JYVÄSKYLÄ STUDIES IN COMPUTING 46

Janne V. Kujala

On Computation in Statistical Models with a Psychophysical Application

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa (S212)
joulukuun 16. päivänä 2004 kello 12.

Academic dissertation to be publicly discussed, by permission of
the the Faculty of Information Technology of the University of Jyväskylä,
in Auditorium S212, on December 16, 2004 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2004

On Computation in Statistical Models with a Psychophysical Application

JYVÄSKYLÄ STUDIES IN COMPUTING 46

Janne V. Kujala

On Computation in Statistical Models with a Psychophysical Application

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2004

Editors
Tommi Kärkkäinen
Department of Mathematical Information Technology, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

URN:ISBN 9513920402
ISBN 951-39-2040-2 (PDF)

ISBN 951-39-2017-8 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä
and ER-Paino Ky, Lievestuore 2004

ABSTRACT

Kujala, Janne V.
On computation in statistical models with a psychophysical application
Jyväskylä: University of Jyväskylä, 2004, 40 p. (+included articles)
(Jyväskylä Studies in Computing
ISSN 1456–5390; 46)
ISBN 951-39-2040-2
Finnish summary
Diss.

We consider two approaches to computation in statistical models: explicit computation within discretized models and approximate computation using Monte Carlo simulation. We present new computational techniques on both of these approaches and apply them to Bayesian adaptive estimation of psychometric functions. The main result is a new estimation procedure for two-dimensional psychometric functions. We motivate a color perception modeling problem and apply the new estimation procedure to it. We shortly discuss the experimental results.

Keywords: Cellular automata, Markov Chain Monte Carlo, Bayesian adaptive estimation, Psychophysics

Author Janne V. Kujala
Department of Mathematical Information Technology
University of Jyväskylä
Finland

E-mail: jvk@iki.fi

Supervisors Dr. Tuomas J. Lukka
Professor Pekka Neittaanmäki
Department of Mathematical Information Technology
University of Jyväskylä
Finland

Reviewers Professor John Canny
Computer Science Division
University of California, Berkeley
USA

Professor Heikki Haario
Department of Information Technology
Lappeenranta University of Technology
Finland

Opponent Professor Keijo Ruotsalainen
Department of Electrical Engineering
Mathematics Divison
University of Oulu
Finland

ACKNOWLEDGEMENTS

I am very grateful to my supervisors Dr. Tuomas Lukka for scientific guidance and Prof. Pekka Neittaanmäki for encouragement and support throughout the course of my postgraduate studies and research.

I would like to thank Prof. Antti Penttinen for discussions and insightful comments at many stages of this work and Prof. Pertti Saariluoma for discussions regarding the psychophysical applications of this work.

I am grateful to Prof. John Canny and Prof. Heikki Haario for reviewing the manuscricpt.

For financial support I thank the Deparment of Mathematical Information Technology, Agora Center and the COMAS Graduate School (each of University of Jyväskylä), University of Jyväskylä, and the Ellen and Artturi Nyyssönen Foundation.

Finally, I would like to express my gratitude to my family and friends for their support.

Jyväskylä, December 2004
Janne V. Kujala

CONTENTS

ABSTRACT **3**

ACKNOWLEDGEMENTS **5**

CONTENTS **7**

1 INTRODUCTION **9**

2 DISCRETE SYSTEMS **11**
2.1 Number-conserving cellular automata 11
2.2 Analytical solutions 12

3 MARKOV CHAIN MONTE CARLO **15**
3.1 Exploring Bayesian posterior distributions 15
3.2 Population-based methods for speeding up MCMC 16

4 BAYESIAN ADAPTIVE ESTIMATION **19**
4.1 Psychophysics 19
4.2 The next dimension 20

5 PSYCHOPHYSICAL APPLICATION **22**
5.1 Rendering recognizably unique textures 22
5.2 Color discrimination experiment 24
5.3 Sensory noise models of chromatic crispening 27
5.4 Bayesian model selection 29
5.5 Discussion 32

6 CONCLUSIONS **35**

REFERENCES **37**

YHTEENVETO (FINNISH SUMMARY) **40**

INCLUDED ARTICLES **41**

LIST OF INCLUDED ARTICLES

[A] J. V. KUJALA AND T. J. LUKKA, *Solutions for certain number-conserving deterministic cellular automata*, Physical Review E 65(2), 026115, 2002.

[B] T. J. LUKKA AND J. V. KUJALA, *Using Genetic Operators to Speed up Markov Chain Monte Carlo Integration*, Monte Carlo Methods and Applications 8(1), 51–71, 2002.

[C] J. V. KUJALA AND T. J. LUKKA, *Bayesian Adaptive Estimation: The Next Dimension*, submitted to journal, 2004.

[D] J. V. KUJALA AND T. J. LUKKA, *Rendering recognizably unique textures*, Proceedings of the 7th International Conference on Information Visualization, pp. 396–405, London, July 2003.

1 INTRODUCTION

It is relatively easy to come up with statistical or physical models for many different problems. However, the use and evaluation of such models is often intractable due to limited computational resources. The main theme of this thesis is the optimization of computational models for practical implementation. We consider discretized models where the problem space consists of discrete cells as well as continuous models which need to be somehow approximated in order to be implemented within a computer. While all the problems studied in this thesis are quite different from each other, we will apply similar techniques in their solutions; we will make use of the symmetries inherent in the definition of the models and apply mathematical reformulations to reduce the computational complexity of their implementation.

In our first problem, motivated by the unexpected behavior of the system observed in [8], we study an (unrealistic) cellular automaton model of highway traffic. The cellular automaton yields a simple discrete statistical model of the flow of cars as a function of the density of the cars. However, computation of a flow diagram through a simulated system is excruciatingly slow. In [A], through several computational and mathematical reformulations, we are able to reach an exact analytical expression of the flow diagram as well as explain the unexpected qualitative features of the system.

The simplest way to implement a continuous model is perhaps by discretization of the problem space. Another approach is Markov Chain Monte Carlo (MCMC) [2], which provides a general purpose tool for sampling a continuous distribution. While the mathematical formulation of the basic Metropolis MCMC algorithm [22] is very general, an efficient implementation depends, to a great extent, on the properties of the system. For this, there is no efficient general procedure (cf. no free lunch theorems in optimization [32]). We consider population based algorithms for speeding up MCMC, a framework on which many new MCMC ideas build on. In this context, we present our independently developed Genetic Operator MCMC algorithm [B], a new method based on an idea from global optimization, and prove its convergence properties [B].

The main practical application of this thesis is Bayesian adaptive estimation of psychometric functions [31], which relate stimulus intensity to observer performance. We apply both the discretization and MCMC approach to this problem.

Using similar computational optimizations as in the above examples we are able to implement a new practical estimation procedure for two-dimensional psychometric functions [C].

Finally, we consider the real-world problem that motivated the work in [C]. We introduce unique background textures, a user interface technique for visualizing document identity [D]. We present a qualitative model of visual perception which provides guidelines for rendering distinguishable textures. To obtain justification for the ad hoc distributions of texture features, we attempt to quantify certain aspects of color perception. We present a color discrimination experiment and apply the new two-dimensional estimation procedure for collecting the data. A Bayesian model of the data justifies certain assumptions of the color distribution of the textures.

Although the statistical models considered in this thesis are different on the surface, they do exhibit similar regularities from an abstract algebraic perspective. For example, we shall see that after certain reformulations both the statistical model of [A] and the discretized model of [C] have similar convolution structure. To exploit these regularities, however, we will use different tools in different models depending on the specific properties of the system.

2 DISCRETE SYSTEMS

In this chapter, we consider some computational and mathematical techniques for discrete models through an overview of [A]. We consider an idealized cellular automaton model of highway traffic. This is an example of a simple discrete system in which exact expressions of the desired statistics are possible, although not simple to derive.

2.1 Number-conserving cellular automata

In this section, we consider one-dimensional *number-conserving* cellular automata, which can be considered as models for one-lane traffic flow.

The road is represented as a string of discrete sites, each having a value of one or zero, representing either a site occupied by a car or an empty site, respectively. The state of the road evolves through discrete time according to certain locally defined rules: the state of one site on the next time step depends only on the states of the nearby sites on the current time step. Being a number-conserving system, the total number of cars remains unchanged during the evolution of the system. Therefore, it is always possible to tell which car went where on the next time step, assuming that the cars do not overtake each other.

In number-conserving cellular automata, the interest is typically in the average *flow* ϕ of the cars as a function of the *density* ρ of the cars, forming what is known as the *fundamental diagram*. For some rules, the average flow depends on the initial configuration of the road in addition to its density. In that case, it is usual to consider the initial state to be random with the states of the individual sites being independently and identically distributed. Typically, the system then evolves to a steady state in which the average flow is well-defined and so the fundamental diagram can be defined as the expected average flow $\bar{\phi}$ given the density.

The traffic rules $\mathcal{R}_{m,k}$ [8], that are the focus of [A], is one example of such rules, see Fig. 1. It can be shown that the average flow of $\mathcal{R}_{m,k}$ is always bounded

$\mathcal{R}_{3,2}$

Time	Road state	Flow (ϕ)
0	`111000011010000`	$11/15$
1	`100011001100010`	$14/15$
2	`000100110001101`	$13/15$
3	`001001000110110`	$13/15$
4	`110010001011000`	$14/15$
5	`001100010100011`	$14/15$

FIGURE 1: The generalized traffic rules $\mathcal{R}_{m,k}$ [8]: each maximal block of at most k cars moves right at most m steps, up to the beginning of the next block, at each time step. An example evolution is shown for $\mathcal{R}_{3,2}$ with cyclic boundary conditions. The density of cars in this example is $\rho = 6/15 = 0.4$ and the average steady state flow is $\phi = 13.5/15 = 0.9$. The expected flow $\bar{\phi}$ plotted in the fundamental diagram is averaged over all (infinite) initial states with the density ρ.

by

$$\phi \leq \min\{m\rho, k(1-\rho)\} \tag{1}$$

and that the bound is exact for $m = 1$ or $k = 1$, yielding a fundamental diagram consisting of two linear pieces. However, when $m, k > 1$, the fundamental diagram exhibits a new non-linear piece, atypical of simple number-conserving rules, see Fig. 2.

To study this unexpected behavior of the system, we want to efficiently compute the expected steady-state flow as a function of the density. However, an explicit brute-force computation requires going through all possible starting states and simulating the system until the steady state is reached, which may take as many as $\mathcal{O}(L^2)$ computation steps for a cyclic starting configuration of length L. In the following, we consider how to reach the same objective more efficiently.

2.2 Analytical solutions

In this section, we discuss the computational and mathematical reformulations in [A] that finally yield an analytical solution of the fundamental diagram of the traffic rules $\mathcal{R}_{m,k}$.

We reformulate the traffic rules as a particle system, in which the initial configuration is represented as a string of certain virtual particles, each comprising of multiple sites of the road representation. Using the virtual particles, we can represent the evolution of the system through certain reactions between these particles, decoupling the linear time from the system and concentrating only on the causal effects of the reactions. This representation allows us to evaluate an initial configuration recursively, in a manner similar to a recursive parsing of a mathematical expression with parentheses. Thus, using a stack-machine algorithm, we can evaluate the initial state of length L in linear $\mathcal{O}(L)$ time, scanning the virtual particle string once from left to right.

Of course, going through all the possible initial configurations is still intractable for any reasonable length L. However, because we have a linear-

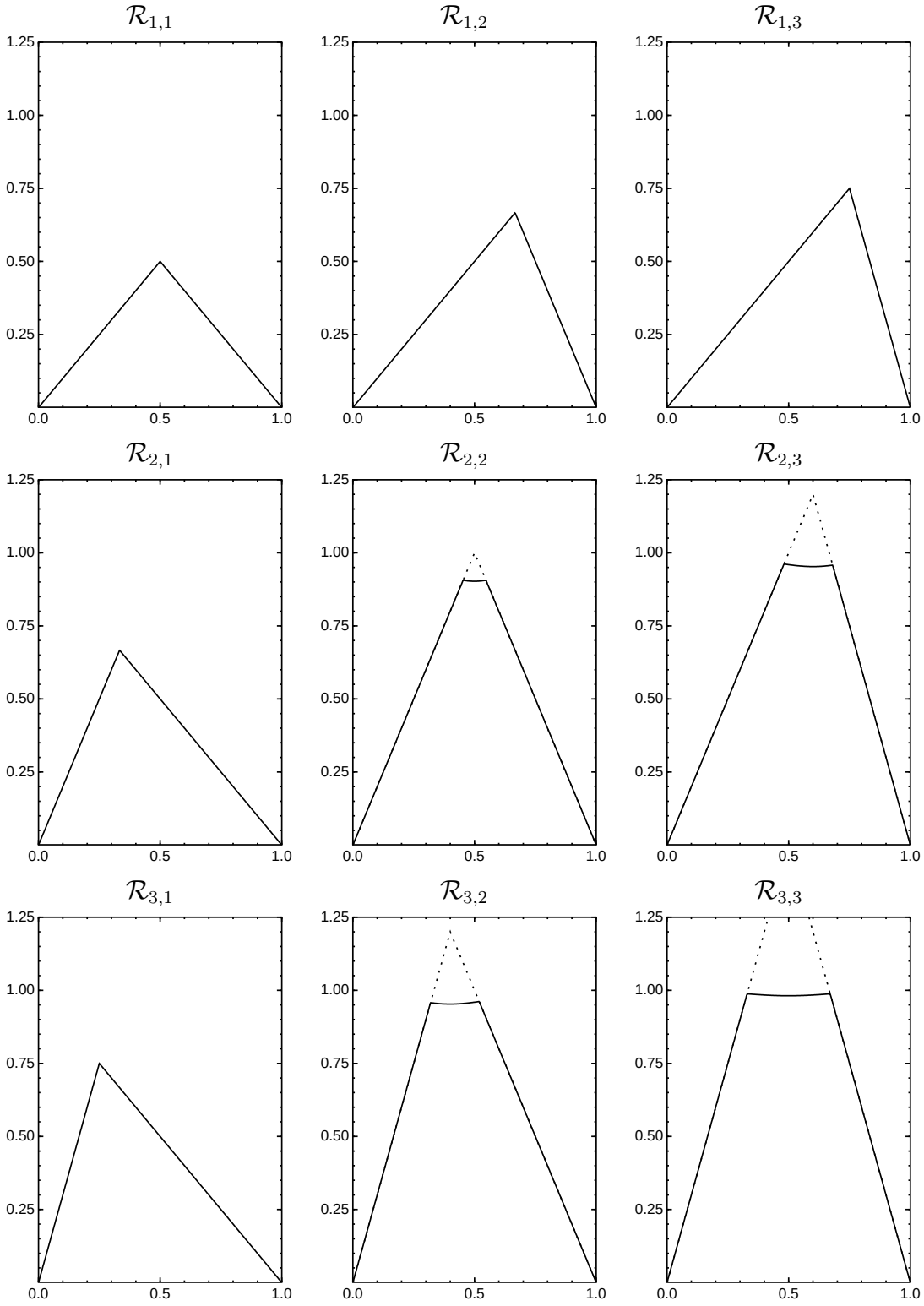

FIGURE 2: Fundamental diagrams for the rules $\mathcal{R}_{m,k}$. For $m = 1$ or $k = 1$, the fundamental diagram is completely determined by the linear bounds given by Eq. (1). However, for $m, k > 1$, the fundamental diagram does not follow the logical pattern (shown dotted in the figure) but is always cut off slightly below flow of one, exhibiting a new, curved phase. In [A], we explain this unexpected behavior and derive an exact analytical expression for the expected flow in the intermediate phase.

scanning algorithm, we can model the evolution of that algorithm statistically. We can model the evolution of the stack machine as a homogeneous Markov chain: the current state of the chain is the state of the stack of the algorithm and the next state depends only on the current state and the upcoming random virtual particle, each of which is independently and identically distributed. If we now assume an infinite length L, the Markov chain will converge to a stationary distribution, which we can use to compute the expected average flow of the steady state in the infinite road.

In order to compute the stationary distribution, we use certain regularities in the distribution of the virtual particles and in the distribution of the stacks of the algorithm during its evolution. By considering the distribution of the stack only at certain points in time (after certain reactions), we note that the sub-stacks of the stack in the stationary distribution are identically distributed. Therefore, we are able to describe the stationary stack distribution by a distribution of stack symbols.

We derive a *convolution* equation relating the probabilities of the different stack symbols in the stationary distribution. Using *generating functions* [11] to reformulate the convolutions as multiplications of formal power series, we are able to solve the system except for one variable. However, we make the unexpected discovery that neither of the two branches of the generating function can yield an everywhere admissible solution. Therefore, the correct solution must *change branches* at some point. Using this requirement, we are able to fully solve the system for any m, k, yielding a polynomial expression relating the expected average flow and the density as well as the exact phase change points in the fundamental diagram. For example, the expected average flow $\bar{\phi}$ for $\mathcal{R}_{2,2}$ is given by the polynomial equation

$$16A^2 + 8A\bar{\phi}^2 - 36A\bar{\phi}^3 + (1 + 27A)\bar{\phi}^4 - \bar{\phi}^5 = 0,$$

where $A = \rho^2(1-\rho)^2$ and ρ is the density, and the phase transitions occur exactly at

$$\rho = 1/2 \pm \left(2\sqrt{2} - 5/2\right)/7.$$

While the result itself has little practical significance, it is the first explicitly known non-linear fundamental diagram and has provided a starting point for the study of critical behavior of number-conserving cellular automata near non-linear phase boundaries [7]. Furthermore, the mathematical and statistical methods used to reach it may be of use in the study of other number-conserving cellular automata as well as other systems. We will use a similar convolution formulation in the practical application of this thesis discussed in Section 4.

3 MARKOV CHAIN MONTE CARLO

In this chapter, we introduce the basic motivation behind Markov Chain Monte Carlo, consider population based methods for speeding up MCMC, and discuss the results of [B]. This is not a complete introduction; we will only consider matters relevant in the applications of this thesis. For general references, see [2, 10, 9].

3.1 Exploring Bayesian posterior distributions

Perhaps the most important application of Markov Chain Monte Carlo is the exploration of a Bayesian *posterior* distribution

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)d\theta} \quad (2)$$

$$\propto p(y \mid \theta)p(\theta) \quad (3)$$

of the unknown parameters θ given the observed data y, where the *prior* distribution $p(\theta)$ incorporates any prior knowledge of the unknowns and the *likelihood* $p(y \mid \theta)$ models the process that is assumed to have generated the data.

Since the expression of the posterior distribution is complex, it is difficult to obtain analytical results for any but the most simple Bayesian models. Therefore, Monte Carlo integration is used instead: under certain conditions, a posterior expectation $\mathbb{E}(f(\Theta))$ can be approximated using a Monte Carlo sample $\{\theta_i\}$ of realizations from the posterior:

$$\mathbb{E}(f(\Theta)) = \frac{1}{L}\sum_{i=1}^{L} f(\theta_i) + o\left(\frac{1}{\sqrt{N}}\right). \quad (4)$$

For most models, it is difficult to draw independent samples from the posterior. Therefore, Markov Chain Monte Carlo is used: a Markov Chain, which converges to the posterior distribution, is constructed. That is, while the chain might only make short moves in the state space, its distribution over a long run will converge to the target distribution. The approximation (4) is still valid for

such a Markov Chain, only its convergence as a function of L is slower by a constant factor depending on the autocorrelation of the chain.

The Metropolis MCMC method [22] constructs the chain as follows: given the current state θ, a proposed next state θ' is drawn from a proposal density $q(\theta, \theta')$ that is symmetric: $q(\theta, \theta') = q(\theta', \theta)$. The new state is accepted with the probability

$$\alpha(\theta, \theta') = \min\left\{1, \frac{p(\theta' \mid y)}{p(\theta \mid y)}\right\}. \tag{5}$$

If the new state is not accepted, the old state remains. Moves to higher density are always accepted, and the random rejections of moves to lower density guarantee that the posterior is a stationary distribution. Under certain mild conditions, the chain will converge to the posterior distribution from any initial state [30].

The speed of the convergence and the autocorrelation of the chain depend on the proposal distribution $q(\theta, \theta')$. If it is too narrow, the chain will move slowly, and if it is too wide, the chain will be stuck for long periods of time due to rejected moves.

The Metropolis method is attractive, because it is very general and because it only needs the unnormalized posterior density given by Eq. (3). However, it may be difficult to choose a good proposal density, and complex posterior distributions can make the computation very slow. There are other basic MCMC methods such as the Hastings[14] generalization of the Metropolis algorithm for asymmetric proposal distributions and Gibbs sampling (see the general references above), but they all have similar limitations. In this thesis, we consider population based methods for speeding up MCMC and apply one such method to a psychophysical Bayesian adaptive estimation problem.

3.2 Population-based methods for speeding up MCMC

In this section, we consider population based methods for speeding up MCMC and present the results of [B] about applying genetic crossing-over without introducing bias.

Many ideas on speeding up MCMC are based on the power space formulation: instead of having only one state θ, the chain has a population of multiple parallel states $\bar{\theta} = (\theta_1, \ldots, \theta_N)$ in the power space with the target distribution consisting of independent copies of the posterior distribution. That way the algorithm has more information available about the shape of the posterior distribution. In this framework, the acceptance probability of the Metropolis method becomes

$$\alpha(\bar{\theta}; \bar{\theta}') = \min\left\{1, \prod_i \frac{p(\theta_i' \mid y)}{p(\theta_i \mid y)}\right\}; \tag{6}$$

the proposal distribution $q(\bar{\theta}, \bar{\theta}')$ can use the information in the population in any way as long as it is symmetric.

The basic Metropolis algorithm can be very slow in multidimensional, multimodal problems, because the chain has to random walk many times between

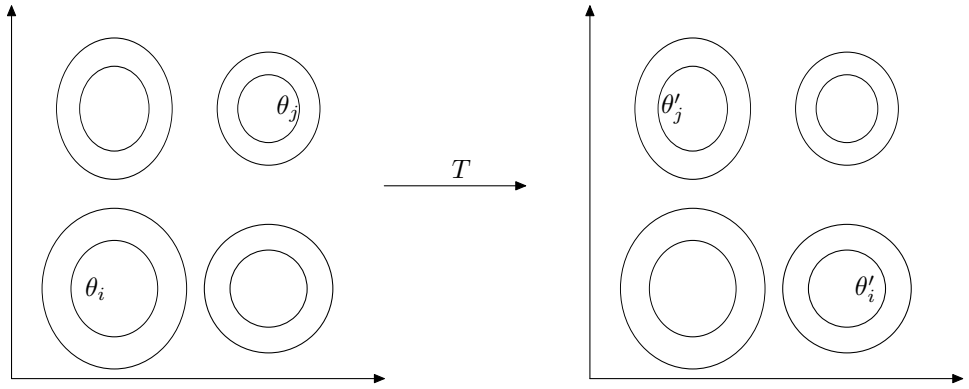

FIGURE 3: Simulating a target distribution with approximately independent components using Genetic Operator MCMC [B]: The underlying method for updating each individual only needs to explore the modes of each component; the crossing-over operators T can then generate all combinations of the modes of the components. In this figure, T swaps the first components of the two randomly chosen individuals.

the different modes in order to yield a representative sample. However, often the modes of a posterior distribution are not arbitrarily distributed but there is some structure. In [B], we propose the use of genetic crossing-over operators in the population MCMC framework to speed up the mixing between the modes. The idea is to allow the chain to jump between different modes of the posterior distribution through recombination of the individuals. Given a crossing-over operator

$$(\theta'_i, \theta'_j) = T(\theta_i, \theta_j)$$

which is its own inverse:

$$(\theta_i, \theta_j) = T(\theta'_i, \theta'_j),$$

the acceptance probability of the crossing over of two randomly chosen individuals i, j becomes

$$\alpha(\bar{\theta}, \bar{\theta}') = \min\left\{1, \frac{p(\theta'_i \mid y)p(\theta'_j \mid y)}{p(\theta_i \mid y)p(\theta_j \mid y)}\right\}. \tag{7}$$

Updating each individual through normal Metropolis transitions and occasionally applying a randomly chosen crossing-over operator T to the population may allow the chain to discover disjoint modes faster, see Fig. 3.

In [B], we prove that this method converges to the target distribution and at least preserves the rate of convergence of the underlying method for updating each individual. We illustrate the speed-up using two examples with crossing-over operators that randomly swap some variables between two individuals. We show that the method speeds up the exploration of the state space in an ideal system with independent components as well as in a system with only approximately independent components.

In the recent trend of applying concepts from evolutionary optimization algorithms in MCMC context there have been many similar ideas [23, 18, 27, 29] (our work in [B] was done independently). One interesting development is the differential evolution sampler [27, 29], which updates each individual by moving

it by the difference of two randomly chosen individuals multiplied by a constant. That way, it may be able to automatically adapt the proposal distribution to the shape of the target distribution.

Another approach gaining interest is the use of *particle filter* [6] algorithms for systems where the data is learned sequentially. A particle filter MCMC algorithm starts from a sample of the prior distribution and goes through a series of target distributions by including the data points one by one to the posterior distribution. At each step, importance resampling is used to transform the sample to the next target distribution and ordinary MCMC transitions are used to avoid degeneracy. The incremental steps eliminate most of the burn-in required to reach each target distribution and can thereby help simulating a complex target distribution. Even if all data is available immediately, the sequential approach may offer a speedup [4]. We will find a particle filter algorithm especially useful in [C], where we apply it to Bayesian adaptive estimation. In that context, the data points are learned one by one and each target distribution is needed to measure the next data point.

4 BAYESIAN ADAPTIVE ESTIMATION

In this chapter, we give a definition of Bayesian adaptive estimation [19, 21] and describe the main practical result of this thesis, a new procedure for estimating two-dimensional psychometric functions [C]. This work is motivated by the psychophysical application in Chapter 5.

Bayesian adaptive estimation is a procedure for Bayesian inference in which the data gathered so far can be used to choose the variable to be observed next. Assuming a prior distribution $p(\theta)$, and a likelihood

$$p(y_{x_1}, \ldots, y_{x_n} \mid \theta) = \prod_i p(y_{x_i} \mid \theta) \tag{8}$$

for the completed, independent, but not identical measurements, the goal is to choose the parameters x_{n+1} of the next measurement so as to gain as much useful information about the unknown Θ as possible. A theoretically good objective function minimizes the expected entropy of the posterior distribution resulting after learning the result of the next measurement $Y_{x_{n+1}}$ [19, 21]. However, it is often not used because of limited computational resources.

4.1 Psychophysics

Psychophysics studies the quantitative relationships between physical stimuli and their psychological perception (see, e.g., [24, 16]). In this context, Bayesian adaptive estimation [31] is used for estimating the parameters of a psychometric function, a function relating the observer performance and stimulus intensity. In a yes–no task, the result y_x of a trial is either correct or incorrect and the psychometric function gives the probability of a correct answer as a function of the stimulus intensity x, see Fig. 4. The Ψ Bayesian adaptive estimation method [17] uses the expected entropy of the posterior distribution as a criterion for choosing the stimulus intensity of the next trial and estimates two parameters of the psychometric function: the intensity threshold α and the slope β; the guessing and lapsing rates γ, δ are assumed to be known. Due to the required computational resources, the Ψ method has not yet been applied to more complex models.

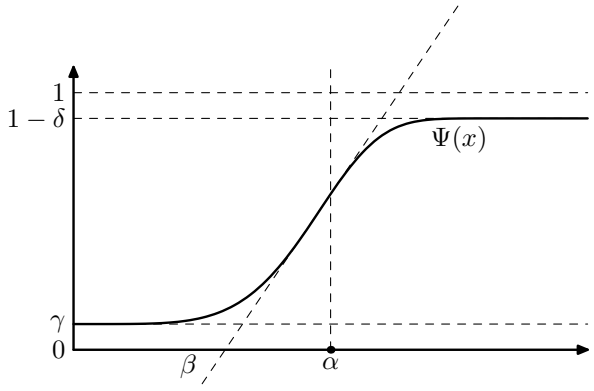

FIGURE 4: The psychometric model: a psychometric function $\Psi_\theta(x)$ gives the probability of a correct answer for a stimulus of intensity x. Experiments are designed to estimate the parameters $\theta = (\alpha, \beta, \gamma, \delta)$ from the trial results.

4.2 The next dimension

In [C], we present two new implementations of the Ψ method enabling its use in connection of more complex models. Instead of minimizing the expected entropy of the posterior distribution, we use the equivalent criterion of maximizing the mutual information between the unknowns Θ and the result of the next trial Y_x. This reformulation reveals a convolution structure in the entropy expressions, allowing for efficient computation using the Fast Fourier Transform, similar to the use of generating functions in solving the convolution equations in [A].

The reformulation also enables an efficient Monte Carlo approximation of the objective function, which only needs a sample of the *current* distribution $p(\theta \mid y_{x_1}, \ldots, y_{x_n})$ for optimizing the placement x_{n+1} of the next trial. We apply a particle filter MCMC algorithm for efficiently computing the Monte Carlo samples at each trial. However, as the target distribution changes at each trial, the MCMC kernel needs to be tuned at each trial to maintain efficiency. In our applications, we have found a method of scaling the coordinate axes at each trial to be sufficient. In this context, adaptive MCMC [13] may also prove to be useful [27, 29].

In [C], we present a model for two-dimensional psychometric functions, allowing estimation of observer performance on a two-dimensional stimulus space, e.g., for measuring the color discrimination threshold around a target color for two-dimensional color differences. The model is based on parameterizing the threshold contour around the target stimulus as an ellipse, see Fig. 5. We apply the new Ψ algorithms to this model, enabling real-time adaptive estimation. Simulations indicate that this new procedure, which we call the 2D-Ψ method, can be much more efficient than estimation based on one-dimensional procedures. Figure 6 shows some experimental results obtained using the new procedure.

In the following, we will motivate a real-world color perception modeling problem and apply the new procedure to it. The results will show that the procedure works in practice and yields consistent results.

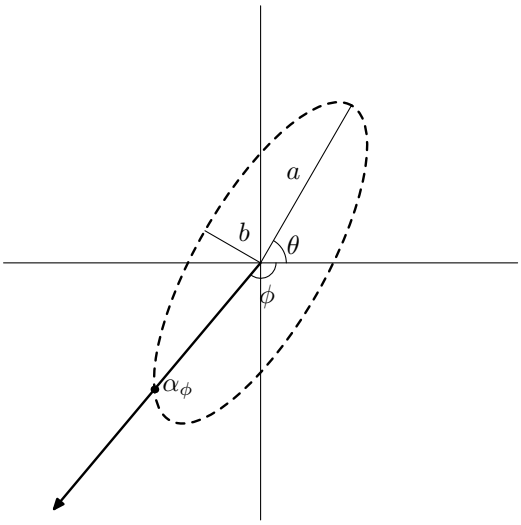

FIGURE 5: The model for two-dimensional stimuli presented in [C]: An ellipse (a, b, θ) gives the threshold α_ϕ for the 1D-model at each angle ϕ. This model is appropriate for many multi-dimensional stimuli such as color differences. It can be readily generalized to more dimensions by using ellipsoids.

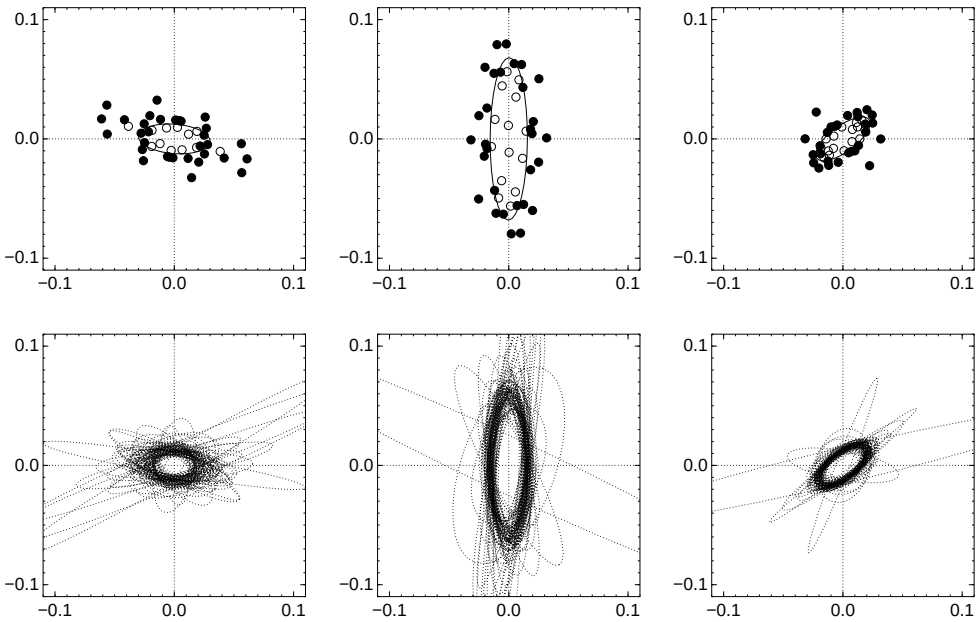

FIGURE 6: One subject's results in three different experimental conditions. The first row shows the adaptively placed trials and their results and the estimated threshold ellipses for each condition. Each trial is plotted as two symmetrically placed circles representing the stimuli to be compared; filled circles indicate successful discrimination. The second row shows a representative sample of size 100 from the posterior distribution of the threshold ellipses given the trial results.

5 PSYCHOPHYSICAL APPLICATION

In this chapter, we consider a new user interface technique [D] which motivates a color perception modeling problem. We will apply the new two-dimensional Bayesian adaptive estimation procedure to this problem. We will shortly discuss different models of the data and consider the practical application of the experimental results.

5.1 Rendering recognizably unique textures

In this section, we discuss the use of *unique background textures* for visualizing document *identity* as proposed in [D] and consider theoretical justification for certain ad hoc choices in the design of the rendering algorithm.

When a user has to work with several similar documents in a hyperstructure where each document can be reached through several routes and when only fragments of documents might be visible, it is easy to get lost or become disoriented (see Fig. 1 in [D]). In [D], we propose that texturing each document with a unique background texture may help the user to instantly recognize each document, even from a fragment. The user should be able to learn the textures of the most often visited documents as per Zipf's law.

For the textures to be useful, they should be easily recognizable and distinguishable from each other. For a document to be recognizable in any context, it is best to have no correlation between the document content and its texture. Therefore, the textures are chosen randomly — but deterministically — using the document identity as a seed.

Thus, making the textures recognizable becomes a problem of choosing a good distribution of textures, a distribution that maximizes the entropy of the textures as transformed by the human visual system. In [D], we present a simple qualitative model of visual perception, which yields guidelines for rendering recognizably unique textures, see Fig. 7. As the model is difficult to quantify, the actual rendering algorithm uses ad hoc distributions based on the general guidelines and trial and error.

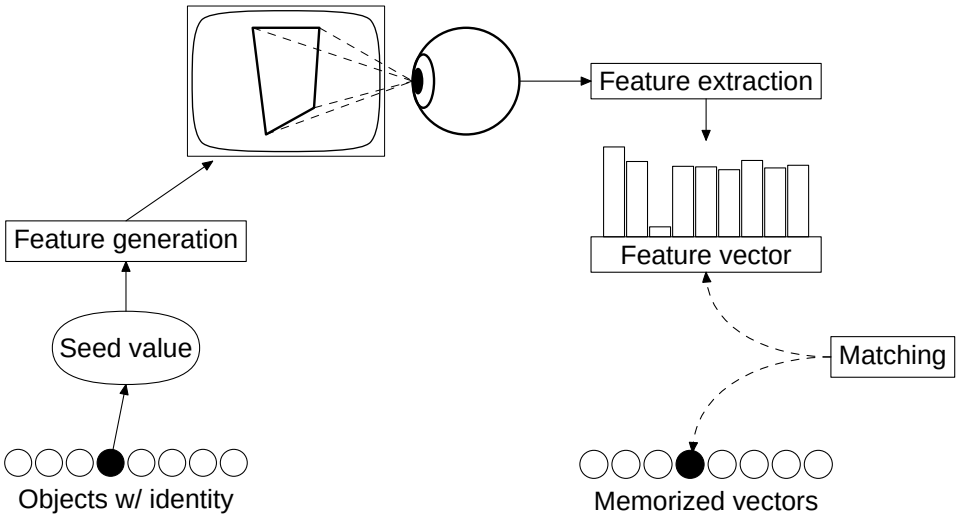

FIGURE 7: The qualitative model of visual perception used to create the rendering algorithm of [D]. ©2003 IEEE, reproduced with permission.

For example, although many features of the textures should be chosen independently of each other (e.g., colors and shapes) for maximum diversity, we have realized that choosing the different colors to be used in one texture independently is not good: each palette would appear just as a mixture of many colors. We found out that it is better to choose similar colors for one texture with higher probability.

We have conducted an initial experiment confirming the recognizability of the textures as compared to plain solid colors. There are many areas in which the computational model could be made more accurate or better justified. However, in this thesis, we will only focus on quantifying certain aspects of color perception, which is relatively orthogonal to the rest of the model.

As the textures are rather complex, we will first simplify the setting as much as possible to make the problem simpler and the results more generally applicable. We will only consider *checkerboard textures* with two different colors. In this context, the goal of rendering recognizably unique textures becomes a problem of defining a distribution of color pairs such that any set of checkerboards colored with such randomly chosen color pairs are maximally recognizable and distinguishable from each other.

The observations of [D] imply that the two colors should not be independently distributed but close to each other in the color space with higher probability. In vision literature, a phenomenon called contrast crispening [33, Sec. 6.3] is known, which means the increase of sensitivity to lightness differences when the intensity of the surrounding color is close to the intensity of the target color. A similar crispening effect has also been observed for certain chromatic changes [28]. These effects could explain why the two colors should be close to each other instead of being independently distributed: the increase in sensitivity for similar colors allows for finer distinguishable variations.

5.2 Color discrimination experiment

In this section, we attempt to provide experimental justification for the color selection scheme used in [D] for the unique background textures. We have designed an experiment that attempts to quantify the perceptual chromatic crispening effects in checkerboard textures with two constant lightness colors. The technical details of the experiment and some initial results are described in [C]. In the following sections, we will focus on the implications of the results.

Stimuli The stimuli were 16x12 fullscreen checkerboards with two constant-lightness colors, c_1 and c_2. The second color was randomly perturbed by a small color difference d, randomly changing the color to either $c_2 + d$ or $c_2 - d$ independently at each square (see Fig. 11 in [C]). For simplicity, the color c_2 was always gray, the origin of the two-dimensional color space.

The subject's task was to decide if a checkerboard picture had two or three different colors indicating whether the perturbation d was visible. The experiment consisted of multiple trials with different perturbations d and different surrounding colors c_1.

The color space was based on the unknown RGB primaries of the monitor. However, the gamma correction was calibrated so as to yield a linear color space.

Procedure The idea of the experiment was to determine the threshold of perceiving the perturbation as a function of the surrounding color c_1 and the direction of the perturbation d. The two-dimensional Bayesian adaptive estimation procedure [C] was used to adapt the perturbation d in order to find an estimate of the elliptical non-discrimination area around the target color c_2 for each condition. 16 different conditions with different surrounding colors were used, each consisting of 25 trials.

Participants The experiment was conducted as a part of a practical course. Two students and one naive subject participated in the experiment.

Results Figure 12 in [C] illustrates the basic finding of the experiment: the shape of the non-discrimination area is pointed towards the surrounding color and it becomes larger as the surrounding color is moved further away in the color space from the target color. This means that the closer the two colors of a checkerboard are to each other, the smaller the differences in the two colors that are visible, supporting the ad hoc color selection scheme of [D].

Figure 8 shows the complete results of one subject. Figure 9 illustrates the posterior threshold distributions as well as the estimated threshold ellipses for the 16 conditions for each subject. These plots are based on MCMC samples from the posterior similar to those used by the MCMC particle filter algorithm in [C], but now we have varied all six parameters of the model, including the guessing and lapsing rates. See Section 5.4 for details of the MCMC runs. The slope and

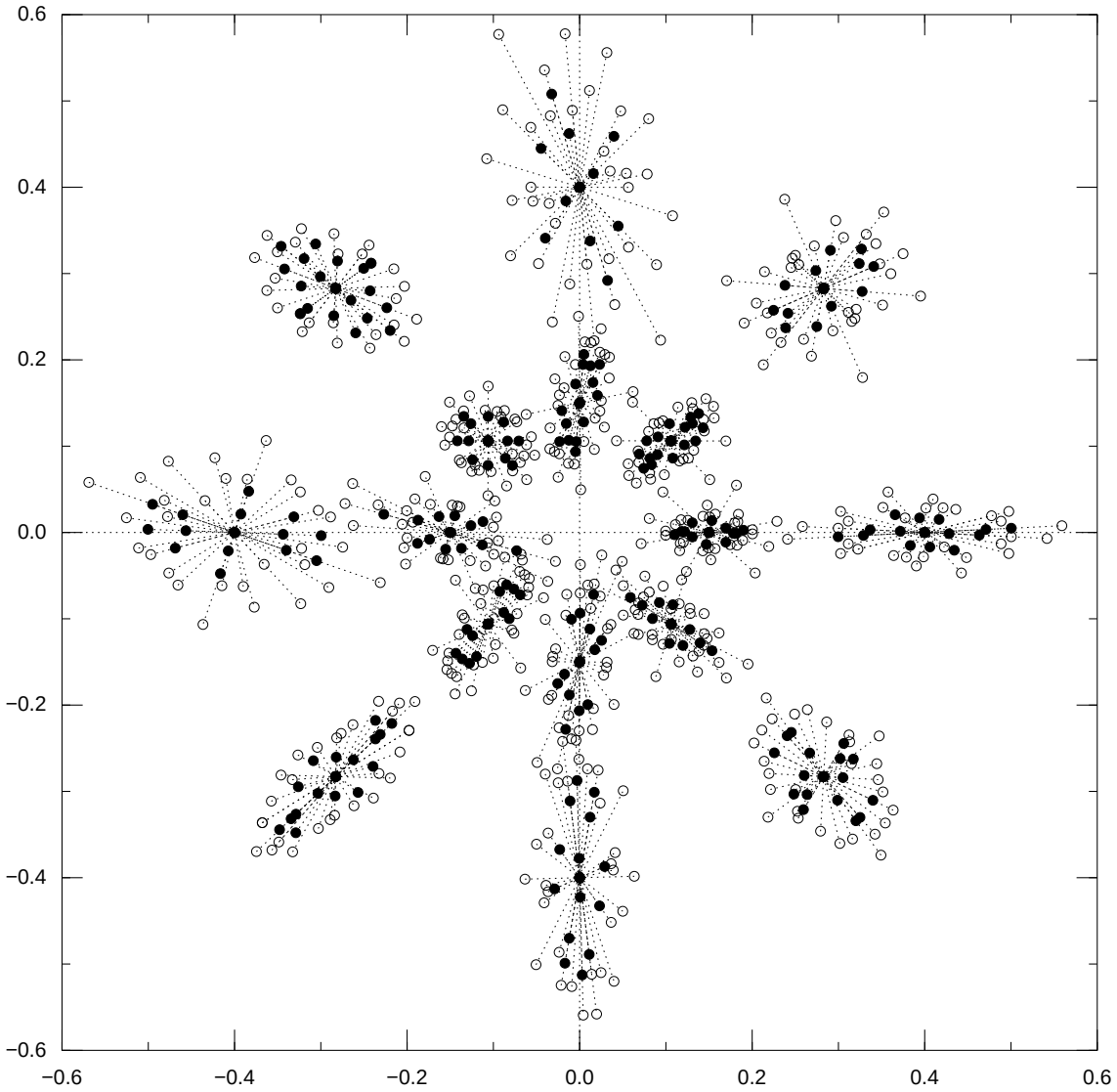

FIGURE 8: One subject's (subj. 3) results in the experiment. The trials of each condition are plotted as circles at $c_1 \pm 2d$ joined by a line segment, that is, each condition is shifted into its own coordinate system centered on the conditioning color and scaled for clarity. Non-filled circles indicate a "yes" response, meaning that the perturbation was visible.

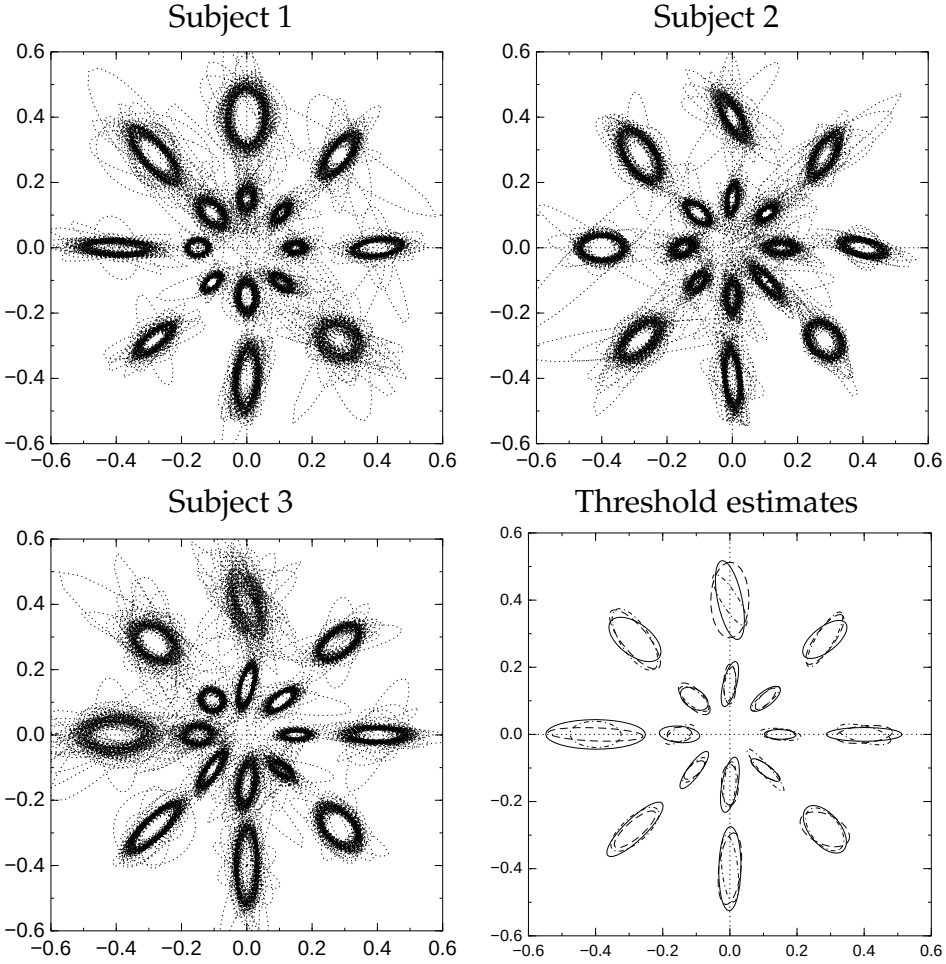

FIGURE 9: The first three plots show representative samples of size 100 from the posterior distributions of the threshold ellipses of all conditions for each subject. The fourth plot shows the mean estimates in one plot with different line types indicating different subjects. The coordinate systems are the same as in the previous figure.

the guessing and lapsing rates are not shown in the plots as they are not well-localized; we consider them as nuisance parameters [10] and, although they are used in the simulation, we in effect always integrate them out. The results of the different subjects seem similar; we will simply concatenate the data of all subjects for modeling the average observer. In the following, we consider different models of the measured data.

5.3 Sensory noise models of chromatic crispening

In this section, we consider different models of the experimental data. We attempt to provide a short quantitative description of the observed phenomena.

We already have a local model for each condition used by the estimation procedure. A naive global model would just assume each condition to be independent of each other, yielding a $16{\cdot}6 = 96$ parameter model (Model 1). However, that model does not generalize the data to unobserved conditions and does not take into account the observed regularities, in particular, the observation that the threshold ellipses seem to consistently point towards the surrounding color.

Before considering more involved models, we note that this observation is actually not entirely true: a linear transformation does not in general map the major axis of an ellipse to the major axis of the image ellipse, and therefore the particular coordinate basis used in the experiment can only by chance yield exactly pointed ellipses. The color coordinates are close to the "canonical" basis, because the scales of the standard RGB coordinates are chosen to be perceptually similar. Still, we do need to somehow include the effect of the coordinate basis into the model.

We shall apply the concepts of general recognition theory [1, 15], a multivariate generalization of the classical signal detection theory [12]. These theories consider noisy observations as random variables taking values in some psychological space where the observer is assumed to have established certain decision boundaries. In this context, suppose that the discrimination of an n-dimensional color difference is based on sensory data of the perturbation $d \in \mathbb{R}^n$ of the target color $c_1 \in \mathbb{R}^n$ under background noise and the noise resulting from the color contrast $x = c_1 - c_2 \in \mathbb{R}^n$ of the surrounding color. Suppose $N \in \mathbb{R}^{n\times n}$ is the covariance matrix of Gaussian background noise and that the color contrast contributes to the total noise with a factor $u \in \mathbb{R}$. Then, the covariance matrix of the total noise is $N + uxx^T$, where xx^T is (four times) the covariance matrix of the color distribution of the texture. Using the one-dimensional Weibull psychometric function (cf. Fig. 4) for modeling the performance for a given signal-to-noise ratio, we can define the probability of perceiving a perturbation d as

$$\Psi(d) = 1 - \delta - (1 - \gamma - \delta)\exp(-r^{10^\beta}), \tag{9}$$

where

$$r^2 = d^T[N + uxx^T]^{-1}d \tag{10}$$

is (the square of) the signal-to-noise ratio, see Fig. 10. This model (Model 2) is

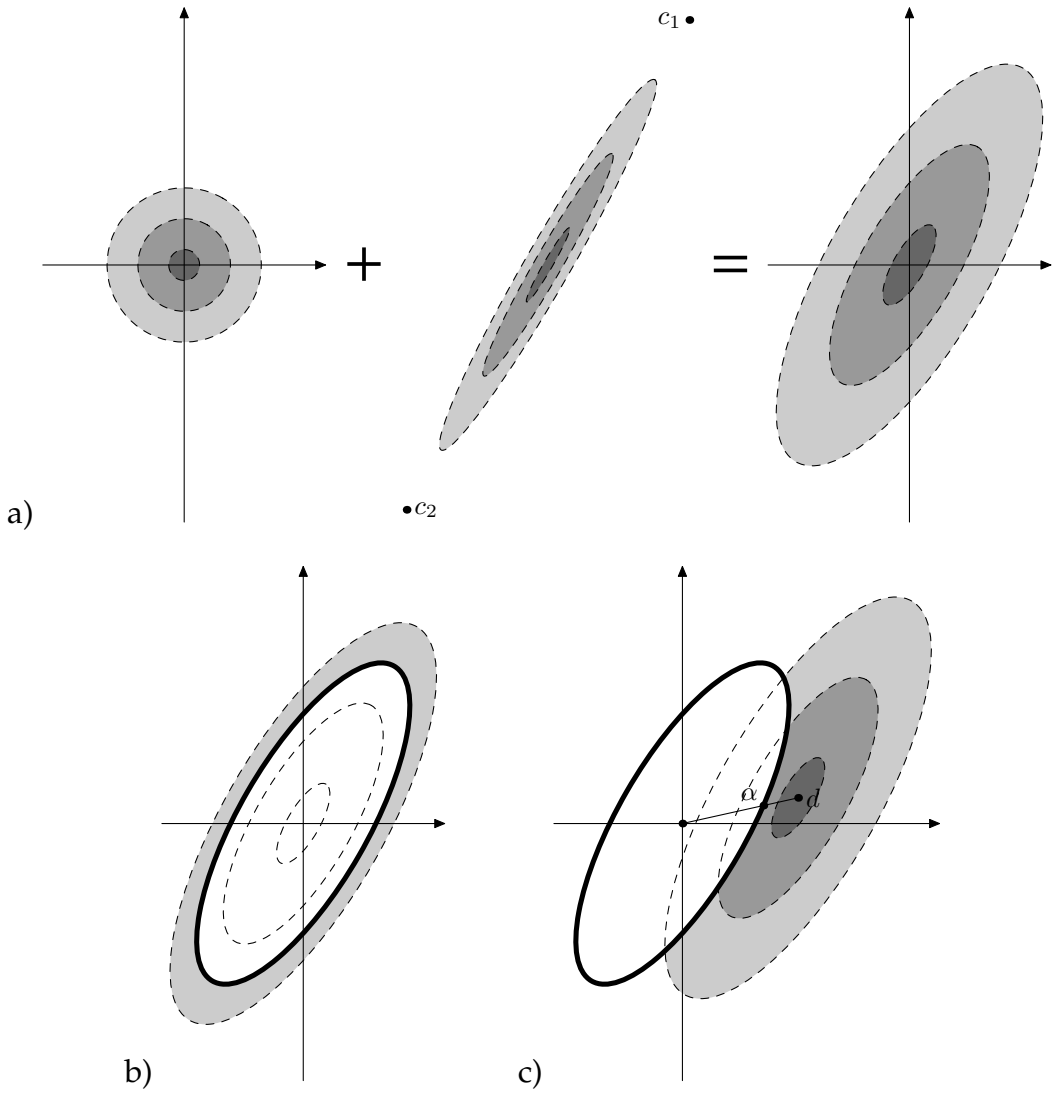

FIGURE 10: A general recognition theory interpretation of the model. a) Observer's sensory noise distribution resulting from the background noise plus the noise from the color contrast $x = c_1 - c_2$. b) Observer's threshold surrounding almost all of the total noise so as to yield a low probability of false positives. The threshold is a closed contour instead of the more typical line as the direction of the signal is not known beforehand. c) Signal d added to the noise. The shaded area outside the threshold yields the probability of perceiving the signal. Assuming Gaussian noise and a low false positive rate, the Weibull psychometric function approximates well the perception probability as a function of the signal-to-noise ratio $\|d\|/\|\alpha\|$.

consistent with the local ellipse model in the sense that for any condition x, this model can match the local model for any parameters.

Suppose $A \in \mathbb{R}^{n\times n}$ decomposes the symmetric noise matrix as $N = AA^T$. Then, the signal-to-noise ratio can be rewritten as

$$r^2 = d^T[N + uxx^T]^{-1}d = (A^{-1}d)^T[I + u(A^{-1}x)(A^{-1}x)^T]^{-1}(A^{-1}d).$$

That is, A represents a linear transformation from the canonical basis, in which the background noise is represented by an identity matrix, to the coordinates used in the experiment. For the two-dimensional experimental data, we parameterize the matrix A with $\log$-size s, $\log$-aspect c and angle θ exactly as in the local model:

$$A = 10^s \begin{bmatrix} 10^{c/2}\cos\theta & -10^{-c/2}\sin\theta \\ 10^{c/2}\sin\theta & 10^{-c/2}\cos\theta \end{bmatrix}.$$

These three parameters are sufficient to parameterize any symmetric noise matrix $N = AA^T$, yielding a total of seven parameters for this model.

Figure 11 illustrates the posterior distributions for the parameters of the two fitted models given the data. The results of the Model 1 are just the independently fitted results also shown in Fig. 9. Model 2 appears to capture the qualitative effects rather well. The deviations from the ellipse orientations predicted by this model can be explained by chance. We have confirmed this by generating replicated observed data using a simulated observer running the same experimental procedure: similar deviations do result randomly. However, there is one apparent shortcoming in this model: the threshold ellipses do not get thicker as the surrounding color moves further away in the color space, as seems to consistently happen in the independently fitted conditions.

Therefore, we add a crosstalk parameter $t \in \mathbb{R}$ modeling the uniform, undirected increase of the noise as the surrounding color moves further away (Model 3):

$$r^2 = d^T[(1 + tx^T N^{-1}x)N + uxx^T]^{-1}d. \tag{11}$$

This can again be written in a canonical basis as

$$r^2 = \hat{d}^T[(1 + t\hat{x}^T\hat{x})I + u\hat{x}\hat{x}^T]^{-1}\hat{d}, \tag{12}$$

where $N = AA^T$, $\hat{d} = A^{-1}d$, and $\hat{x} = A^{-1}x$. The volume of the threshold ellipsoid $r^2 = 1$ for condition x is

$$\pi\sqrt{\det[(1 + tx^T N^{-1}x)N + uxx^T]} = \pi\det(A)\sqrt{(1 + t\|\hat{x}\|^2)^{n-1}(1 + (u + t)\|\hat{x}\|^2)}. \tag{13}$$

The last column in Fig. 11 shows the predictions of this model. It can be seen that the uncertainty is now smaller for the innermost ellipses than without the crosstalk factor; the model now seems to fit the data better. In the following, we compare the goodness of fit and the complexity of these models.

5.4 Bayesian model selection

In this section, we compare the three models of the experimental data using multiple Bayesian model comparison criteria.

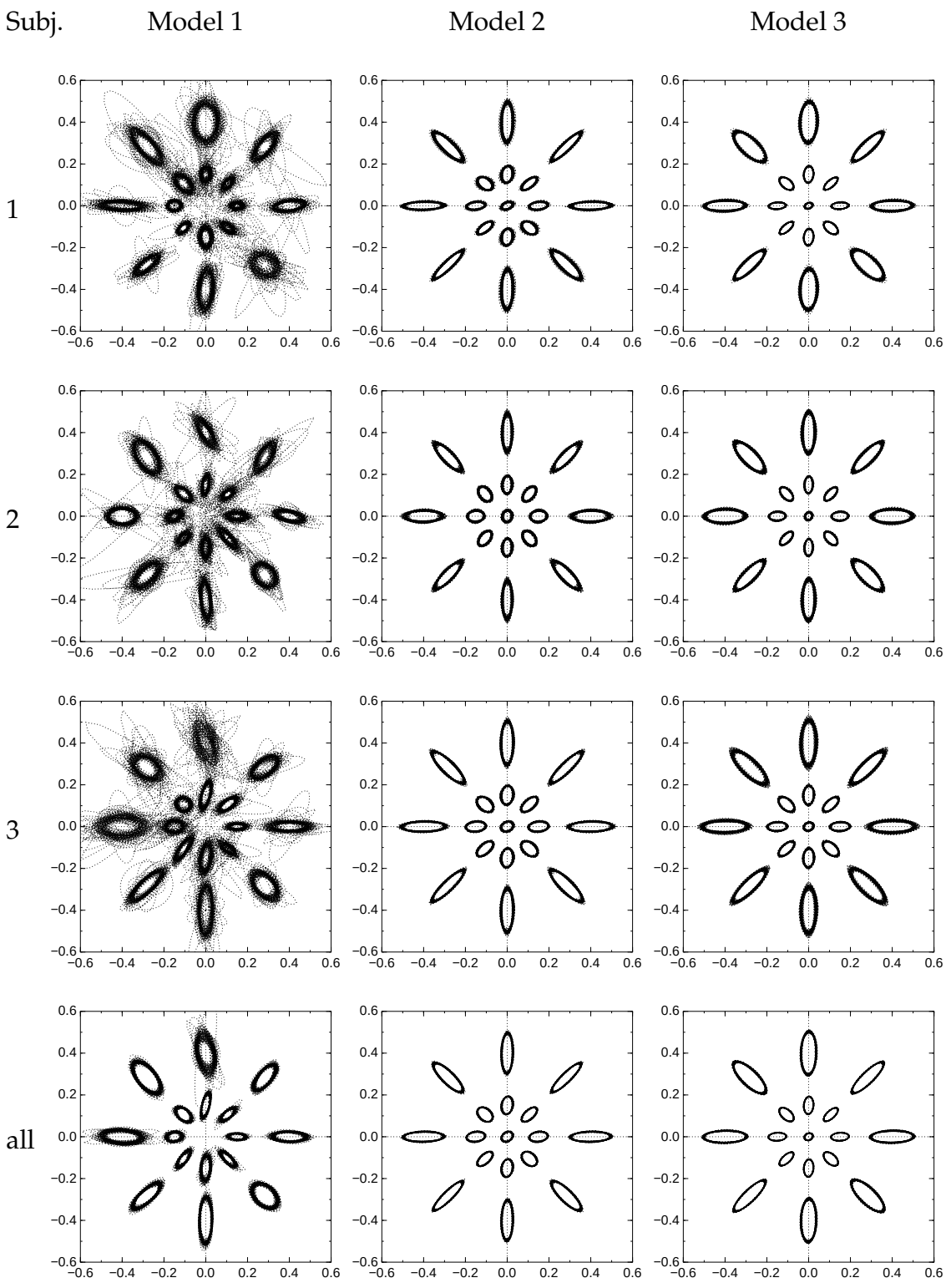

FIGURE 11: Representative samples of size 100 from the posterior threshold distributions given by the three models. The last two models can predict the threshold for unobserved conditions; the prediction for the gray surrounding color is shown as an example.

The canonical criterion for model comparison is the posterior probability of the models:

$$p(M_k \mid y) = \frac{p(y \mid M_k)p(M_k)}{\sum_k p(y \mid M_k)p(M_k)}. \tag{14}$$

Assuming identical prior probabilities $p(M_i)$ of the models, this amounts to comparing the model likelihoods

$$p(y \mid M_k) = \int p(y \mid \theta_k, M_k)p(\theta_k \mid M_k)d\theta_k, \tag{15}$$

where θ_k represents the unknowns of the model M_k. The ratio of two model likelihoods is called a Bayes factor [10, 9]. The model likelihood automatically quantifies the Occam's razor in the sense that the more variables in an overspecified model there are, the smaller the fraction of the volume of θ_k that has a high likelihood, yielding lower model likelihood [20]. In the following, we shall drop the explicit model specification M_k from the equations as the comparison criteria can be evaluated independently for each model.

Although model likelihoods seem like the ideal tool for model comparison, they are highly sensitive to the prior distributions of the models. As we use uninformative priors, the prior ranges of the variables are somewhat arbitrary, and so, although the prior ranges have little effect on the predictions of the model, they do greatly affect the model likelihood.

The model likelihood (15) can be approximated by the harmonic mean of the likelihood of a posterior sample $\{\theta_i\}$ [9]:

$$p(y) \approx \frac{L}{\sum_{i=1}^{L} 1/p(y \mid \theta_i)}. \tag{16}$$

This approximation has the added benefit that extremely unlikely regions of the priors will be disregarded for a practical MCMC sample size and therefore, as a side effect of the approximation error, this estimate is slightly less sensitive to the priors of the models.

Expectation of the deviance

$$D(\theta) = -2\log(p(y \mid \theta)),$$

of the unknown parameters θ of the model can be used as a measure of the misfit. Deviance Information Criterion [26]

$$\text{DIC} = \mathbb{E}(D(\Theta)) + p_D$$

adds a quantity called the effective number of variables

$$p_D = \mathbb{E}(D(\Theta)) - D(\mathbb{E}(\Theta))$$

as an additional penalty of model complexity to the expected deviance. These criteria can be directly computed using the MCMC posterior sample.

Table 1 shows the values of these model comparison criteria for the three models. The uniform priors used for computing these results had the ranges:

$\log_{10}$-size $[-4, 0]$, $\log_{10}$-aspect $[-2, 0]$, angle $[0, \pi]$, $\log_{10}$- noise factor $[-4, 0]$, $\log_{10}$-crosstalk $[-10, 0]$, $\log_{10}$-slope $[-0.1, 0.9]$, guessing and lapsing rates $[0, 0.1]$. These priors were the same for all models, except that Model 2 misses the cross talk factor and Model 1 misses both noise factors and has 16 copies of the other 6 parameters. The Metropolis MCMC algorithm was used with a manually tuned Gaussian proposal distribution for a run of length 220000 for each model and each data set; skipping a burn-in of 20000 steps, every 200th state was saved for a sample of size 1000 for computing the statistics of each run.

Model 3 seems best according to all criteria, except for the one exception of the expected deviance for the concatenated data of all subjects. However, as discussed in [26], an additional penalty of model complexity may be appropriate. Still, there might be a better model with a complexity between that of Model 3 and Model 1, e.g., one with a hierarchical specification [10]. However, Model 3 seems adequate in capturing the main features of the data.

Subj.	Model	$\mathbb{E}(D(\Theta))$	$D(\mathbb{E}(\Theta))$	p_D	DIC	$p(y)$	$\hat{p}(y)$
1	1	318.1	261.7	56.4	374.6	2E-83	8E-89
	2	321.5	315.8	5.8	327.3	4E-74	7E-72
	3	288.0	280.5	7.4	295.4	5E-69	1E-64
2	1	327.3	269.6	57.7	385.0	8E-86	4E-92
	2	337.1	331.6	5.4	342.5	3E-77	5E-75
	3	317.8	310.9	6.9	324.8	8E-75	8E-72
3	1	300.6	247.3	53.4	354.0	5E-82	7E-86
	2	310.4	304.4	5.9	316.3	8E-72	3E-69
	3	297.4	290.2	7.2	304.7	2E-70	8E-67
all	1	874.2	806.5	67.7	942.0	1E-225	2E-215
	2	980.1	973.6	6.5	986.6	3E-218	3E-215
	3	917.7	910.0	7.7	925.4	5E-207	3E-201

TABLE 1: Comparison of the three models using expected deviance $\mathbb{E}(D(\Theta))$, Deviance Information Criterion (DIC), the model likelihoods $p(y)$, and the model likelihoods computed using the harmonic mean estimate $\hat{p}(y)$.

5.5 Discussion

A *perceptually uniform* color space [5] means a color space in which a just perceptible difference is of similar magnitude in coordinates everywhere in the space. Thus, a uniform distribution over this space maximizes the probability that two randomly chosen colors are distinguishable. We can analogously define a perceptually uniform density of *color pairs* as the inverse of the non-discrimination volume around each color pair, the volume in which the color pairs are not distinguishable from the target pair (Fig. 12). This distribution approximately[1] max-

[1]The distribution is only approximately optimal as it disregards boundary effects and changes of the scale within the non-discrimination volumes. However, the smaller the non-discrimination volumes, the smaller these effects.

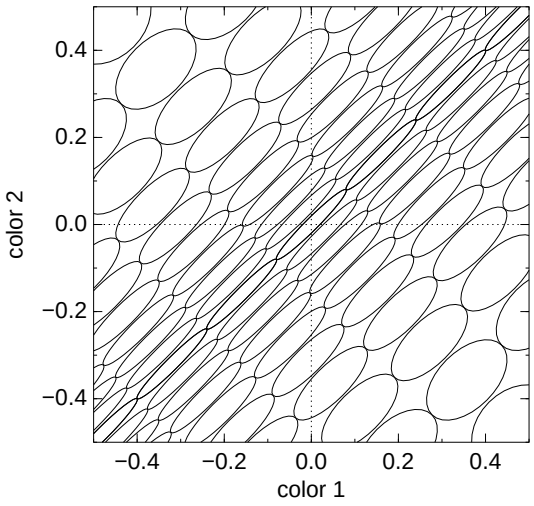

FIGURE 12: An illustration of a perceptually uniform distribution of (one-dimensional) color pairs. Each ellipsoid represents the non-discrimination volume around a color pair. In the perceptually uniform distribution each non-discrimination ellipsoid has approximately the same probability.

imizes the probability that any two randomly chosen pairs are distinguishable.

However, the experiment only measured the sensitivity to changes in one color of the pair, and only at certain points in the space of color pairs. Furthermore, it is difficult to define what kind of distinguishability is desired (side-by-side comparison, memory comparison, etc.) and so the discrimination criterion of the experiment might not be the most ideal for the use of unique background textures. Nevertheless, we expect a qualitatively similar crispening effect to be observed in any discrimination task.

We will attempt to derive a perceptually uniform color space based on the available observed data. To enable the generalization of the experimental data, we assume that the non-discrimination volume $V(c_1, c_2)$ around the color pair (c_1, c_2) can be decomposed as

$$V(c_1, c_2) = \frac{1}{2} A_1(c_1 - c_2) A_2, \tag{17}$$

where A_1 denotes the (measured) non-discrimination area for contrast changes on the constant average color plane $\{\, (c_1', c_2') : c_1' + c_2' = c_1 + c_2 \,\}$ and A_2 denotes the (unmeasured) non-discrimination area for changes in the average color on the constant contrast plane $\{\, (c_1', c_2') : c_1' - c_2' = c_1 - c_2 \,\}$. This decomposition assumes that the axes of the non-discrimination ellipsoid lie on these planes so that its volume can be accurately decomposed into the areas of the intersection ellipses.

Additionally, the decomposition assumes that the color space is uniform with respect to translations, allowing us to generalize the results measured for $(c_1, 0)$ to all of the space of color pairs. This can be justified by the adaptive effects of the eye: even though the target color c_2 was always gray in the experiment, it did not appear gray: chromatic adaptation shifted its appearance towards the opposite color of the surrounding color. Thus, we can assume that the observed phenomena do not depend on the target color being gray.

Finally, to justify A_2 being a constant, we need to make one more assumption in addition to the uniformity: we assume that the discrimination of a change in the average color of a texture does not depend on the color contrast of the texture. That is, we do not expect high contrast to mask out otherwise visible changes in the average color. While this assumption is perhaps only approximately true, it may be appropriate for the application because the viewing conditions of a texture can cause a shift in the apparent color of the texture, masking out any small differences in the sensitivity to changes in the average color.

The decomposition equation (17) allows us to compute the relative volumes of all non-discrimination ellipsoids based only on the measured sensitivity to contrast changes. Estimating the parameters of Model 3 from the concatenated data of all subjects and substituting them into the non-discrimination area formula (13), we obtain the perceptually uniform distribution of color pairs (c_1, c_2) given by the density

$$p(c_1, c_2) \propto \frac{1}{A_1(c_1 - c_2)A_2} \propto \frac{1}{\sqrt{1 + 0.017(1)d^2 + 0.000018(3)d^4}}, \tag{18}$$

where $d = \|A^{-1}(c_1 - c_2)\|$ and the inverse of the matrix

$$A = \begin{bmatrix} 0.0090(9) & -0.0037(9) \\ 0.005(1) & 0.0070(6) \end{bmatrix} \tag{19}$$

maps the coordinates to the canonical basis[2]; the numbers in parentheses indicate the standard error of the last decimal of the numerical value. In the canonical basis, a color difference of one unit approximately corresponds to a just noticeable difference, assuming that there is no noise from the surrounding colors. Thus, the canonical basis is a linear approximation to a perceptually uniform color space.

Note that the assumptions made for generalizing the results do disregard several possible phenomena such as:

- The effects of color categories (blue, green, etc.) within which the thresholds may be enlarged due to the subject remembering only the color category and not the fine changes within the category (cf. [25]).
- A possible decrease of sensitivity for colors c_1 and c_2 very close to each other due to the small color contrast preventing visual segregation of the squares in the checkerboard (cf. [3]).
- Nonlinearities in the perceptual color space (cf. non-linear uniform color spaces in [5]).

Nevertheless, the assumptions seem justified for a first approximation of a perceptually uniform space of color pairs, considering that the finer details could possibly differ between different observers or between different discrimination criteria.

We do not consider the validation of the results any further, as it is beyond the computational focus of this thesis.

[2]The canonical basis is actually not uniquely determined by the data: for any orthogonal matrix R, $R^TR = I$, the matrix AR in place of A yields a rotated version of the canonical basis, but the same perceptually uniform density $p(c_1, c_2)$.

6 CONCLUSIONS

We have considered computational and mathematical tools for computation in discrete and continuous statistical models. We have found mathematical reformulations useful in reducing the computational complexity of many models. We have exploited symmetries in the models, such as the geometric properties of the symbol distributions in [A], the assumed near decomposability of the distribution in [B], and the linear transformation invariance of the model in [C]. Often regularities manifest themselves as convolutions, which we have computed efficiently as multiplications by the use of generating functions in [A] and by the use of Fast Fourier Transforms in [C]. We have used a Markov Chain as a modeling tool in [A] as well as considered the general toolbox of Markov Chain Monte Carlo methods in [B] and [C].

The main practical result is a Bayesian adaptive estimation procedure for two-dimensional psychometric functions [C], which has been made possible by the recent increases in the available computational power and the algorithmic improvements motivated by [A] and [B].

We have considered the real-world problem of efficiently rendering recognizably unique textures [D]. We have applied the new psychophysical procedure [C] to quantifying certain chromatic crispening effects salient in the perception of these textures. The experimental data provides justification for the ad hoc choices in the design of the rendering algorithm. Finally, we have built a simple Bayesian model of the data and shown how it can be used to define a perceptually uniform space of color pairs.

Author's contribution Here, I report on my role in the work done in cooperation with my supervisor Dr. Tuomas Lukka.

In [A], the symbolic stack machine model and the series of mathematical reformulations yielding the final result were mostly due to me. Dr. Lukka motivated the problem and demonstrated the fundamental properties of the system.

The idea of applying genetic operators in MCMC as well as the first version of the manuscript [B] came from Dr. Lukka. I formalized the idea mathematically, proved the convergence results, implemented the examples, and related it with the relevant literature.

In [C], I formulated the model for two-dimensional psychometric functions, came up with the reformulation of the objective function, discovered the FFT algorithm, implemented the experiment, and did the proofs in the appendix. Dr. Lukka came up with the particle filter MCMC algorithm.

The use of unique background textures in [D] was Dr. Lukka's idea. We worked together on the implementation, during which the idea was considerably reshaped and the often counterintuitive principles were discovered. I implemented and conducted the recognizability experiment.

The ongoing work on quantifying color perception was done together with Dr. Lukka and with input from Prof. Pertti Saariluoma.

REFERENCES

[1] F. Gregory Ashby and James T. Townsend. Varieties of perceptual independence. *Psychological Review*, 93(2):154–179, 1986.

[2] Julian Besag, Peter Green, David Higdon, and Kerrie Mengersen. Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10(1):3–66, 1995.

[3] R. M. Boynton, M. M. Hayhoe, and D. I. A. MacLeod. The gap effect: Chromatic and achromatic visual discrimination as affected by field separation. *Optica Acta*, 24:159–177, 1977.

[4] Nicolas Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3):539–551, 2002.

[5] Colorimetry. CIE Publication 15.2, 2nd Edition, 1986.

[6] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[7] Henryk Fukś. Critical behaviour of number-conserving cellular automata with nonlinear fundamental diagrams. *Journal of Statistical Mechanics: Theory and Experiment*, P07005, 2004.

[8] Henryk Fukś and Nino Boccara. Generalized deterministic traffic rules. *International Journal of Modern Physics C*, 9(1):1–12, 1998.

[9] Dani Gamerman. *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman & Hall, London, 1997.

[10] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, London, 1995.

[11] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, second edition, 1994.

[12] David M. Green and John A. Swets. *Signal Detection Theory and Psychophysics*. Peninsula Press, Los Altos, CA, 1966.

[13] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[14] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

[15] Helena Kadlec and James T. Townsend. Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, 36(3):325–374, 1992.

[16] Stanley A. Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8):1421–1455, 2001.

[17] Leonid L. Kontsevich and Christopher W. Tyler. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16):2729–2737, 1999.

[18] Faming Liang and Wing Hung Wong. Evolutionary Monte Carlo: Applications to C_p model sampling and changepoint problem. *Statistica Sinica*, 10(2):317–342, 2000.

[19] D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

[20] David J.C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

[21] David J.C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992.

[22] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[23] James W. Myers, Kathryn Blackmond Laskey, and Tod S. Levitt. Learning Bayesian networks from incomplete data with stochastic search algorithms. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-1999)*, pages 476–485, 1999.

[24] Louis Narens and Rainer Mausfeld. On the relationship of the psychological and the physical in psychophysics. *Psychological Review*, 99(3):467–479, 1992.

[25] B. A. C. Saunders and J. van Brakel. Are there non-trivial constraints on colour categorization (with discussion). *Behavioral and Brain Sciences*, 20(2):167–228, 1997.

[26] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4):583–639, 2002.

[27] Malcolm J. A. Strens, Mark Bernhardt, and Nicholas Everett. Markov Chain Monte Carlo sampling using direct search optimization. In *Proceedings of the Twentieth International Conference on Machine Learning*, pages 602–609. Morgan Kaufmann, 2002.

[28] Hiroshi Takasaki. Chromatic changes induced by changes in chromaticity of background of constant lightness. *Journal of the Optical Society of America*, 57(1):93–96, 1966.

[29] Cajo J. F. ter Braak. Genetic algorithms and Markov Chain Monte Carlo: Differential Evolution Markov Chain makes Bayesian computing easy. Technical report, Biometris, Wageningen University and Research Centre, April 2004.

[30] Luke Tierney. Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22(4):1701–1762, 1994.

[31] Bernhard Treutwein. Adaptive psychophysical procedures. *Vision Research*, 35(17):2503–2522, 1995.

[32] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.

[33] Günter Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & sons, second edition, 2000.

YHTEENVETO (FINNISH SUMMARY)

Tarkastelemme kahta lähestymistapaa laskentaan tilastollisissa malleissa: täsmällistä laskentaa diskretisoiduilla malleilla ja likimääräistä laskentaa Monte Carlo -simuloinnilla. Esitämme uusia laskennallisia tekniikoita molempiin lähestymistapoihin ja sovellamme niitä psykometristen funktioiden bayesläiseen mukautuvaan mittaamiseen. Päätulos on uusi mittausmenetelmä kaksiulotteisille psykometrisille funktioille. Johdattelemme värinäön mallintamisongelman ja sovellamme uutta mittausmenetelmää siihen. Pohdimme koetuloksia lyhyesti.