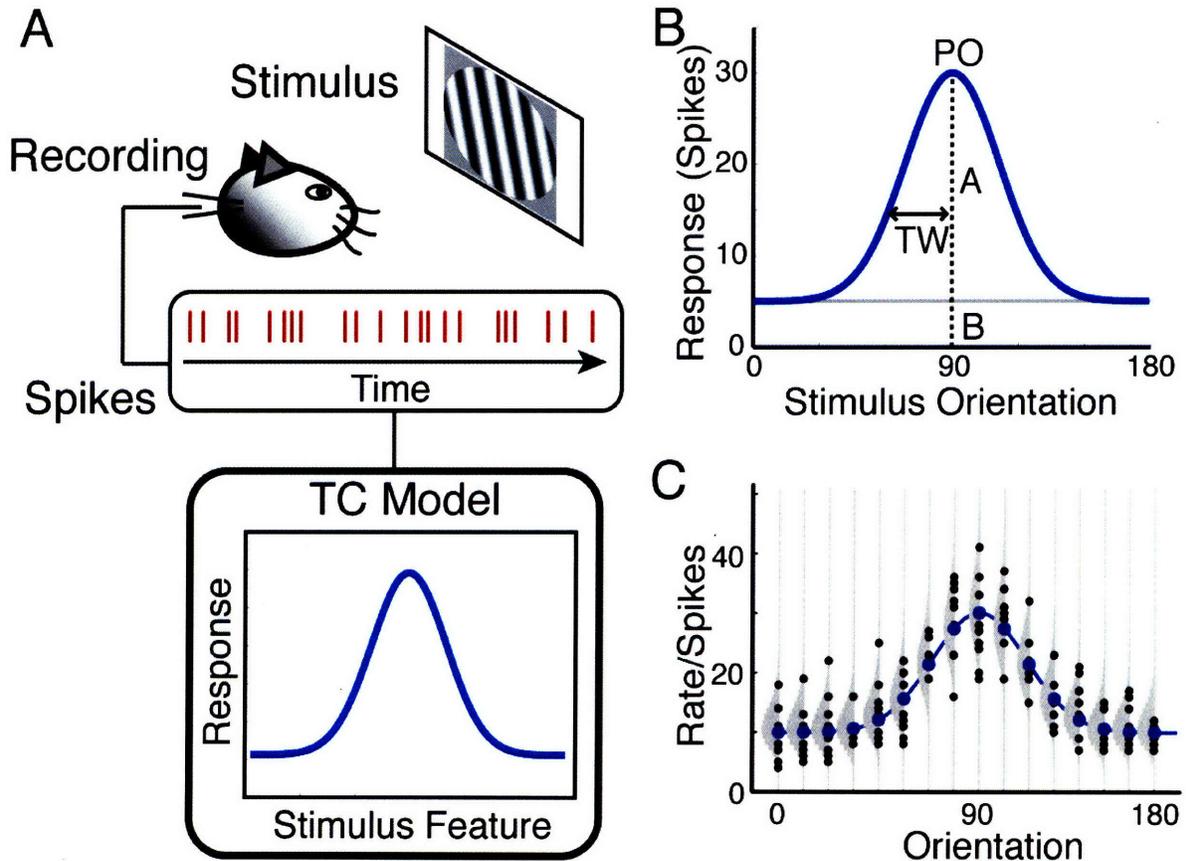


Figure 3.1: Conceptual introduction to Bayesian modeling of tuning curve data

A) A schematic illustrating the setup of the sensory neuroscience experiments for which the analyses discussed here are applicable. An animal, either anesthetized or awake, is presented with stimuli which vary along a single feature dimension. Neuronal responses to these stimuli are recorded, and then described using tuning curve functions. B) An example tuning curve model, the circular Gaussian function. The function can be described by the values of four parameters: the baseline (B), amplitude (A), preferred orientation (PO), and tuning width (TW). C) A schematic of the generative process assumed by the hierarchical statistical model. A neuron exhibits tuning described by a particular tuning curve function. The firing rate in response to a stimulus of a particular value is determined by the value of this tuning curve function when evaluated at this location. The response for a particular trial is then generated according to the assumed probability model, here a Poisson process.

panel A in Figure 3.1 on page 30) and such tuning curves are probably the most frequently used tool to describe the response properties of neurons. In these experiments, researchers often wish to answer questions such as the following:

1. Is the recorded neuron selective for the stimulus property in question? For example, a number of studies have asked what percentage of neurons in visual cortex are tuned to orientation (e.g., Maldonado et al. 1997; Ohki et al. 2005).
2. If the neuron is selective, how can its selectivity be described quantitatively? For example, many studies describe how well a neuron is tuned to the orientation of a stimulus (Carandini and Ferster, 2000; Ringach et al., 2002). Also, how much uncertainty remains in this quantification after all of the evidence from the data has been taken into account? Often studies aim at giving error bounds on estimated neuronal properties. For example, it has long been a debate if neurons have sharper orientation tuning late in response than earlier in the response (Gillespie et al., 2001; Ringach et al., 1997). Answering such questions demands reliable margins of certainty (Schummers et al., 2007).
3. Given two or more hypotheses concerning the functional or qualitative form of the neuron's selectivity, for which of these hypotheses does the data provide the most evidence (Swindale, 1988)? For example, it has been asked if orientation-selective cells in monkey striate cortex are tuned in the shape of a so-called Mexican hat (Ringach et al., 1997).
4. Do the neuron's selectivity properties change after an experimental manipulation, either qualitatively or quantitatively? For example, several studies have asked whether adaptation, pharmacological intervention, or attention affect the tuning of neural responses to visual stimuli (Dragoi et al., 2001; McAdams and Maunsell, 1999; Nelson et al., 1994; Sharma et al., 2003; Somers et al., 2001), while others have posed similar questions in the motor domain (Gandolfo et al., 2000; Li et al., 2001).

A range of techniques have been developed for the analysis of tuning properties over many decades (Hubel and Wiesel, 1959; Hubel and Wiesel., 1962). Early researchers noticed that significant insight into the functional properties of individual visual neurons could be gained by producing a simple plot of the number of spikes elicited as a function of, say, stimulus orientation. This plot, the so-called tuning curve, is the standard tool used for quantifying neural response properties. For asking various scientific questions, it became desirable to quantify these response properties in order to analyze neural responses across different animal

species and brain areas, and as a function of experimental manipulations. To this end, well-established statistical techniques such as maximum likelihood estimation (MLE) were applied (e.g. Swindale, 1988). These approaches have allowed principled answers to a large number of scientific questions. However, the statistical power of these methods is limited, and some of the above questions are very hard to answer with these techniques. In this paper we develop a set of methods that can conveniently answer the above questions with high statistical power.

We describe a Bayesian hierarchical modeling framework which allows these questions to be posed in a principled way, as well as a Markov chain Monte Carlo (MCMC) sampling process for performing the statistical inference procedures necessary to answer these questions. The approach provides a statistically rigorous setting which can be applied to a wide range of models of neural tuning, and overcomes many of the limitations of other widely used methods. While the statistical methods described here (i.e., hierarchical models and sampling-based inference) have been successfully applied in a range of scientific fields (Gelman et al., 2004; MacKay, 2003; Metropolis et al., 1953; Robert and Casella, 2004), we believe that they should be more widely used in the analysis of physiological data (but see Gillespie et al., 2001; Sahani and Linden, 2003). One major barrier to their adoption has been the difficulty in implementing MCMC sampling routines which are both appropriate and computationally efficient. To alleviate this obstacle, we make available a software toolbox which will allow researchers to easily and routinely apply all of the methods discussed here in the analysis of their own data.

3.2 Methods

The spiking patterns of neurons will depend both on the properties of a stimulus and on the tuning properties of the neuron. In a typical experiment, the spikes are observed and stimuli known, and we want to ask which tuning properties are exhibited by the recorded neuron. In Bayesian terms, we want to ask which tuning properties are most consistent with the measured data, given assumptions about their functional form (e.g. Gaussian or sigmoid). These Bayesian models assume an explicit generative process that defines how the observed firing patterns are (statistically) related to both the stimuli and the underlying tuning properties of the recorded neurons. We then use numerical methods to optimally estimate the values of the tuning properties, as well as the uncertainty which remains (i.e., the error bars).

TC Function	Formula
Constant	$R_c(S; b) = b$
Linear	$R_{lin}(S; b, a) = b + aS$
Sigmoid	$R_{sig}(S; b, a, x_o, c) = b + a \frac{1}{1 + e^{-c(S - x_o)}}$
Gaussian	$R_g(S; b, a, \mu, \sigma^2) = b + a \exp \left\{ -\frac{(S - \mu)^2}{2\sigma^2} \right\}$
Circular Gaussian	$R_{cg}(S; b, a, \mu, \sigma^2, P) = b + a \sum_{k=-\infty}^{\infty} \exp \left\{ -\frac{(S + kP - \mu)^2}{2\sigma^2} \right\}$

Table 3.1: Example tuning curve functions

3.2.1 Hierarchical Model

The key feature of our analysis method is that it describes a (hypothetical) probabilistic relationship of how the parameters of a chosen tuning curve (TC) function affect the neuronal responses to stimuli collected in an experiment (see panels B & C in Figure 3.1 on page 30). Given experimental data, such as spike counts; knowledge of the stimuli presented over the course of the experiment; and an assumption about the functional form of the TC, we want to know how likely is each set of tuning curve parameters. Some example tuning curve functions are shown in table 3.1.

In statistical terms, this means that we need to estimate the parameters of the tuning function $\theta = \{p_1 \dots p_K\}$ (such as tuning width and preferred orientation) from the observed neural responses \mathbf{x} and stimuli \mathbf{S} (both vectors of length N , the number of experimental trials). We also assume that we may have some prior knowledge $\Pr(p_j | \phi_j)$ about the properties of these parameters, as characterized by the (possibly vector-valued) hyperparameter ϕ_j . In most of the cases discussed here, we usually assume that the prior distributions are uniformly distributed on some reasonable range.

For concreteness, we specify a full hierarchical model for a setting in which the tuning

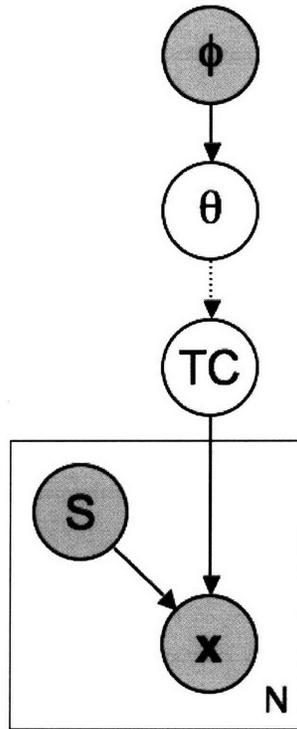


Figure 3.2: A graphical model of the tuning curve data analysis

curve function is a circular Gaussian with a periodicity of 180 degrees, the tuning curve parameters have uniform priors, and the data are spike counts assumed to be generated by a Poisson process:

$$\begin{aligned}
 \phi_1 &= \{a_1, b_1\} \\
 \phi_2 &= \{a_2, b_2\} \\
 \phi_3 &= \{a_3, b_3\} \\
 \phi_4 &= \{a_4, b_4\} \\
 B &\sim U(a_1, b_1) \\
 A &\sim U(a_2, b_2) \\
 PO &\sim U(a_3, b_3) \\
 TW &\sim U(a_4, b_4) \\
 x_i &\sim \text{Poisson}(R_{cg}(S_i; B, A, PO, TW, 180))
 \end{aligned}$$

A graphical model depicting all models of this form is shown in Figure 3.2. For such

Probability Model	Data Type
Poisson	Spike counts
Negative Binomial	Spike counts
Additive Gaussian	Fluorescence measurements (confocal & 2-photon microscopy)
Multiplicative Gaussian	Spike rates

Table 3.2: Probability models for different types of physiological data

models, we can apply Bayes' rule to derive the posterior distribution of the TC parameter variables:

$$\begin{aligned}
 \Pr(p_1 \dots p_K | x, S, \phi_1 \dots \phi_K) &\propto \Pr(x | p_1 \dots p_K, S, \phi_1 \dots \phi_K) \prod_{j=1}^K \Pr(p_j | \phi_j) \\
 &= \prod_{i=1}^N \Pr(x_i | R_{TC}(S_i; p_1 \dots p_K)) \prod_{j=1}^K \Pr(p_j | \phi_j) \quad (3.1)
 \end{aligned}$$

R_{TC} is the value which results from evaluating the TC function for the particular stimulus value S_i (i.e., the stimulus presented on trial i), given particular values of the parameter variables. The term on the left hand side of the equation is the joint posterior distribution of the parameter values, while the two terms on the right hand side are the data likelihood and the prior distributions of the parameters, respectively.

This model can be adapted to suit a particular experimental context in two main ways. Firstly, the TC function should be chosen to reflect as accurately as possible the hypothesized selectivity properties of the neuron in question. For example, for neurons in primary visual cortex, a sigmoid function might be chosen when the stimulus contrast is being manipulated (e.g., Sclar et al., 1989), while a circular Gaussian (CG) function might be chosen when stimulus orientation is varied (e.g., McAdams and Maunsell, 1999). Secondly, an appropriate probability model should be chosen to describe the relationship between the TC model output and the observed data. For example, a Poisson probability model might be appropriate when the data consists of spike counts, while a normal (i.e., Gaussian) model might be preferred when the data consists of continuous values such as fluorescence readings from confocal or 2-photon microscopy.

3.2.2 MCMC Sampling

Because the TC functions commonly used to model neural responses can impose complex nonlinear relationships between the parameters of interest, it is usually impossible to compute the joint posterior in equation 3.1 analytically. Sampling methods such as MCMC offer a powerful and flexible set of tools by which to overcome these limitations (Liu, 2002; Metropolis et al., 1953; Robert and Casella, 2004). These methods can be used to compute an approximation of the joint posterior parameter distribution.

Sampling methods operate by performing a search through parameter space. In this regard, they are similar to optimization techniques such as least-squares fitting methods. Unlike standard optimization algorithms, which only move “uphill” from less optimal solutions to better ones, sampling methods move both uphill and downhill to explore the parameter space. The resulting random walk through parameter space is biased such that the amount of time spent in any region of the space is proportional to the probability density of the parameter values corresponding to that region. As a result, MCMC sampling returns an approximation of the joint probability distribution of the model parameters, given the data, any prior information, and assumptions about the form of the model (see panel A in Figure 3.3 on page 37).

To sample the TC parameter values, we employ a simple Metropolis sampler that follows the basic process outlined in Chapter 2. Specifically, we sample for each parameter p_i by starting with the current sampled value, $p_i^{(t)}$, and proposing a new value p_i^* . While this proposal can in principle be generated in a number of ways, we will use normally distributed proposals whose mean is the current sample value, i.e. $p_i^* \sim N(p_i^{(t)}, \sigma_0)$. Next, the probability in equation 3.1 on the preceding page is computed for the two different values of p_i , yielding the following two quantities:

$$\begin{aligned}\pi_i^* &\propto \Pr(p_1^{(t)} \dots p_i^* \dots p_K^{(t)} | \mathbf{x}, \mathbf{S}, \phi_i \dots \phi_K) \\ \pi_i^{(t)} &\propto \Pr(p_1^{(t)} \dots p_K^{(t)} | \mathbf{x}, \mathbf{S}, \phi_i \dots \phi_K)\end{aligned}$$

If π_i^* is greater than $\pi_i^{(t)}$, then the proposal is accepted. If π_i^* is less than $\pi_i^{(t)}$, then it might still be accepted according to the usual Metropolis acceptance rule.

It can be shown that sampling according to this procedure will result in parameter samples whose values approximate the joint posterior distribution of the parameters, given the data (Metropolis et al., 1953). The samples can therefore be used as a proxy for the complete joint parameter distribution, and questions about this true distribution can be answered by referring to the samples. Panel B in Figure 3.3 on page 37 shows the results of sampling

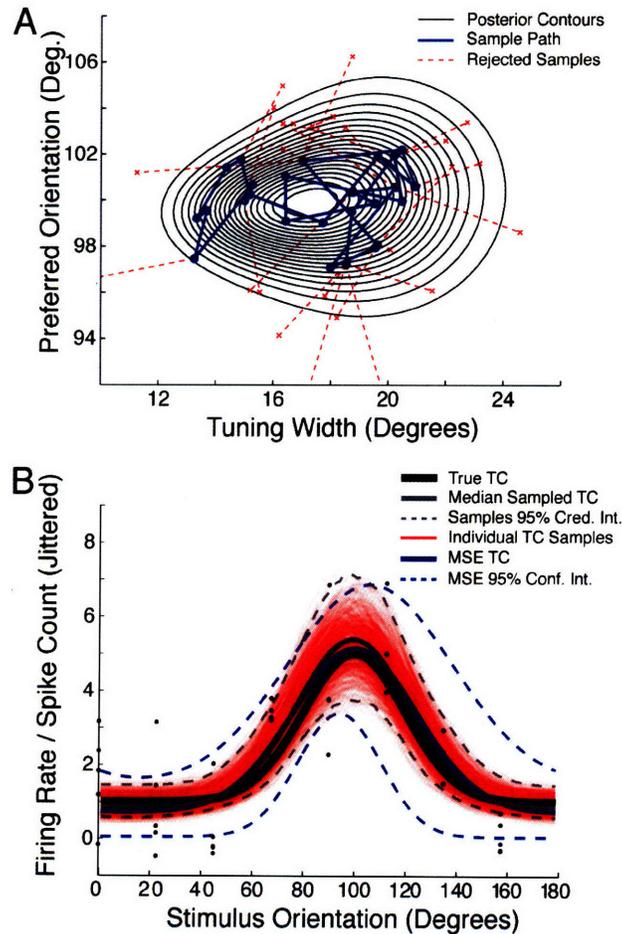


Figure 3.3: Methods for sampling from tuning curve data

A) Metropolis sampling proceeds by making a biased random walk on the surface defined by the joint posterior probability density of the model parameters. Shown here are two dimensions of this surface for a circular Gaussian tuning curve function; the preferred orientation and tuning width parameters are shown, while the amplitude and baseline parameters are omitted for clarity. The grey lines are the contours of uniform density. The blue path shows the trajectory of the samples, with rejected proposals shown in red. Note that rejected samples always represent “downhill” moves, while accepted samples can be either uphill or downhill. If enough samples are collected, the number of samples per region will converge to match the height of the density contours in the region; in other words, the samples will approximate the distribution. B) Each sample corresponds to a different circular Gaussian tuning curve. Shown here are sampled tuning curves (light red lines) which are obtained by applying the generative model to simulated spike count data (black dots) generated from a known tuning curve (solid black line). Both the sampling method (solid grey line) and MSE-based nonlinear optimization (solid blue line) perform well in recovering ground truth in this case. However, the 95% error bounds generated from the samples (dashed grey lines) are narrower than those obtained from optimization, indicating that the hierarchical model and sampling approach is able to make more efficient use of the data. In addition, plotting the individual samples provides a nuanced understanding of the range of possible tuning curves which could have produced the observed data.

using a circular Gaussian tuning curve function on simulated data, for which ground truth is known.

It should be noted that optimization techniques, such as those which use the mean squared error (MSE), also make (implicit) assumptions about the joint probability distribution of the model parameters. In particular, such methods assume that the parameters are normally distributed, an assumption which may not be appropriate in many cases. Panel B in Figure 3.3 on page 37 also shows the fit obtained with such a method.

In the discussion which follows, we assume that we have applied an MCMC procedure like the one just described to obtain M samples for each of the K tuning curve model parameters: $\mathcal{P} = \left\{ \{p_1^{(1)} \dots p_1^{(M)}\} \dots \{p_K^{(1)} \dots p_K^{(M)}\} \right\}$.

3.2.3 Bayesian Model Selection

Once we use Bayesian sampling methods to estimate tuning functions we can straightforwardly use these methods to also ask if one model is better at explaining the data than another model. To compare different models in this way, it is necessary to compare how well each of them can explain the data. In a Bayesian setting, this implies integrating over the specific parameter values (in eq. 1). Bayesian model selection relies on the ratio of the probabilities of each model holding true, given the experimentally observed data (MacKay, 2003):

$$\frac{\Pr(M_2|\mathbf{x})}{\Pr(M_1|\mathbf{x})} = \frac{\Pr(M_2) \Pr(\mathbf{x}|M_2)}{\Pr(M_1) \underbrace{\Pr(\mathbf{x}|M_1)}_{\text{Bayes Factor}}} \quad (3.2)$$

Here, M_1 and M_2 are two models to be compared, and \mathbf{x} is the observed data, as before. The first term on the right-hand side is the prior preference for one model over the other. The second term, called the Bayes factor, is the ratio of the likelihood of the observed data, given each model (i.e., with all model parameters integrated out). This Bayes factor measures how much support the data provides for one model relative to the other. For example, if this quantity is 10, then it would be reasonable to say that the data provides evidence that model 2 is 10 times as likely to explain the data as model 1. In general, it can be very difficult to actually perform the necessary integration to obtain the Bayes Factor. For models of the complexity described here, however, very reasonable approximations of this integral can be obtained directly from the MCMC samples. One reasonable approximation of this quantity is simply the harmonic mean (Kass and Raftery, 1995) of the sampled probabilities:

$$\Pr(\mathbf{x}|M) \approx \left(\frac{1}{m} \sum_{j=1}^m \Pr(\mathbf{x}|p_1^{(j)} \dots p_K^{(j)}, M)^{-1} \right)^{-1} \quad (3.3)$$

While this quantity is easy to compute from the output of the MCMC sampling routine, it is known to be numerically unstable in some cases. To test whether this instability is relevant in the present application, we performed simulations in which the number of samples was increased to a very large number; these are described in more detail in the results section.

3.3 Results

We demonstrate how these general methods can be applied by experimentalists to answer a range of scientific questions which frequently arise in the analysis of physiological data.

3.3.1 Estimating parameter values

In many experiments, one of the first challenges confronting an experimentalist is to quantify the tuning properties of each recorded neuron. It is important to compute not only the most probable response properties, but also the range of responses which are plausible given the observed data, also known as the error bars. We can pose this question more precisely by asking, for each parameter of the chosen TC function, what range of parameter values is consistent with the set of observed neuronal responses? For example, what is the plausible range of preferred orientations for an orientation selective neuron? Because the collection of MCMC samples constitutes an approximation of the joint distribution of the parameter values, questions of this type can be answered directly from these samples. To obtain, say, a 95% credibility interval for parameter p_i , the samples $\{p_i^{(1)} \dots p_i^{(M)}\}$ are first sorted from lowest to highest to obtain the rank order $\langle p_i^{[1]} \dots p_i^{[M]} \rangle$. Next, the lowest and highest 2.5% of the samples are discarded. For example, if 1000 samples were collected, then the lowest and highest 25 would be discarded. The remaining extreme values define the desired error margin (see panels A & B in Figure 3.4 on page 40). In the 1000-sample example, these would be the 26th and 975th sample from the ranked list. Importantly, this method does not need to make the assumption that the distribution is Gaussian.

3.3.2 Visualization of potential tuning curves

The presented approach comes with the additional advantage of visualizing the set of potential tuning curves that are compatible with the observed data. This visualization can be a useful

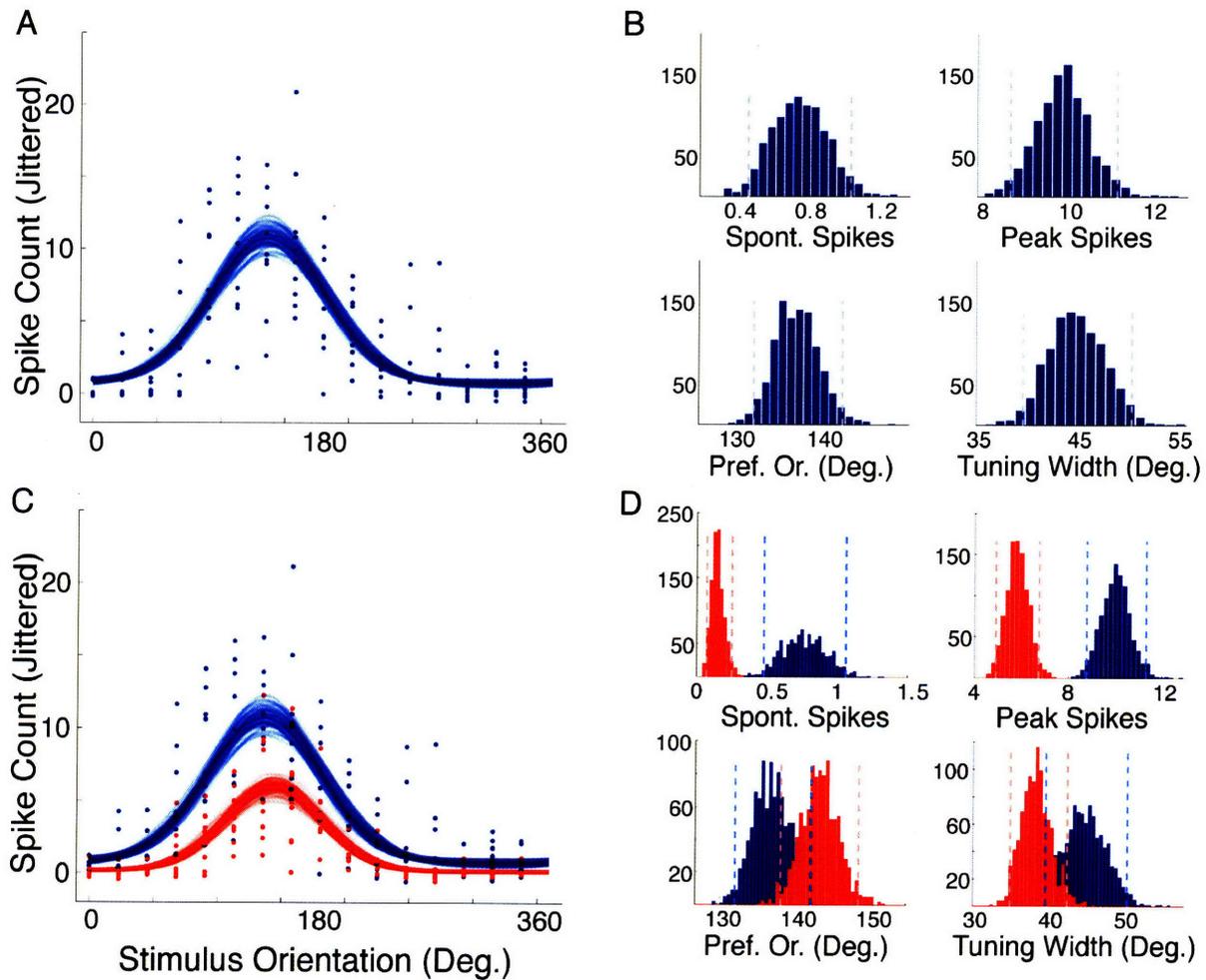


Figure 3.4: Parameter estimates from samples on tuning curve data

A) Spike counts (dots) and sampled tuning curves (lines) from a neuron in anesthetized mouse visual cortex, recorded in response to gratings of different orientations. On each trial, a different orientation was chosen (x-axis) and the number of spikes emitted in response was recorded (y-axis). B) Parameter sample histograms and credibility intervals for the sampled tuning curve parameters. Each panel displays the samples for a different parameter: the spontaneous spike count, the number of spikes emitted in response to the preferred orientation, the identity of the preferred orientation, and the tuning width of the response. Dashed gray lines indicate the 95% credibility intervals for each of these parameter values. C) Spike counts and sampled tuning curves of the same cell, recorded under both control (blue) and adapted (red) conditions. The tuning curves suggest that the cell has undergone a significant change in its response properties. D) Histograms of each of the sampled tuning curve parameter values and their corresponding credibility intervals, similar to B, but here comparing the values obtained in the control and adapted conditions. Both the baseline firing and amplitude at preferred orientation have changed significantly between the two conditions, as indicated by the fact that the credibility intervals are disjoint. The preferred orientation and tuning width parameters are not significantly different between the two conditions, because the intervals are overlapping. For all example data shown here, extracellular recordings were performed in mouse visual cortex, using methods similar to those described in (Wang et al., 2006).

tool for understanding the quantitative power that is provided by the data from a single cell (as used in Figure 3.4 on page 40 and Figure 3.5 on page 42).

3.3.3 Assaying quantitative changes in response properties

Quantifying tuning properties using the preceding method is often merely the first step in answering scientific questions about physiological data. Frequently, we wish to know whether a particular manipulation has resulted in a significant change in tuning properties. For example, it might be asked whether visual neurons change their orientation selectivity properties in response to attentional influences, sensory pattern adaptation, or perceptual learning (Dragoi et al., 2001, 2000a; McAdams and Maunsell, 1999). In these kinds of experiments, neuronal responses would be measured in both the control and test conditions, and MCMC samples would be obtained separately for these two sets of data. In order to determine whether the data supports a conclusion of significant changes in tuning, credibility intervals are computed for each of the two conditions as above. To ascertain whether a significant shift has occurred, it is sufficient to observe whether these two intervals overlap. If there is no overlap (i.e., the intervals are disjoint), then the data supports the conclusion that the corresponding response property has changed between the two conditions (see Figure 3.4 on page 40, panels C & D). If the two intervals do overlap, then this conclusion is not supported. This negative result is consistent with two possibilities. The underlying response property might really be unaffected by the experimental manipulation, which is a true negative result. Another possibility, however, is that the tuning property has indeed changed but not enough data has been collected to reduce uncertainty to the point where this change can be reliably detected. In neither of the cases will this method wrongly report a change.

3.3.4 Assessing selectivity to the stimulus feature using model selection

In many cases, it is of primary scientific interest to determine whether a neuron is actually selective for the stimulus property which is being varied in the experiment. For example, many but not all neurons in striate cortex are selective for particular stimulus orientations (Maldonado et al., 1997). Before drawing conclusions about the quantitative details of each neuron's orientation-dependent response, it is necessary to assess whether the cell is selective in the first place.

Bayesian model selection (BMS) can be used to determine how much evidence is provided by the data for one response model over another (MacKay, 2003). In order to determine whether a cell is indeed selective for a stimulus property, BMS can be employed to compare

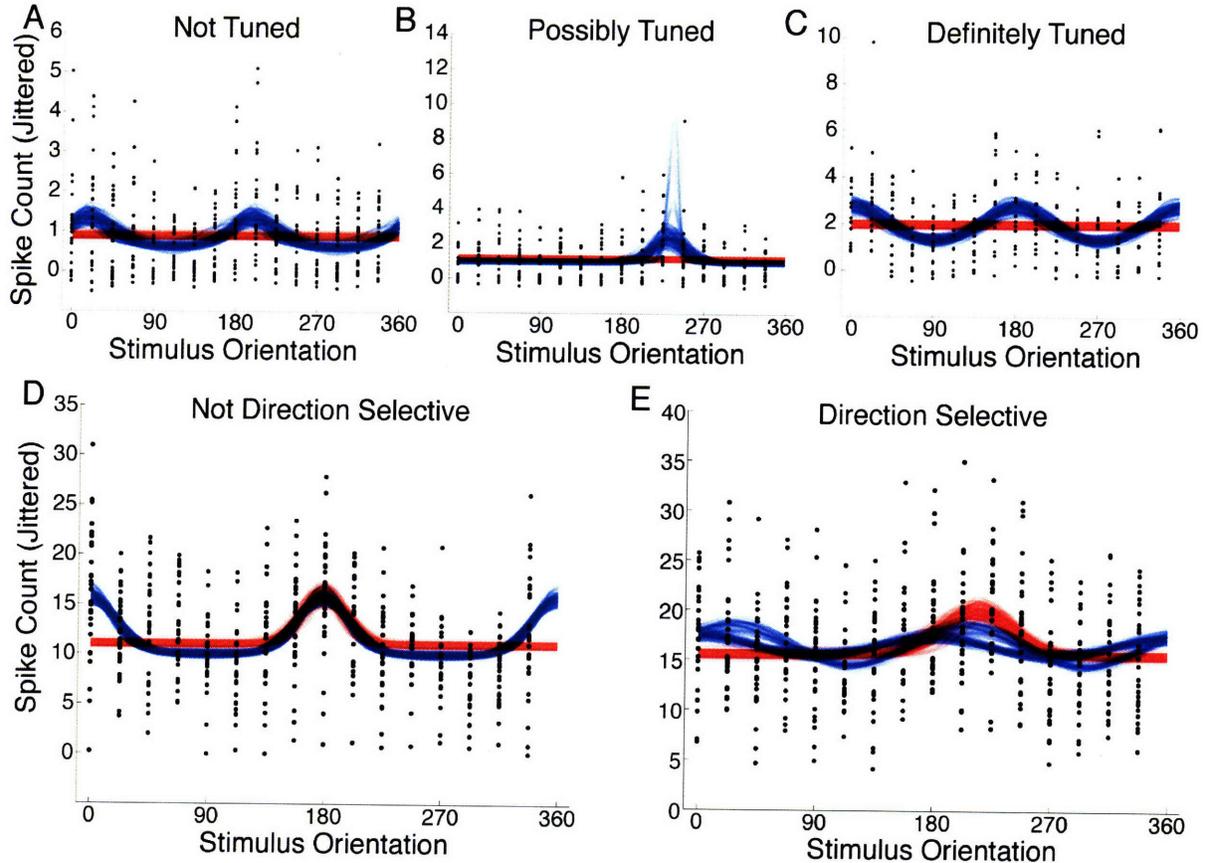


Figure 3.5: Model selection for tuning curve data

A) Spike counts (dots) and sampled tuning curves (lines) for an example cell which is not significantly tuned, as indicated by the fact that the evidence provided by the data for a tuned model (blue lines) is less than that for an untuned model (red lines). B) As in A, except that the data provide marginally more support for the tuned model. C) As in A, except that the data provide overwhelming support for the tuned model over the untuned model. D) As in A-C, except that two different tuning models are being compared. Blue lines correspond to samples from a circular Gaussian TC model with periodicity of 180 degrees, and red lines indicate a circular Gaussian with periodicity of 360 degrees. For the cell shown here, the data provide more evidence for the 180-degree model, indicating that the cell is not significantly tuned for the direction of motion of the stimulus. E) As in D, except that the data for this neuron provide more evidence for the 360-degree periodicity model, indicating that the cell is selective for the direction of motion of the stimulus.

models with two different TC functions: one in which the response varies with the stimulus value, and another in which the response is assumed to be insensitive to the particular stimulus (see Figure 3.5 on page 42, panels A-C). BMS is performed by computing the Bayes Factor (BF), which is the ratio of probability assigned to the data by each model, integrating over all possible parameter values.

A key property of BMS is that it appropriately penalizes models with more degrees of freedom. This “regularization” ensures that models with extra parameters are not favored merely because they are more expressive, which is a well-known complication in model comparison procedures (MacKay, 2003).

3.3.5 Comparing different selectivity models

A number of additional scientific questions can be also posed in terms of Bayesian model selection. For example, some neurons in primary visual cortex of the cat are direction-selective, meaning that their responses to a grating of a particular orientation depend on the direction of motion of the grating. Other cells are not direction selective, and their responses are not modulated by the direction of motion of the grating. BMS can be used to distinguish between these two types of cells. In this case, two different TC functions would be used, each of which corresponds to one of the two hypotheses. The first is a circular Gaussian function with a periodicity of 180 degrees; this function represents a non-direction selective response, because its values are the same for stimuli of opposite direction. The second TC function is a circular Gaussian with a periodicity of 360 degrees, which represents responses for only one direction of motion (see Figure 3.5 on page 42, panels D & E). When BMS is applied in this setting, using the same methods as described previously, its output indicates the strength of evidence provided by the data that a particular cell is tuned for direction of motion.

3.3.6 Performance of the harmonic mean estimator

It is generally impossible to compute the Bayes Factor directly because, like the integration operation necessary to directly perform Bayesian inference, this computation takes time which is exponential in the number of parameter dimensions in the model. It is therefore necessary to employ an appropriate estimation technique to approximate the Bayes Factor. We employ the harmonic mean estimator described in the methods (Kass and Raftery, 1995). While it is easy to compute the value of the estimator from the log likelihood scores of MCMC samples, this method is known to have some limitations that cause it to be numerically unstable. For this reason, we have performed a number of analyses to check that the estimator provides reliable answers when applied to the tuning curve models.

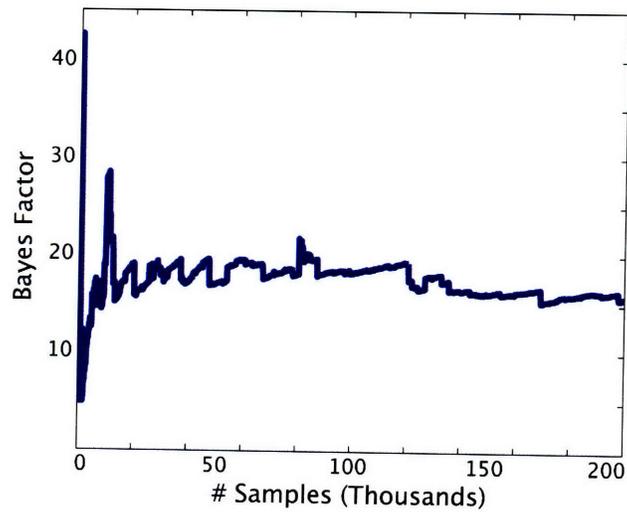


Figure 3.6: Performance of the harmonic mean estimator

This plot shows the Bayes Factor computed using the harmonic mean estimator, as a function of the number of MCMC samples collected. In this case, circular Gaussian and constant tuning curve models were compared for their ability to explain a simulated data set. It shows that, while relatively large jumps in the estimate do occur even after many samples have been collected, the estimator is able to reliably and correctly indicate that the data is tuned for orientation. Plots such as this one can easily be computed when analyzing real data as well, and may be useful in determining the model selection process is producing reliable results.

Figure 3.6 shows an example of such an analysis, which demonstrates that the harmonic mean estimator undergoes relatively large jumps in value even after many samples have been taken into account. However, these jumps do not change the basic conclusion that the cell strongly prefers a tuned model. If the estimated Bayes Factors were more marginal, though (in, say, the range between 5 and 10), these jumps might affect the overall conclusion of the model selection. In such borderline cases, it is probably better not to rely on the output of this estimator.

Obtaining accurate estimates of the marginal likelihood - the quantity that lies at the heart of the Bayes Factor calculation - is a famously difficult problem in computational statistics. The harmonic mean estimator used here has provided adequate performance in the context of assessing the evidence from data in support of different tuning curve models, and it is the only widely accepted method that is easily and directly computable from the standard output of MCMC samplers. It will be important in future work, however, to further validate the performance of this estimator, and also to evaluate alternative estimation methods. Indeed, improved methods are an active area of research, and any future advances can be readily applied to the tuning curve analysis framework (see, for example, Chib and Jeliazkov, 2001).

3.4 Comparison with the Bootstrap

The methods described in this chapter comprise a complete and principled framework for the analysis of tuning curve data, and they are sufficient to answer virtually all scientific questions which can be posed in the context of such experiments. Over the years, however, physiologists have adopted a number of other techniques to provide answers to some of these questions. While an exhaustive review of these alternatives is outside the scope of this thesis, we will discuss one alternative approach in detail: the combination of nonlinear least-squares optimization for tuning curve fitting and the bootstrap resampling methods for assessing uncertainty about the tuning curve parameter values.

3.4.1 The Bootstrap

The bootstrap method is a very general and powerful approach to statistical inference that relies on the basic operation of creating surrogate data sets by resampling from the observed data (Efron and Tibshirani, 1986). Given a set of (presumed independent) observations $\mathbf{x} = \{x_1 \dots x_n\}$, a single bootstrap sample $\mathbf{x}^{(b)}$ is created by sampling, with replacement, from the original collection of data until the sample contains the same number of data points as the original set. For example, $\mathbf{x}^{(1)} = \{x_1, x_1, x_3, x_4, x_4\}$ is a valid bootstrap sample when

$n = 5$.

The bootstrap can be used to compute the uncertainty (i.e., error bars) about any summary of the observed data, $T(\mathbf{x})$. The general bootstrap procedure is as follows:

1. Construct B bootstrap samples from the original observations, as described above
2. For each sample b , compute $T(\mathbf{x}^{(b)})$
3. Construct confidence intervals from the sample summaries

Because $T(\mathbf{x})$ can be essentially any summary of the data, the bootstrap is extremely general, and can be used to assess the accuracy of almost any data-derived function. In addition, the nonparametric form of the bootstrap described here does not make any assumptions about the probability distribution that generated the data. The parametric bootstrap is a related method that inserts a step in which the parameter values for the hypothesized distribution are estimated from each bootstrap sample; the resulting confidence intervals are then estimated from these sampled parameter values.

3.4.2 Nonlinear least-squares optimization

Optimization is a process by which to identify the values that maximize (or minimize) the value of a certain function:

$$\arg \max_x h(x)$$

The optimization literature is vast, and we will not attempt to even summarize it here. Because optimization methods have been studied intensively for many decades, and because they are used in so many different fields of science and engineering, very high quality implementations are available. For the purpose of this comparison, we will therefore use the implementations provided by the MatLab optimization toolbox.

3.4.3 Combining the bootstrap and nonlinear optimization

Our complete method for assessing the confidence intervals on the tuning curve parameters combines nonlinear optimization for curve fitting, and the bootstrap for assessing the degree of uncertainty about those estimates:

1. Draw B bootstrap samples from the observed data
2. For each bootstrap sample, use constrained nonlinear optimization with a squared-error objective function to find the tuning curve parameter values that best fit the data

3. For each tuning curve parameter of interest, use the estimate provided by the optimization routine on each bootstrap sample to construct confidence intervals

This procedure results in a set of samples that can be compared directly to those samples obtained from the Metropolis procedure.

Before proceeding, we note that this alternative method incorporates a set of assumptions that are comparable to those encoded in the Bayesian generative model. The least squares objective function used in the optimization step constitutes an implicit noise model for the data, corresponding to an assumption of Gaussianity. In addition, the optimization constraints - which are necessary to ensure the convergence of the optimization routine - contain similar information to the uniform prior distributions imposed on the TC parameters in the Bayesian model. I argue that the full generative model makes these assumptions more explicit, and more open to criticism. The generative approach also affords greater flexibility to modify the assumptions when appropriate.

3.4.4 Comparing the methods

Figure 3.7 shows the results of applying both Bayesian statistical inference and the bootstrap sampling described above to a simulated cell when many trials are available. The two methods give nearly identical results: both recover the true tuning curve with similar amounts of residual uncertainty. This is not surprising, given that the data in this case, while sufficiently unambiguous to indicate a general form for the solution, nevertheless display enough variability to require substantial error margins.

The advantage of the Bayesian approach becomes clear when less data is available from particular neuron. Figure 3.8 compares the two approaches as applied to a simulated cell from which a relatively small number of experimental trials are available. In this case, the Bayesian statistical inference approach indicates its increased uncertainty more gracefully than the bootstrap method. This is because the bootstrap, based on the process of drawing with replacement from the data, runs into trouble when only a small pool of data is available from which to resample.

While this example may seem extreme in the small amount of data that is collected, it should be noted that the amount of time available to record from a given neuron is often quite limited. This limitation is compounded in experiments in which it is necessary to characterize the neuron's response to a variety of different stimuli; in such cases, it is highly desirable to assess tuning properties with as few trials as possible. In other cases, estimating the tuning properties is merely an initial step that is used to guide further stimulus presentations. For example, the experimenter may wish to quickly assay the preferred orientation and tuning

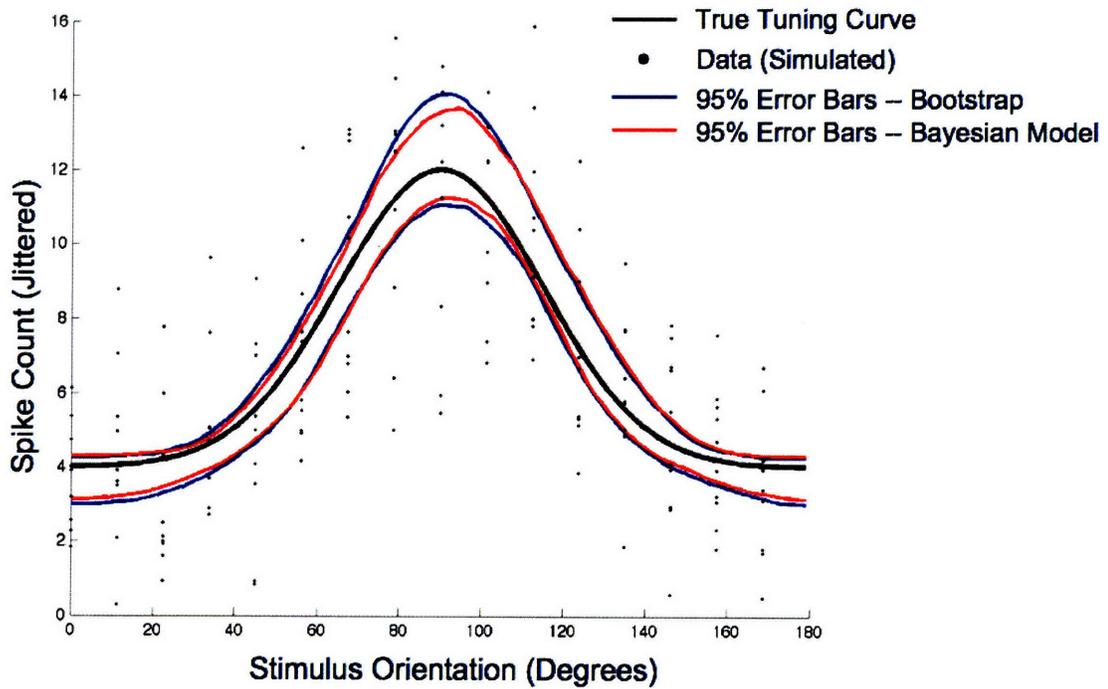


Figure 3.7: Comparison 1 between sampling and the bootstrap

This figure compares MCMC sampling and the bootstrap, when applied to simulated data (black dots) generated by a Poisson process from a circular Gaussian tuning curve (solid black line). In this case, both methods return similar error margins, and both succeed in recovering the true tuning curve.

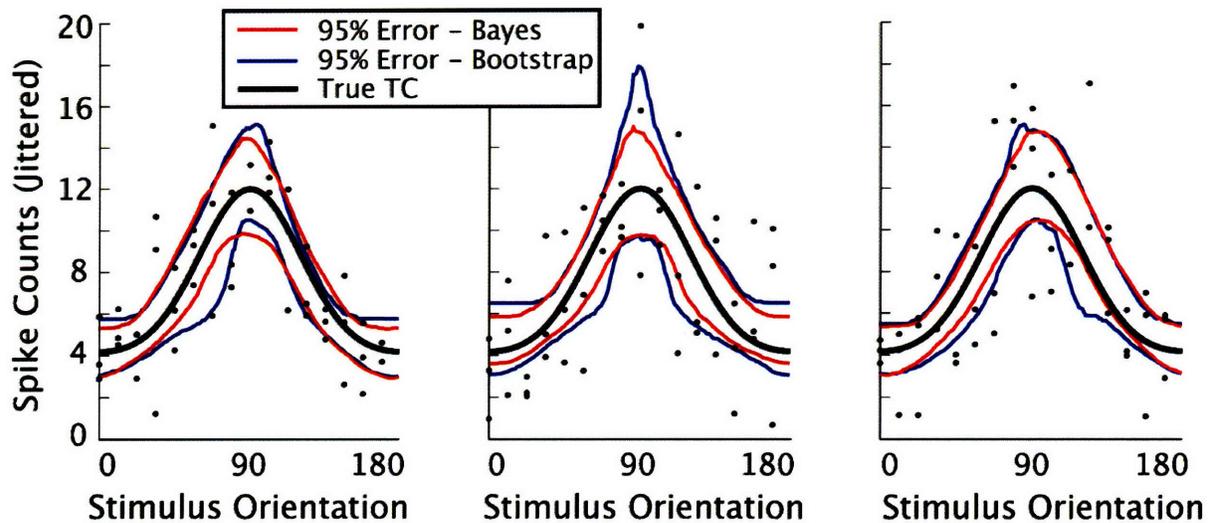


Figure 3.8: Comparison 2 between sampling and the bootstrap

When less data is available, the bootstrap becomes unstable and may provide less satisfying answers. The Bayesian approach is able to indicate its greater degree of residual uncertainty in a more graceful fashion. Shown above are the results from analyzing three separate simulated data sets, all generated from the same underlying model.

width of an orientation-selective neuron, so that later trials can be guaranteed to strongly drive the neuron. In such settings, it is of paramount importance to assess tuning properties as quickly as possible.

As a final example, we note that, in some cases, the experimenter is unable to present all of the stimulus values which would have been optimal in order to unambiguously characterize the tuning properties of a particular neuron. Figure 3.9 shows yet another simulated neuron in which stimuli of different spatial frequencies are presented to characterize the cell's tuning along that particular stimulus dimension. In this case, only a handful of spatial frequencies were presented, and the evidence provided by the resulting data proves to be ambiguous. By including prior information about the peak firing rate of the neuron (implemented as a normally-distributed prior distribution centered at the peak observed response), the Bayesian approach is able to recover plausible tuning curves from this impoverished data set. It is important to note that the experimenter must justify the prior knowledge that is used, either by recourse to the literature, past experimental experience, or data from other sources. Bayesian inference does not provide a magic resolution to such scientific arguments, but it does provide a consistent and rigorous method by which to incorporate information from multiple sources.

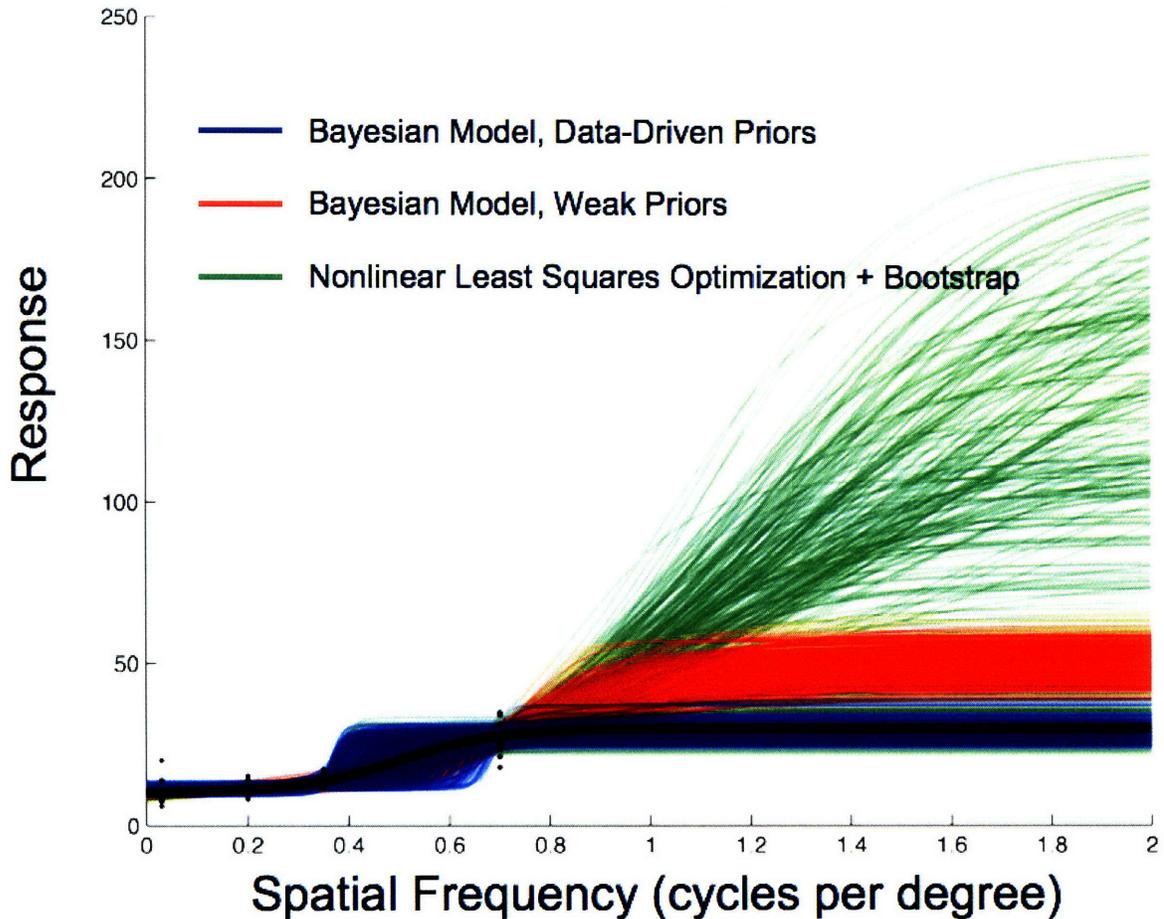


Figure 3.9: Inclusion of prior knowledge to improve TC estimates

In some cases, the experimenter is unable to present all of the stimulus values that would ideally serve to sufficiently identify the tuning properties of a given neuron. In this simulated case, only four spatial frequencies were presented, and the resulting data is consistent with a wide range of tuning curves (red and green lines indicate results from Bayesian inference and bootstrap sampling, respectively). Bayesian statistical inference offers the option of incorporating additional prior knowledge, however. Shown in blue are the tuning curve samples achieved under a model in which the peak firing rate of the neuron is assumed to be approximately the same as the maximum firing rate observed over the course of the experiment. This simple assumption is enough to vastly constrain the range of plausible tuning curves, and permits the analysis to recover the true tuning curve (black line).

3.5 The BayesPhys Toolbox

A primary goal of this chapter is to demonstrate that Bayesian modeling and sampling methods are broadly applicable, and can be made widely accessible to experimental neuroscientists. To this end, the methods described here have been implemented in a MatLab toolbox called BayesPhys. This toolbox allows all of these analyses to be performed with minimal effort. In particular, the user does not need to implement the Metropolis sampler, the tuning curve functions, or the posterior distribution calculations.

The toolbox is described in detail in Appendix A.

3.6 Discussion

Here we have shown how methods based on MCMC can help answer frequently-occurring questions in the realm of electrophysiology. We have shown that the same method can be used to estimate properties of neural tuning, obtain bounds on their values, examine if they change from one condition to the next and ask which model best explains the data. These methods are easy to implement and have the potential to significantly improve practical data analysis.

Throughout this paper we used language that relates to the estimation of the tuning properties of sensory neurons. However, the same set of methods can be used for the analysis of movement data where neural firing patterns depend on movement decisions instead of sensory stimuli (e.g., Gandolfo et al., 2000; Georgopoulos et al., 1986; Li et al., 2001).

Over several decades, researchers have accumulated a range of standard analytical techniques that address questions similar to those that we ask. For example, the orientation selectivity index (OSI) is a standard method of assessing the degree to which a neuron is selective for the orientation of stimuli (Dragoi et al., 2000; Sharma et al., 2003). However, OSI does not readily allow for the determination of whether a change in tuning is statistically significant, or if the neuron is reliably selective at all. The same value of OSI may be reached in two different ways. Firstly, in a situation where few spikes were used to calculate the OSI, random fluctuations may produce some apparent orientation selectivity. Alternatively, the same OSI value may result from a situation where a large number of spikes were obtained and the neuron is significantly but very weakly tuned to orientation. The way this problem is usually avoided is by setting thresholds for both the value of the OSI and the amount of data available. Our method provides an answer to the same question that has a strong statistical motivation.

As another example, most analyses are based on an implicit assumption of Gaussian

distributions, an assumption which is inappropriate for some common forms of neuronal data (e.g., spike counts). The hierarchical model described here can incorporate a range of probability models, such as the Poisson and negative binomial distributions in addition to Gaussian. This modeling flexibility also means that the framework described here can be directly applied to data other than spike counts, including fluorescence measurements from confocal and two-photon microscopy. There are also a wide range of easily applicable TC functions, including sigmoids, bilinear, von Mises functions, rectified cosines, and so on.

A primary advantage of the method described here is its ability to provide estimates of the uncertainty which remains concerning derived quantities, such as the preferred orientation or width of a tuning curve. These quantities, while ultimately resting on the solid foundation of a rigorously defined probabilistic model, can nevertheless be defined in terms of complex, nonlinear mathematical relationships. This capability, that of estimating in a principled way the statistical distribution of TC parameters already familiar to experimentalists, should enable many researchers to apply these methods and answer questions about their data in a natural way.

3.7 Next Steps

The hierarchical model introduced in this chapter, while applicable to the analysis of data from a wide range of experiments, is nevertheless quite limiting in at least two key ways.

The first limitation imposed by the model flows directly from the nature of tuning curves: such analysis is only applicable when a single stimulus feature is varied at a time. In fact, sensory neurons are known to be simultaneously selective for multiple stimulus features. As but one example, individual neurons in the mammalian primary visual cortex can be tuned for angular orientation, spatial frequency, direction of motion, color, and other stimulus properties. The tuning curve, while useful for understanding the response properties of neurons to single stimulus features in isolation, is not sufficiently expressive to explore the full richness of neuronal responses.

The general framework described here can be readily extended to more complex response models, which take into account the simultaneous selectivity of neurons to multiple stimulus features. For example, we have performed preliminary analyses of intracellular data collected in cat area 17 complex cells in order to determine the number of simple cell subunits that contribute to the complex cell's response (data not shown). These analyses indicate that complex cell responses may reflect input from a greater number simple-cell subunits than are suggested by standard models (Adelson and Bergen, 1985); i.e., there may be evidence for six or more simple-cell inputs, rather than just four (in agreement with Rust et al., 2005). This

type of cascaded linear-nonlinear-stochastic model represents a fully Bayesian reframing of the classical LNP model (see Chapter 6).

Another significant limitation of the tuning curve model is that it does not include any notion of temporal dynamics, either from the stimulus or due to the intrinsic physiological properties of the neuron or the network in which it is embedded. It is clearly critical to understand these dynamics in order to adequately describe and interpret neuronal behavior.

The next chapter presents a model, developed to answer a particular scientific question, which explicitly incorporates the dynamics of both the stimulus and the neuronal responses. By doing so, this model is able to provide detailed answers about the fine-scale temporal responses of individual neurons. This added explanatory power comes at a cost, however: this model will be more complex, requiring both more parameters and a more sophisticated sampling strategy in order to approximate the posterior distribution. It is important to note, though, that this new model will make use of most of the same concepts as the humble tuning curve model described here.

Chapter 4

Analysis of Orientation Tuning Dynamics

In a recent experiment, the responses of cat primary visual neurons were recorded in response to a stimuli which consisted of “oriented noise”: movies in which each frame contained a grating whose orientation was randomly chosen. A primary scientific question was whether the tuning properties of the neurons—the preferred orientation and the degree of orientation selectivity—remained constant over the timecourse of the neuronal response, or whether the data provided evidence that these properties changed between early in the response and later. To answer this questions, we devised a hierarchical model which describes how stimuli elicit spiking responses from neurons with particular tuning properties, and we apply Bayesian sampling methods including Metropolis and Gibbs sampling to to infer the posterior distribution of the model parameters, given the recorded data. The model also includes a prior assumption of smoothness in the neuronal responses. We find that a subset of recorded neurons did indeed show evidence of shifts in tuning properties. We also compare the Bayesian analysis techniques to a standard analysis framework combining nonlinear least-squares optimization and bootstrap sampling, and we find that the Bayesian methods offers increased statistical power and accuracy. In short, Bayesian methods provide more accurate parameter estimates with a given amount of data than the approach typically employed by physiologists in the analysis of such data.

4.1 Introduction

The previous chapter introduced the use of Bayesian modeling and sampling methods for the analysis of neuroscientific data. While that setting, the analysis of tuning curve data, is applicable to a wide range of experiments and scientific questions, there are nevertheless

significant limitations to its applicability. In this chapter, we introduce a model and an accompanying sampling-based inference strategy which overcome one of the primary limitations by explicitly modeling the temporal dynamics of neuronal responses.

This model was designed and implemented to answer a specific, tightly-focused set of scientific questions. While the orientation selectivity of neurons in mammalian primary visual cortex has been well studied since the beginning of the modern era of physiological experimentation (Hubel and Wiesel, 1959; Hubel and Wiesel., 1962), a number of questions have nevertheless persisted about the details of these responses, especially with regard to how they depend on cell type, laminar location, and location in the orientation map by which orientation selectivity is organized in primary visual cortex of many mammalian species. In particular, it has not been understood whether different cells, by virtue of their location and role in cortical circuits, display different behaviors over the time course of their orientation-selective responses. Are these responses stable over the tens of milliseconds over which they persist, or do they evolve significantly between the early portion of the response and its later phases? Does the preferred orientation stay the same, or does it shift over time? Likewise, does the degree of selectivity remain the same, or does it instead broaden or sharpen? Answering such questions has the potential to shed significant light on the functional properties of the networks in which such cells are embedded, and to provide crucial constraints for theories of the development of cortical selectivity and perception.

Answering these questions requires a model which describes the dynamics of neuronal responses, and an inference procedure which can determine, in a principled way, which apparent changes are actually significant in the presence of the variability in the brain, and the noise which inevitably accompanies neurophysiological recordings. In this chapter, we present a hierarchical model which fits this requirement, along with a sampling-based statistical inference method for determining whether changes in tuning properties from one time period to the next are significant.

A key feature of this model is that includes significant prior assumptions about the dynamics of the tuning curve parameters over the time course of the response. In particular, the model assumes that these parameters evolve “smoothly”: the parameter values at subsequent latencies are not independent of one another, but are instead correlated. This prior draws on our knowledge of the nature of neuronal systems, in which responses to stimuli over very short periods of time (one the order of 10-20ms) are known to be correlated. The model and inference procedure make use of this information to overcome much of the noise inherent in this data set, and this robustness will prove crucial in allowing the model to successfully answer the scientific questions highlighted above.

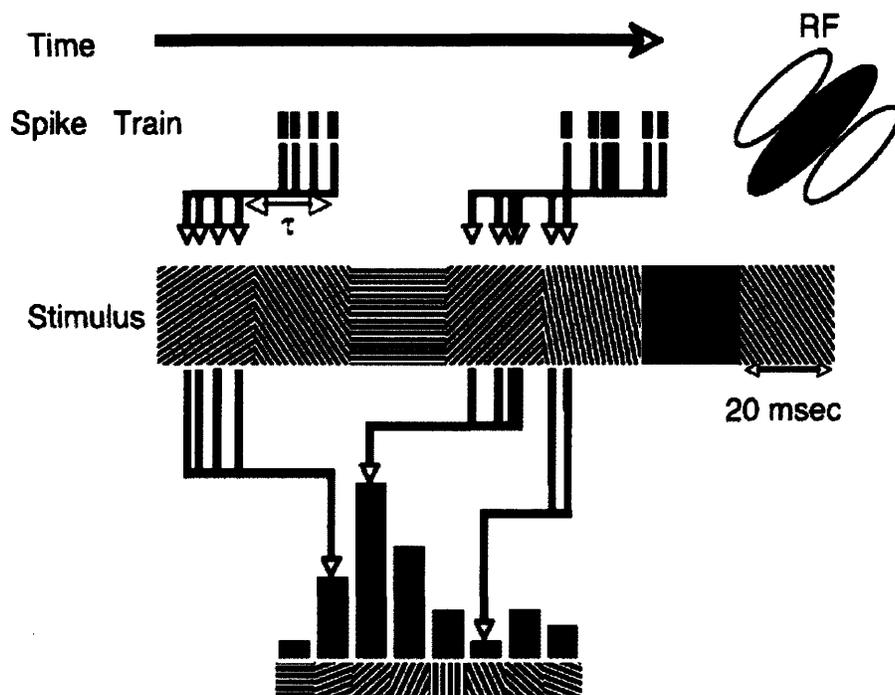


Figure 4.1: Schematic of the reverse correlation experimental setting.

A schematic description of the experimental paradigm used in this section, as well as the reverse correlation analysis method typically employed to analyze such white-noise stimuli. These stimulus is composed of movies, each frame of which contains a grating with a randomly chosen orientation. Neurons in cat primary visual cortex are recorded using standard electrophysiological techniques. In order to perform the reverse correlation analysis, a counter is incremented each time a spike is preceded at a certain latency (shown as τ) by a given stimulus orientation; this procedure results in a “receptive field” whose dimension are stimulus latency and orientation. From Schummers et al., 2007.

4.1.1 The experimental paradigm and standard analysis

The experiment, described in detail in Schummers et al. (2007) and depicted schematically in Figure 4.1, was conducted as follows. Individual neurons in cat primary visual cortex were isolated and recorded using standard electrophysiological techniques. Stimuli consisted of movies, each frame of which contained a grating with a randomly chosen orientation; in addition, some movie frames consisted of a blank (gray) field. Spikes were recorded, with special care taken to ensure the temporal alignment of the stimulus presentations and the resulting spike trains. The electrode depth was used to estimate whether each neuron lay in the superficial or deep layers of cortex. In addition, optical imaging techniques were used to ascertain the approximate location in the cortical orientation map, either a “pinwheel” at which orientation selectivity varied rapidly, or a “domain” in which orientation selectivity remained relatively constant (Bonhoeffer and Grinvald, 1993). Finally, standard tests were

Algorithm 1 The stimulus triggered average

```

Let T = the number of latency bins
Let S = the number of stimulus frames presented
Let N = the number of different stimulus orientations presented
Initialize R as a zero-valued matrix of dimensions N by T
For i = 1 to S
  For j = 1 to T
    Increment R(i,j) by the number of spikes observed

```

performed to determine whether each neuron was a “simple” cell or a “complex” cell (Kato et al., 1978).

Spike data collected under such reverse-correlation paradigms are typically analyzed using techniques such as the stimulus triggered average (STA, described in Algorithm 1), or its close relative, the spike triggered average. This technique is very useful in visualizing each neuron’s response to noise stimuli, as exemplified in Figure 4.2. Inspection of the STA can help to ascertain whether the neuron in question is selective for orientation, and provide basic information about the duration of the response and the identity of the preferred stimulus orientation.

4.1.2 Previous work

The present study was not the first to attempt an answer to these questions. In pioneering work over a decade ago, Ringach et al. (1997; 2002) employed a very similar paradigm in macaque monkeys. They observed a wide range of response, including neurons whose responses seemed to develop “Mexican hat” tuning, whose responses broadened or sharpened over the time course of the response, and whose preferred stimulus drifted from one orientation to another. While suggestive, however, these researchers were unable to ascertain whether these purported changes were statistically significant, or whether they instead reflected the variability inherent in neuronal responses and the noise associated with all neuronal recording techniques. A main obstacle in their analysis was the inability of techniques such as the STA to quantify the uncertainty which remains about their conclusions.

This previous study employing white noise stimuli in the orientation domain, as well as the present study, are both part of a large body of research devoted to the exploration of neuronal systems through noise stimuli (see, for example, Citron and Emerson, 1983; Korenberg and I.W.Hunter, 1990; Marmarelis and Marmarelis, 1978; Victor and Knight, 1979). Such noise stimuli are often employed because of their ability to efficiently explore the very large space of possible stimuli which might elicit responses from a given neuron. It is well known, however,

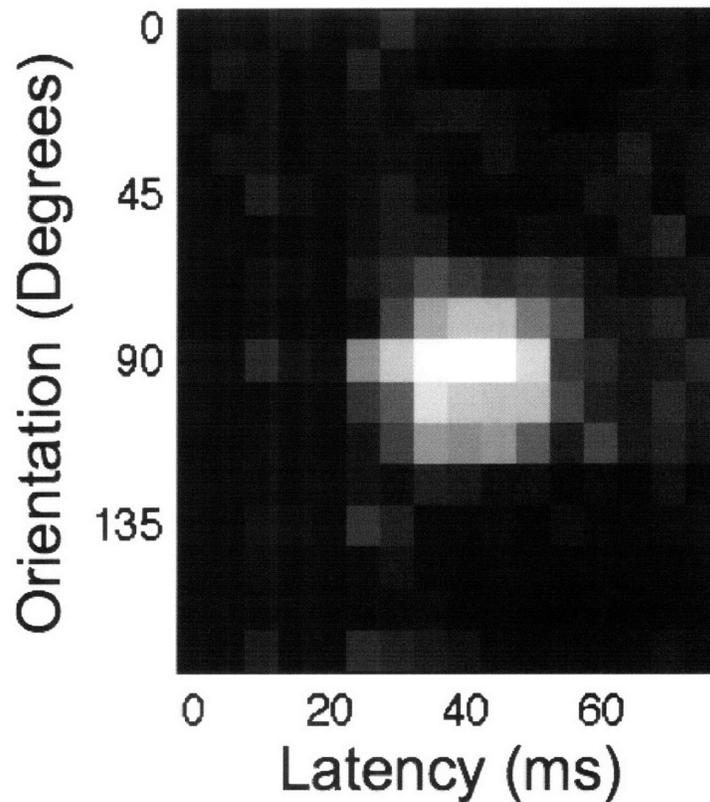


Figure 4.2: Stimulus triggered average example

White noise stimuli, such as the oriented noise stimulus used in this experiment, are typically analyzed using the stimulus triggered average (STA, or its close relative, the spike triggered average). In applying this analysis to the oriented noise protocol, a two-dimensional table indexed by stimulus latency and stimulus orientation is created, with each element in the table initialized to zero; this table is visualized here. For each stimulus frame, the relevant table entries are incremented by the appropriate amount every time spikes are recorded at a given latency after frame onset. The STA shown here depicts the responses of a highly selective cell which strongly prefers orientations around 90 degrees, and whose responses are largely confined to the period between 30 and 50 milliseconds after the frame is presented.

that one weakness of these methods is their inability to accurately describe the behavior of neurons as the responses become increasingly nonlinear. As but one example, reverse correlation techniques on their own cannot be used to successfully probe complex cells in primary visual cortex; these cells must instead be modeled by so-called “cascade” models, which include at least two stages of linear and nonlinear processing. Another shortcoming of reverse-correlation techniques, directly relevant to the scientific questions posed here, is that they do not provide a mechanism by which to estimate the uncertainty which remains about their response estimates.

4.2 Methods

To describe the dynamics of the orientation response, we build on the tuning curve model from the previous chapter. Rather than using a single tuning curve to describe the response, however, we employ a series of tuning curves, each dedicated to describing the response properties of the neuron at a particular delay after the stimulus onset. While the complete mathematical description of the model becomes quite involved, it is conceptually straightforward. As depicted schematically in Figure 4.3, the model is best visualized as a generative process by which tuning curve parameters, changing smoothly over time, produce an orientation-by-latency receptive field. This receptive field, when stimulated by the oriented noise stimulus, can be used to determine the instantaneous firing rate of the cell in responses to any sequence of stimulus frames.

The model is shown as a DGM in Figure 4.4. Unlike the simple tuning curve model, the orientation dynamics model relies on an informative prior to overcome the noise inherent in the experimental data. As explained more rigorously below, we assume that the tuning properties of the cell evolve smoothly over the time course of the response. This assumption prevents the inference process from asserting spurious changes in tuning due to noise in the observations.

The basic premise of the model is that each time a particular stimulus frame is presented, the firing rate of the neuron is modulated in a characteristic, time-lag-dependent way over the subsequent time window. The instantaneous firing rate of the cell is a summation of the influence of the preceding stimulus frames whose windows “cover” the current time point.

4.2.1 Hierarchical model description

Formally, the model is completely characterized by a specification of the joint probability distribution of the random variables denoting the union of the model parameters and the

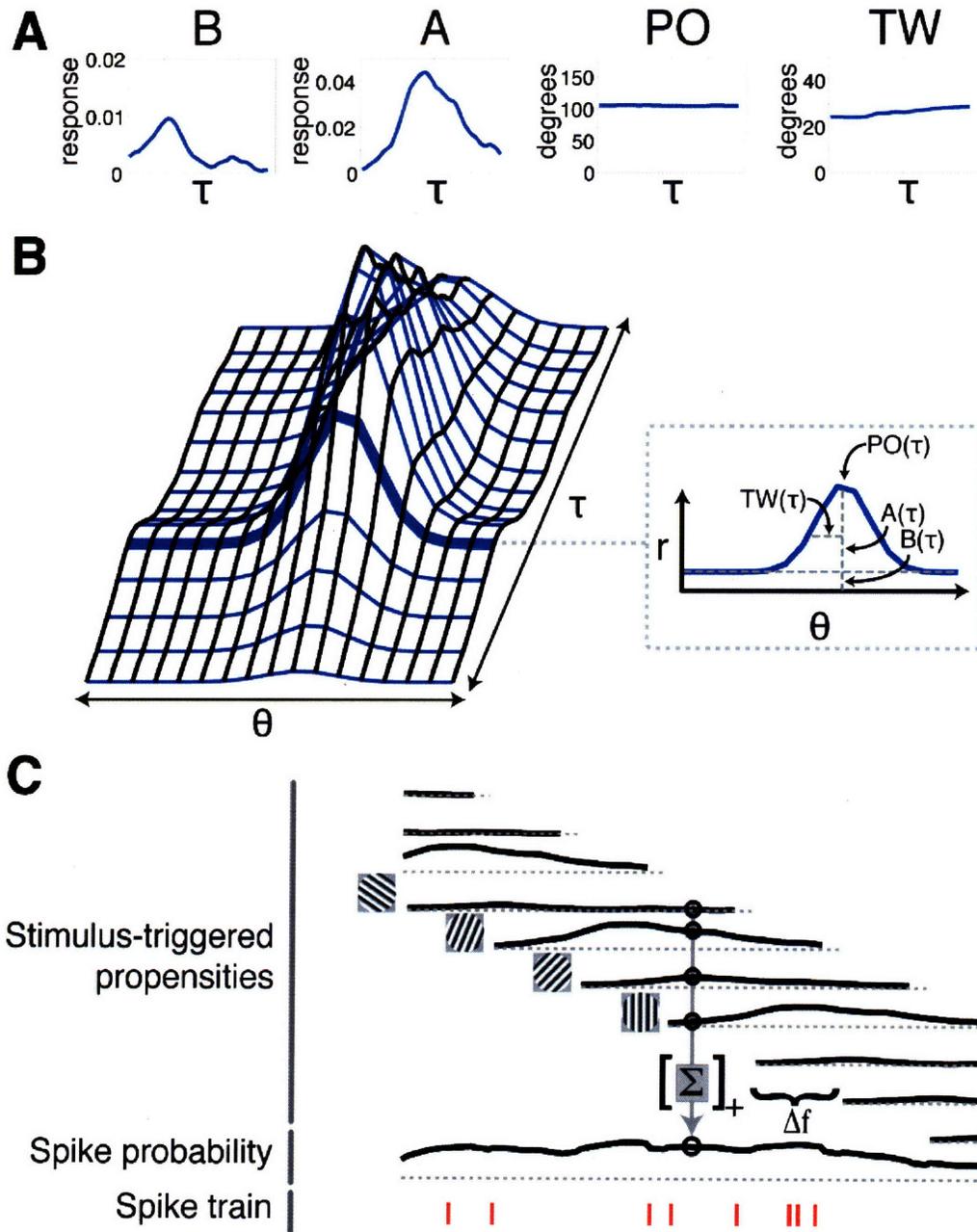


Figure 4.3: Schematic of the orientation dynamics analysis

A schematic of the hierarchical model used to analyze the oriented noise data. The spikes are binned at 5ms, with each 5ms latency period modeled by a different circular-Gaussian tuning curve (each parametrized by its baseline level, **B**, amplitude, **A**, preferred orientation, **PO**, and tuning width, **TW**). As shown in panel A, it is assumed that these tuning curve parameters are generated by a zero-mean Gaussian random walk (i.e., brownian motion), a process which *a priori* favors smooth parameter changes over widely changing ones. These tuning curves together define an orientation-by-latency receptive field (panel B). In order to compute the instantaneous spiking propensity at each moment in the stimulus presentation, the latency-specific responses from each of the preceding stimulus frames are summed (and rectified if necessary). This process results in a firing rate, from which spikes are generated according to a Poisson process (panel C).

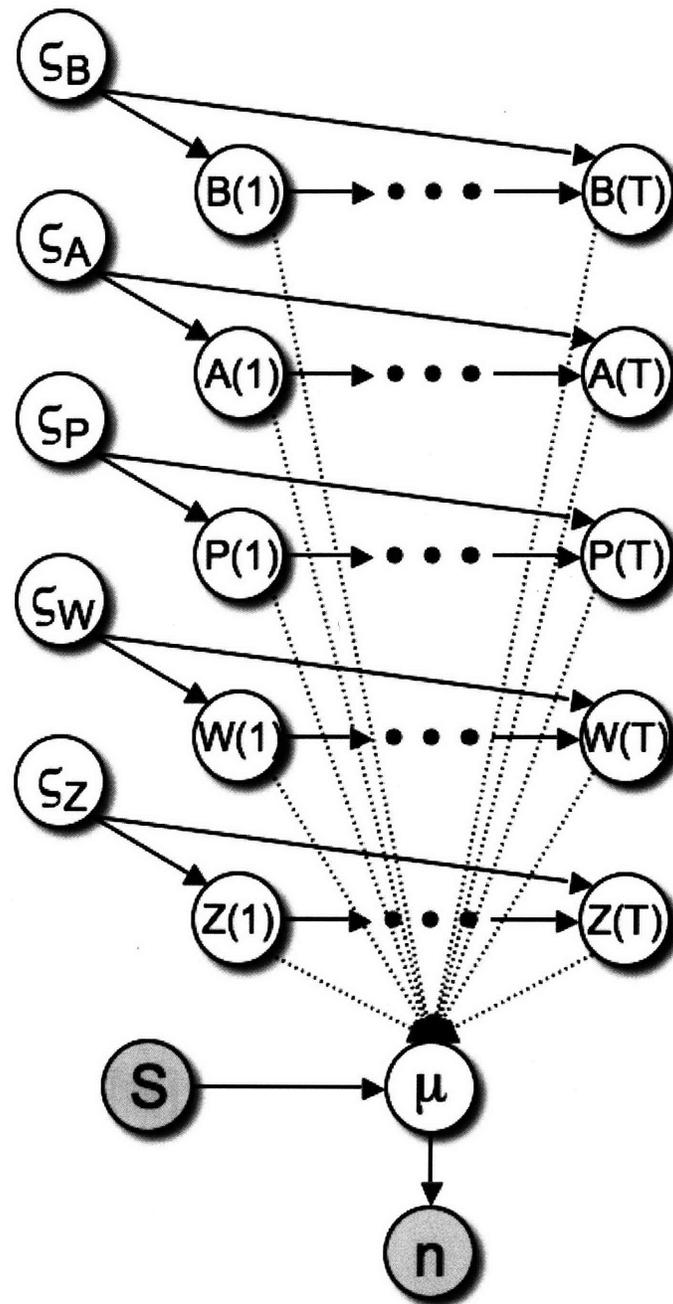


Figure 4.4: Hierarchical model of orientation dynamics analysis

A DGM corresponding to the orientation dynamics data analysis. The hyperparameters ζ determine the degree of smoothness which is exhibited by each of the four tuning curve parameter vectors $B(\tau)$, $A(\tau)$, $P(\tau)$, and $W(\tau)$, and also the blank response vector $Z(\tau)$. These tuning curve parameter vectors, together define an orientation-by-latency receptive field; the RF can be used, along with the stimulus vector $S(t)$, to compute the putative firing rate $\mu(t)$. This firing is assumed to generate the observed spike counts $n(t)$ via a Poisson process.

observed data. We use the time-indexed vectors $n(t)$ to denote the spike counts, and $S(t)$ the stimulus sequence; both vectors are of the same length M (i.e., the duration of the recording session). The spike counts and stimulus sequence are both observed. There are also four latency-indexed vectors to represent the tuning curve parameters: the baseline value of the tuning curves, $B(\tau)$; their amplitudes, $A(\tau)$; their preferred orientations, $P(\tau)$; and their tuning widths, $W(\tau)$. An additional, latency-indexed vector $Z(\tau)$ represents the neuron's response to the blank stimulus frames. To round out the model parameters, there are five scalar-valued variables that represent the degree of smoothness governing each of the tuning curve parameter types: ς_B^2 , ς_A^2 , ς_P^2 , ς_W^2 , and ς_Z^2 . In the discussion to follow, we use the symbol λ to represent the entire set of tuning curves parameter types, $\{B, A, P, W, Z\}$; furthermore, we use γ as a placeholder variable to represent a single element of this set.

As suggested by the DGM in Figure 4.4, the joint probability density of these random variables factorizes as follows:

$$\Pr(n(t), S(t), \lambda(\tau), \varsigma_\lambda^2) = \Pr(n(t)|\mu(t)) \Pr(\mu(t)|\lambda(t), S(t)) \Pr(S(t)) \Pr(\lambda(\tau)|\varsigma_\lambda^2) \Pr(\varsigma_\lambda^2)$$

The factors will be specified in turn.

Data Likelihood

The orientation selectivity of the neuron at each 5 ms time bin after stimulus presentation is assumed to follow the form of a CG curve. In addition, the response to the blank stimulus is determined directly from the $Z(\tau)$ parameter vector. That is,

$$\Delta\mu(S, \tau) = \begin{cases} B(\tau) + A(\tau) \sum_{j=-\infty}^{\infty} \exp\left\{-\frac{(S-P(\tau)-j180)^2}{2W^2(\tau)}\right\} & , S \in \omega \\ Z(\tau) & , S = \text{BLANK} \end{cases} \quad (4.1)$$

The maximum delay T is a value chosen to be appropriate to the response duration of each cell. The estimated instantaneous firing rate $\tilde{\mu}(t)$ is then the sum of all active stimulus-induced spiking propensities (as computed via equation 4.1):

$$\tilde{\mu}(t) = \left[\sum_{i=1}^{\lceil T/\Delta f \rceil} \Delta\mu(S(t - i\Delta f), T - i\Delta f + k) \right]_+$$

Here, the $[\cdot]_+$ notation indicates that the firing rate is half-rectified in order to ensure that it is non-negative; this rectification step is rarely triggered in practice and serves mainly to

ensure that the Poisson likelihoods are validly specified. The quantity $\tilde{\mu}(t)$ is a deterministic function of the tuning curve parameters, and its realization as a random variable is therefore (trivially) formulated as:

$$\Pr(\mu(t)|\lambda(t), S(t)) = \delta(\mu(t), \tilde{\mu}(t))$$

The likelihood of the observed spike train can then be evaluated directly from $\mu(t)$, as a series of conditionally independent Poisson trials:

$$\Pr(n(t)|\mu(t)) = \prod_{i=1}^M \frac{\mu(i)^{n(i)} e^{-\mu(i)}}{n(i)!}$$

Tuning curve parameters

Each of the four tuning curve parameter vectors, plus the blank response parameters (together comprising the set λ) is assumed to follow a discrete-time random walk from one latency to the next. This random walk is drawn from a zero-mean normal distribution, and is therefore equivalent to classical brownian motion; i.e.,

$$\begin{aligned} \Delta\gamma(\tau) &\equiv \gamma(\tau) - \gamma(\tau - 1) \\ &\sim N(0, \varsigma_\gamma^2) \end{aligned} \tag{4.2}$$

Because these step sizes $\Delta\gamma$ are independently drawn from this distribution, the probability of observing a particular parameter time course is proportional to the product over the relevant Gaussian likelihoods:

$$\Pr(\gamma|\varsigma_\gamma) \propto \prod_{\tau=2}^T \exp \left\{ -\frac{(\Delta\gamma(\tau))^2}{2\varsigma_\gamma^2} \right\}$$

Hyperprior parameters

As will be seen in the next section describing the sampling methods used for this model, it is often easier to perform inference when random variables are defined using a conjugate prior. For this reason, the variances of the random walks ς_γ are assumed to be distributed as a scaled inverse Chi-squared random variable with $T - 1$ degrees of freedom:

$$\varsigma_\gamma^2 \sim Inv - \chi^2(T - 1, s_{\Delta\gamma}^2) \tag{4.3}$$

where $s_{\Delta\gamma}^2$ is the empirical variance of the $\Delta\gamma$ values described in equation 4.2. This is one formulation of an “uninformative” prior over the variance of normally distributed data. A primary advantage of this formulation is that there are no arbitrary, fixed parameters governing the degree of smoothness of each tuning curve parameter type. Rather than being fixed *a priori*, these hyperparameters are learned directly from the data. In other words, the amount of smoothness inferred by the model is automatically derived by the inference process to match the actual evidence provided by the experimental observations for each neuron.

4.2.2 Inference and sampling

The orientation dynamics model requires a more sophisticated inference strategy than the simple tuning curve model. This added complexity is necessitated by several factors. First, the dynamics model has many more parameters than the tuning curve model, approximately 100 for a typical cell, rather than 4 or 5; effectively locating and exploring the high-probability regions of this high-dimensional space is a more difficult prospect. Second, the smoothness prior on the tuning curve parameters imposes strong correlations between the variables. In short, points in parameter space that respect this smoothness constraint will be more probable than points in the space which do not, and this fact should be reflected in the sampling strategy. Finally, the hyperparameters ς_γ governing the smoothness priors are defined using a conjugate distribution, which means that their distributions can be inferred using an efficient Gibbs sampling approach. For this reason, the overall sampling strategy will be a hybrid of Metropolis and Gibbs methods, rather than the pure Metropolis approach employed in the tuning curve model.

Gibbs sampling the smoothness parameters

In order to employ Gibbs sampling for a particular random variable, it is necessary to derive the conditional distribution of that random variable, given all of the other variables in the model. The graphical model shown in Figure 4.4 indicates that each smoothness parameter ς_γ is conditionally independent of all other model variables, given the tuning curve parameters $\gamma(\tau)$. That is,

$$\Pr(\varsigma_\gamma^2 | n(t), S(t), \lambda(\tau), \varsigma_\lambda^2) = \Pr(\varsigma_\gamma^2 | \gamma(\tau))$$

Furthermore, as shown in equation 4.3, the smoothness parameters are distributed according to an inverse chi-square distribution whose parameters are a function of the tuning curve parameters. Because it is possible to draw samples directly from this distribution, it is therefore straightforward to apply Gibbs sampling for inference on these parameters.

Metropolis sampling the tuning curve parameters

For the tuning curve parameters $\lambda(\tau) = \{B(\tau), A(\tau), P(\tau), W(\tau), Z(\tau)\}$, we must rely on Metropolis sampling. This immediately raises the question of an appropriate proposal distribution. A naive approach would use the same method as in the simple tuning curve model: propose samples for each tuning curve parameter independently from a normal distribution centered at the current sample value (i.e., $\gamma(i)^{(t)} \sim N(\gamma(i)^{(t-1)}, \sigma_0^2)$). Unfortunately, such a strategy leads to very inefficient sampling, largely because of the presence of the smoothness prior.

To overcome these difficulties, the key is to employ a proposal distribution which respects the constraints imposed by the smoothness prior. In particular, we wish to generate proposed samples which make large moves in the parameter space, but which nevertheless result in high acceptance rates. In short, the answer is to use a proposal distribution which is itself “smooth”. While there are doubtless many ways to achieve this goal, we make use of the discrete cosine transform, a frequency transform which shares many properties of the Fourier transform but which is more convenient when working with purely real numbers.

To propose a new sample for the parameter vector $\gamma(\tau)$, we first use the DCT to transform the vector into its frequency representation. We next perturb those elements of the transformed vector corresponding only to the low frequencies via normally-distributed values. Finally, we perform the inverse DCT to recover the (perturbed) vector in the original basis. The result of this process is a proposed sample of the tuning curve parameter vector which is smoothly deformed from the previous sample (see Figure 4.5 for an illustration).

4.3 Results

As shown in Figure 4.6, we demonstrate that a subset of the recorded neurons do indeed undergo statistically significant shifts in their tuning properties over the time course of their response to the oriented noise stimuli. These shifts occur in both the identity of the preferred orientation, as well as the degree of sharpness of orientation tuning.

Overall, we determined that a substantial number of cells show statistically significant changes in tuning curve parameters over a period of the response between the peak of the response, and τ_s later in the response. Twenty-six percent (15/58) of cells showed a significant shift in preferred orientation. The mean size of these shifts was 12 degrees, with a standard deviation of 6 degrees; the largest observed shift was 24 degrees. Seventeen percent of cells (10/58) showed a sharpening of tuning, and 7% (4/58) showed a broadening. When observed, these changes in tuning width typically continued to increase in magnitude throughout the later portion of the response. The mean magnitude of the tuning width changes (including

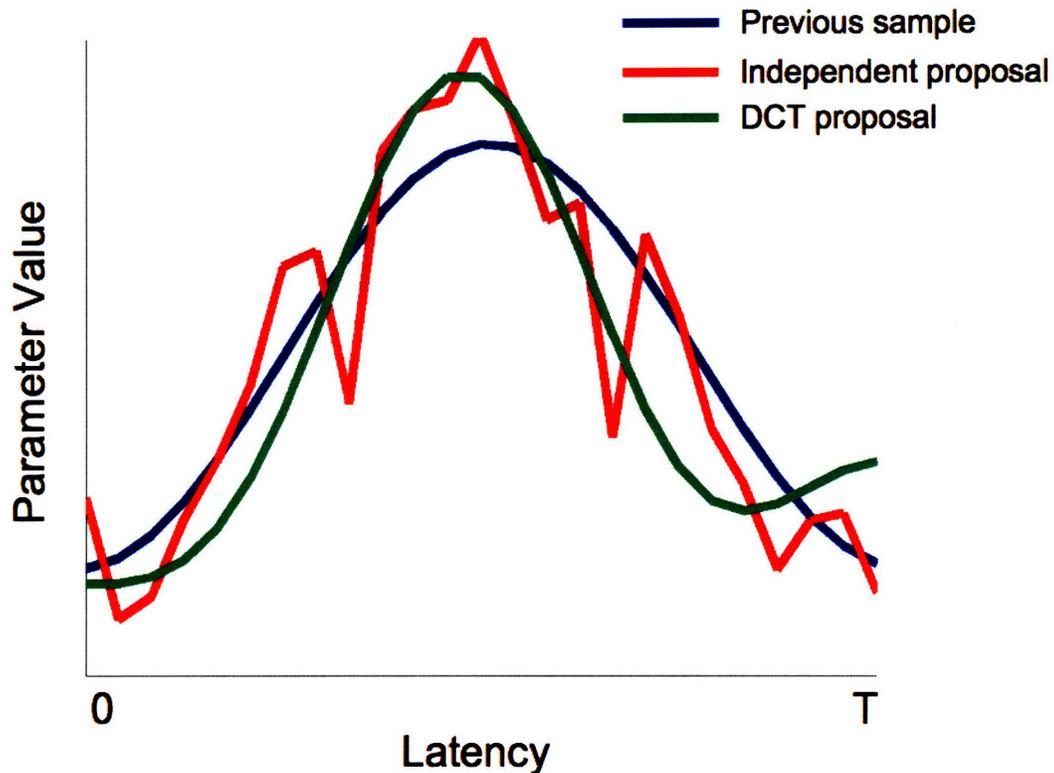


Figure 4.5: Discrete cosine transform proposals

The current tuning curve parameter vector is shown in blue, along with proposals generated by two different proposal distributions. The red line shows a proposal in which each element of the original vector is independently perturbed by a value drawn from a normal distribution. The green line shows a proposal based on the DCT method described in the text. Note that both proposals differ from the original sample by the same total amount; the DCT proposal, however, will probably have a much higher probability of being accepted, because (like the previous sample) it respects the smoothness prior.

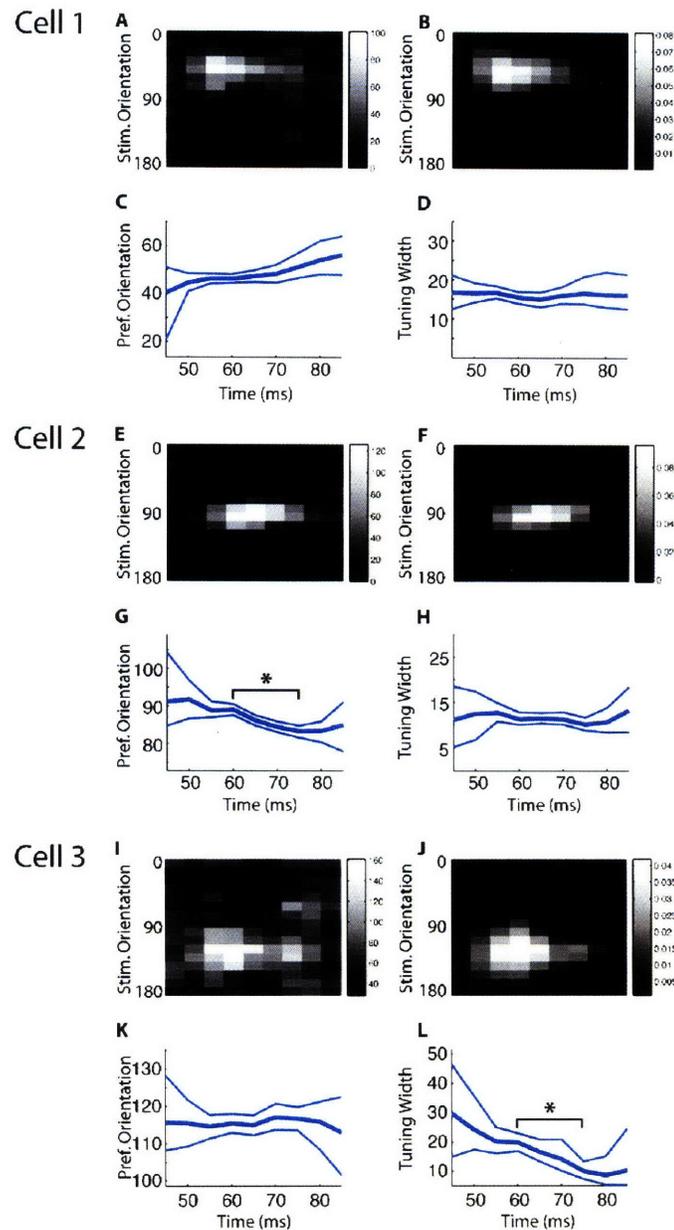


Figure 4.6: Results from the orientation dynamics analysis

A–D. Results from a cell which does not show evidence for a shift in either preferred orientation or tuning width. (A) The stimulus-triggered average response of the cell to stimuli of different orientations. The values shown indicate the total number of spikes which were observed to follow each stimulus orientation by a particular time lag (binned at 5 ms). (B) The expected response of the cell, as a function of stimulus orientation and time lag, as estimated by MCMC sampling procedure. Plotted values indicate the estimated change in firing rate that is induced by each stimulus orientation at a particular time lag. (C) The time course of the preferred orientation parameter of the Bayesian model, showing the expected value (thick line) and 95% error margins (thin lines). (D) The time course of the tuning width parameter of the Bayesian model, again showing the expected value and 95% error margins at each time lag. (E–H) A cell which undergoes a shift in preferred orientation (in G) over the time course of its response; conventions as in (A–D). (I–L) A cell which sharpens its tuning (in L) over the time course of its response; conventions as in (A–D).

both broadening and sharpening) was 12 degrees, with a standard deviation of 6 degrees. Overall, 38% (22/58) of cells showed some significant change in tuning curve parameters. Thus, a substantial minority of cells showed significant changes in either tuning width, preferred orientation, or both. The changes were typically small, but reliable, indicating that subtly different inputs during different phases of the responses can be reliably detected with our stimulation protocol and analysis tools.

In addition, a number of neurons which, at first glance, appear to undergo shifts in tuning properties were found by the inference procedure to actually provide no significant evidence for such shifts. For these cells, the apparent shifts were best explained as the result of variability and noise. While not perfect in its ability to distinguish between these cases, the model presented here is a significant improvement over other techniques which have been employed to answer such questions in the past. The following section explores this issue more fully, and provides a direct comparison with other methods.

4.4 Comparison with the Bootstrap and Optimization

As with the basic tuning curve model and sampling strategy, it is important to compare the orientation dynamics model and inference methods presented in this chapter to the standard approaches used to interpret neuronal responses to noise stimuli. In this section, we describe such an alternative approach, based again on the bootstrap, in combination with nonlinear least-squares curve fitting methods.

We use simulated data to perform the comparison. This allows us to compare the results of the various methods with ground truth, and thereby evaluate the quality of the answers obtained from each approach.

4.4.1 Bootstrap and curve fitting procedure

The bootstrap procedure is predicated upon a resampling process, in which each bootstrap sample is constructed by sampling with replacement for the original collection of (presumed independent) experimental trials. In the current experiment, the oriented noise movies were played between 5 and 10 times for each neuron. We will define each presentation of the complete stimulus ensemble as a single trial. We note that it would be inappropriate to define each movie frame as a trial, because the response to these frames are certainly not independent under any reasonable approximation.

We perform two different bootstrap procedures. In the first, the raw STAs are used; in the second, the STAs are smoothed along the latency axis using a simple boxcar filter.

The bootstrap procedure is as follows:

1. Construct the stimulus-triggered average separately for each of the N trials, using the method described above.
 - (a) In the variant of the procedure which includes smoothing, the per-trial STAs are smoothed along the latency axis by simple convolution with an acausal filter.
2. To construct each of B bootstrap samples, select N of the stimulus-triggered averages with replacement.
3. For each of these B samples, for each latency bin up to T , use standard constrained nonlinear least squares optimization methods to compute the circular Gaussian tuning curve which best describes the resampled data.

The result of this procedure is, for each tuning curve parameter type, a $B \times T$ matrix of samples from which various summaries can be computed, including median parameter values and confidence intervals. These samples are directly comparable to those obtained from MCMC sampling.

4.4.2 Comparison Results

The smoothness prior allows the Bayesian model to be sensitive to changes in tuning properties, while still maintaining superior robustness to noise. We demonstrate this by applying both methods to two simulated neurons. In both cases, we collect a few thousand simulated spikes, an amount of data which is consistent with many of the cells in the real data set; we find that the stimulus-triggered average plots which results from these simulated data are broadly similar to those from the actual data (as shown in figures 4.7 and 4.8).

The first simulated neuron, shown figure 4.7, does not undergo shifts in either preferred orientation or tuning width over the time course of its response. The Bayesian method is successful in recovering the true tuning values. The alternative method, however, has some difficulties. Using the raw (i.e., unsmoothed) STAs, the bootstrap is able to just barely confirm that there is no change in preferred orientation. The smoothed version, on the other hand, detects a spurious shift in preferred orientation. This pattern repeats itself in the tuning width estimation: the unsmoothed version correctly indicates that there is no evidence for a shift, though indicating very large uncertainty in the parameter values, while the smoothed version again detects a shift in tuning which is not actually present.

The superior results obtained by the Bayesian approach can largely be attributed to the smoothness prior. The prior allows the model to “share” information between latency bins,

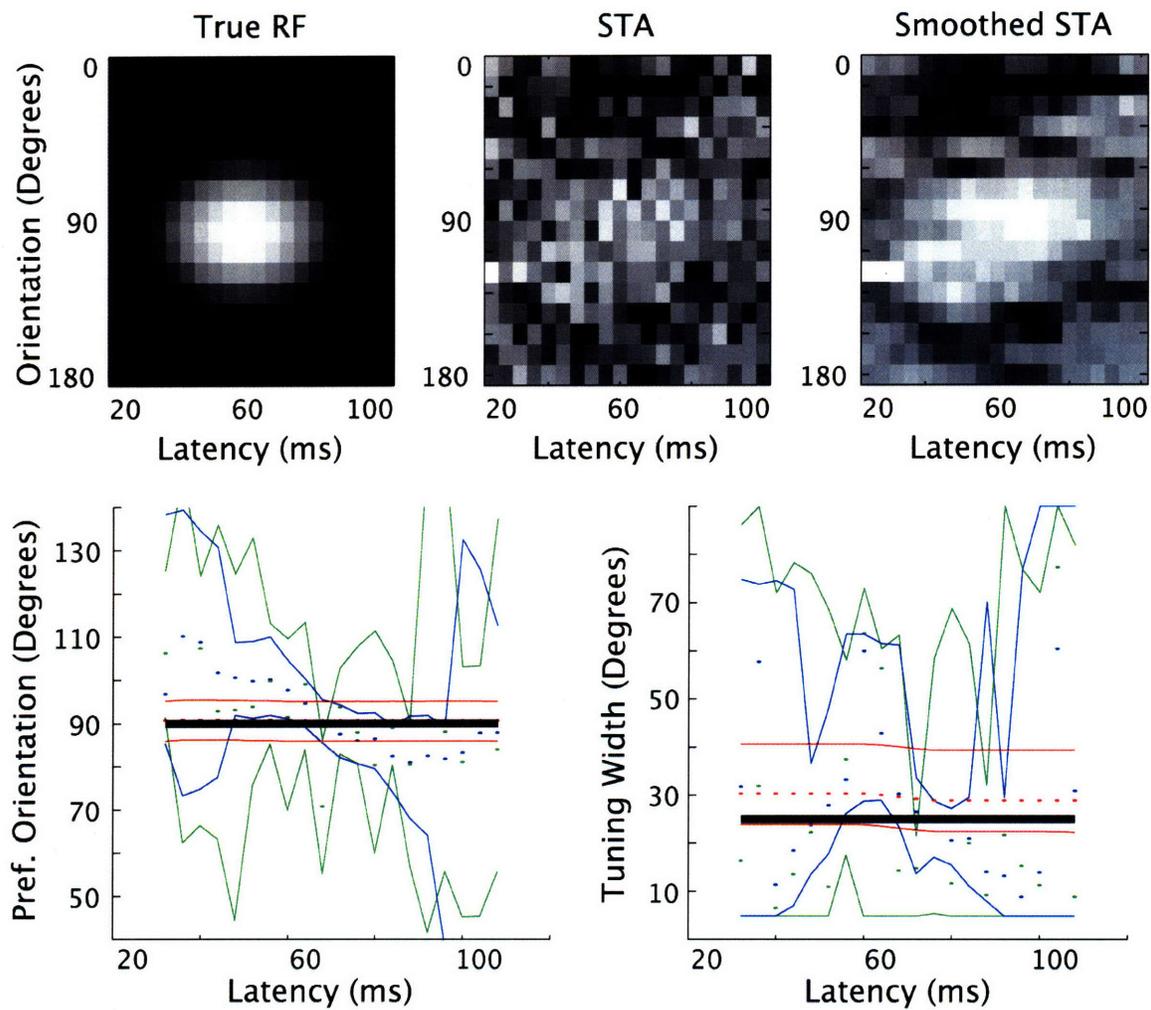


Figure 4.7: Simulated neuron with no tuning shift

Shown here are results from the various methods, as applied to a simulated cell that does not undergo a shift in tuning properties (either preferred orientation or tuning width) over the time course of its response. Shown along the top are the true orientation-by-latency receptive field of the cell, the raw stimulus-triggered average (STA) computed from several thousand simulated spikes, and a smoothed STA produced by convolving the raw STA with a linear filter. The bottom panels show the results obtained from fitting the various models to the simulated data. True parameter values are shown in black, MCMC sampling results in red, bootstrap samples from the raw STA in green, and bootstrap samples from the smoothed STA in blue. In all cases, dots indicate the median sampled value at each latency bin, and lines denote the 95% error margins from the samples. The MCMC sampling results succeed in recovering the true tuning curve parameter values, while both sets of bootstrap results incorrectly infer tuning changes in one or both cases.

so that those bins which provide strong evidence for particular tuning parameters can help to constrain those bins for which the evidence is less clear. The unsmoothed alternative does not allow any such sharing, and is therefore highly susceptible to the noise in the unsmoothed STA. The smoothed version fails for a different reason: while it does allow some measure of sharing between latency bins, the amount of sharing is dictated beforehand. In other words, this method is unable to trade off between the evidence provided by the data and the constraints from the prior. It is exactly this sort of dynamic tradeoff which the Bayesian model and inference procedure enable.

A different set of problems arise with the second simulated neuron, which does undergo a shift in preferred orientation. Again, the Bayesian approach is largely successful in detecting the shift. In this case, the alternative applied to the unsmoothed data is able to detect this shift, albeit with wider error margins than those provided by the Bayesian method. The smoothed version, however, is less successful; it fails to recover the actual time course of the shift in preferred orientation. This is because the smoothing partially obscures the shift in tuning.

It is important to note that virtually any smoothing protocol, if chosen and applied to the data before inference is performed, will help in some cases and hurt in others. This is because such an approach is not sensitive to the strength of evidence provided by the actual data from the neuron. The Bayesian approach, on the other hand, learns how much smoothing is appropriate for each neuron via hyperparameter inference, and this knowledge is combined with evidence from the data. The result is a process which is able to automatically adjust to the particular behavior of each analyzed cell, without human intervention.

4.5 Discussion

The model discussed here is the first that we are aware of to employ MCMC sampling methods to infer the tuning properties of individual neurons to noise stimuli. We believe that this general approach is likely to be useful in a number of related settings, and Chapter 6 provides more information about how the setting might be extended.

In this context, a primary advantage of sampling based inference is that the methods do not restrict models to those cases in which the posterior, or likelihood, scoring function exhibits particular properties, such as convexity. This was especially important in this model, because the scientific questions were clearly posed in terms of quantities such as “preferred orientation” and “tuning width”. The most direct way to answer such questions, therefore, was in terms of a model that was parametrized in terms of such quantities. This parametrization – in particular, the use of tuning curve functions such as the circular Gaussian – imposes

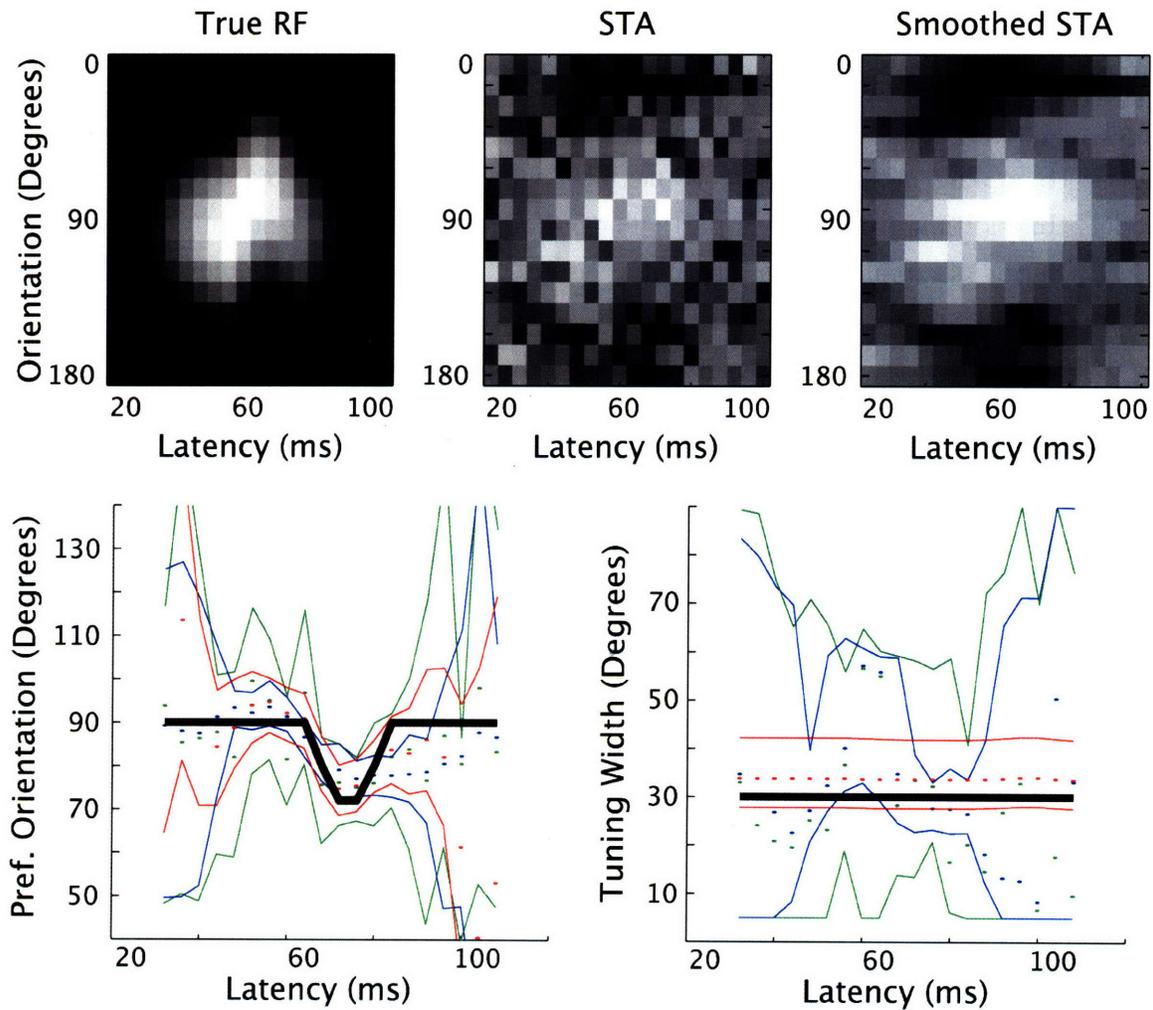


Figure 4.8: Simulated neuron with change in preferred orientation

Shown here are results from the various methods, as applied to a simulated cell that displays a transient shift in preferred orientation, shown by the black line in the lower-left panel. All conventions are as in Figure 4.7. The MCMC sampling results succeed in recovering the true tuning curve parameter values in both cases. The bootstrap results from the unsmoothed STA also recover the shift in preferred orientation, though with larger error margins. The smoothed bootstrap sampling is not able to correctly characterize the shift in preferred orientation; it has effectively “oversmoothed” the data.

complex nonlinear relationships among the random variables in the model. This, in turn, prevents the straightforward application of an alternative framework such as the generalized linear model (Truccolo et al., 2005).

We note that Sahani and Linden (2003) employed a generative model with similar structure to analyze neuronal responses in auditory cortex. While similar to this previous model in some respects, the present effort differs from their approach in at least two important ways. First, we employ sampling-based inference methods, rather than evidence-optimization techniques. Sampling methods are more flexible, and may be applicable to a wider range of models. In addition, we employ a particular parameterization of the stimulus-response function (in terms of circular Gaussian tuning curves). This use of prior knowledge about the nature of the neuronal responses brings two advantages: firstly, by reducing the degrees of freedom in the model, it allows for more statistical power to be gained from the limited amount of data available. Second, the parameterization used in the model is particularly suited to answering the scientific questions at hand - i.e., whether or not the preferred orientation and tuning width of the neuronal responses change between the early and late portions of the response. A more generally-framed model such as the one used by Sahani and Linden would not have provided such direct answers to these questions.

Chapter 5

Neuronal Adaptation as Bayesian Inference

Neurons in the sensory system exhibit changes in excitability occurring over many time scales. These changes will produce noise in postsynaptic neurons and thus perceptual errors. However, postsynaptic neurons may use adaptation to counteract the changes of presynaptic neurons. Here we present a model that formalizes how neurons could optimally adapt to minimize the influence of changing presynaptic neural properties on their outputs. This model explains a range of physiological data from experiments which have measured the overall properties and detailed timecourse of sensory adaptation in the early visual cortex. We also show how a range of experimentally measured short-term plasticity phenomena can be understood as a near optimal solution to this adaptation problem. We thus provide a link between high level computational problems, the properties of cortical neurons, and synaptic physiology.

5.1 Introduction

Neurons in the sensory system adapt at many time scales. This adaptation occurs in response to many different patterns of sensory stimulation, ranging from simple to complex. For example, many studies have noted that the strength of synaptic transmission, as measured by EPSC magnitude, decreases rapidly with strong drive (see, for example, Galaretta and Hestrin, 1998; Beck et al., 2005). At longer time scales, other studies have examined the changes in orientation and spatial frequency tuning which result from the presentation of adapting stimuli over seconds or minutes (Saul and Cynader, 1989b,a; Dragoi et al., 2000b; Kohn and Movshon, 2003, 2004). It is an interesting question whether these adaptation

phenomena, spanning time scales separated by several orders of magnitude, can be accounted for by a single explanatory principle. To answer this question, we present a simple, normative model which explains many of these seemingly disparate physiological results by asking what the sensory system should do to best reach its computational goals. We also demonstrate that this model is consistent with biologically plausible mechanisms of synaptic plasticity.

Our approach is motivated by the fact that the excitabilities of individual neurons fluctuate over time scales ranging from milliseconds to hours. Such fluctuations can be readily observed in experimental recordings in which spike trains are collected over many trials in response to the same repeated stimulus. Consider a post-synaptic neuron which is receiving input from many pre-synaptic neurons whose levels of excitability are fluctuating in this manner. To reliably signal the presence of a stimulus feature, the post-synaptic neuron must adjust the gain of its synapses to “cancel out” such input fluctuations. Failure to perform these gain corrections will lead to erroneous inferences about the stimulus identity.

Short-term adaptation phenomena, those occurring over tens or hundreds of milliseconds, have often been explained as neural or synaptic “fatigue” (Tsodyks and Markram, 1997). Under this interpretation, adaptation can be viewed as a failure by the system to achieve the proper response due to the temporary depletion of some chemical resource at the synapse. While this explanation may hold true from the point of view of the biological mechanism, it does not clarify whether adaptation serves a functional or computational purpose, nor identify what that purpose might be. We show that the synaptic changes exhibited by our model are consistent with those predicted by a more detailed, fatigue-based model of synaptic plasticity. This suggests that processes such as neurotransmitter depletion, rather than constituting a failure of the system to achieve a set goal, may instead directly serve a high-level computational principle and therefore implement a key functional role.

Adaptation to sensory patterns typically occurs over longer time scales, ranging from seconds to minutes. Here we focus on so-called “repulsive” adaptation phenomena, in which perception is shifted away from an adapting stimulus. Examples include the tilt after-effect (TAE) (Gibson, 1933; Gibson and Radner, 1937) and the motion after-effect (MAE) (Anstis et al., 1998; Mather et al., 1998). A leading normative explanation for the existence of such psychophysical phenomena is that the physiological changes underlying these perceptual shifts serve to optimize the amount of information transmitted by the sensory system (Wainwright, 1999; Fairhall et al., 2001; Stocker and Simoncelli, 2005). The modest gains which are observed to occur in information transmission, however, are unlikely to counterbalance the relatively large perceptual errors which can occur under these adaptation paradigms.

We propose an alternative normative theory of repulsive adaptation which explains these phenomena as the natural outcome of a sensory system that is attempting to perform optimal

inference. In our model, repulsive perceptual effects are understood as the result of rational errors which are committed by a sensory system which must make certain assumptions about the statistics of its inputs. Such assumptions are necessary for any system to be able to perform efficient, fast inference about the identity of sensory stimuli in response to ambiguous input signals. When these assumptions are violated, for example by laboratory adaptation stimuli that are rare in the natural environment, phenomena such as the tilt and motion after-effects are the result. While it may be difficult to distinguish between the two normative theories experimentally, they do differ significantly in their predictions. In particular, our model predicts that suppressive adaptation due to over-stimulation will occur more quickly than facilitatory adaptation resulting from under-stimulation. This asymmetry derives from the assumption of sparsity in neuronal drives under natural stimulus conditions (Vinje and Gallant, 2000).

Underlying our analysis of repulsive adaptation phenomena is the assumption that the stimuli which lead to effects such as the TAE and MAE are in fact very rare in the natural environment. These unusual stimuli fool the perceptual system by mimicking a situation in which the excitability of some neurons has been increased (and others possibly decreased). We therefore propose that the physiological and perceptual phenomena of sensory adaptation stem from the fundamental ambiguity which exists between the intensity of sensory stimuli and the level of excitability of the neurons which process these signals.

We demonstrate that a very simple network model of orientation tuning that includes synaptic gain changes based on this normative model provides a very natural explanation for the adaptation-induced shift in orientation tuning observed in studies such as Dragoi et al. (2000). We also show that this model, which is rooted in the principle of optimal inference, replicates the short-term synaptic dynamics predicted by the fatigue-based model of Tsodyks and Markram (1997). This observation suggests that biophysical mechanisms such as neurotransmitter depletion might serve the nervous system's high-level computational goals.

5.2 Methods

5.2.1 Model of Gain Dynamics

We assume that a post-synaptic neuron seeks to estimate fluctuations in its inputs that occur over time scales ranging from a few milliseconds to minutes. To accomplish this goal, the neuron must estimate the excitability of its inputs at each of these time scales. We refer to these estimates of excitability as gains. Throughout the following, we use $M = 10$ time scales

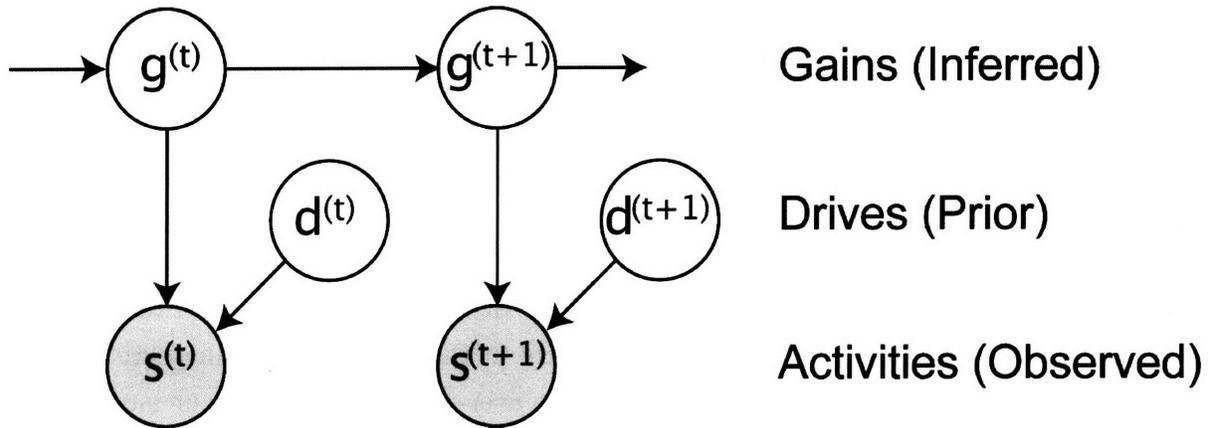


Figure 5.1: A DGM of the adaptation model

$\tau_1 \dots \tau_M$, linearly spaced on a log scale from a few milliseconds (for τ_1), up to 10's of minutes (for τ_M). The temporal dynamics of the M gains are modeled as Brownian motion with zero-mean drift, with a variance from one time step to the next that is inversely proportional to the time scale:

$$Q_j = \frac{1}{\tau_j}$$

The individual gain values decay towards 1 at rates which ensure that the total variance associated with each of the time scales is the same. We further assume that synaptic drives are sparsely distributed under natural stimulus conditions (Vinje and Gallant, 2000), meaning that high drive values are much less common than low ones. We formalize this assumption by placing an exponential prior distribution on the drives, i.e., $\Pr(d) = e^{-d}$. Lastly, we assume that the synaptic input at each time step is the product of this sparsely-distributed true drive and the total current gain estimate:

$$s^{(t)} = d^{(t)} \times \sum_{j=1}^M g_j^{(t)}$$

Given the exponential prior distribution on d , and the assumption that $d = s/g$, we can employ a variable transformation and marginalize over d to find the following likelihood distribution for s :

$$\Pr(s|g) \propto \sum_{j=1}^M \frac{1}{g_j} \exp \left\{ -\frac{s}{g_j} \right\} \quad (5.1)$$

This exponential likelihood distribution incorporates our sparseness assumption, and indicates that high input values are much less common, and therefore more informative, than low

input values. This asymmetry will influence the dynamics of the gain estimates by allowing gain increases to be detected more quickly and with less uncertainty than gain decreases.

5.2.2 Gain Estimation

We have described a statistical model whose conditional independence structure is equivalent to that of the state space model, a standard framework for describing the dynamics of normally distributed variables. The likelihood distribution for s that we derive from our sparseness assumption is non-Gaussian, however. We therefore seek a local Gaussian approximation of the likelihood, which will allow us to use the standard inference machinery of a Kalman filter to estimate the synaptic gains in response to an any series of stimulus values.

We form a local Gaussian approximation to the log of the likelihood distribution for s (Equation 5.1) by computing the second-order Taylor expansion about $g^{(t-1)}$ (see B for details of the derivation). Given this approximation, we can use a standard Kalman filter to estimate the mean and variance of the gains at each time step, given our observations of the synaptic inputs and our prior assumptions of the drives which produced them. In practice, we find this approximation to be quite stable.

5.2.3 Short-Term Synaptic Plasticity

We use this model of gain dynamics to model the short-term changes in strength of a single synapse. We compare the resulting gain changes to a model based on neurotransmitter depletion (Tsodyks and Markram, 1997). The input level is randomly varied every 500ms over a uniform range, and the M gains are estimated using the dynamics model. The overall gain estimate is determined by the sum of the individual gains:

$$G^{(t)} = \sum_{j=1}^M g_j^{(t)}$$

5.2.4 Orientation Tuning Adaptation

Network Model We also apply the model of gain dynamics to a simple network model of orientation tuning in primary visual cortex. In this model, a single postsynaptic cell receives input from N presynaptic cells. Each of the N synapses has a full complement of M gains, and the gains at each synapse i are estimated independently in response to the inputs observed at each time step, $s_i^{(t)}$. The activity level of the postsynaptic cell at time t is determined by its gain-modulated inputs:

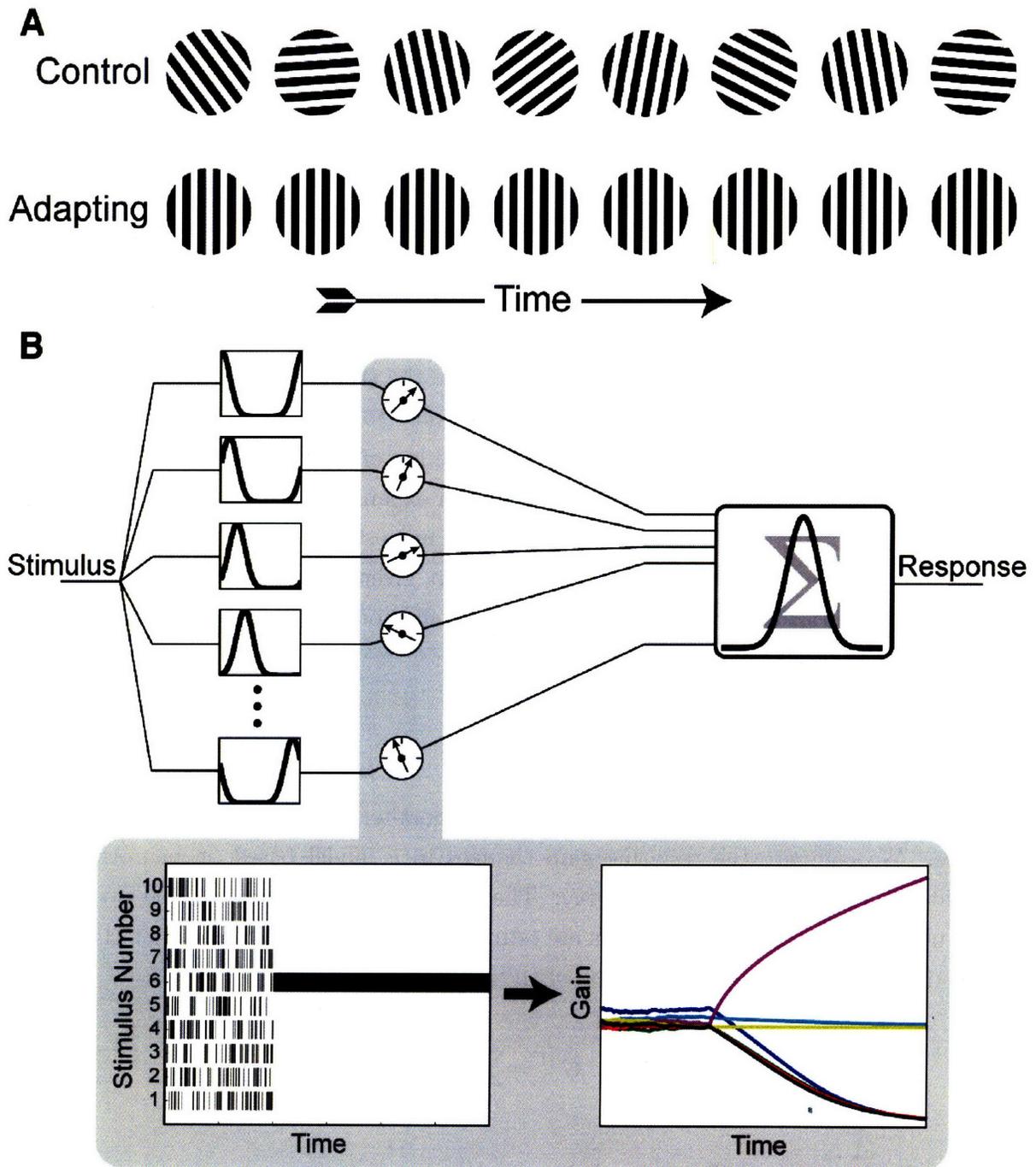


Figure 5.2: The network model of orientation adaptation

A) The stimuli presented during the control (upper) and adaptation (lower) epochs of the sensory adaptation simulations; these are meant to replicate the stimuli used in Dragoi et al. (2000). B) A schematic of the network model used to simulate orientation adaptation. A population of presynaptic neurons, each with its own orientation tuning curve, reacts to the stimulus. These presynaptic activities are then modulated by the gains, weighted, and summed to produce the postsynaptic response. While the synaptic weights remain constant, the gains are updated over time according to the dynamical model.

$$R_{post}^{(t)} = \sum_{i=1}^N \frac{w_i}{G_i^{(t)}} s_i^{(t)}$$

Here, $G_i^{(t)}$ is the total gain estimated at time t for input i . Note that the postsynaptic cell is completely linear, as its response computed as a weighted sum of its inputs at each time step. The w_i 's are the fixed components of the synaptic strengths, and follow a circular Gaussian tuning curve profile:

$$w_i = b + \sum_{k=-\infty}^{\infty} \exp \left\{ \frac{i/N - PO + 180k}{2TW^2} \right\}$$

The presynaptic cells are also assumed to have circular Gaussian tuning curves, with preferred orientations evenly distributed every $180/N$ degrees. In this simple model, the presynaptic cells do not have their own input gains and therefore do not undergo adaptation.

At each time step, the stimulus causes each of the presynaptic cells to become active according to its tuning curve. This stimulus drive is in addition to a lower, spontaneous level of activity b which is always present on all of the pre-synaptic neurons. This input profile is a simple approximation of the orientation-tuned inputs which are known to feed neurons in primary visual cortex. This simple model omits some prominent features of real orientation-selectivity networks, including nonlinear responses and recurrent connections.

Stimuli Each orientation adaptation simulation is divided into two epochs (Figure 5.2A). In the first “control” epoch, the stimulus orientations are drawn uniformly and independently and presented for one second each (i.e., white noise). In the second “adaptation” epoch, a single stimulus orientation is chosen and is presented continuously for the duration of the simulation. During this adaptation period, the presynaptic cell whose preferred orientation is closest to the adaptation stimulus is maximally active, and all of the other inputs are active to a lesser degree. The exact degree of activation is determined by the circular Gaussian tuning curve associated with each presynaptic cell. A schematic of the complete model is shown in Figure 5.2B.

5.3 Results

We use the dynamical model to explain adaptation data containing changes over time scales ranging from tens of milliseconds to several minutes. In all cases, we use the same set of time scale parameters.

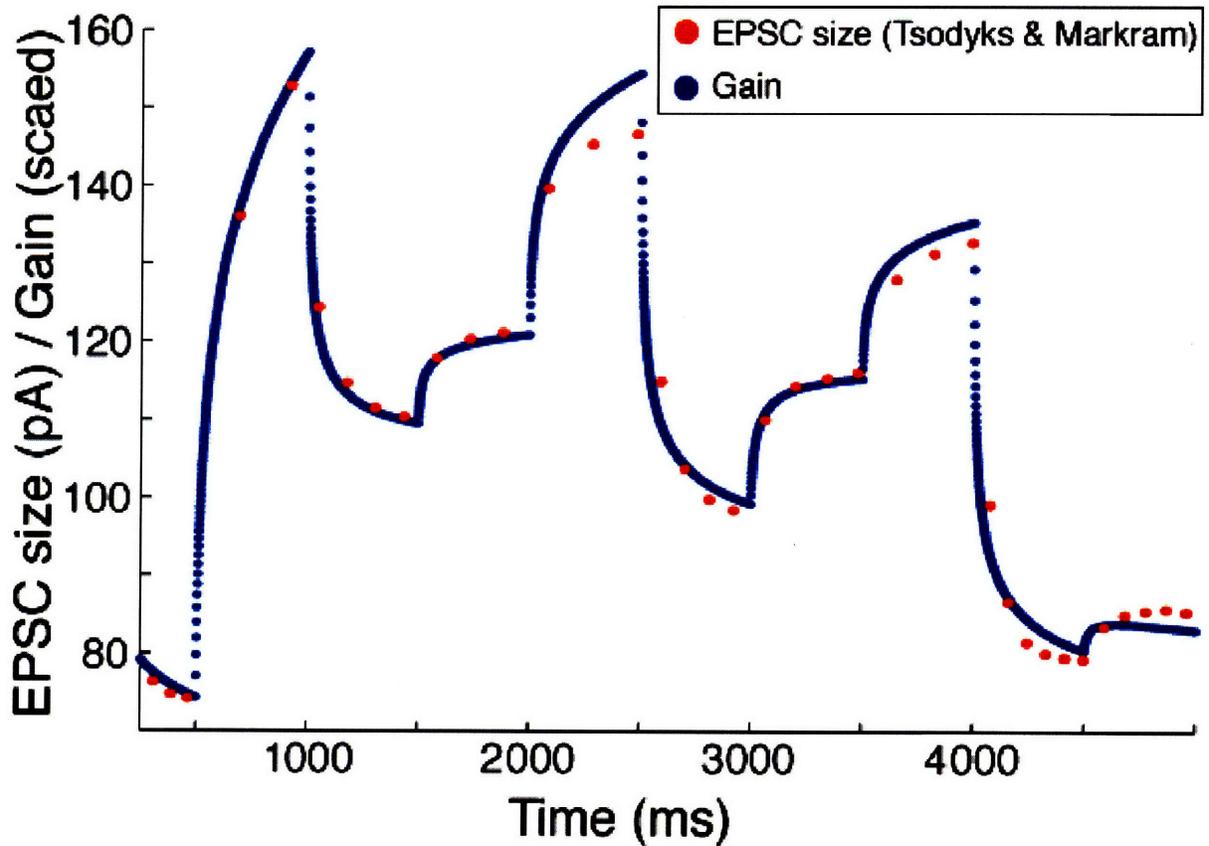


Figure 5.3: Comparison with Tsodyks and Markram

The model proposed by Tsodyks & Markram (1997) allows for a closed-form calculation of successive EPSC magnitudes, given an arbitrary time-varying input. Here we compare the EPSC magnitudes predicted by this model (red dots) with the gains predicted by our dynamical model (blue). The close match indicates that the changes in EPSC magnitude are consistent with those that would be produced by a mechanism in service of an optimal adaptation principle.

5.3.1 Short-Term Synaptic Plasticity

Tsodyks and Markram (1997) employ a dynamic model of synaptic plasticity to explain a host of physiological data about the properties and dynamics of signal transmission at intracortical synapses. They propose a simple model of neurotransmitter cycling, and they express the depletion and subsequent recovery processes using a system of differential equations. The parameters of this model are directly interpretable in terms of biophysical properties, which the authors were able to fit to biologically plausible values. Given these parameter values, their model produced behavior that closely matched EPSC magnitudes obtained using intracellular recording techniques. In light of this close match, we will compare the behavior of our model directly with theirs, using the behavior of their model as a proxy for the experimental

data.

Figure 5.3 shows that our model, based on the principle of optimal adaptation in the face of unpredictable changes in neural excitability, produces changes in synaptic strength which closely match those predicted by Tsodyks and Markram when driven by a randomly varying input. As long as a sufficient number of time scales are included in the model (i.e., spanning several order of magnitude), these results are quite robust. Obtaining the close match shown between the two models is largely a matter of linearly scaling the output of the normative model so that its arbitrary units are aligned with the biophysically realistic units employed by the mechanistic model. That is, their more detailed mechanistic model exhibits behavior that is entirely consistent with our model, even though our model was designed to implement a high-level computational goal and does not implicate any particular biological explanation. This result suggests that low-level biological processes such as transmitter depletion may embody, and may even have evolved to implement, the inferential goals of the nervous system.

5.3.2 Orientation Tuning Adaptation

Recent physiological experiments indicate that orientation-selective neurons in primary visual cortex of the cat undergo repulsive shifts in their tuning curves when adapted to oriented gratings (Dragoi et al., 2000). Similar results were earlier found after adaptation to particular spatial frequencies (Saul and Cynader, 1989a). Interestingly, these physiological adaptation effects were observed to follow a rather complex temporal profile. If the adapting stimulus lay on one flank of the tuning curve, then the response at this “near flank” would be quickly reduced, followed somewhat later by an increased response on the far flank. Furthermore, this far-flank facilitation often resulted in an increase in the magnitude of the response at the (shifted) peak of the tuning curve. In other words, the response profile was not merely translated away from the adapting stimulus, but instead underwent a variety of effects which occurred on at least two separate time scales (see Dragoi et al., 2000, Figures 1 & 2). While the overall effect was therefore one of repulsion, the dynamics of the tuning shift displayed additional subtleties.

Our model explains both the repulsive shift observed in the orientation tuning curve after adaptation, as well as the separate time scales which govern the effects on the two flanks (Figure 5.4). This separation of time scales results from the assumption of sparse activities. Intuitively, the fact that the adapting stimulus is being overdriven compared to the control condition is apparent to the system almost immediately. This means that the gain on the adapted input will begin increasing almost immediately after the adapting phase begins. In contrast, the fact that the non-adapted inputs are being under-driven takes longer to become

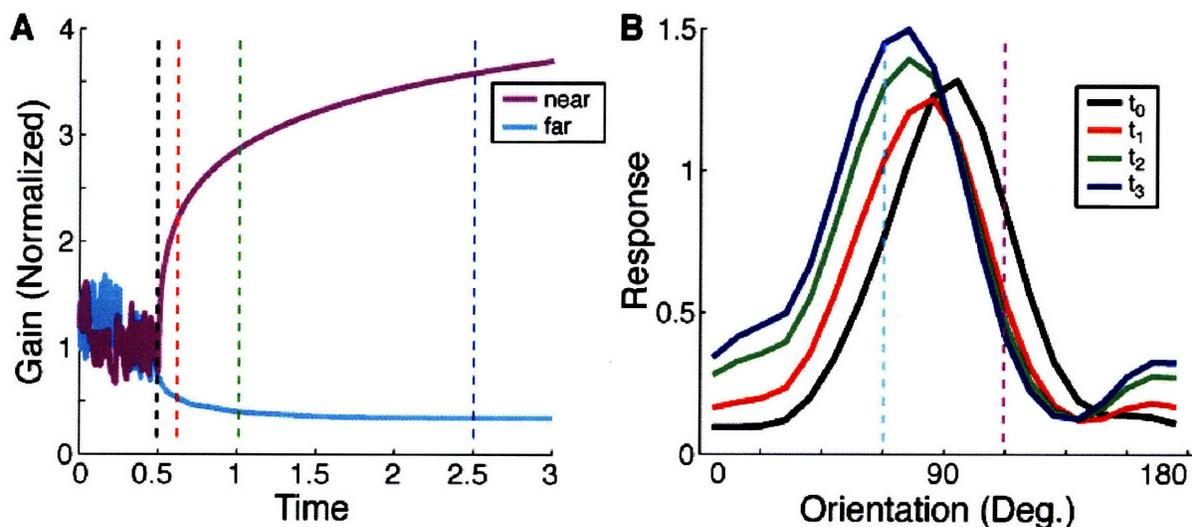


Figure 5.4: Repulsive orientation adaptation

A) The time course of two sets of gains during a period of adaptation to a single stimulus. The gains correspond to presynaptic inputs whose preferred orientations are on the near (magenta) and far (teal) flanks of the control tuning curve. The dashed lines indicate the time points at which the control tuning curve was measured (black), as well as three successive adapted tuning curves (red, green, and blue). B) The control tuning curve (black), as well as three successive adapted tuning curves corresponding to the time points indicated in A.

apparent to the system, and the decrease of this gain therefore happens more slowly. This is because the inputs are expected a priori to be relatively silent, given the assumption of sparse firing.

This separation of the two time scales can equivalently be explained by the properties of the likelihood distribution of $s(t)$, as shown in the second-order properties captured by the shape of the local Gaussian approximation. In particular, the second derivative of the log likelihood is linear in $s(t)$, which means that the variance of the Gaussian approximation will scale with the inverse of $s(t)$. Large values of $s(t)$, corresponding to high firing rates, will therefore result in a lower variance on the estimate of the gain. Low firing rates will conversely result in higher variance, and therefore greater uncertainty about the gain. Again, this asymmetry is a direct result of the system's assumptions about the sparse statistics of its inputs.

Dragoi et al. (2000) observed a striking heterogeneity in the physiological response of individual neurons in cat primary visual cortex. A natural question to ask, and one which their paradigm was not able to address, is what accounts for this variation in response from one cell to the next. We performed two simulation runs in which the parameters governing the synaptic gain changes were kept constant. The only parameters that were varied between

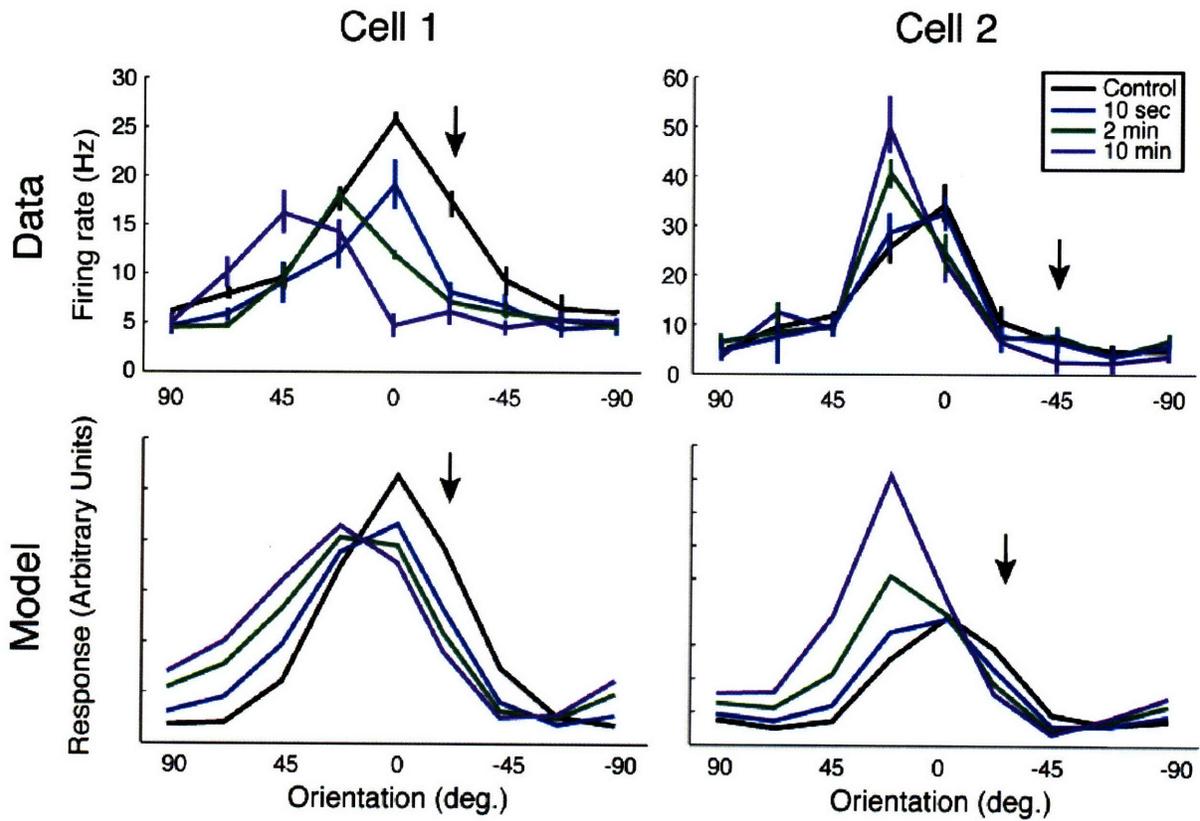


Figure 5.5: Comparison with Dragoi et al.

A, B) Control and adapted tuning curves from two real example neurons, adapted from Dragoi et al. (2000). C, D) Control and adapted tuning curves from two model neurons. In these two cases, the network tuning parameters are slightly different, but the dynamics parameters are the same.

the two runs were those which determined the tuning properties of the adapting neuron (which were adjusted to produce control tuning curves which approximately matched the real example neurons); in addition, the tuning widths of the input neurons were randomly perturbed.

As shown in Figure 5.5, we determined that the qualitative differences between the observed tuning curve shifts could be explained solely by differences in the tuning properties of the model neurons. In particular, it is not necessary to posit different gain change dynamics to account for this heterogeneity. Given that real neurons are known to receive inputs with varying tuning properties (Schummers et al., 2002; Marino et al., 2005), this is a plausible explanation for the variation in responses observed *in vivo*.

5.4 Discussion

5.4.1 Other Normative Models

Our model explains adaptation phenomena as the result of the nervous system's effort to achieve a computational goal, and it is therefore a normative model. Several previous efforts have offered explanations for adaptation which emphasize its role in optimizing the transmission of sensory information (Wainwright, 1999; Brenner et al., 2000). While our approach uses a different explanatory principle, that of computational stability, these norms are not mutually exclusive. That is, maximal information transmission and computational stability might simultaneously influence the processing of sensory information, and it is possible that their relative dominance changes as a function of the specific environmental context.

In a related effort, Stocker and Simoncelli (2005) explain sensory adaptation as the result of dynamic reallocation of a scarce computational resource. They demonstrate that the existence of repulsive adaptation is consistent with the apportionment of greater signal-to-noise ratio to the adapting stimulus, and a complementary reduction for non-adapting stimuli. This approach assumes that the persistent presence of the adapting stimulus is interpreted by the nervous system as a clue that the statistics of the environment have changed, and that the internal representation should therefore be adjusted to better accommodate this new environmental context. This description can be contrasted with our model in several ways. Firstly, under this alternative view, shifts in neuronal tuning properties due to adaptation constitute a net gain of function, rather than a rational error as predicted by our model. In addition to this difference in interpretation of the function of adaptation, it should also be noted that our model requires fewer arbitrary assumptions about the computational architecture of adaptation than that of Stocker and Simoncelli. In particular, their model

assumes that a particular computational resource - signal to noise ratio - is scarce, and is allocated according to some process. Our model, on the other hand, relies on the much more general principle of estimating input levels; the resulting changes in synaptic strength can be computed locally, and may be implemented by a number of biological mechanisms (such as receptor trafficking). Finally, our model is able to make detailed (and falsifiable) predictions about the dynamics of these tuning changes, in contrast with the static model suggested by Stocker and Simoncelli.

5.4.2 Relationship to Mechanistic Models

Other recent models have proposed more detailed mechanistic explanations of the physiological and perceptual aspects of sensory adaptation. Teich and Qian (Teich and Qian, 2003, 2006) identify the parameter changes which are necessary for several well-studied mechanistic models of orientation selectivity to replicate experimentally observed tuning shifts. This more detailed treatment is consistent with our approach, and may describe a possible network-level implementation of our abstract model. However, it does not address the origin or computational utility of the physiological shifts.

Similarly, Jin et al. (2005) perform a detailed analysis of the relationship between adaptation-induced changes in the population activities of a recurrent network model and the strength of perceptual illusions such as the TAE. Again, they do not describe the underlying computational cause of these physiological effects.

Together with these previous studies, our model suggests a unified explanation of these sensory adaptation effects. The nervous system, in an effort to achieve stability in its perceptual inferences in the face of a changing environment, mistakes atypical adapting stimuli for changes in its internal properties. This leads to changes in physiological response profiles and, finally, to perceptual illusions such as the tilt after-effect and the waterfall illusion.

5.4.3 Relationship to Kording et al., 2007

The model we propose here solves the gain inference problem using a Bayes-optimal estimation procedure, following a general approach recently applied to adaptation in the motor system (Kording et al., 2007). While the structure of the underlying model of gain estimation is similar to the approach used in that earlier work, there are at least two important differences. First, the meaning imparted to the gains in the two models is completely different. Because Kording et al. restricted their model to the motor system, they were able to employ a straightforward interpretation of the gains in terms of motor performance. In the context of individual neurons in the sensory system, however, the interpretation of the gains

is more subtle; in our model, the gains refer to the adapting neuron's estimate of the overall excitability of each of its inputs. A second important distinction is our inclusion of a sparseness prior on the statistics of these neuronal inputs; this additional assumption necessitates the use of the local Gaussian approximation to the likelihood. Finally, we assume that there are separate and independent gain estimators for each input, and that the behavior of the adapting neuron reflects the integration, linear or otherwise, of these different processes.

5.4.4 Limitations and Future Work

The model described here does not take into account many of the known properties of neural activity, and cannot explain all of the physiological recordings obtained under adaptation paradigms. For example, the inputs in the model are currently independent of one another, and the synaptic gains are adapted independently. Such a setting ignores the correlations which are known to exist in the firing patterns of neural populations, and such correlations may be crucial in explaining other aspects of physiological adaptation phenomena. In addition, the simple network model that we use to simulate orientation tuning omits many of the properties which such thalamocortical networks are known to exhibit, and which more detailed models do include (Ben-Yishai et al., 1995; Somers et al., 1995). For example, our network model is completely linear and contains no recurrent connections; both of these features are thought to contribute strongly to the nature of orientation tuning. We think it likely that a more complex network model which incorporated these and other properties would be more successful in replicating the detailed, quantitative behavior observed in tuning shifts under adaptation. Such a model, however, would necessarily be significantly more complex, and its dynamical behavior would be much more difficult to interpret.

The principle that underlies our normative model strongly suggests that neurons able to maintain computational stability in the face of fluctuating input levels should, as a population, allow for better performance in the decoding of sensory signals - at least when the statistics of these inputs conform to the expected statistics. In particular, a population of neurons that adapt according to the normative model described here should allow for a more accurate decoding of, say, stimulus orientation, even when the inputs are corrupted by fluctuations occurring on multiple time scales. Such a test is an important future part of the future work necessary to further validate and support the theory.

Chapter 6

Conclusion and Future Directions

In Chapter 1, I claimed that Bayesian modeling and inference methods would offer advantages in two areas of computational neuroscience: the analysis of physiological data, and the interpretation of how individual neurons process sensory information. I have presented three case studies which support this claim; each has illustrated how the proper quantification of uncertainty is central to the proper resolution of the corresponding statistical inference problem. Chapters 3 and 4 presented two applications of Bayesian modeling and inference methods for the analysis of neurophysiological data. Both of these projects succeeded in providing answers to pressing scientific questions faced by neurophysiologists in recent experiments. In Chapter 6, I demonstrated how the Bayesian model of perception, which has been developed in recent years by number of researchers, can be applied to the interpretation of the behavior of individual neurons.

Based on these results, I believe that the combination of structured prior knowledge (as captured in appropriate hierarchical models) and effective sampling-based inference provide a strong foundation for the design and execution of a range of innovative and potentially informative experiments. In this final chapter, I discuss some possible directions for these future studies.

6.1 Mainstreaming sampling approaches

A primary goal of this thesis is to demonstrate that data analysis methods based on the Bayesian statistical framework, leveraging hierarchical models and a range of Bayesian inference methods, can offer significant advantages in answering scientific questions. I believe that these methods can and should become part of the standard data analysis toolkit for neuroscience. Of course, some models and inference methods are more readily accessible than others. While relatively complex models such as the orientation dynamics application

discussed in Chapter 4 will probably continue to require significant expertise in order to be successfully implemented and applied to a given data set, simpler models such as the tuning curve application discussed in Chapter 3 can and should be integrated into standard practices. The goal of the BayesPhys toolbox described in Appendix A is to speed the adoption of this framework.

The present effort is not unique in its effort to expand the application of hierarchical models and sophisticated statistical inference techniques to the analysis of neurophysiological data. For example, Sahani and Linden (2003) employ a framework that is generally similar to that which is described in Chapter 4 to estimate the transfer functions of neurons as they respond to noise stimuli. These authors also find that smoothness and sparseness assumptions improve the ability of the model to predict future neuronal responses. Another line of research has focused on understanding the behavior of model neurons whose membrane voltage dynamics implement a stochastic version of the standard leaky integrate-and-fire model (Paninski et al., 2004; Paninski, 2004). As a final example, Truccolo et al. (2005) describe a very general framework by which neural spike trains can be predicted by spike history, the firing of other neurons in the local network, and the stimulus. By situating their analysis in the generalized linear model (GLM) framework (and thus proving the convexity of the corresponding likelihood function), they are able to make use of standard, highly efficient iterative algorithms for maximum likelihood estimation.

None of these other efforts use sampling-based inference, however. This is probably due to a common conception, often expressed informally but rarely tested empirically, that sampling methods are too slow to be applied successfully to sophisticated models. Consequently, much of the effort has been focused on either placing models into settings that allow for standard optimization methods to be applied, or to apply approximations such as variational methods. These other approaches are certainly valuable, and in any case there are certainly enough outstanding problems in the field to accommodate a number of different strategies. One goal of this thesis has been to show that sampling methods, when correctly applied, can provide results with reasonable amounts of computation time on standard desktop computers. Furthermore, sampling methods allow for greater flexibility in model specification than alternative inference approaches, because the probability landscape does not need to meet requirements such as convexity.

6.2 Generalizing the model

Many researchers, including some of those mentioned in the previous section, have noted the wide applicability and significant predictive power of functional models based on the so-called

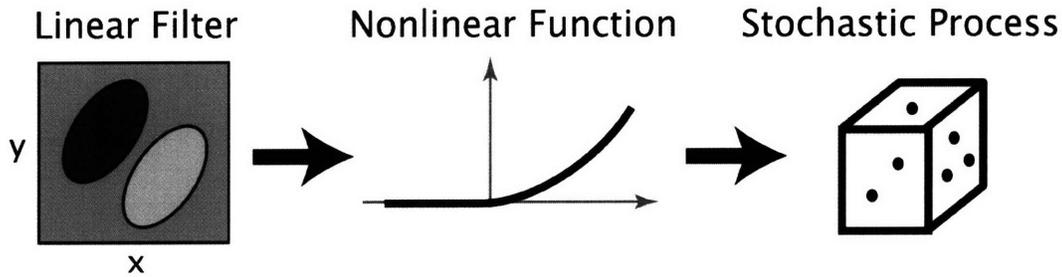


Figure 6.1: A schematic of the linear-nonlinear-stochastic model

A schematic representation of the basic linear-nonlinear-stochastic functional model. A linear filter first weights the stimulus. The filter output is then passed through a nonlinear function. Finally, the output of this function is used to generate neuronal activity. The model may be extended by combining multiple linear filters, each with its own nonlinear function, or by assuming additional complexity in the stochastic process. Adapted from Simoncelli et al. (2004).

linear-nonlinear-stochastic (LNS) framework. Such models comprise a linear filtering stage, a nonlinear function, and finally a stochastic activity-generation process. Over time, this basic formulation has been extended to include multiple cascaded linear-nonlinear stages, as well as more sophisticated stochastic processes. As these models grow in complexity, however, they eventually contain many parameters; given the limited amounts of data that can be obtained from typical recording experiments, it may become difficult to infer the likely values for all of these model dimensions.

Linear filters, when not constrained to any particular functional form, are high-dimensional objects; especially when combined with nonlinear output functions in LNS model, the optimal linear filters are difficult to estimate with acceptably small amounts of residual uncertainty. As demonstrated in this thesis, there is often much to be gained by imposing structure on the response models (such as particular tuning curve functions); this assumed structure reduces the number of parameters, at the cost of only considering solutions that lie within the chosen model space. This reduction in the number of hypotheses entertained is unavoidable in practice, however, when fitting complex models using limited amounts of data - indeed, it is an advantage of Bayesian methods that they make this reduction explicit. While such parametrized models can be difficult to solve (because the optimization landscape may have local minima), sampling methods offer a promising way forward.

While good hypotheses exist about the functional properties of primary visual neurons, middle visual areas such as V4 and MT remain more mysterious. Stimulus selectivities have been partially described, but the overall response models of these neurons have not been fully characterized. Cascaded linear-nonlinear-stochastic models provide a promising

framework, but fitting such models is difficult. The general setting described here, using sampling methods to infer parameter values, offers a way forward. Primary challenges include identifying good stimulus protocols and developing good proposal distributions to allow the sampling to proceed with reasonable computational requirements.

6.3 Adaptive Experiments

Many physiological experiments are severely constrained by the amount of recording time that is available to characterize a given neuron's response properties. In most cases, a neuron can be recorded for at most a few hours; in practical terms, the time available for stimulus presentation is limited to tens of minutes. Additional difficulties are posed by the noise inherent in many recording techniques, which leads experimenters to present stimuli a number of times in order to be able to average over many trials.

In most sensory experiments, stimulus ensembles are largely determined beforehand. In other experiments, the ensemble is adjusted along one or more dimensions (orientation, spatial frequency, etc.) to evoke optimal responses from the neuron. These strategies, however, are not optimally efficient. Ideally, later stimuli should be presented which maximize the amount of information gained about the neuronal responses, conditioned on the responses to earlier stimuli (see, for example, O'Connor et al., 2005).

As a simple example, imagine an experiment in which the goal is to characterize the orientation response bandwidth of a primary visual neuron. An initial set of trials should be used to roughly characterize the neuron's tuning curve. At this point, the analysis methods of Chapter 3 can be applied to obtain an initial estimate of the tuning curve parameters. Once this initial estimate is obtained, the key point is that additional presentations of certain stimulus values will provide more evidence about the tuning curve parameters than others. In this case, loosely speaking, stimulus values on the flanks of the tuning curve will be more informative; see Figure 6.2 for a simple illustration of this process. We note that the full benefit of this approach will become apparent in more complex models for which it is not obvious from simple inspection which stimulus values will provide the most benefit.

While a further development of the theory and methods underlying such adaptive paradigms is beyond the scope of this thesis, this principle can be applied in more complex settings as well. In cortical areas and stimulus protocols for which the response models have many dimensions, such adaptive paradigms may represent the only method by which the response properties of neurons can be adequately characterized in the limited recording time available.

Adaptive experimental paradigms rely on the ability to quickly and reliably fit models to data. While the sampling approach is not unique in its ability to do so, it does provide an

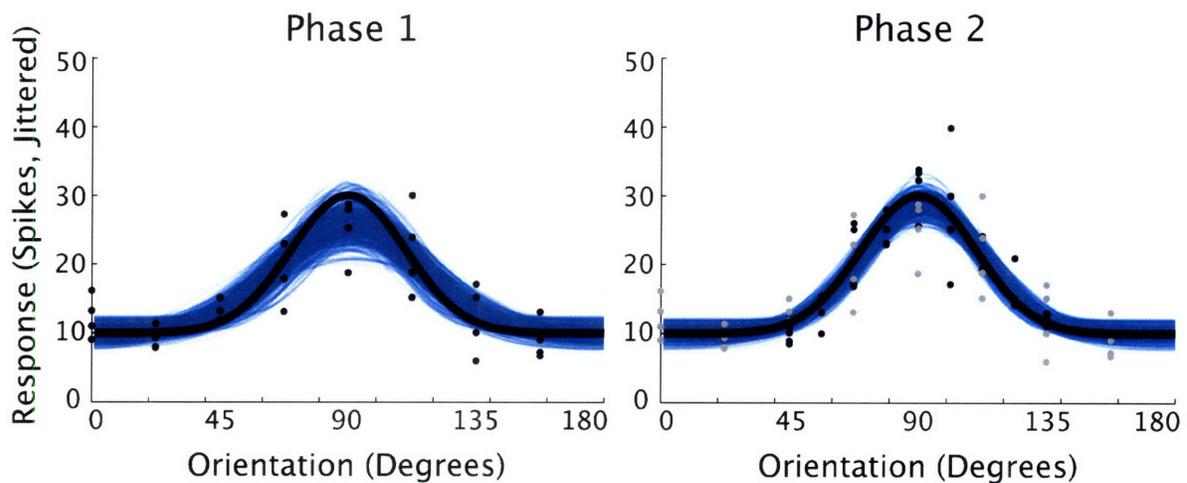


Figure 6.2: A simulated experiment using adaptive stimuli

A simple example using simulated data to illustrate how an adaptive stimulus protocol aid in the characterization of neuronal responses. Shown at left are the results of fitting a circular Gaussian tuning curve model to data obtained from an initial phase of stimulus presentations (true tuning curve show as a thick black line, observed spike counts as black dots, and sampled tuning curves as blue lines). Noting that additional stimulus presentations on the flanks of the initial tuning curve will provide the most evidence about the response properties of the neuron, the next phase of stimulus presentations are concentrated on these values. Shown at right are the final results obtained from this next phase (previous responses shown in gray, new responses shown in black).

especially flexible and robust approach.

6.4 Conclusion

Recent years have seen a number of philosophical debates about the merits and applicability of Bayesian statistical methods, relative to other approaches. This thesis has demonstrated that, whatever the outcome of these arguments, hierarchical probabilistic models and Bayesian statistical inference methods are effective at answering questions about physiological data and explaining the behavior of the nervous system. I hope that the research described here will help these approaches move into the mainstream of computational neuroscience.

Appendix A

BayesPhys Usage

This appendix describes how to use the BayesPhys MatLab toolbox, a software package which implements all of the tuning curve analysis methods described in Chapter 3.

A.1 Preparing stimuli and data for analysis

The main sampling function, `tc_sample`, requires that the stimuli and the neuronal responses be provided as equal-length vectors. Each element of the vector corresponds to a single trial; the stimulus vector contains the stimulus value presented on that trial (for example, the angular orientation), and the response vector contains the measured neuronal activity (for example, the number of spikes elicited by the stimulus).

A.2 Performing sampling

All sampling is performed via the `tc_sample` function. Given the stimulus values, neuronal responses, TC function, and noise model, this function performs MCMC sampling on the appropriate TC parameters and returns a structure containing the results. The arguments to `tc_sample` are as follows:

1. The vector of stimulus values
2. The vector of response values
3. A string indicating the name of the TC function to use, e.g., 'circular_gaussian_180'
4. A string indicating the name of the noise model to use, e.g., 'poisson'
5. A structure containing any necessary options, such as the number of samples to collect

When this function is executed, it will provide output indicating that it is performing two sampling runs, one “burn-in” run to calibrate the proposal distributions, and then a final run to collect samples. Typical output therefore looks as follows:

```
>> S=tc_sample(x,y,'circular_gaussian_180','poisson',opts);
Burnin
Sampling
```

A.3 Interpreting sampling results

The `tc_sample` function returns a structure containing the parameter samples, some basic summaries, and also the log probability scores associated with each sample:

```
>> S
S =
      P1: [102x1 double]
P1_median: 5.1441
      P1_CI: [4.5031 5.8740]
      P2: [102x1 double]
P2_median: 9.7283
      P2_CI: [7.7713 11.6021]
      P3: [102x1 double]
P3_median: 92.1375
      P3_CI: [88.0366 96.1815]
      P4: [102x1 double]
P4_median: 20.4554
      P4_CI: [16.4116 25.3014]
log_prior: [102x1 double]
log_llhd: [102x1 double]
log_post: [102x1 double]
```

The meaning of each TC parameter depends on the TC function that was used in the sampling process. For example, in the case of the circular Gaussian function, P1 is the baseline response, P2 is the peak amplitude response, P3 is the preferred stimulus value, and P4 is the tuning width.

These basic results can be used in a number of ways. The samples can be directly examined to recover the plausible range of tuning values (by plotting a histogram, for example).

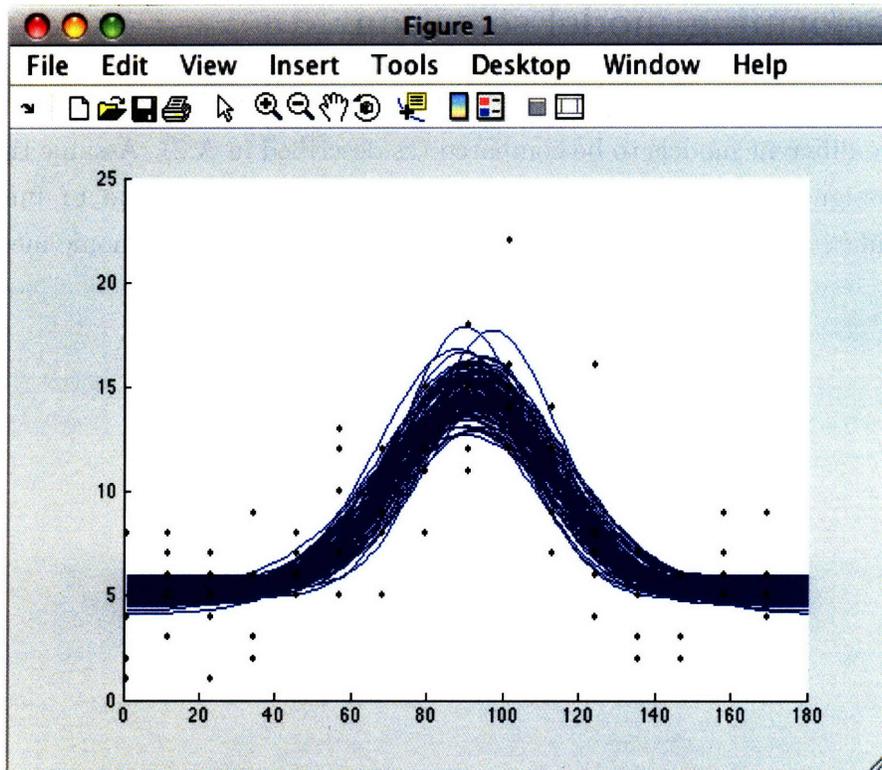


Figure A.1: BayesPhys results example

Example tuning curve samples (blue), plotted with the response data (black dots)

Median values and 95% credibility intervals are automatically precomputed, and provided in the results structure.

To compare the parameter values obtained under two different experimental conditions, first compute samples under each condition separately. Then use, the `compare_samples` function to estimate the fraction of samples obtained under one condition that are greater than those from the other condition. For example, the following call shows that the fourth parameter is greater in the first condition, with more than 95% probability:

```
>> compare_samples(S1,S2,4)
ans =
    0.9590
```

The tuning curve samples can also be plotted, along with the data (shown in Figure A.1):

```
>> plot_cg_samples(S,180)
>> plot(x,y,'.k')
```

A.4 Performing model selection

To perform model selection using Bayes Factors, first perform sampling on the same data using the two different models to be compared (as described in A.2). Assume these sampling results are assigned to variables named `S1` and `S2`. Then, the `compute_bf` function can be used to calculate an approximation of the Bayes Factor using the harmonic mean estimator.

Appendix B

Gaussian Approximation of Synaptic Activity

In order to apply the Kalman filter to estimate the gains, we need to compute a local Gaussian approximation of the likelihood $\Pr(s|g)$. We differentiate the logarithm of the likelihood once to obtain:

$$\begin{aligned} h'(g) &= \frac{d \log(\Pr(s|g))}{dg} \\ &= \frac{-1}{g} + \frac{s}{g^2} \end{aligned}$$

Differentiating a second time yields:

$$\begin{aligned} h''(g) &= \frac{d^2 \log(\Pr(s|g))}{dg^2} \\ &= \frac{1}{g^2} + \frac{2s}{g^3} \end{aligned}$$

The local Gaussian approximation about g therefore has mean $\mu = g - \frac{h'(g)}{h''(g)}$ and variance $\sigma^2 = -\frac{1}{h''(g)}$. This mean and variance can then be used by the standard Kalman update equations to obtain a new estimate of g , given a new input value s .

Bibliography

- Adelson, E. and Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2.
- Anstis, S., Verstraten, F. A. J., and Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences*, 2(3):111–117.
- Beck, O., Chistiakova, M., Obermayer, K., and Volgushev, M. (2005). Adaptation at synaptic connections to layer 2/3 pyramidal cells in rat visual cortex. *Journal of Neurophysiology*, 94:363–376.
- Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci U S A*, 92(9):3844–8.
- Bonhoeffer, T. and Grinvald, A. (1993). The layout of iso-orientation domains in area 18 of cat visual cortex: optical imaging reveals a pinwheel-like organization. *Journal of Neuroscience*, 13:4157–4180.
- Brenner, N., Bialek, W., and de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702.
- Carandini, M. and Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *J Neurosci*, 20(1):470–84.
- Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis-hastings output. *Journal of the American Statistical Association*, 96:270–281.
- Citron, M. C. and Emerson, R. C. (1983). White noise analysis of cortical directional selectivity in cat. *Brain Research*, 279:271–277.
- Dragoi, V., Rivadulla, C., and Sur, M. (2001). Foci of orientation plasticity in visual cortex. *Nature*, 411(6833):80–6.

- Dragoi, V., Sharma, J., and Sur, M. (2000a). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, 28(1):287–98.
- Dragoi, V., Sharma, J., and Sur, M. (2000b). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, 28(1):287–98.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1:54–75.
- Fairhall, A. L., Lewen, G. D., Bialek, W., and de Ruyter Van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, 412(6849):787–92.
- Galaretta, M. and Hestrin, S. (1998). Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neuroscience*, 1:587–594.
- Gandolfo, F., Li, C., Benda, B. J., Schioppa, C. P., and Bizzi, E. (2000). Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc Natl Acad Sci U S A*, 97(5):2259–63.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall.
- Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419.
- Gibson, J. (1933). Adaptation, aftereffect and contrast in the perception of curved lines. *J Exp Psychol*, 16:1–31.
- Gibson, J. and Radner, M. (1937). Adaptation, after-effect and contrast in the preception of tilted lines: I. quantitative studies. *J Exp Psychol*, 20:453–467.
- Gillespie, D. C., Lampl, I., Anderson, J. S., and Ferster, D. (2001). Dynamics of the orientation-tuned membrane potential response in cat primary visual cortex. *Nat Neurosci*, 4(10):1014–9.
- Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *J. Physiol. (Lond.)*, 148:574–591.
- Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.*, 160:106–154.
- Jaynes, E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

- Jin, D. Z., Dragoi, V., Sur, M., and Seung, H. S. (2005). Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex. *J Neurophysiol*, 94(6):4038–50.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45.
- Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kato, H., Bishop, P., and Orban, G. (1978). Hypercomplex and simple/complex cell classifications in cat striate cortex. *Journal of Neurophysiology*, 41:1071–1095.
- Knill, D. C. and Richards, W., editors (1996). *Perception as Bayesian Inference*. Cambridge University Press.
- Kohn, A. and Movshon, J. A. (2003). Neuronal adaptation to visual motion in area mt of the macaque. *Neuron*, 39(4):681–91.
- Kohn, A. and Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque mt neurons. *Nat Neurosci*, 7(7):764–72.
- Kording, K. P., Tenenbaum, J. B., and Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat Neurosci*, 10(6):779–86.
- Korenberg, M. J. and I.W.Hunter (1990). The identification of nonlinear biological systems: Wiener kernel approaches. *Annals of Biomedical Engineering*, 18:629–654.
- Kunz, M., Trotta, R., and Parkinson, D. R. (2006). Measuring the effective complexity of cosmological models. *Phys Rev D*, 74.
- Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Comput Neural Syst*, 9:R53–R78.
- Li, C. S., Padoa-Schioppa, C., and Bizzi, E. (2001). Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron*, 30(2):593–607.
- Liu, J. (2002). *Monte Carlo Strategies in Scientific Computing*. Springer.
- MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

- Maldonado, P. E., Godecke, I., Gray, C. M., and Bonhoeffer, T. (1997). Orientation selectivity in pinwheel centers in cat striate cortex. *Science*, 276(5318):1551–5.
- Marmarelis, P. N. and Marmarelis, V. Z. (1978). *Analysis of physiological systems: the white noise approach*. Plenum Press.
- Mather, G., Verstraten, F., and Anstis, S. M. (1998). *The motion aftereffect : a modern perspective*. MIT Press.
- McAdams, C. J. and Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *J Neurosci*, 19(1):431–41.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Nelson, S., Toth, L., Sheth, B., and Sur, M. (1994). Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science*, 265(5173):774–7.
- O’Connor, K., Petkov, C., and Sutter, M. (2005). Adaptive stimulus optimization for auditory cortical neurons. *Journal of Neurophysiology*, 94:4051–4067.
- Ohki, K., Chung, S., Ch’ng, Y. H., Kara, P., and Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433(7026):597–603.
- Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network*, 15(4):243–62.
- Paninski, L., Pillow, J. W., and Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput*, 16(12):2533–61.
- Ringach, D. L., Hawken, M. J., and Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387(6630):281–4.
- Ringach, D. L., Shapley, R. M., and Hawken, M. J. (2002). Orientation selectivity in macaque v1: diversity and laminar dependence. *J Neurosci*, 22(13):5639–51.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, 2nd edition.
- Roweis, S. and Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Comput*, 11(2):305–45.

- Rust, N., Schwartz, O., Movshon, J., and Simoncelli, E. (2005). Spatiotemporal elements of macaque v1 receptive fields. *Neuron*, 46:945–956.
- Sahani, M. and Linden, J. F. (2003). Evidence optimization techniques for estimating stimulus-response functions.
- Saul, A. B. and Cynader, M. S. (1989a). Adaptation in single units in visual cortex: the tuning of aftereffects in the spatial domain. *Vis Neurosci*, 2(6):593–607.
- Saul, A. B. and Cynader, M. S. (1989b). Adaptation in single units in visual cortex: the tuning of aftereffects in the temporal domain. *Vis Neurosci*, 2(6):609–20.
- Schummers, J., Cronin, B., Wimmer, K., Stimberg, M., Martin, R., Obermayer, K., Koering, K., and Sur, M. (2007). Dynamics of orientation tuning in cat v1 neurons depend on location within layers and orientation maps. *Frontiers in Neuroscience*, 1(1):145–159.
- Sclar, G., Lennie, P., and DePriest, D. D. (1989). Contrast adaptation in striate cortex of macaque. *Vision Res*, 29(7):747–55.
- Sharma, J., Dragoi, V., Tenenbaum, J. B., Miller, E. K., and Sur, M. (2003). V1 neurons signal acquisition of an internal representation of stimulus location. *Science*, 300(5626):1758–63.
- Simoncelli, E. P., Paninski, L., Pillow, J., and Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In Gazzaniga, M., editor, *The cognitive neurosciences*.
- Smith, A., Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer.
- Somers, D., Dragoi, V., and Sur, M. (2001). *Orientation selectivity and its modulation by local and long-range connections in visual cortex*, volume 15 of *Cerebral Cortex*, pages 471–520. Academic Press.
- Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci*, 15(8):5448–65.
- Stocker, A. and Simoncelli, E. (2005). *Sensory adaptation within a bayesian framework for perception*. MIT Press.
- Swindale, N. (1988). Orientation tuning curves: empirical description and estimation of parameters. *Biological Cybernetics*, 78(1):45–56.

- Teich, A. F. and Qian, N. (2003). Learning and adaptation in a recurrent model of v1 orientation selectivity. *J Neurophysiol*, 89(4):2086–100.
- Teich, A. F. and Qian, N. (2006). Comparison among some models of orientation selectivity. *J Neurophysiol*, 96(1):404–19.
- Trotta, R. (2007). Applications of bayesian model selection to cosmological parameters. *Monthly Notices of the Royal Astronomical Society*.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol*, 93(2):1074–89.
- Tsodyks, M. V. and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci U S A*, 94(2):719–23.
- Victor, J. D. and Knight, B. W. (1979). Nonlinear analysis with an arbitrary stimulus ensemble. *Q. Appl. Math.*, 37:113–136.
- Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–6.
- Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Res*, 39(23):3960–74.
- Wood, F., Goldwater, S., and Black, M. J. (2006). A non-parametric bayesian approach to spike sorting. In *IEEE Eng Med Biol Soc*, pages 1165–1168.
- Yuille, A. and Kersten, D. (2006). Vision as bayesian inference: analysis by synthesis? *Trends in Cognitive Science*, 10:301–308.