On the Computational Role of the Simple Cells in Early Vision

by

Tim R. Pattison, B.E. (Hons.), B.Sc.

Thesis submitted for the degree of

Doctor of Philosophy

The University of Adelaide
Faculty of Engineering
Department of Electrical and Electronic Engineering

October, 1993

Awarded 1994

TABLE OF CONTENTS

Abstract vii

Declaration ix

Acknowledgments xi

List of Figures xiii

List of Tables xv

Glossary xvii

1 Introduction 1

1.1 Background and Motivation 1

1.1.1 The Simple Cell Receptive Field Profile 1

1.1.2 Neural Networks for Sensory Signal Processing 2

1.1.3 Summary 3

1.2 Overview 4

1.3 Intended Audience 5

2 The Gabor Function Model of Simple Cell Receptive Field Profiles 7

2.1 Introduction 7

2.2 Simple Cell Receptive Field Profiles 7

2.2.1 Spatial Receptive Field Profile 7

2.2.2 Spectral RFP 11

2.2.3 Spatial Linearity 12

2.2.4 Spatiotemporal RFP 15

2.2.5 Binocular RFP 18

2.2.6 Summary 20

2.3 Gabor Function Models 20

2.3.1 Spatial RFP 20

2.3.2 Spectral RFP 26

2.3.3 Optimal Joint Localisation 29

2.3.4 Spatiotemporal RFP 33

2.3.5 Binocular RFP 35

2.3.6 Summary 35
2.4 Conclusion 36

3 On the Computational Role of the Simple Cells **37**
3.1 Introduction 37
3.1.1 Bottom-up vs. Top-down 38
3.1.2 Qualified Completeness 39
3.1.3 Filtering and Decomposition 40
3.1.4 Verification of Bottom-up Theories 42
3.1.5 Hierarchical Processing 42
3.1.6 Summary 44
3.2 Feature "Detectors" 44
3.2.1 Nonlinear Detectors 44
3.2.2 Linear Matched Filters 45
3.2.3 Summary 48
3.3 Directional Spatial Derivatives 48
3.3.1 Retino-Cortical Derivative Operators 48
3.3.2 Fractional Derivatives 51
3.3.3 Discriminant Functions 51
3.3.4 Gaussian Derivatives 53
3.3.5 Summary 53
3.4 Spatial Frequency Analysis 54
3.4.1 Introduction 54
3.4.2 The Gabor Expansion 55
3.4.3 Do Simple Cells Perform a Gabor Decomposition? 58
3.4.4 Weyl-Heisenberg Frames 60
3.4.5 Discrete Window Fourier Transform 61
3.4.6 Efficient Coding Through Gabor Expansion 62
3.4.7 Summary 65
3.5 Wavelet-like Analysis 65
3.5.1 Introduction 65
3.5.2 Discrete Wavelet Transform 67
3.5.3 Do Simple Cells Perform a Discrete Wavelet Transform? 69
3.5.4 Gabor "Wavelet" Expansion 70
3.5.5 Efficient Coding Through Wavelet-Like Analysis 71
3.6 Applications of Gabor Functions 72
3.7 Conclusion 73

4 Neural Networks for Non-Orthogonal Image Decomposition **75**
4.1 Introduction 75
4.2 LSE Image Decomposition Problem 76
4.2.1 Matrix Formulation 76
4.2.2 Squared Reconstruction Error (SRE) Minimisation 77
4.2.3 Practical SRE Minimisation 78
4.3 Neural Networks for SRE Minimisation 81
4.3.1 Daugman (1988a) 82
4.3.2 Wang & Yan (1992) 85
4.3.3 Cohen & Shawe-Taylor (1990) 87
4.3.4 Cichocki & Unbehauen (1992) 88
4.4 Neural Networks for Minimum Norm SRE Minimisation 90
4.4.1 Culhane et al. (1989) 90
4.4.2 Yan (1991b) 91
4.4.3 Pece (1992) 93
4.5 Conclusion 93

5 Single-Layered Neural Networks for Decomposition **95**
5.1 Introduction 95
5.2 Pattison (1992) 95
5.3 Additional Leakage Term 98
5.4 Preconditioning 99
5.4.1 Sensitivity to Weight and Derivative Round-Off Errors 99
5.4.2 Diagonal Preconditioning 101
5.4.3 Accelerating Convergence 105
5.5 Examples 108
5.6 Conclusion 109

6 Neural Network for Bound-Constrained Quadratic Optimisation **111**
6.1 Introduction 111
6.2 Unconstrained Quadratic Optimisation 112
6.3 Bound-Constrained Quadratic Optimisation 113
6.4 Equilibrium Point Analysis 117
6.4.1 Multiple Constrained Minima 119
6.5 Convergence Analysis 120
6.6 Related Neural Networks 122
6.6.1 The Hopfield Network 122
6.6.2 Generalised Brain-State-In-A-Box (GBSB) Network 123
6.6.3 Continuous-time GBSB Network 124
6.7 Optimisation Strategy 125

6.7.1 Description & Comparisons 125
6.7.2 Examples 127
6.7.3 Constraint Generalisation 133
6.8 Exponential Stability 134
6.8.1 Theoretical Results 134
6.8.2 Practical Considerations 135
6.9 Preconditioning 137
6.10 Bound Constrained SRE Minimisation 140
6.11 Conclusion 142

7 Do The Simple Cells Perform Image Decomposition? 145
7.1 Introduction 145
7.2 Biological Implementation of Decomposition Networks 146
7.2.1 Plausibility Considerations 146
7.2.2 Corticofugal Feedback 151
7.2.3 Intracortical Feedback 153
7.2.4 Summary 161
7.3 Predicted and Identified Spatiotemporal RFPs 162
7.3.1 Introduction 162
7.3.2 Theoretical RFPs 163
7.3.3 Practical Identification 168
7.3.4 Summary and Discussion 172
7.4 Discussion 173
7.5 Conclusion 174

8 Conclusions 177
8.1 Overview 177
8.2 Summary and Conclusions 177
8.3 Discussion 179
8.4 Contributions 180

A Spatial and Spectral RF Investigations 181
A.1 Spatial RF 181
A.2 Spectral RF 181

B RFP Identification Using Impulses 185

C Gabor Function Applications in Image Coding and Analysis 189

D Stability in Nonlinear Networks **193**

D.1 Boundedness of Solutions of the BCP Network 193

D.2 Convergence Proof for CGBSB Network 193

D.3 Uniform Delay on Lateral Connections of BCP Network 195

E Hebbian Weight Development **197**

Bibliography **199**

ABSTRACT

The simple cells in feline and primate primary visual cortex are involved in the coding and early processing of spatiotemporal information acquired binocularly from the visual field. Each simple cell can be viewed as an approximately linear device characterised by its receptive field profile (RFP), a spatially reversed version of its spatiotemporal impulse response function.

The Gabor function model of the simple cell RFP is evaluated, and the recent controversy concerning the relevance to early vision of its achievement of the lower bound on joint spatial and spectral spread dictated by the Weyl-Heisenberg Uncertainty Principle is illuminated. In an investigation of the multi-dimensional signal processing performed by the simple cells, image processing and coding schemes which might explain the observed variety of simple cell spatial RFPs are reviewed. These schemes are classified into the categories of *filtering* and *decomposition*, according to whether the RFP is used as the kernel of a spatial filter, or as an expansion function whose coefficent is to be calculated for the visual image.

Artificial neural networks (ANNs) which find the least-squares solution to the set of linear equations posed by the image decomposition problem are critically reviewed, and a single-layered, linear recurrent ANN is proposed for this task. The linear neural activation function used by this network is then replaced by a more biologically plausible, piece-wise linear, saturating nonlinearity, and the resultant globally stable network is shown to effect the optimisation of more general (semi)definite quadratic forms subject to bound constraints on the optimisation variables. Although biologically plausible, these networks, when used as models of simple cell processing, are found to predict simple cell spatiotemporal RFPs whose spatial component differs in general from the chosen expansion functions. It is concluded that the simple cell spatial RFPs are *not* used as visual expansion functions, but rather as the kernels of (possibly position-dependent) spatial filters, as is suggested by their definition.

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University or other tertiary institution, and to the best of the author's knowledge and belief contains no material previously published or written by another person, except where due reference has been made in the text.

Should the thesis be accepted for the award of the Degree, the author hereby consents to this copy, when deposited in the University Library, being made available for loan and photocopying.

Signature

Date October 15, 1993

ACKNOWLEDGMENTS

I gratefully acknowledge the assistance and support of my supervisor, Professor R.E. Bogner.

I would particularly like to thank Dr Abdesselam ("Salim") Bouzerdoum for his assistance, patience, encouragement and friendship, all of which have been vital to the completion of this thesis.

My thanks go also to Dr B.R. Davis and Dr R. Merchant for many helpful discussions, and Professors W. Levick and T. Caelli for their encouragement.

Most importantly, I thank my fiancée Cathy for her patience, encouragement and companionship which have seen me through the darkest hours.

This work was supported by a Postgraduate Research Fellowship from the Communications Division (CD) of the Electronics Research Laboratory (ERL) of the Defence Science and Technology Organisation (DSTO), Salisbury, South Australia.

Support was also provided by the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), The Levels, Pooraka, South Australia, of which the author is a member.

T.R.P

LIST OF FIGURES

2.1 Alternative descriptions of the spatial receptive field 9
2.2 Spatiotemporal RFPs 17
2.3 2D GRGF and its Fourier Transform magnitude 23

3.1 Matched filter kernels for edges and bars 46
3.2 Derivative masks from 1D Lagrange interpolating polynomials 50
3.3 Derivative masks from 1D fractional discriminant functions 52
3.4 Phase-space lattice for Gabor expansion 56
3.5 Spatial and spectral lattices for 2D Gabor expansion 58
3.6 Spatial and spectral lattices for wavelet-like 2D representation 67
3.7 Phase-space lattice for discrete wavelet transform 68

4.1 Network proposed by Daugman (1988a) 83
4.2 Corrected version of Daugman's network 84
4.3 Network proposed by Wang & Yan (1992) 86
4.4 Generative Back-Propagation Network 87
4.5 Network proposed by Cichocki & Unbehauen (1992) 88
4.6 Condensed version of network proposed by Cichocki & Unbehauen (1992) 89
4.7 Generalisation of network proposed by Culhane et al. (1989) for DFT and DHT 91
4.8 Network proposed by Yan (1991a) 92
4.9 Network proposed by Pece (1992) 94

5.1 Relaxation network proposed by Pattison (1992) 95
5.2 Comparison of symmetrical and asymmetrical preconditioning 104
5.3 Effect of symmetric diagonal preconditioning on convergence rate 107
5.4 Sequence of reconstructed images for unconstrained network 109

6.1 Linear recurrent network for unconstrained quadratic optimisation 113
6.2 Bound constraint enforcement functions 114
6.3 Neuron models for bound-constrained quadratic optimisation 115
6.4 Bound-constrained steepest descent — global minimum inside boundaries 129
6.5 Bound-constrained steepest descent — global minimum on boundary . . 129
6.6 Effect of preconditioning on the search direction 131
6.7 Comparison of GBSB, CGBSB and BCP optimisation strategies 132

6.8 Accelerated convergence through preconditioning 139
6.9 Sequence of reconstructed images for BCP network 141
6.10 Sequence of reconstructed images for BCP network 141
6.11 Sequence of reconstructed images for BCP network 142

LIST OF TABLES

6.1 Parameters for Example 6.1 120
6.2 Parameters for Example 6.2 128
6.3 Network equilibria for problems in Table 6.2 128
6.4 Parameters for Example 6.3 130
6.5 Network equilibria for Example 6.3 130
6.6 Parameters for Example 6.4 131
6.7 Network equilibria for Example 6.4 132
6.8 Effects of preconditioning in Example 6.3 139

A.1 Spatial analyses of simple cell RFPs 182
A.2 Spatial Fourier analyses of simple cell RFPs 183

C.1 Gabor function applications. 190
C.2 Gabor function selection. 191
C.3 Systematic parameter variation by canonical representation schemes . . . 192

GLOSSARY

$\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \ldots$	Vectors are denoted by bold typeface
$\mathcal{A}, \mathcal{B}, \mathcal{C} \ldots$	Sets are denoted by caligraphic typeface, except as noted below
$\mathbb{R}, \mathbb{Z}, \mathbb{C}$	The sets of real, integer and complex scalars
$\mathbb{R}^n, \mathbb{Z}^n, \mathbb{C}^n$	The sets of real, integer and complex n-dimensional vectors
$\mathbb{R}_+^n, \mathbb{Z}_+^n$	The sets of positive real and integer n-dimensional vectors
$f: \mathcal{A} \rightarrow \mathcal{B}$	Function f maps $x \in \mathcal{A}$ onto $f(x) \in \mathcal{B}$
$f: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{C}$	Function f maps $x \in \mathcal{A}$ and $y \in \mathcal{B}$ onto $f(x, y) \in \mathcal{C}$
$L^2(\mathbb{R}^n)$	The set of square-integrable functions $f: \mathbb{R}^n \rightarrow \mathbb{C}$
$(a, b]$	The set of real numbers $x: a < x \leq b$
$[a, b] \subset \mathbb{Z}$	The set of integers $i: a \leq i \leq b$
$\{a, b, c\}$	The set with elements a, b and c
$[a, b, c]^T$	The column vector with elements a, b and c
$\mathcal{F}\{\cdot\}$	The Fourier transform
$\int_{\mathcal{A}} \cdot \, d\boldsymbol{x}$	Shorthand for $\int \ldots \int_{\mathcal{A}} \cdot \, dx_1 \ldots dx_n$

ANN	Artificial Neural Network
BCP	Bound Constraint Projection network
BLUE	Best Linear Unbiassed Estimator
CGBSB	Continuous Generalised Brain-State-in-a-Box network
DFT	Discrete Fourier Transform
DHT	Discrete Hartley Transform
DWT	Discrete Wavelet Transform
DWFT	Discrete Window Fourier Transform
FSWF	Feedforward Spatial Weighting Function
GABA	Gamma-aminobutyric Acid
GE	Gabor Expansion
GRGF	Generalised Real-valued Gabor Function
GBSB	Generalised Brain-State-in-a-Box network
LGN	Lateral Geniculate Nucleus (of the Thalamus)
LIP	Lagrange Interpolating Polynomial
LSE	Least Squared Error
LWF	Line Weighting Function
RANN	Recurrent Artificial Neural Network

RF	Receptive Field
RFP	Receptive Field Profile
RGF	Real-valued Gabor Function
SRE	Squared Reconstruction Error
V1	Primary Visual Cortex (Striate Cortex, Brodmann's Area 17)
1D,2D	One/Two Dimension(al)

Chapter I

INTRODUCTION

1.1 Background and Motivation

1.1.1 The Simple Cell Receptive Field Profile

With the exception of the retina, the primary visual cortex has been perhaps the most intensively studied area of the mammalian early visual system. Its functional proximity to the visual input and its accessibility in the occipital lobe have made it a particularly attractive target for the visual electrophysiologist. Much of our knowledge about the electrophysiology of the striate cortex, as the primary visual cortex is alternatively known, comes from experiments on feline and primate subjects. Although some differences do exist between their respective visual cortices (Crawford et al., 1990), the presence in both species of cells exhibiting similar functional characteristics suggests that these characteristics contribute to the solution of generic visual problems which are common to the respective visual environments of these somewhat different animals. Single-cell recordings from the primary visual cortex have shown that neurons in this region are involved in the encoding and early processing of spatial, temporal and stereoscopic information. Their implication in fine spatial vision (Crawford et al., 1990), in particular, and their consequent potential as a source of biological inspiration for image preprocessing in the field of pattern recognition, have made them especially interesting to the machine vision community.

The *simple cells* were so named because unlike the responses of the remainder of the primary visual cortical cells from which the pioneering team of Hubel & Wiesel (1962) recorded, the responses of these cells to spatially extended visual stimuli could be largely predicted from their responses to small circular spots of light at various positions in the visual field. This observation is at least superficially suggestive of the mathematical property of *superposition*, which is characteristic of linear systems. Linear systems are in general more mathematically tractable, and consequently better understood, than the more general class of nonlinear systems. It is perhaps, therefore, the promise of a predominantly linear characterisation of the simple cells which has motivated a number of visual electrophysiologists and computational neuroscientists to concentrate their efforts on the investigation and modelling of the multi-dimensional (i.e. spatiotemporal) signal processing performed by these cells.

The *receptive field profile* (RFP) of a visual cortical cell is a function which describes the linear weighting applied by that cell to the visual stimulus to produce the observed response. More formally, the RFP is the kernel of a linear integral operator which, when applied to the visual stimulus over the domain of its definition, yields the output of that cell. As this definition suggests, the RFP is closely related to the impulse response function of the cell; the distinction lies simply in the conceptually convenient sign-reversal of the spatial variable. Its domain of definition corresponds to that over which the stimulus is described; for completeness this should be the spatiotemporal domain, although frequently the temporal dimension, and occasionally the second spatial dimension, are omitted for convenience. The stimulus position is expressed in eye-centred spherical coordinates (with the range dimension omitted). Consequently a cell with binocular input has one RFP for each eye; these will be collectively referred to as the *binocular* RFP of the cell. It will also often prove convenient to describe an RFP in the Fourier frequency domain, and, by extension, to refer to this as the *spectral* RFP.

The accuracy with which the response of a simple cell can be reproduced by the linear model of that cell's processing inherent in the definition of an RFP is dependent on the approximation involved in the neglect of the nonlinear terms, if any, which contribute to this processing. A qualitative investigation of the extent of this approximation is presented in Section 2.2, along with an examination of the validity or otherwise of the omission of one or more of the stimulus dimensions.

1.1.2 Neural Networks for Sensory Signal Processing

The response of a simple cell to a visual stimulus is in fact the cumulative result not only of the processing performed by neurons in the afferent retino-geniculo-cortical visual pathway up to and including that simple cell, but also of intra- and possibly extra-striate feedback influences on that cell. Evidence for intra-striate feedback, in particular, onto the simple cells is reviewed in Chapter 7. A natural framework within which to model the neural processing which contributes to the simple cell output, and hence also its RFP, is provided by the field of *artificial neural networks* (ANNs), and in particular *recurrent* artificial neural networks (RANNs). Inspired at least initially by their biological counterparts, ANNs are densely interconnected networks of computationally simple and relatively slow processors called *artificial neurons*. These networks are characterised chiefly by their fine-grained parallel distributed processing (Rumelhart & McClelland, 1986a; Rumelhart & McClelland, 1986b), in which the constituent artificial neurons operate largely in parallel and the computational load on a network is finely distributed between its processors. RANNs are simply ANNs in which the neural interconnection patterns give rise to feedback loops, which are generally thought to exist in the striate cortex (Douglas & Martin, 1991).

Artificial neural networks have received considerable attention in recent years from the field of signal processing (Juang et al., 1991). ANNs provide a generic architecture suited to the fine-grained parallel implementation of a large number of conventional signal processing algorithms. This property, combined with their architectural similarity to the neural substrate on which they are hypothetically implemented, makes them well suited to the modelling of the multi-dimensional signal processing performed by early sensory systems. The addition of a learning rule by which the interconnection strengths of an ANN are updated furthermore admits the neural implementation of adaptive signal processing algorithms.

The ANN framework not only offers potential neural realisations of models of early sensory signal processing, but also, by facilitating the consideration of practical constraints on the biological implementation of such models, assists in their formulation and refinement. For example, attempts to understand the limitations on the neural transmission of information in such networks have resulted in theories concerning the transformation of sensory messages to reduce their statistical redundancy (Attneave, 1954; Barlow, 1959; Barlow, 1961) and thereby achieve an efficient coding of the sensory input. Through the introduction of an unsupervised learning rule, the transformation effected by a redundancy-reducing network can also be made adaptive to the changing statistics of the sensory environment (Linsker, 1988; Linsker, 1989; Linsker, 1990; Földiák, 1989; Földiák, 1992), a property which would be essential for the survival of the organism. The principle of redundancy reduction has proven influential not only in subsequent theories of biological sensory coding (Atick & Redlich, 1990b; Atick & Redlich, 1991; Atick & Redlich, 1992), but also in the fields of image coding and compression.

1.1.3 Summary

Recurrent artificial neural networks (RANNs) have been chosen as a useful tool for the modelling and investigation of the multi-dimensional signal processing performed by the simple cells in the mammalian early visual system. The simple cells have been selected over other cortical cells because of their approximate linearity. The linear component of simple cell processing of the visual stimulus is characterised by the receptive field profile (RFP). The primary visual cortex has been chosen because of the comparative abundance of single-cell electrophysiological recordings from this area, which are needed to infer the properties of the RFPs of individual cells[1], and the controversiality of the existence of simple cells in higher cortical areas (Henry, 1993).

[1]In contrast, inferences drawn from psychophysical and area-level electrophysiological observations invariably concern large multi-cell "units" such as psychophysical channels or cortical columns.

1.2 Overview

This thesis reviews models of the simple cell RFP and its variation over the simple cell population, and uses artificial neural networks to investigate the multi-dimensional signal processing role of the RFP in the formation of a cortical representation of the visual image.

A critical review of simple cell RFP models is presented in Section 2.3. This review centres around the Gabor function model and a simple, real-valued generalisation thereof referred to here as the *generalised real-valued Gabor function* (GRGF); in justification of this approach, it will be argued that the latter is the only model proposed to date which exhibits sufficient degrees of freedom to fully describe the simple cell RFP[2]. Alternative models are discussed primarily where they address shortcomings identified in the GRGF model. The Gabor function model has attracted considerable attention since its proposal by Daugman (1980) and Marčelja (1980), most of which has focused on its achievement of the lower bound set by the Weyl-Heisenberg Uncertainty Principle (Weyl, 1932) on the joint spread or "uncertainty" of a function in the spatial and spatial frequency domains. Reviewed in Section 2.3, the recent empirical success of the GRGF model of the simple cell RFP has bolstered suggestions by Marčelja (1980) and Daugman (1985) that this property indicates a role for the simple cells in the joint spatial and spectral localisation of the stimulus (although see Wechsler (1990)). The electrophysiological foundations of this hypothesis and the recent controversy it has attracted are criticised in Section 2.3.3.

Variation of the spatial RFP over the simple cell population is addressed in Chapter 3. Theories of simple cell processing which postulate a systematic variation of various parameters of the spatial RFP model over the population are examined for their consistency with the observed variety of simple cell RFPs. Consideration is restricted to those schemes which not only employ a suitably realistic model of the simple cell RFP, but also make a serious attempt to account for the variety of spatial RFPs observed. Among the schemes addressed are the Gabor and Wavelet Expansions, and the Wavelet and Window Fourier Transforms. A brief overview of recent machine vision and image processing applications of schemes involving the GRGF RFP model is also presented.

Theories concerning the variation of RFP parameters across the population, and which assume a linear model of simple cell processing, are divided in Chapter 3 into the two classes of *filtering* and *decomposition*. The filtering hypothesis of simple cell processing, to which theories in the former category subscribe, postulates that the spatial RFP acts as the kernel of a spatial filter which is applied to the image. In contrast, the decomposition hypothesis of simple cell processing, to which theories in the latter category subscribe, postulates that the spatial RFP acts as an expansion function whose

[2]This observation was made originally by Jones (1991) in justification of the decision by Jones & Palmer (1987a) to fit only the GRGF model to their electrophysiologically identified simple cell RFPs. It will become evident, however, that this observation remains current.

coefficient, in the corresponding expansion of the image, is signalled by the simple cell. By viewing the expansion functions as features to be detected, and the coefficients as measures of their relative presences in the visual image, the decomposition hypothesis can be seen to be closely related to the earlier feature-detection hypothesis of simple cell processing, which is discussed in Section 3.2. Although it is argued that the filtering hypothesis appears more consistent with the definition of the simple cell spatial RFP, recurrent artificial neural networks are reviewed in Chapter 4 and developed in Chapters 5 and 6 which, if implemented in the early visual system, could calculate the coefficients required by decomposition theories. The biological plausibility of these networks is assessed in Chapter 7, and the spatiotemporal simple cell RFPs which they predict are analysed for their consistency with the chosen expansion functions.

Although primarily motivated by the desire to produce biologically plausible models of image decomposition by the simple cells, the review and development in Chapters 4–6 of RANNs for the minimisation of the squared reconstruction error (SRE) has important benefits for the least-squares solution of linear systems of equations. These include: the demonstration that several RANNs which were previously though to provide only approximate solutions to the SRE minimisation problem in fact provide exact solutions to a regularised form of the problem; the development of a diagonal preconditioning strategy to improve the stability of a proposed single-layered linear RANN in the presence of weight implementation and signal propagation errors; the development of a single-layered nonlinear RANN capable of minimising a positive (semi)definite quadratic error function subject to bound constraints on the optimisation variables, a problem which arises in several engineering disciplines (Moré & Toraldo, 1991); and the recognition that this single-layered network may serve as a biologically plausible implementation of various resistive networks used in the discrete-space solution of other regularised problems in early vision.

1.3 Intended Audience

It may already have become apparent that the intended reader is expected to be conversant with elementary neuroanatomy, neurophysiology and visual electrophysiology. He or she should also be competent in linear algebra and the theory of dynamical systems, although a number of the concepts relevant to the latter will be introduced as they are needed. In order to head off criticism from experts in either field, an attempt has been made to maintain a certain rigour in the treatment of both the biological and mathematical content. However this could only be achieved at the expense of its readability to practitioners of the complementary discipline. The author therefore requests the reader's indulgence in acknowledging, in cases where one or the other appears to be deficient, the inevitable trade-off between rigour and readability faced in the presentation of this work.

Chapter II

THE GABOR FUNCTION MODEL OF SIMPLE CELL RECEPTIVE FIELD PROFILES

2.1 Introduction

This chapter opens with a brief introduction to the concept of a "receptive field" (RF) and a summary of the receptive field characteristics which define a "simple" cell. A linear characterisation of the simple cell receptive field known as the "receptive field profile" (RFP) is then presented, and the limitations of such a characterisation examined. The extension of the RFP to account for the temporal and binocular behaviour of the simple cell RF in addition to its spatial properties is presented, and the validity or otherwise of the consideration of the spatial component of the receptive field profile in isolation from its temporal and binocular components is discussed. Finally, subject to the reservations which arise from this discussion, the "generalised real-valued Gabor function" (GRGF) defined in this chapter is investigated as a model of the simple cell spatial RFP. Special emphasis is placed on an analysis of the elegant series of identification experiments performed by Jones & Palmer (1987b; 1987; 1987a), and their subsequent evaluation of the GRGF model.

2.2 Simple Cell Receptive Field Profiles

2.2.1 Spatial Receptive Field Profile

The *receptive field* (RF) of a neuron whose axon forms part of the vertebrate optic nerve was defined by Hartline (1938) as that region of the retina whose illumination would elicit a response in the axon. In a series of experiments which pioneered the electrophysiological exploration of the visual cortex, Hubel & Wiesel (1959; 1962; 1965; 1968) generalised this term to cover visual cortical neurons and to refer more generally to that region of the visual field over which a stimulus could influence the firing of the neuron (Bishop & Henry, 1972). Using stationary flashing stimuli consisting of positive contrast spots or bars on a background of uniform luminance, and recording extracellularly with a microelectrode from single neurons in feline and later primate (monkey) primary visual cortex (V1), they determined the spatial *organisation* of the receptive field of each recorded neuron by mapping out the sign of the change in firing rate (increase or decrease) following

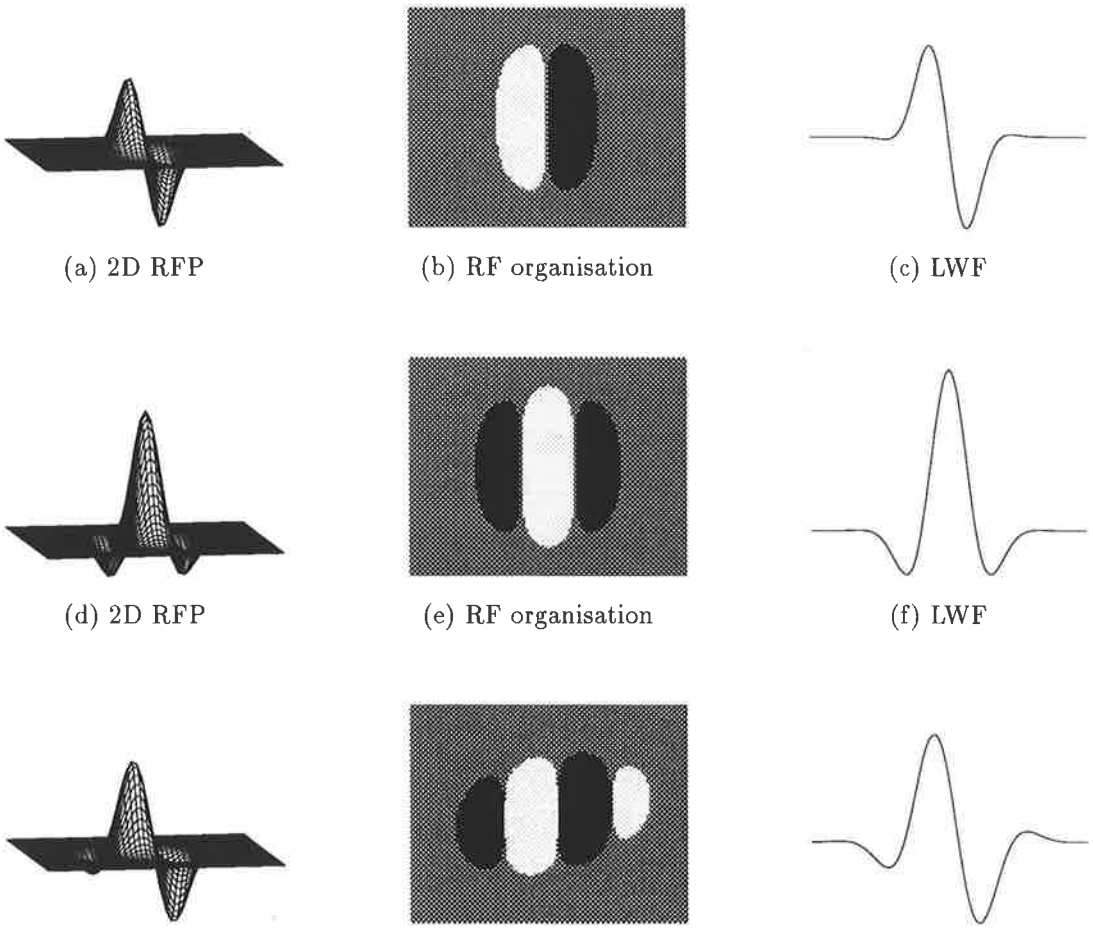
stimulus onset as a function of the two dimensional (2D) spatial position of the stimulus. Visual cortical cells were classified into one of the three categories *simple*, *complex* and *hypercomplex* on the basis of characteristic properties of their receptive fields, which for the simple cells were (Hubel & Wiesel, 1962):

- RF organisation consisting of distinct excitatory and inhibitory subfields.
- Summation of responses to stimuli presented concurrently in the same subfield.
- Antagonism between excitatory and inhibitory subfields.
- Predictability of responses to stationary or moving stimuli from the RF organisation.

Henry (1977) proposed a derivative RF taxonomy in which these defining characteristics were refined for stationary stimuli, and extended to characterise responses to both moving edges and bars of increasing elongation. Notwithstanding the differences between the resultant *S-cell* classification and its simple cell precursor[1], the term *simple cell* will in accordance with common practice be used interchangably to refer to either cell category.

Hubel & Wiesel found that the RF of each simple cell was organised into parallel elongated excitatory and inhibitory *subfields* alternating in the direction perpendicular to their axis of elongation and separated by straight-line borders, as shown in Figures 2.1(b), (e) and (h). A subfield was designated *excitatory* — or later *on-excitatory* — if the onset of a bright stimulus in that region produced an increase in the firing rate of the neuron, and *inhibitory* if stimulus onset produced a suppression of the firing rate (*on-inhibitory*) or stimulus offset produced an elevation of the firing rate (*off-excitatory*). In subsequent investigations — the relevant details of which are summarised in Appendix A.1 — the resultant map of RF organisation was augmented to include a measure of the efficacy of a standard stimulus in each position. In two dimensions, this augmented map of stimulus efficacy as a function of visual angle — examples of which are shown in Figures 2.1(a), (d) and (g) — is known as the spatial *receptive field profile* (RFP) of the cell. However for practical reasons, the integral of the RFP along the direction of subfield elongation — known as the 1D *line weighting function* (LWF) — has been preferred by many investigators. The LWF is obtained by recording, for each position along the axis of subfield alternation, the strength of the neuron's response to a long contrasting bar oriented parallel to the axis of subfield elongation. Three examples of simple cell LWFs are illustrated in Figures 2.1(c), (f) and (i). Simple cells with between 1 and 8 RF subfields have been reported, although those with 2 and 3 subfields — such as those

[1]The interested reader is referred to Henry (1985) and White (1989, p. 115-121) for further discussion on the relationship between these two classifications.



(a) 2D RFP (b) RF organisation (c) LWF

(d) 2D RFP (e) RF organisation (f) LWF

(g) 2D RFP (h) RF organisation (i) LWF

Figure 2.1: The 2D RFPs of typical simple cells (after Jones & Palmer (1987b)) having (a) 2, (d) 3 and (g) 4 subfields respectively. The corresponding subfield organisations are illustrated in (b), (e) and (h) respectively, with inhibitory regions shown dark, and excitatory regions bright. Corresponding line weighting functions are shown in (c), (f) and (i) respectively.

illustrated in Figures 2.1(b) and (e) — appear to constitute the majority — in excess of 70% — of recorded simple cells (Glezer et al., 1989)[2].

Only recently have full two dimensional RFP analyses been undertaken. Employing the more sophisticated *reverse correlation* system identification technique and small rectangular stimuli more closely approximating the ideal spatial impulses[3], Jones & Palmer (1987b) correlated the number of output spikes emitted by a given simple cell in successive 50ms time bins with the position of the spatially localised stimulus on a 16×16 grid spanning the cell's spatial RFP. At the beginning of each 50ms interval, a new stimulus was chosen at random from the 512 possible stimuli (positive or negative contrast in one of 256 grid positions). The correlation of the response with a given stimulus was continued over a number of subsequent stimulus presentation intervals, and the resultant *correlogram* averaged over repeated presentations of the same stimulus. Separate correlograms were compiled for positive and negative contrast stimuli, and then subtracted to produce a final 2D RFP for each post-stimulus time bin. In accordance with estimates of the order of 50ms (Ikeda & Wright, 1975b) to 60ms (Hamilton et al., 1989) for the latency of simple cell response to retinal stimulation, the RFP for the 50-100ms time bin was invariably the most pronounced, but was otherwise qualitatively similar to those for subsequent time bins[4].

Their subsequent mathematical analysis of the spatial properties of the resultant RFPs revealed details which invalidate a number of pre-existing generic models of the 2D simple cell RFP. Notably, while most contemporary models — and indeed some proposed since then — relied heavily on precise odd and even symmetry of the RFP, earlier findings based on LWFs that the majority of simple cell RFPs are in fact *asymmetric* (Kulikowski et al., 1980; Field & Tolhurst, 1986) were supported by the 2D analysis of Jones & Palmer (1987b). Furthermore their finding that Cartesian separability[5] of the RFP was clearly violated in a number of cases called into question the common approach of modelling the LWF and extending the model to 2D by multiplying the 1D model by a window — such as a Gaussian — in the orthogonal direction.

These shortcomings were addressed by Jones & Palmer (1987b) using a generalised form of the 2D *real-valued Gabor function* (RGF) model having 9 free parameters. A χ^2

[2]It is difficult however to draw any firm conclusions from these or any other electrophysiological findings about relative frequencies in the population. Sampling bias (Robson, 1983) may result from — among other things — a tendency of microelectrode techniques to record preferentially from larger cells and those with certain geometries (Anderson et al., 1990, p. 215), and from a preponderance of recordings from cells in particular cortical layers if RF properties are unevenly distributed between the layers.

[3]Spatial impulses are used in the identification of the first order Volterra kernel (Schetzen, 1980).

[4]although see McLean & Palmer (1989), and the discussion thereof in Section 2.2.4

[5]A function $f(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{R}^n$ is said to be *Cartesian separable* if there exist functions $f_i(x_i) \quad i \in \{1 \ldots n\} \subset \mathbb{Z}_+$ such that $f(\boldsymbol{x}) = \prod_{i=1}^{n} f_i(x_i)$. In the present case $n = 2$, $\{f_1, f_2\}$ are functions of the independent spatial variables $\{x_1, x_2\}$, and the function $f(\boldsymbol{x}) = f_1(x_1) f_2(x_2)$ is Cartesian separable.

test on the residual error — assuming a Gaussian spatial noise model — after a least-squares fit using the generalised RGF (GRGF) model showed that this model captures essentially all the necessary degrees of freedom of the simple cell RFP. Given by the real and imaginary parts of a Gabor function — the product of a Gaussian and a sinusoid — the RGF was proposed independently in 1D by Marčelja (1980) and in 2D by Daugman (1980) as a model of the simple cell LWF and RFP respectively. The optimal joint localisation of the 2D Gabor function in the spatial and spatial-frequency domains (Daugman, 1985) forms the foundation for its recent popularity in computational theories of vision. Following the present introduction to simple cell receptive fields, which is intended to be as far as possible model-free, the GRGF and rival models of the simple cell RFP are compared in Section 2.3 on the basis of their ability to account for the above and other experimental results. The relevance of the joint localisation property to biological and computational vision is critically evaluated in Section 2.3.3.

2.2.2 Spectral RFP

Several years after the early experiments of Hubel & Wiesel, Campbell & Robson (1968) showed that for low to moderate contrasts the detectability by a human subject of a spatial grating having a sine, square, rectangular or saw-tooth wave luminance profile was determined solely by the amplitude of the fundamental spatial Fourier component of the grating, and that the non-sinusoidal gratings could not be distinguished from sinewave gratings until their contrast was sufficient to cause the higher Fourier components to exceed apparently independent thresholds. These results suggested the presence in the visual system of "linearly operating independent mechanisms selectively sensitive to limited ranges of spatial frequencies". Blakemore & Campbell (1969) showed that these psychophysical spatial frequency *channels* were also selective for stimulus orientation, suggesting an analysis of the visual image in terms of its constituent *2D* spatial frequencies. A possible cortical locus of these channels was soon identified by Campbell et al. (1969), who showed using grating stimuli that cells in feline primary visual cortex were selectively sensitive to a band of spatial frequencies as well as to the orientation of the grating.

The band-pass tuning properties of the response or contrast sensitivity function of the simple cells, in particular, in both primate and feline primary visual cortex for spatial frequency magnitude, orientation and phase (usually with the other two stimulus variables held constant) have since been elaborated by a number of researchers. The relevant features of a representative sample of such experiments are summarised in Appendix A.2. However, the spatial Fourier harmonics present in non-sinusoidal stimuli — such as bars, edges or square-wave gratings — several of which may lie within the spatial frequency

(magnitude) pass-band of the cell, tend to exaggerate the apparent low-frequency response or sensitivity of the cell (Pollen & Ronner, 1982), leading to reports that simple cells are for example much more narrowly tuned for spatial frequency than for the width of a single bar (Albrecht et al., 1980). This phenomenon could influence not only the magnitude tuning curves obtained using such stimuli, but also those for orientation, since the orientation tuning for each of the stimulus harmonics could potentially differ in both bandwidth and optimal orientation (Daugman, 1983). However, with the exception of the investigation by Jones et al. (1987), magnitude and orientation tuning curves have generally been obtained only at the optimal setting of the other parameter. As noted by Daugman (1980), extrapolation of these results to non-optimal settings requires the assumption of polar separability of the 2D *spectral RFP*, defined by analogy with the spatial RFP as a plot of the amplitude of the fundamental Fourier component of the temporal response to a drifting or temporally modulated sinewave grating against the 2D spatial frequency (magnitude and orientation) of the grating.

To examine the 2D spectral RFP of the simple cells whilst avoiding the assumption of polar separability, Jones et al. (1987) used the method inherent in the above definition, with stimuli drawn from an ensemble of drifting sinusoidal gratings in which spatial frequency was distributed evenly over a 16×16 approximately Cartesian grid spanning the cell's responsive range. The majority of the resultant spectral RFPs were markedly polar inseparable, invalidating previous independent investigations of spatial frequency magnitude and orientation cited above. Their results were found to support a model of the simple cell spectral RFP based on the Fourier transform of a GRGF, as would be expected — given the good fit provided by the GRGF to the spatial RFP — if the simple cell could be treated as a linear device. The validity of this linearity hypothesis is examined in the following section.

2.2.3 Spatial Linearity

Movshon et al. (1978b) initiated a reconciliation between previous spatial and spectral characterisations of the simple cells by demonstrating good qualitative agreement between the line weighting function of each simple cell as predicted by the inverse Fourier transform of the spectral magnitude tuning curve — assuming odd or even spatial symmetry in the absence of Fourier phase measurements — for an optimally oriented sinusoidal grating stimulus, and that determined directly using stationary bar stimuli of the same orientation. Experimentally derived spatial phase information was later incorporated by Andrews & Pollen (1979), who again showed qualitative agreement between the predicted and experimental line weighting functions, with the exception that additional subfields beyond the measured spatial RFP could be inferred from the spectral data. Using bars, edges and gratings drifting at or near the optimal stimulus velocity, Glezer et al. (1980)

and later Kulikowski & Bishop (1981b) — using experimentally determined and estimated phase respectively — confirmed this agreement in the spatial domain, and found similar agreement in the spectral domain. Although clearly open to criticism concerning their presumption of polar separability of the spectral RFP, these results provide preliminary evidence for predominantly linear spatial behaviour in the simple cells.

Deferring consideration of the temporal and binocular behaviour of simple cells until Sections 2.2.4 and 2.2.5 respectively, an idealised monocularly driven simple cell is spatially linear if its output $r(t)$ can be expressed mathematically as[6]

$$r(t) = \int_{\mathcal{V}} w(\boldsymbol{x}) s(\boldsymbol{x}, t)\, d\boldsymbol{x} \tag{2.1}$$

where $s : \mathcal{V} \times \mathbb{R} \to \mathbb{R}$ is the stimulus contrast as a function of both position $\boldsymbol{x} \in \mathbb{R}^2$ — measured as a visual angle from the optical axis — in the visual field $\mathcal{V} \subset \mathbb{R}^2$, and time $t \in \mathbb{R}$, and $w : \mathbb{R}^2 \to \mathbb{R}$ is the 2D spatial RFP of the cell[7]. However, the response of a simple cell to a drifting sinewave grating approximates a half-wave rectified sinewave (see e.g. Henry (1985)), an observation which is strongly suggestive of approximately linear spatial summation over the RFP followed by the application of a threshold nonlinearity. Such a model is commonly used to explain experimental observations (see e.g. Jones et al. (1987)). A nonlinearity $f : \mathbb{R} \to \mathbb{R}$ of this type can be incorporated into (2.1) to yield

$$r(t) = f\left(\int_{\mathcal{V}} w(\boldsymbol{x}) s(\boldsymbol{x}, t)\, d\boldsymbol{x} \right)$$

In Appendix B, it is shown that for the case of the ideal halfwave rectification function this nonlinearity can be rendered transparent by an idealisation of the reverse correlation technique of Jones & Palmer (1987a) described in Section 2.2.1, revealing the underlying spatial behaviour of the cell. Another type of nonlinearity which could be exhibited within a simple cell RF is a nonlinear function $p : \mathbb{R} \to \mathbb{R}$ prior to spatial summation such that (neglecting for the moment the rectification nonlinearity f)

$$r(t) = \int_{\mathcal{V}} w(\boldsymbol{x}) p\left(s(\boldsymbol{x}, t)\right)\, d\boldsymbol{x}$$

By subtracting the negative contrast ("dark") correlogram from that obtained using positive contrast ("light") stimuli, the reverse correlation technique also eliminates the effects of spontaneous activity and any even-order terms in the Taylor series expansion of p, which produce the same response for light and dark stimuli (McLean & Palmer, 1989). The experimental approximation of this idealised reverse correlation technique may not however completely eliminate the influences of the even-order terms of the rectification

[6] Bold face lower-case type is henceforth used to denote vectors $\boldsymbol{v} \in \mathbb{R}^n$ and the notation $\int_{\mathcal{V}} d\boldsymbol{v}$ is shorthand for the component-wise integration $\int \cdots \int_{\mathcal{V}} dv_1 \cdots dv_n$. In (2.1) for example, where $n = 2$, $\boldsymbol{v}$ is the visual angle $\boldsymbol{x} \in \mathbb{R}^2$.

[7] The reader unfamiliar with the functional notation used in this exposition is referred to the brief explanation provided in the glossary.

nonlinearity, in addition to which the real threshold *contrast response function*[8] (Albrecht & Hamilton, 1982) f of the cell may — contrary to the assumption made in Appendix B — contain finite odd-order terms which would influence the obtained RFP.

Nevertheless, after fitting the RGF model[9] to the simple cell 2D RFP obtained using this reverse correlation technique and the corresponding 2D spectral RFP obtained in the manner described above, Jones & Palmer (1987a) found reasonable quantitative agreement between the 2D spatial and spectral receptive field characterisations[10]. Although arguably open to criticism on the basis of the presumption of a particular RFP model — despite rigorous statistical testing of the residual (Jones & Palmer, 1987a) — this observation rules out the possibility of significant higher odd-order terms in Taylor series expansions of both the nonlinearity p prior to spatial summation and the rectification nonlinearity f. However, since it does not rule out even-order terms, it provides only qualified support for the tentative conclusion of spatial linearity drawn from the earlier 1D comparisons of spatial and spectral RFPs.

By way of an independent test of the spatial linearity of the simple cells, Tolhurst & Dean (1987) investigated the applicability of the *Principle of Superposition*, which states that for a linear system the response to the simultaneous presentation of two stimuli should be the sum of the individual responses to the stimuli presented separately. The in- and counter-phase sinusoidal modulation of two optimally oriented bars presented in adjacent subfields of opposite sign produced reasonable agreement with the relationship between stimulus contrast and simple cell response amplitude predicted by taking into account both the thresholding (half-wave rectification) behaviour of the cell and the difference in temporal phase[11] between the responses to bars presented separately. In particular, all simple cells exhibited approximately linear summation of excitatory inputs; some departure from spatial linearity was however observed for inhibitory input. The Principle of Superposition had been tested earlier by Henry et al. (1978) with more equivocal results. Whilst superposition was found to hold after allowing for output thresholding when two rectangular spots were drifted either individually or jointly — with fixed spatial offset — across the receptive field in the preferred direction, this was not the case when two optimally oriented bars were flashed individually and then simultaneously in the receptive field. Despite the fact that for lower contrast levels the suprathreshold response to the dual stimulus was found to increase approximately linearly with stimulus

[8]The contrast response function describes the response of the cell as a function of stimulus contrast. Whilst in the present model the saturation of this function with increasing contrast is attributed to the output nonlinearity f, it is likely that at least some of this saturation is in fact attributable to nonlinear mechanisms prior to the presumed linear summation stage.

[9]The RGF model is described in detail in Section 2.3.1.

[10]although see Sections 2.3.2 and 2.3.3

[11]See Section 2.2.4 for a discussion on the variation of temporal response phase across the simple cell RF.

contrast even when linear superposition did not apply, the discrepancy with their own results was attributed by Tolhurst & Dean (1987) to response saturation (Albrecht & Hamilton, 1982) at the comparatively higher contrast levels used.

A purely linear model of the simple cell spatial RFP is predicated upon the assumption of balanced, antagonistic and spatially coextensive on- and off-type input to each subfield. Known as the *push-pull* model (Palmer et al., 1991), this arrangement is necessary to produce an equal response of opposite sign to stimuli of positive and negative contrasts, as required by the Principle of Superposition. Due to half-wave rectification of the response, however, the postulated inhibitory input to a subfield is not revealed by experiments performed around the response threshold, such as those of Jones & Palmer (1987b) and Jones et al. (1987). When the latter investigation of the 2D spectral RFP was repeated by Palmer et al. (1991) in the presence of a second uncorrelated stimulus used to elevate the mean firing rate of the cell, a spatial-frequency dependent suppression of this mean was produced by drifting sinusoidal grating stimuli at spatial frequencies to which the cell had been previously found to be unresponsive. A similar result was reported by Bonds (1992) in independent investigations of orientation and spatial frequency tuning, while Ramoa et al. (1986) demonstrated an orientation- but not (spatial) frequency-sensitive suppression of pharmacologically-elevated simple cell activity by previously ineffective stimuli. Palmer et al. (1991) argued that their observations were consistent with a modified form of the push-pull model in which the excitatory and inhibitory inputs are neither precisely balanced nor spatially coextensive. This model is at least qualitatively consistent with electrophysiological results suggesting that the strengths of antagonistic inputs — such as on-excitatory and off-inhibitory inputs — to a given subfield are frequently unbalanced (Heggelund et al., 1983; Heggelund, 1986a; Tolhurst & Dean, 1987), that an inhibitory input region centred on an excitatory subfield may not be precisely coextensive with that subfield (Ferster, 1988) and that one or other type of input is completely lacking for some subfields (Glezer et al., 1982; Heggelund et al., 1983). Taken together, these results suggest that the spatial processing performed by each simple cell consists not only of the dominant linear component inferred by Tolhurst & Dean (1987), but also spatially nonlinear components. In modelling the simple cell as a spatially linear device, one should therefore not lose sight of the approximation involved in neglecting these nonlinear terms[12].

2.2.4 *Spatiotemporal RFP*

The linear model presented in (2.1) to describe the spatial RFP of a simple cell predicts a response $r(t)$ to the abrupt presentation of a stationary stimulus $s(\boldsymbol{x}, t) = s_{\boldsymbol{x}}(\boldsymbol{x})u(t)$

[12]The interested reader is referred to Henry (1985) for further discussion on early results concerning the spatial linearity and nonlinearity of the simple cells.

which is given by

$$r(t) = u(t) \int_{\mathcal{V}} w(\boldsymbol{x}) s_{\boldsymbol{x}}(\boldsymbol{x})\, d\boldsymbol{x}$$

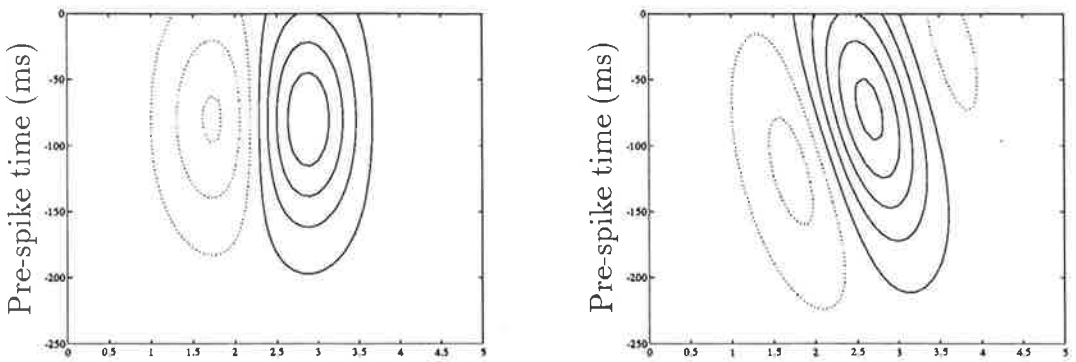
where $u(t)$ is the unit step function and $s_{\boldsymbol{x}}(\boldsymbol{x})$ is the stimulus contrast as a function of the visual angle $\boldsymbol{x}$ alone. The predicted response, which is proportional to the unit step function, fails to capture the temporal behaviour of real simple cells, which exhibits both a finite delay (Ikeda & Wright, 1975a), and a time course (and temporal frequency response) reminiscent of a low-pass or band-pass temporal filter (Ikeda & Wright, 1975b). Neurophysiologists refer to such responses, and by extension the cells exhibiting them, as *sustained* or *X-like* and *transient* or *Y-like* respectively. Furthermore simple cells are known to receive X- or Y-like inputs (or possibly both) from neurons in the *lateral geniculate nucleus* (LGN) of the *thalamus* (see e.g. White (1989, p. 136)) — through which the majority of feedforward input to primary visual cortex passes on its way from the retina — and each of these inputs may be of either on- or off-type. Finally, Hubel & Wiesel (1962) noted that simple cells in the feline primary visual cortex were often significantly more responsive to a stimulus if it was drifted across the receptive field, and that the response varied with both the direction and velocity of the stimulus. The velocity tuning curves of the simple cells for fixed stimulus size have since been elaborated by a number of researchers (see e.g. Movshon (1975)), and experiments designed to elaborate simple cell spatial and spectral RFPs routinely use stimuli moving at the optimal velocity (magnitude and direction). A proper consideration of the temporal characteristics of the simple cell is therefore likely to be vital to a full understanding of its computational role.

If we retain a linear model for simplicity, temporal behaviour can be incorporated by defining a *spatiotemporal RFP* $w{:}\mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ such that

$$r(t) = \int_{-\infty}^{t} \int_{\mathcal{V}} w(\boldsymbol{x}, t-\tau) s(\boldsymbol{x}, \tau)\, d\boldsymbol{x}\, d\tau \tag{2.2}$$

A major consequence of this model is that the notion of a spatial RFP is no longer strictly defined unless w is spatiotemporally separable, in which case it can be expressed as $w(\boldsymbol{x}, t) = w_{\boldsymbol{x}}(\boldsymbol{x}) w_t(t)$ where $w_{\boldsymbol{x}}(\boldsymbol{x})$ and $w_t(t)$ encapsulate independently the spatial and temporal behaviour respectively. A separable spatiotemporal RFP is shown in Figure 2.2(a). This property was found to hold for only 24 of the 52 simple cells tested by McLean & Palmer (1989), who used a reverse correlation technique similar to that of Jones & Palmer (1987b) but with finer (1ms) temporal resolution to determine the apparent spatial RFP of the cell as a function of the pre-spike stimulus presentation time. The resultant 2D spatiotemporal plot may be viewed as a 2D (1 spatial and 1 temporal dimensional) section through the 3D spatiotemporal RFP along the (spatial) axis of subfield alternation.

For the remaining 28 of the 52 simple cells studied by McLean & Palmer (1989) — all of which were directionally selective — this plot revealed that the RF subfields drift with approximately uniform velocity in the preferred direction of stimulus motion with



(a) Separable (b) Inseparable

Figure 2.2: Spatiotemporal RFP $w(x, -t)$ (after McLean & Palmer (1989, Figs. 1B & E)) showing for each position along the axis of subfield alternation and for each pre-spike time, the probability that a bright (solid contours) or dark (dotted contours) stimulus presented at that spatiotemporal position will elicit a spike at time $t = 0$. (a) Separable. (b) Inseparable.

decreasing pre-spike presentation time, as illustrated in Figure 2.2(b). The drift velocity of these subfields, which agreed well with both the preferred direction and velocity of the simple cell, provides a measure of the degree of Cartesian inseparability of the spatiotemporal RFP: the higher the subfield drift velocity, the less valid the notion of a discriminable spatial RFP. In accordance with this observed spatiotemporal inseparability, Reid et al. (1987) showed that velocity-sensitive behaviour in some simple cells could be largely explained by a model based on linear spatial summation of subfield inputs whose temporal phase advanced (delay decreased) linearly in the preferred direction of stimulus motion. Measurements of the spatiotemporal *phase transfer function* of simple cells in both monkey and cat were similarly shown by Hamilton et al. (1989) to be largely consistent with the inseparable *linear quadrature* model of Watson & Ahumada (1983; 1985).

Linear spatiotemporal RFP models did not however account completely for the velocity sensitivity observed by Reid et al. (1987) in simple cells. Furthermore, in the study by McLean & Palmer (1989), 8 of the 24 cells having separable spatiotemporal RFPs were still found to exhibit a preference for one direction of motion. This latter observation was supported by Emerson & Citron (1988; 1989), who used reverse correlation techniques to identify the first- and second-order *spatiotemporal Wiener kernels* for the simple cell (see e.g. Schetzen (1980)). A considerable direction selective component in the second-order kernel was identified by performing a second-order reverse correlation of the neural spike train with the spatiotemporal stimulus, which consisted of 16 optimally

oriented bar stimuli spaced uniformly along the axis of subfield alternation and whose contrasts were randomly and independently selected every 16ms from three possible values: zero, and equal positive and negative contrasts. This result indicates a direction selective interaction between stimuli at small spatiotemporal offsets, such that following the presentation of a bar the cell would prefer the subsequent presentation of a second bar of the same contrast sign at a position whose spatial offset in the preferred direction increases linearly with the temporal offset. The preferred direction predicted from such nonlinear mechanisms is generally in agreement with that predicted from the first-order Wiener kernel (Emerson & Citron, 1989; Reid et al., 1987) — viz. the spatiotemporal RFP.

Thus in addition to Cartesian inseparability of the spatiotemporal RFP, a simple cell may also exhibit spatiotemporal nonlinearity. For further discussion of the likely importance of nonlinearities in spatiotemporal vision, and the limitations of linear analysis in this context, the interested reader is referred to Regan (1991).

2.2.5 *Binocular RFP*

Contrary to the implicitly monocular treatment of the simple cell presented so far, Hubel & Wiesel (1962) noted that most if not all simple cells could be driven to some extent by input to either eye. The provocative observation by Hubel & Wiesel (1962) that the left- and right-eye receptive fields for a given simple cell occupied corresponding positions on the two retinae led Barlow et al. (1967) and Pettigrew et al. (1967) to postulate and investigate the involvement of cells in feline primary visual cortex in the computation of binocular disparities and hence — in the case of horizontal disparities — of stereoscopic depth. In the feline primary visual cortex, Pettigrew et al. (1967) showed that individual simple cells were tuned to a range of horizontal disparities, while Barlow et al. (1967) demonstrated that the optimal disparity varied significantly between neurons[13]. The stereoscopic depth tuning properties of primary visual cortical neurons have since been elaborated by a number of researchers for both monkey (see e.g. Poggio et al. (1988), Poggio (1980)) and cat (see e.g. LeVay & Voigt (1988)).

If the inputs from the left and right eyes are assumed to be processed independently by linear spatiotemporal mechanisms prior to linear combination of the binocular inputs, the output $r(t)$ of the simple cell is given by

$$r(t) = \int_{-\infty}^{t}\int_{\mathcal{V}_L} w_L(\boldsymbol{x}, t-\tau)s_L(\boldsymbol{x},\tau)\, d\boldsymbol{x}\, d\tau + \int_{-\infty}^{t}\int_{\mathcal{V}_R} w_R(\boldsymbol{x}, t-\tau)s_R(\boldsymbol{x},\tau)\, d\boldsymbol{x}\, d\tau \qquad (2.3)$$

where $s_L : \mathcal{V}_L \times \mathbb{R} \to \mathbb{R}$ and $s_R : \mathcal{V}_R \times \mathbb{R} \to \mathbb{R}$ represent the 3D stimulus as viewed from the left and right eyes respectively, $w_L : \mathcal{V}_L \times \mathbb{R}_+ \to \mathbb{R}$ and $w_R : \mathcal{V}_R \times \mathbb{R}_+ \to \mathbb{R}$ are the corresponding monocular spatiotemporal RFPs, and $\mathcal{V}_L, \mathcal{V}_R \subset \mathbb{R}^2$ are the left- and right-eye visual fields.

[13] although they did not differentiate between simple, complex and hypercomplex cell types.

The pair $\{w_L, w_R\}$ will be referred to as the *binocular spatiotemporal RFP*[14]. Evidence for linearity of binocular summation has been provided by Ohzawa & Freeman (1986) and Freeman & Ohzawa (1990b; 1990a), who showed that the dependence of the response amplitude on the spatial phase offset between dichoptically presented drifting sinusoidal gratings at the optimal orientation and spatial frequency was approximately sinusoidal, and exhibited a single phase at which response was (almost) completely suppressed. This result was largely confirmed by Hammond (1991), with the exception that a few simple cells were found to be largely insensitive to the interocular phase shift.

Ohzawa & Freeman (1986) showed good agreement between their experimental results and a linear model in which the spatial LWFs were identical for the left and right eyes, but offset by the optimal horizontal disparity for that cell. This model is in broad agreement with the observations that the left and right eyes have similar spatial frequency and orientation tuning (Freeman & Ohzawa, 1990a; Skottun & Freeman, 1984) — although see Hammond & Pomfrett (1990) — RFP organisation (Hubel & Wiesel, 1962), and LWFs (Maske et al., 1984). However, by plotting the 2D RFPs for left- and right-eye monocular input using reverse correlation techniques similar to those of Jones & Palmer (1987b), Freeman & Ohzawa (1990b) demonstrated that the left and right eye spatial RFPs frequently differed, and in some cases the spatial ordering of the subfields along the common axis of alternation was completely reversed. They suggested that this observation was more consistent with a model such as that of Nomura et al. (1990)[15] in which the alternating pattern of subfields — windowed in both eyes by a function such as a Gaussian centred at zero disparity — was phase shifted by the optimal disparity of the cell. However collocation of the windowing function was not explicitly tested.

To describe the relative strength of the inputs from the two eyes, Hubel & Wiesel (1962) proposed the *ocular dominance index*, according to which cells receiving exclusive input from the contralateral or ipsilateral eye — as determined using monocular stimulation — were assigned the extreme values 1 and 7 respectively, while intermediate values signified somewhat coarsely the relative dominance of the two inputs. Monocular experiments on the simple cells — such as those cited in previous sections — conventionally present stimuli to the dominant eye. However, Hubel & Wiesel (1962) found that some cells could only be activated by binocular input, suggesting nonlinear interaction between inputs from the two eyes. Furthermore Freeman & Ohzawa (1990b) found that simple cells which appeared exclusively monocular under monocular stimulation showed clear evidence of input from the supposedly silent eye during binocular stimulation. The

[14]The spatial component of this characterisation differs from the *binocular receptive field* defined by Ohzawa & Freeman (1986), which is produced by the summation of the spatial *RFPs* of the left and right eyes. Their model therefore fails to account for features which, either through an interocular difference in the viewing perspective or through experimental manipulation, are not visible through one of the two eyes.

[15]To be discussed further in Section 2.3.4.

strength of this input varied little as the relative contrast of the stimulus to that eye was varied over a 10-fold range, indicating the presence of independent gain control mechanisms for inputs to the two eyes. This latter observation suggests that monocularly measured ocular dominance is largely irrelevant to the binocular operation of the cell.

A more complete characterisation of the simple cell as a binocular device would involve the determination of the binocular spatiotemporal Wiener kernels of the simple cell for 5D stimuli expressible in terms of the three variables: left- and right-eye positions $\boldsymbol{x}_L, \boldsymbol{x}_R \in \mathbb{R}^2$ and time t. Although insufficient to account for the temporally non-stationary effects of contrast gain control, such an analysis may reveal more complex details such as tuning for (possibly oblique) motion in depth (see e.g. Regan et al. (1990)), and thereby indicate new directions of investigation.

2.2.6 Summary

In the foregoing discussion, the spatial, spectral, spatiotemporal and binocular receptive field profiles of the simple cells have been described, and the extent to which the linear characterisation of the cell inherent in the notion of an RFP is valid has been briefly addressed. It is concluded that with the appropriate reservations, the simple cell may be viewed to a first approximation as a linear device characterised by its binocular spatiotemporal RFP.

2.3 Gabor Function Models

2.3.1 Spatial RFP

Marčelja (1980) demonstrated a strong resemblance between simple cell LWFs and the one-dimensional (1D) *real-valued Gabor functions* (RGFs), which form the real and imaginary parts of the 1D *Gabor functions*. Given by the product of a Gaussian and a sinusoid, the 1D Gabor functions were named in honour of Gabor (1946), who showed how they can be used in the representation of 1D signals. The Gabor functions were generalised by Daugman (1980; 1985), Kulikowski et al. (1982) and Watson & Ahumada (1983) to two dimensions, by Heeger (1987) to 3D, and by MacLennan (1991) to n dimensions, yielding the n-dimensional Gabor function $g{:}\mathbb{R}^n \rightarrow \mathbb{C}$

$$g(\boldsymbol{x}) \triangleq a \cdot \exp\{-\pi \|S^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\} \cdot \exp\{j2\pi \boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0)\} \tag{2.4}$$

where ^{T} denotes the transpose operator. Gabor functions are parameterised by the set $\{\boldsymbol{x}_0, S, \boldsymbol{u}_0, a\}$, where $\boldsymbol{x}_0 \in \mathbb{R}^n$ is the location of the Gaussian centre, $S \in \mathbb{R}_+^{n\times n}$ is a diagonal matrix whose ith diagonal entry s_{ii} is proportional to the standard deviation of the Gaussian along the principal axis aligned with the ith coordinate axis, $\boldsymbol{u}_0 \in \mathbb{R}^n$ is the

(oriented) spatial frequency vector in cycles per unit length, and $a \in \mathbb{R}$ is an amplitude scaling factor. In two dimensions, for example,

$$\boldsymbol{x}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \qquad S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \qquad \boldsymbol{u}_0 = \begin{bmatrix} u_x \\ u_y \end{bmatrix}$$

with the *aspect ratio* $\lambda \triangleq \min\{\frac{s_x}{s_y}, \frac{s_y}{s_x}\}$. The explicit notation $g(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, a)$ will be reserved for situations in which it is useful to emphasise this parameterisation.

The real and imaginary parts of $g(\boldsymbol{x})$ are given by

$$\Re\{g(\boldsymbol{x})\} = a \cdot \exp\{-\pi \|S^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\} \cdot \cos\{2\pi \boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0)\} \tag{2.5a}$$

$$\Im\{g(\boldsymbol{x})\} = a \cdot \exp\{-\pi \|S^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\} \cdot \sin\{2\pi \boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0)\} \tag{2.5b}$$

respectively, and will be referred to as the *real-valued Gabor functions*[16] (RGF). While the principal axes of the Gaussians in (2.5) are aligned with the coordinate axes, the sinusoid may have arbitrary orientation $\arg\{\boldsymbol{u}_0\}$, with the result that the RGF is not in general Cartesian separable. In a number of the 2D simple cell RFPs identified by Jones & Palmer (1987b), lack of Cartesian separability was evident as a progressive displacement, across the RF, of the subfields in the direction of subfield elongation (see e.g. Jones & Palmer (1987b, Fig. 2F)), as illustrated in Figure 2.1(h).

A simple generalisation of the Gabor function which is necessary to accommodate the observed variety of simple cell 2D spatial RFPs involves the addition of two extra degrees of freedom: a phase shift $\phi \in \mathbb{R}$ of the sinusoid relative to the Gaussian centre $\boldsymbol{x}_0$ (Kulikowski et al., 1980; Watson & Ahumada, 1983; Field & Tolhurst, 1986); and a rotational angle $\boldsymbol{\theta} \in \mathbb{R}^{n-1}$ between the principal axes of the Gaussian and the coordinate axes (Daugman, 1985; Jones & Palmer, 1987b, both in 2D). The resultant functions — which will henceforth be referred to as *generalised Gabor functions* — are parameterised by the set $\{\boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \phi, a\}$ with $\boldsymbol{\theta}, \phi$ in radians, and are given by

$$g(\boldsymbol{x}) \triangleq a \cdot \exp\{-\pi \|S^{-1}R_{-\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\} \cdot \exp\{j[2\pi \boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0) + \phi]\} \tag{2.6}$$

where $R_{-\boldsymbol{\theta}}$ is the operator (matrix) performing a rotation through the angle $-\boldsymbol{\theta}$. In two dimensions, for example,

$$R_{-\boldsymbol{\theta}} = \begin{bmatrix} cos(\theta_z) & -sin(\theta_z) \\ sin(\theta_z) & cos(\theta_z) \end{bmatrix}$$

Except where indicated otherwise, $g(\boldsymbol{x})$ is henceforth used to denote the *generalised* version of the Gabor function. The real and imaginary parts of the generalised Gabor

[16]The term *Gabor elementary function* was reserved by Stork & Wilson (1990) for these real-valued functions, but is generally used by others either interchangeably with the term *Gabor function* (Daugman, 1985) or to refer without distinction to either a Gabor function or its real and imaginary parts (Marčelja, 1980). The new term *real-valued Gabor function* has therefore been coined to avoid the potential confusion arising from these conflicting conventions.

function are given by

$$\Re\{g(\boldsymbol{x})\} = a\cdot\exp\{-\pi\|S^{-1}R_{-\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\}\cdot\cos\{2\pi\boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0)+\phi\} \quad (2.7a)$$

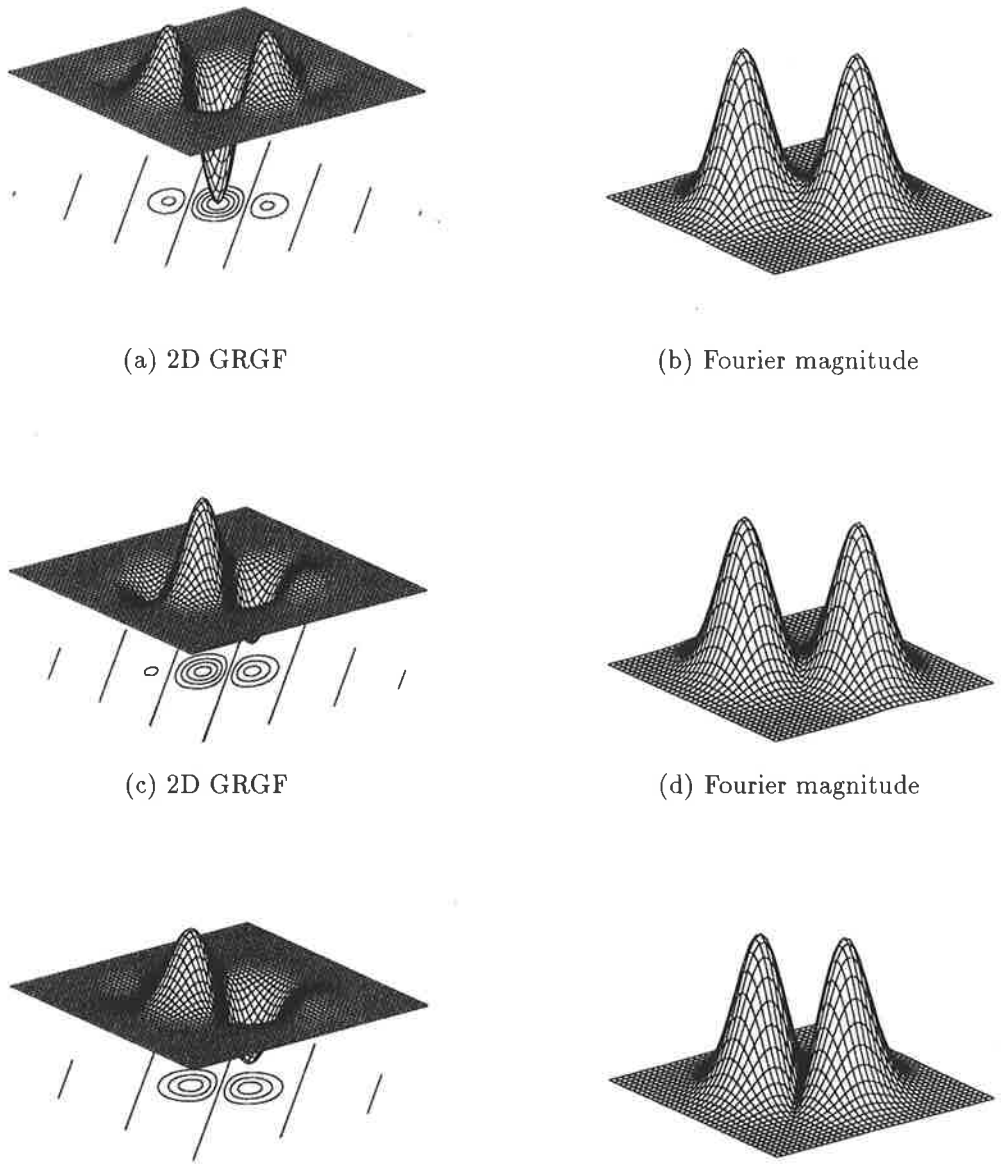
$$\Im\{g(\boldsymbol{x})\} = a\cdot\exp\{-\pi\|S^{-1}R_{-\boldsymbol{\theta}}(\boldsymbol{x}-\boldsymbol{x}_0)\|_2^2\}\cdot\sin\{2\pi\boldsymbol{u}_0^T(\boldsymbol{x}-\boldsymbol{x}_0)+\phi\} \quad (2.7b)$$

respectively, and are related simply by a 90° phase shift in the sinusoidal term. Without loss of generality, it is therefore only necessary when fitting this model to a real spatial RFP to consider the *generalised RGF* (GRGF) $w(\boldsymbol{x}) \triangleq \Re\{g(\boldsymbol{x})\}$, with $w:\mathbb{R}^n\rightarrow\mathbb{R}$ having ϕ as a free parameter. Examples of a 2D GRGF are presented in Figures 2.3(a), (c) and (e).

The necessity of the above generalisation of the RGFs to arbitrary Gaussian orientation is most readily appreciated through an examination of the 2D *spectral* RFP, which is the topic of Section 2.3.2. The need for a non-zero phase shift is evident for cases in which the axis of subfield alternation in the spatial RFP shows approximate alignment with a principal axis — the ith say — of the overall RF window, yet the RFP still fails to exhibit the odd or even symmetry about the hyperplane $x_i = [\boldsymbol{x}_0]_i$ — perpendicular to that principal axis — which is predicted by the two RGF models in (2.5) respectively. Clear examples of such asymmetric RFPs are presented by Jones & Palmer (1987a, Figs. 2E & 4B). Ironically however, despite the conclusion by Jones & Palmer (1987b) that "most simple receptive fields are neither even symmetric nor odd symmetric", their report is not infrequently cited in support of models requiring these symmetries (see e.g. Koenderink & van Doorn (1990a)), and in image analysis applications — reviewed in Chapter 3 — in which only the original (odd- and even-symmetric) RGFs are used.

The frequent *asymmetry* of simple cell RFPs — noted earlier by Kulikowski et al. (1980) and Field & Tolhurst (1986) in connection with the LWF — limits the generality of most existing models of the simple cell spatial RFP. Many such models are motivated by the computational goals — discussed further in Sections 3.2 and 3.3 respectively — of detecting lines and edges in the visual image and of calculating directional spatial derivatives of the image contrast (although see e.g. Atick & Redlich (1990a)). According to the latter approach, the antisymmetric and symmetric RFPs with two and three subfields respectively — illustrated in Figures 2.1(d) and (e) and commonly labelled as "edge and line detectors"[17] — are viewed as resulting from the application of discrete-space approximations of the first- and second-order directional derivative operators respectively to the photoreceptor RFP, which is frequently modelled as a Gaussian. The resultant *Gaussian derivative* model (Gaussian-windowed Hermite polynomials) which was first proposed and evaluated by Young (1985) in 1D, and later proposed independently by Martens (1990) and Koenderink & van Doorn (1990a; 1990b) in 2D, as a model of the LWF and RFP respectively, relies on these canonical symmetries. As will be seen in Section 3.3 however, the generic *fractional discriminant function* model of Hungenahally

[17]Although for reasons discussed later in Section 3.2 this terminology should be discouraged.



(a) 2D GRGF (b) Fourier magnitude

(c) 2D GRGF (d) Fourier magnitude

(e) 2D GRGF (f) Fourier magnitude

Figure 2.3: Mesh and contour plots of a 2D GRGF with the common parameter values (unit length given by 1 grid interval): $\boldsymbol{x}_0 = [0,0]^T$ (centre of grid), $\arg\{\boldsymbol{u}_0\} = 45°$, $s_{maj} = 17.7$, and $\lambda = 0.667$. Additional parameters are (a) $|\boldsymbol{u}_0| = 0.0509$, $\theta_z = 45°$, & $\phi = 180°$; (c) $|\boldsymbol{u}_0| = 0.0509$, $\theta_z = 30°$, & $\phi = 60°$; and (e) $|\boldsymbol{u}_0| = 0.0382$, $\theta_z = 45°$, & $\phi = 90°$. (b),(d),(f) Magnitude of the Fourier transform of the RGFs illustrated in (a),(c),(e) respectively. Note that only the odd-symmetric (sine-phase) GRGF exhibits zero magnitude at spatial DC (centre of plot).

et al. (1992; 1993) avoids this pitfall. Nevertheless all such directional derivative RFP models proposed to date exhibit Cartesian separability, although this property need not hold if for example the Gaussian derivative model were to be extended to the case of anisotropic Gaussians.

Attempts to relate the RGF model of the simple cell spatial RFP to the theories of edge and line detection and directional spatial derivatives may have contributed to the tendency to use only the odd- and even-symmetric RGF models, despite the obvious modelling advantage conferred by the unused extra degree of freedom. Although aware of the existence of asymmetric RFPs, Marčelja (1980) cited among other things the evident utility of extracting lines and edges from the visual image as justification for his 1D RGF model. Similarly Sakitt & Barlow (1982) proposed an economical scheme for the cortical representation of the visual image in terms of odd- and even-symmetric Gabor functions, which were nevertheless referred to as "edge and bar detectors". Paler & Bowler (1986) noted the similarity between the odd- and even-symmetric RGFs having 2 and 3 subfields[18] and the Canny edge detector (Canny, 1986) and second Gaussian derivative respectively. Pollen & Ronner (1983) on the other hand preferred these canonical symmetries[19] on the basis of the mistaken belief that they are required for the optimally efficient Gabor function decomposition of the image[20], once again despite being aware that some simple cells exhibit an asymmetrical RFP. This view was probably motivated by the localised Fourier analysis hypothesis of simple cell spatial information processing as discussed in Section 3.4, since it is conventional — but not necessary — in Fourier analysis to use sinusoids in strict sine and cosine phase.

Although Jones & Palmer (1987a) later fitted the GRGF model to their data, it is important to note that the conclusions drawn by Jones & Palmer (1987b) were — apart from the assumption of spatial linearity — independent of any particular model of the simple cell RFP. This is an important point, since any conclusions drawn after the fitting of a particular model would necessarily be dependent on the appropriateness of that model. However, the lack of RF symmetry and of Cartesian separability clearly visible in some of the simple cell RFPs identified by Jones & Palmer (1987b) was also reflected in the parameters obtained by Jones & Palmer (1987a) after fitting the GRGF model to the RFPs of 36 simple cells. In particular, the angle $|\arg\{\boldsymbol{u}_0\}-\theta_z|$ between the sinusoid and the nearest principal axis of the Gaussian — which is by definition in the range $[0°, 45°]$ — exceeded 20° for 7 of the 25 cells for which parameters were tabulated. Furthermore, the phase angle $|\phi|$ was distributed approximately uniformly over the range $[0°, 90°]$, as opposed to clustering around 0° and 90°, which would be required — assuming Cartesian

[18] i.e. a Gaussian with relatively small standard deviation along the axis of subfield alternation, so that only two or three subfields are prominent.

[19] although see Pollen et al. (1985).

[20] See Section 3.4.2 for further details.

separability — to support the hypothesis that simple cells fall into the canonical odd- and even- symmetries. This latter result supports the similar finding reported by Field & Tolhurst (1986) for the simple cell LWF. The aspect ratio λ for the cells tested ranged from 0.23 (strongly elongated) to 0.92 (almost round), with no absolutely preferred value.

The accuracy of the fit provided by the GRGF model was later confirmed by Palmer et al. (1991) using 1ms temporal resolution for the prespike time and selecting the time bin at which maximum response was achieved, allaying possible concerns that the RFPs obtained by Jones & Palmer (1987b) using 50ms time bins might be unrepresentative of the instantaneous spatial structure in cases of spatiotemporally inseparable RFPs due to blurring caused by subfield drift. However it should be pointed out that neither group observed simple cells with more than three subfields. Although this finding is not surprising given the relative scarcity of such cells (Glezer et al., 1989), the relatively small sample size used (36), and the average retinal eccentricity of the recordings (Jones & Palmer, 1987b), it means that the GRGF model — although able to account at least qualitatively for an arbitrary number of subfields — is as yet untried on simple cell 2D RFPs having 4 or more subfields.

Problems with the GRGF Model

The principal objection to the GRGF model of the simple cell spatial RFP is that it has infinite spatial extent or noncompact support (MacLennan, 1991) — albeit with rapidly diminishing weighting — and consequently an infinite number of zero crossings due to the sinusoid (Stork & Wilson, 1990), as is evident from the zero-level contours in the GRGF plots of Figures 2.3(a), (c) and (e). In contrast, the maximum number of subfields recorded to date for a simple cell is 8 (Glezer et al., 1989). In reality however, no cell can have a monocular RFP which exceeds the visual field for the corresponding eye, so that truncation of any RFP model at the edge of the visual field is a practical necessity. In view of the exhorbitant "wetware" (neural hardware) cost of providing feedforward input via the LGN to each simple cell from every photoreceptor[21], evolution might be expected to prefer a more parsimonious solution in which the support of real RFPs would in fact be significantly smaller than this again, especially in cases where the (measurably) non-zero portion of the RFP subtends only a small portion of the visual field. However, the recordings of Jones & Palmer (1987b) did not extend beyond a few standard deviations of the allegedly Gaussian window, and would almost certainly have required a prohibitive number of stimulus cycles to reveal a non-zero weighting at this distance, so that the exact behaviour of the window would be difficult to reveal experimentally.

A second potential objection to the GRGF model is that it fails to account for the end-stopping or *hypercomplex* property exhibited by many cells classified as simple according

[21] although lateral interactions between simple cells and feedback from subsequent cortical layers could also contribute to the RFP. This possibility is discussed further in Chapter 7.

to the criteria of Henry (1977) (Orban, 1991). "End-stopping" refers to the progressive suppression of simple cell response (often in excess of 50%) as the length of an optimally oriented bar is increased beyond its optimal value. To account for this behaviour using a linear RF model, one would need to postulate the existence of inhibitory flanks at either or both ends of the each excitatory RF subfield, a feature which cannot be accommodated by the GRGF RFP model. Nevertheless, since end-stopping was not observed in the identification experiments of Jones & Palmer (1987b), the possibility exists that this phenomenon is a consequence of even-order spatial nonlinearities which are transparent to the reverse-correlation identification technique and excluded by the linear definition of the RFP. It remains however to demonstrate that simple cells exhibiting the hypercomplex property were not systematically (albeit possibly inadvertently) excluded from the sample tested by Jones & Palmer (1987b).

2.3.2 *Spectral RFP*

The Fourier transform $G{:}\mathbb{R}^n \rightarrow \mathbb{C}$ of the generalised Gabor function — where $G(\boldsymbol{u}) \triangleq \mathcal{F}\{g(\boldsymbol{x})\}$ and $\mathcal{F}$ denotes the Fourier transform (FT) — is given by

$$G(\boldsymbol{u}) = a \cdot \exp\{-\pi \|SR_{-\boldsymbol{\theta}}(\boldsymbol{u} - \boldsymbol{u}_0)\|_2^2\} \cdot \exp\{j[2\pi(\boldsymbol{u} - \boldsymbol{u}_0)^T \boldsymbol{x}_0 + \phi]\}$$

Using the notation $G(\boldsymbol{u}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \phi, a)$ once again to emphasise the parameterisation, the Fourier transform $W{:}\mathbb{R}^n \rightarrow \mathbb{C}$ of the GRGF $w(\boldsymbol{x})$ can be expressed as

$$W(\boldsymbol{u}) \triangleq \mathcal{F}\{\Re\{g(\boldsymbol{x})\}\} = G(\boldsymbol{u}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \phi, \frac{a}{2}) + G(\boldsymbol{u}; \boldsymbol{x}_0, S, -\boldsymbol{u}_0, \boldsymbol{\theta}, -\phi, \frac{a}{2}) \qquad (2.8)$$

The Fourier transforms of the RGFs in (2.5) are special cases of (2.8) in which $\phi = 0°$ and $\phi = 90°$ respectively and $\boldsymbol{\theta} = \mathbf{0}$. $W(\boldsymbol{u})$ consists of two Gaussians centred at $\boldsymbol{u}_0$ and $-\boldsymbol{u}_0$ and modulated by sinusoids having phase ϕ and $-\phi$ respectively at the Gaussian centres, and "frequency" $\boldsymbol{x}_0$. The magnitude of $W(\boldsymbol{u})$ in (2.8) can be derived using the fact that $|W(\boldsymbol{u})|^2 = W(\boldsymbol{u})W^*(\boldsymbol{u})$ — where $|\cdot|$ and * denote the complex magnitude and conjugate respectively — and by noting that

$$W^*(\boldsymbol{u}) = G(\boldsymbol{u}; -\boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, -\phi, \frac{a}{2}) + G(\boldsymbol{u}; -\boldsymbol{x}_0, S, -\boldsymbol{u}_0, \boldsymbol{\theta}, \phi, \frac{a}{2})$$

Three examples of the Fourier magnitude of a GRGF are presented in Figures 2.3(b), (d) and (f).

The comprehensive investigation of the magnitude of the simple cell 2D spectral RFP undertaken by Jones et al. (1987) revealed not only the polar inseparability of the RFP, but also in some cases its Cartesian inseparability, evident as a lack of radial alignment of any principal axis of the spectral window in the 2D spectral plane. However contrary to the predictions of horizontal or vertical alignment derived from the RGF model, these axes were mostly in *approximate* radial alignment, indicating the need for the rotational

angle $\boldsymbol{\theta}$ introduced in (2.7). These observations are once again independent of any particular model of the 2D spectral RFP.

The ability of the RGF model of the spatial RFP to describe also the spatial frequency tuning properties of the simple cells was demonstrated in 1D by Marčelja (1980) for spatial frequency magnitude, and suggested in 2D by Daugman (1980) for 2D spatial frequency (magnitude and orientation). Visual inspection by Jones & Palmer (1987a) of the error after the least-squares fit of the GRGF model to the Fourier magnitude data obtained by Jones et al. (1987) revealed no obvious residual spectral structure, providing support for $W(\boldsymbol{u})$ — the FT of the GRGF — as a model of the simple cell spectral RFP. A comparison of the GRGF parameters estimated from fits to the spatial and spectral RFPs of each of 25 simple cells also showed broad agreement between parameter estimates. The two principal exceptions to this general agreement were poor agreement of the phase estimates and a consistent tendency for the estimates of the Gaussian space constants s_{ii} obtained from the spectral data to be larger than those obtained in the spatial domain. The latter phenomenon is discussed in Section 2.3.3.

Problems with the GRGF Model

The comparison by Jones & Palmer (1987a) between the estimates of the sinusoidal phase parameter ϕ obtained from the spatial and spectral RFPs for individual simple cells however revealed little correlation between these two estimates, and a tendency for the spectral estimate to lie closer to or — in the case of 6 of the 25 cells for which data was tabulated — exactly at 90°. This same effect was noted also in the monkey by Hawken & Parker (1987), who fitted the 1D GRGF model to the spatial frequency contrast sensitivity tuning curve. They pointed out that the Gabor function model tended to consistently over-estimate the contrast sensitivity at low spatial frequencies, and that $\phi = 90°$ gave the best least-squares fit probably because it provided the sharpest low-frequency roll-off of all the GRGFs. Furthermore $\phi = 90°$ is the only phase which predicts the zero response to spatial DC (uniform illumination) commonly reported for simple cells (Stork & Wilson, 1990); compare for example Figures 2.3(b) and (d) with Figure 2.3(f).

In the case of Hawken & Parker (1987), who measured the contrast required to elicit a mean response which was 2 standard deviations above the spontaneous activity of the cell, it is possible that the rate of low-frequency roll-off was overestimated due to the nonlinear effects noted by Palmer et al. (1991). In particular, if in addition to the sinusoidal response modulation expected for such a stimulus the grating also suppressed the *mean* activity in a frequency-dependent manner, the stimulus contrast required to elicit the specified mean response could well have been overstated at low frequencies, and hence the contrast sensitivity underestimated. Although Movshon et al. (1978a) reported qualitative agreement between tuning curves obtained for the same cell

using both the fundamental Fourier component of the temporal response and the grating contrast required to elicit the smallest discernible response, the response criterion used by Hawken & Parker (1987) may have been slightly higher, enhancing the effect of the nonlinearity. Nevertheless, this explanation accounts for neither the poor agreement of the spatial and spectral phase estimates observed by Jones & Palmer (1987a), who used the fundamental of the temporal response to obtain the spectral RFP, nor for the consistent overestimation — based on the spatial RFP — of the spectral width parameters. Thus it is appears more likely that nonlinear mechanisms are acting to limit the extent of the spectral RFP even in the absence of the elevated mean firing rate used by Palmer et al. (1991), and it is therefore not possible to discount the GRGF as a model of the *spatial* RFP of the simple cell on the basis of the above spectral observations. Such mechanisms would be consistent with the model-independent observation by Westheimer (1984) that the line weighting functions of simple cells rarely exhibit a sufficient number of subfields to account for the narrowness of the observed spatial frequency tuning curve.

Nevertheless, despite the apparent scarcity of models of the simple cell spatial and spectral 2D RFPs which exhibit sufficient degrees of freedom to even warrant an attempt at fitting them to the experimental data (Jones, 1991), the investigation by Jones & Palmer (1987a) is open to the criticism that no alternative models were tried (Stork & Wilson, 1990). The Fourier transform of an alternative model of the 2D spatial RFP was shown by Hawken & Parker (1987) to account in a much more satisfactory manner than the 1D GRGF model for the fall-off of the contrast sensitivity function of the simple cell at low grating spatial frequency. According to the *Difference of Offset Difference-of-Gaussian* (DOODOG) model proposed by Hawken & Parker (1987) in 1D — but readily generalisable to 2D — and based on a suggestion originally by Hubel & Wiesel (1962), each RF subfield consists of input from the LGN having the familiar centre-surround *difference-of-Gaussians* (DOG) RFP (see e.g. Wright & Ikeda (1973)). The overall simple cell RFP results from the linear combination of such RFPs with appropriate sign and spatial offset, and — in the 2D case — alignment to produce the corresponding elongated RF subfields. Cartesian inseparability is readily introduced in the 2D generalisation by weighting the inputs along the direction of subfield elongation differently for each subfield. However, despite the appeal of the spatial domain version of this model — which stems from its more direct reflection of the RF properties of the geniculocortical inputs to the simple cells — it has been shown that such a model cannot alone account for the 2D spatial frequency tuning properties of the simple cells (Webster & de Valois, 1985; Wörgötter & Koch, 1991). Furthermore, no direct fit of this model to the spatial RFP of the simple cell was attempted, and given the established lack of residual structure after such a fit for the 2D GRGF, there is no reason to believe that the DOODOG model might provide a better description of the simple cell spatial RFP.

2.3.3 *Optimal Joint Localisation*

Weyl's (1932) *Uncertainty Principle* — which was also derived independently by Gabor (1946) in the context of communication theory — imposes a fundamental limit on the extent to which the energy distributions $|\psi(\boldsymbol{x})|^2$ and $|\Psi(\boldsymbol{x})|^2$ of any 1D function $\psi : \mathbb{R} \to \mathbb{C}$ and its Fourier transform $\Psi : \mathbb{R} \to \mathbb{C}$ can be simultaneously concentrated or localised in the spatial and spatial-frequency domains respectively. This principle was extended by Daugman (1985) to functions of two dimensions, and later generalised independently by Wechsler (1990) and MacLennan (1991) to functions $\psi : \mathbb{R}^n \to \mathbb{C}$ in n dimensions and their Fourier transforms $\Psi : \mathbb{R}^n \to \mathbb{C}$. The generalised Uncertainty Principle may be stated as

$$\Delta x_i \Delta u_i \geq \frac{1}{4\pi} \qquad \forall i \in \{1 \ldots n\} \subset \mathbb{Z}_+ \tag{2.9}$$

with

$$\Delta x_i \triangleq \sqrt{\frac{\int_{-\infty}^{\infty} (x_i - \overline{x}_i)^2 |\psi(\boldsymbol{x})|^2 \, d\boldsymbol{x}}{\int_{-\infty}^{\infty} |\psi(\boldsymbol{x})|^2 \, d\boldsymbol{x}}} \tag{2.10a}$$

$$\Delta u_i \triangleq \sqrt{\frac{\int_{-\infty}^{\infty} (u_i - \overline{u}_i)^2 |\Psi(\boldsymbol{u})|^2 \, d\boldsymbol{u}}{\int_{-\infty}^{\infty} |\Psi(\boldsymbol{u})|^2 \, d\boldsymbol{u}}} \tag{2.10b}$$

where x_i, u_i are the ith components of the spatial and spatial frequency vectors $\boldsymbol{x}, \boldsymbol{u} \in \mathbb{R}^n$ respectively, u_i is in cycles per unit length, and

$$\overline{x}_i \triangleq \frac{\int_{-\infty}^{\infty} x_i |\psi(\boldsymbol{x})|^2 \, d\boldsymbol{x}}{\int_{-\infty}^{\infty} |\psi(\boldsymbol{x})|^2 \, d\boldsymbol{x}}$$

$$\overline{u}_i \triangleq \frac{\int_{-\infty}^{\infty} u_i |\Psi(\boldsymbol{u})|^2 \, d\boldsymbol{u}}{\int_{-\infty}^{\infty} |\Psi(\boldsymbol{u})|^2 \, d\boldsymbol{u}}$$

are the centroids of the distributions obtained by integrating the original energy distributions over all other dimensions $j \neq i$. Δx_i and Δu_i provide measures of the spatial and spectral *spread* of the energy distribution of the function in the ith direction, while the product of their inverses is a corresponding measure of its *joint localisation*. The necessary but insufficient condition

$$\prod_{i=1}^{n} \Delta x_i \Delta u_i \geq \left(\frac{1}{4\pi}\right)^n$$

for (2.9) is sometimes presented as an alternative statement of this generalised Uncertainty Principle (see e.g. Wechsler (1990)).

Weyl (1932) showed that in one dimension (1D), the class of functions for which the equality in (2.9) holds consists of those expressible as the product of a Gaussian and a sinusoid, which would later come to be known as the Gabor functions. Satisfaction by the Gabor functions of the equality in (2.9) for all i was asserted by Kulikowski et al. (1982) for the 2D case, and later proven by Daugman (1985) for 2D and MacLennan (1991) for

nD. A phase shift ϕ, although not explicitly considered in these proofs, does not affect this optimal joint localisation property of the nD Gabor functions since the sinusoidal term vanishes when the complex modulus is computed in (2.10). Furthermore while uniqueness of this property to the nD Gabor functions was not demonstrated, it follows simply from the 1D case by noting that in order to satisfy the ith equality in (2.9), the 1D energy distribution formed by integrating over all other dimensions $j \neq i$ must itself be that of a Gabor function (possibly with $u_0 = 0$). The requirement for satisfaction of the equality in (2.9) for all i therefore uniquely specifies the set of generalised nD Gabor functions for which the principal axes of the Gaussian are aligned with the coordinate axes.

The Gabor functions represent a continuum in the inevitable trade-off between spatial and spectral localisation, with the pure spatial and spectral (Fourier) domain representations constituting its two extremes. An important advantage, according to Daugman (1989b), of a representation of the visual scene using the 2D Gabor functions is that it therefore

> ... facilitates the extraction of local 2D spectral information (texture, scale, axes of modulation) without loss of information about 2D location or metrical relationships.

permitting for example the spatial segmentation of the visual image into regions defined by distinct textural (spectral) signatures, with optimal spatial localisation of the texture boundaries in each spectral band.

However, as indicated by Stork & Wilson (1990) for the 1D case, the GRGFs do *not* exhibit the optimal joint localisation permitted by (2.9). This is a consequence of the fact — demonstrated by Papoulis (1968, p. 197) for the 2D case — that any *real-valued* function minimising the localisation about the hyperplane $x_i = 0$ is Cartesian separable and has a *Gaussian* form along the other coordinate axes; simultaneous optimisation by a real function of the joint localisation about all hyperplanes $x_i = 0$ is therefore only achieved by a *Gaussian* whose principal axes are aligned with the coordinate axes. Motivated by the sketch of a proof provided by Gabor (1946), Stork & Wilson (1990) claimed to have shown that the 1D functions achieving *local* minima of the product of the spatial and spectral spreads defined in (2.10) are the derivatives of a Gaussian, of which the Gaussian itself is a special case. Their proof has since been criticised on two counts. Yang (1992) pointed out that the functions which satisfy the first-order conditions for a local optimum are Gaussian-windowed Hermite polynomials in which the standard deviation of the Gaussian is greater by a factor of $\sqrt{2}$ than that of the Gaussian derivatives mistakenly proposed by Stork & Wilson (1990). More importantly, however, Yang (1992) argued that both Gabor (1946) and Stork & Wilson (1990) had failed to check the second-order conditions for a local optimum, and based on empirical

evidence provided by Klein & Beutter (1992), he concluded that these functions in fact constituted saddle points for the joint uncertainty measure. Klein & Beutter (1992) furthermore showed that under a loose restriction on the class of permissible spatial RFP functions, the Gaussian-windowed Hermite polynomials, with the exception of the Gaussian itself, give local *maxima* of the joint uncertainty. Thus in addition to the objections to the derivative-of-Gaussian model raised in Section 2.3.1, the first order derivative exhibits a joint spread which exceeds the theoretical minimum by at least an order of magnitude (Stork & Wilson, 1990, Fig. 3), and this factor increases as the derivative order or the standard deviation of the Gaussian increases. Nevertheless, the GRGFs exhibit similarly poor joint localisation.

In order for the optimal joint localisation property of the Gabor functions to have any relevance to the early visual processing performed by the simple cells, it is necessary to show that the simple cell layer might feasibly implement *complex-valued* Gabor functions. Given the already-established resemblance between the GRGFs and the RFPs of the simple cells, the only feasible[22] scheme by which this might be achieved involves the direct implementation of the real and imaginary parts of each generalised Gabor function by a pair of simple cells whose spatial RFPs are respectively described by the GRGF pair

$$w(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi \quad , a) \tag{2.11a}$$

$$w(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi + \tfrac{\pi}{2} \quad , a) \tag{2.11b}$$

in spatial phase quadrature. To allow for the approximate half-wave rectification exhibited by simple cells, this basic scheme was augmented by Pollen & Ronner (1981; 1982; 1983) — for the case $\phi = 0$ — to include a second pair of simple cells in antiphase to the first with RFPs given respectively by

$$w(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi + \pi \quad , a) \tag{2.12a}$$

$$w(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi + \tfrac{3\pi}{2} \quad , a) \tag{2.12b}$$

Pollen & Ronner (1981) reported evidence for pairs of adjacent feline simple cells showing the quadrature relationship required by both (2.11) and (2.12) between their preferred phases at the optimal grating spatial frequency (magnitude and phase). Approximate phase quadrature was observed for all simple cell pairs for which two distinct responses could be isolated from a single-electrode recording. The same procedure was also used by Foster et al. (1983) to demonstrate the existence of antiphase simple cell pairs having approximately odd-symmetric RFPs. In both cases, some leeway was permitted

[22]The only conceivable alternative involves the alternate signaling of the real and imaginary parts of the inner product of the RFP with the image by a single simple cell. Justification for such a scheme might be derived from the observed translation of the pattern of subfield alternation with pre-spike time, as described in Section 2.2.4. However, since any such scheme would work only for stationary or near-stationary stimuli, this alternative will not be afforded further consideration.

in the comparison of optimal orientation ($\pm 5°$), spatial frequency (± 0.25 octaves) and phase ($\pm 10°$), and the spatial frequency tuning curves were not quantitatively compared. However, if — as these observations suggest — the scheme proposed by Pollen & Ronner (1981) is indeed implemented by the simple cells, high precision in the necessary relationships between both the phase and non-phase parameters of adjacent simple cells should not be expected, since according to the *Principle of Sloppy Workmanship* (Huggins & Licklider, 1951; Grzywacz & Yuille, 1990) — restated by MacLennan (1992a) as the *Robustness Principle* — the success of any computational scheme involving real neurons should not rely heavily on precision of mathematical detail. Nevertheless, the small number of cell pairs identified by these two groups as having the required phase relationships (12 and 4 respectively), the atypicality of the recording situation in which these relationships are observed (Pollen & Ronner, 1981), and the lack of independent confirmation of these results collectively raise doubts concerning the ubiquity of their observations (Stork & Wilson, 1990). It is notable therefore that in his recent defence of the biological implementation of complex-valued Gabor functions, Daugman (1993) sought and provided no new evidence in favour of the existence of the requisite quadrature-phase simple cell pairs.

As established earlier however, achievement of the optimal joint localisation permitted by (2.9) also requires alignment of the Gaussian component of each Gabor function with the coordinate axes, a requirement which conflicts with the observation by Jones et al. (1987) of approximate *radial* alignment of the simple cell spectral RFP. In general no single rotation of the coordinate axes can be found which allows simultaneous optimal joint localisation for all biologically relevant generalised Gabor functions. Thus even if we assume that the simple cells implement complex-valued generalised Gabor functions — by the above scheme or any other — they cannot collectively exhibit the optimal joint spatial and spectral localisation permitted by (2.9).

Assuming that the virtues of optimal joint spatial and spectral localisation extolled by Daugman (1989b) are nonetheless desirable, then rather than discounting the GRGF as a model of the simple cell RFP, the observed orientation dependence of the joint localisation measure in (2.9) points to a deficiency in the component-wise measure defined in (2.10) of the spread of the energy distribution. Introduced by Daugman (1985) and adopted by both Wechsler (1990) and MacLennan (1991), this generalised measure is the square root of the second moment of the energy distribution of the function about the hyperplane $x_i = 0$ ($u_i = 0$). The relevance of the corresponding joint localisation measure to biological vision has been questioned by Stork & Wilson (1990), who presented several alternative 1D localisation measures for which there appears to be no less *apriori* biological justification, and which are not optimised by the GRGFs. The generalisation of such alternative 1D localisation measures to nD should however avoid a component-wise definition — such as that in (2.10) — of the spread of the energy distribution, since

any such localisation measure will suffer from the same orientation dependence as that exhibited by the particular uncertainty principle in (2.9). The component-wise nD extension by Wechsler (1990) of the *joint entropic uncertainty* proposed by Leipnik (1959), for example, is optimised only by the nD Gabor functions with unrotated Gaussians.

Spatial localisation of the RFP according to some suitable localisation measure is likely to be important in reducing the biological cost of providing neural interconnections to support it. However, the results of Palmer et al. (1991) discussed in the previous section suggest that spectral localisation is enhanced by *nonlinear* mechanisms, which are not amenable to analysis using localisation measures based on the Fourier transform of the spatial RFP. This nonlinear enhancement of the spectral localisation was confirmed by Jones & Palmer (1987a) and Palmer et al. (1991), who plotted the effective areas $\Delta x_1 \Delta x_2$ and $\Delta u_1 \Delta u_2$ occupied by the best fitting GRGFs for the spatial and spectral RFPs respectively of the same simple cell, and found that many cells exhibited considerably *better* joint localisation than the theoretical optimum dictated by (2.9). This observation, which is difficult to discount simply on the basis of experimental or data-fitting errors since the joint localisation of any given GRGF is at least an order of magnitude *worse* than this theoretical optimum (Stork & Wilson, 1990, Fig. 3), is therefore especially damning for joint localisation analyses based on assumptions of spatial linearity.

2.3.4 *Spatiotemporal RFP*

The 2D spatial RFP obtained by Jones & Palmer (1987b) and Palmer et al. (1991) for each simple cell was strictly the integral of the 3D spatiotemporal RFP over the prespike time bin — of width 50ms and 1ms respectively — for which the strongest overall response was obtained. These time "slices" however constitute an incomplete characterisation of the linear behaviour of the simple cell, since they reveal nothing about the temporal structure of the spatiotemporal RFP, and for those simple cells which exhibit Cartesian inseparability of their spatiotemporal RFP, provide an inadequate description of even the spatial dependence. Nevertheless since these time "slices" are well characterised by the 2D spatial GRGF model, this model is commonly used as a starting point for a more complete description of the 3D spatiotemporal RFP of the simple cell.

The most direct extension of the nD spatial GRGF is the $n+1$ dimensional GRGF $w(\boldsymbol{x})$ whose spatiotemporal argument $\boldsymbol{x} \equiv [\boldsymbol{x}', t]^T \in \mathbb{R}^{n+1}$ is the concatentation of the spatial variable $\boldsymbol{x}' \in \mathbb{R}^n$ and the temporal variable $t \in \mathbb{R}$. First suggested by Adelson & Bergen (1985) for one spatial dimension ($n = 1$) and $\phi = 0$, the corresponding spatiotemporal model is in general Cartesian inseparable[23], exhibiting the type of oriented spatiotemporal subfields demonstrated by McLean & Palmer (1989) for a section of the

[23]although the corresponding spatiotemporal Gabor function $g(\boldsymbol{x})$ in its complex form is Cartesian separable.

3D spatiotemporal simple cell RFP along the axis of subfield alternation. This property is not for example shared by the Cartesian separable Gaussian derivative RFP model proposed by Martens (1990). The spatiotemporal GRGF model has since been extended to $n = 2$ spatial dimensions by Heeger (1987), and to arbitrary spatial dimensions by MacLennan (1991), in both cases for $\phi = 0$. To date however, no attempt appears to have been made to fit this model to the simple cell spatiotemporal RFP.

Unlike the ideal velocity-selective filter, which is a plane passing through the origin in the spatiotemporal frequency domain, the GRGF is tuned to a Gaussian-shaped region of the spatiotemporal frequency domain. Since any quadrature-phase pair (2.11) of spatiotemporal GRGFs approximate a Hilbert transform pair (Adelson & Bergen, 1985), the outputs of a quadrature pair may be squared and added to calculate an approximation to the *motion energy* (Adelson & Bergen, 1985) of the visual scene in the spatiotemporal frequency band to which the pair is tuned. The approximate stimulus velocity can then be inferred from the outputs of the corresponding Gabor motion energy filters (Heeger, 1987; Grzywacz & Yuille, 1990). The existence of such quadrature-phase spatiotemporal GRGF pairs amongst the simple cells — which has however yet to be established — would also lend credence to the invocation by MacLennan (1991) of the 3D spatiotemporal Uncertainty Principle (2.9) in justification of the possible use of these 3D GRGFs by the simple cells. However, this principle remains unable to account for GRGFs having Gaussians not aligned with the coordinate axes.

A major problem with the GRGF model of the spatiotemporal RFP of the simple cell is that it is *non-causal*, so that the current output of the cell is dependent on future inputs. However, the effect of truncation at the plane $t = 0$ to ensure causality will be negligible provided the temporal centre of the GRGF is located sufficiently far into positive time. An alternative model of the simple cell spatiotemporal RFP which does not violate the causality constraint is that of Watson & Ahumada (1983; 1985), who used a GRGF and a gamma function respectively as the spatial and temporal components of a separable spatiotemporal filter. Whilst this filter is not by itself direction selective, the linear combination of the outputs of the corresponding filter and its spatiotemporal Hilbert transform — whose spatial dependence is approximated by a 2D spatial GRGF in quadrature phase to the first — results in an output which is both direction selective and appropriately tuned to spatiotemporal frequency. The separable filter produced by the Hilbert transform of the original is however non-causal, and truncation at $t = 0$ is once again required. With this modification, the two separable filters and the resultant inseparable filter are respectively plausible models of the spatiotemporal RFPs of direction-symmetric and direction selective simple cells.

2.3.5 *Binocular RFP*

The GRGF model of the binocular RFP of the simple cell proposed by Nomura et al. (1990) uses the pair of 1D GRGFs

$$w_L(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi \quad , a) \tag{2.13a}$$

$$w_R(\boldsymbol{x}; \boldsymbol{x}_0, S, \boldsymbol{u}_0, \boldsymbol{\theta}, \quad \phi+\delta\phi \quad , a) \tag{2.13b}$$

— whose parameters are identical except for a sinusoidal phase shift $\delta\phi$ — to describe the left- and right-eye receptive field profiles of a given simple cell. In particular, the Gaussians of the two RFPs are centred at the same retinal coordinates in the left and right eyes — i.e. at zero retinal *disparity* — while the phase shift $\delta\phi$ is used to control the depth — relative to the current point of fixation — to which the cell is optimally tuned. The phase shift is therefore observable as an interocular shift of the RF subfields relative to the common Gaussian, in general agreement with the experimental observations of Freeman & Ohzawa (1990b). With appropriate choices of the parameters $\delta\phi$, u_0, s and the output threshold, this model was found to account well for the different types of depth tuning reported by Poggio & Fischer (1977) and revised by Poggio et al. (1988) — viz. tuned excitatory (TE), tuned inhibitory (TI), near, far, tuned near, and tuned far — as well as the corresponding degree of ocular dominance. This model also accounts in a natural way for the intermediate types of depth tuning reported by LeVay & Voigt (1988), suggesting that the above depth tuning categories may represent a somewhat artificial division of a continuum formed by the continuous variation of the parameters in (2.13).

In addition to successfully accounting for the tuning of simple cells to stereoscopic depth via *horizontal* binocular disparity, the above phase-shifted GRGF model of the binocular spatial RFP also predicts — given the approximately uniform distribution of orientation preference amongst the simple cells — the additional tuning of cells in primary visual cortex to vertical disparities as observed by Barlow et al. (1967). However, the computational role of vertical disparity sensitivity in early vision remains unclear.

2.3.6 *Summary*

It was argued in Section 2.2 that with the appropriate reservations, the simple cell may be treated to a first approximation as a linear device characterised by its binocular spatiotemporal RFP. In the present section, it has been shown that the monocular spatial form of this RFP is well described by a two-dimensional generalised real-valued Gabor function (GRGF), and that this model can be naturally extended to incorporate both the spatiotemporal and binocular behaviour of the cell, facilitating an understanding of the computational role of the simple cell in these augmented domains. Poor agreement between two key parameters derived from the Fourier transform of the GRGF which

best describes the spatial RFP and the best-fit GRGF for the experimentally determined spectral RFP is at least in part attributable to the action of nonlinear mechanisms, and does not in itself invalidate the GRGF model of the spatial RFP. The component-wise nD extension of Weyl's (1932) 1D Uncertainty Principle commonly used as justification for the complex-valued Gabor function RFP model has been shown to be inapplicable for realistic choices of the GRGF parameters, thereby releasing the modeller from the need to maintain the unrealistic assumption of infinite spatiotemporal support for the RFP.

2.4 Conclusion

Visual stimuli presented within the receptive field of the simple cell are in general subjected to nonlinear binocular spatiotemporal processing, the linear component of which is characterised by the binocular spatiotemporal RFP. Whilst the approximation involved in neglecting the second- and higher-order terms of a complete nonlinear characterisation of this processing has yet to be quantified, it has been argued here that the RFP accounts at least qualitatively for a number of the experimental observations concerning simple cell processing of visual stimuli. The monocular spatial RFP, to which (as will be seen in the next chapter) consideration is frequently restricted, is strictly speaking only defined in the case where the chosen monocular spatiotemporal RFP is Cartesian separable into temporal and spatial components. Nevertheless, to the extent that such a characterisation is valid, the GRGF is arguably the best, and certainly the most extensively and accurately tested, model of the simple cell spatial RFP proposed to date.

Chapter III

ON THE COMPUTATIONAL ROLE OF THE SIMPLE CELLS

3.1 Introduction

In the previous chapter, it was established that a simple cell may to a first approximation be treated as a linear device characterised by its binocular spatiotemporal RFP. This characterisation suggests a role for the simple cells in the processing and encoding of information regarding the spatial form, motion and stereoscopic depth of the visual stimulus. What is lacking, however, is a unified theory of simple cell processing which provides both a realistic model of the binocular spatiotemporal RFP, and an account of the variation of the model parameters over the simple cell population. Indeed experimentally testable candidates for the first essential ingredient of such a theory are conspicuously absent. The limited and largely qualitative nature of experimental evidence in favour of RFP models which address even the binocular spatial or monocular spatiotemporal domains furthermore suggests that consideration should in the mean time be restricted to the monocular spatial domain, where the GRGF model enjoys comparatively strong experimental support (Jones et al., 1987).

Motivated initially by theories postulating a role for the simple cells in a local Fourier analysis of the visual image, the GRGF model of the monocular spatial RFP presented in Section 2.3.1 has been appropriated by a number of competing theories of the computational role of the simple cells, ranging from edge- and line-detection, through the computation of spatial derivatives, to multiresolution image analysis. In the search for a realistic and unified theory of simple cell processing, the present discussion is restricted primarily to theories which use either the GRGF or a similarly realistic model of the monocular spatial RFP of the simple cell, and which could conceivably account for the experimentally observed variety of spatial RFPs.

An important proviso on the restriction of attention to the spatial domain is the fact that the isolated consideration of monocular spatial processing to the exclusion of the binocular and temporal domains may, as noted in the previous chapter, overlook potentially important features of simple cell processing, such as the encoding of motion in depth. However it should be noted in mitigation that at least some proportion of the simple cells may be either monocular or motion-insensitive up to a first-order (linear)

characterisation. The extent of overlap of these two populations (if indeed they form distinct populations at all) and the relative contribution of this overlap to the overall simple cell population is however unclear. In general therefore, theories which result from purely spatial descriptions of the simple cells are at best incomplete, and await future extension to the binocular spatiotemporal domain.

3.1.1 Bottom-up vs. Top-down

The requirement that any suitable theory of simple cell processing should provide an accurate account of experimentally observed RFPs and their variety is in line with the analytic approach to early vision characteristic of the nascent field of *computational neuroscience*, according to which theories of vision involving the simple cells, for example, should be based on an accurate empirical determination of what they compute (Palmer et al., 1991). Such theories often assume little or nothing about the subsequent processing of the simple cell outputs other than that the representation of the retinal image at this level should be perceptually *complete* (Daugman, 1990; Geisler & Hamilton, 1986), so that no spatial contrast information which is known to be used by higher level perceptual processes is removed by the simple cells. This approach can be described as *bottom-up*, in the sense that theories of higher level processing are dependent on accurate knowledge of the processing performed by earlier stages in the visual pathway.

An alternative *top-down* or *synthetic* approach to vision proposed and strongly advocated by Marr (1982) requires that one first decide *what is to be computed* by a particular visual subsystem and only then decide *how* it might be computed by the available neural hardware implicated in that visual task. In the interests of parsimony, RFP models resulting from such hypotheses concerning, for example, the role of the simple cells in a given computational task generally exhibit the minimum number of degrees of freedom necessary to fulfil the purported role. The simple cell model used by the *Boundary Contour System* (Grossberg et al., 1989; Shapley et al., 1990) is for example odd symmetric and varies only in its preferred orientation. Since the number of degrees of freedom required — and in some cases permitted — by RFP models resulting from the top-down approach falls far short of the number observed by Jones & Palmer (1987b) in real simple cells, these models provide at best an incomplete explanation of the computational role of the simple cells in early vision. Different top-down hypotheses may also lead to either irreconcilable models of the simple cell RFP or irreconcilable schemes for the systematic variation of certain RFP parameters over the simple cell population, so that the associated computational schemes could only be subserved in the same visual system by two or more mutually exclusive or only partially overlapping populations of simple cells. Since the aim of the present exposition is to examine theories which attempt to provide a realistic and unifying account of the form and variety of simple cell spatial RFPs, most

top-down models are therefore automatically excluded from consideration[1].

3.1.2 *Qualified Completeness*

The requirement of most bottom-up theories that the representation of the visual image by the simple cells be complete is necessarily subject to limitations imposed by the mechanisms by which the visual scene is converted into a retinal and subsequently cortical image. We assume for simplicity an array of photoreceptors whose RFPs are described by the function $h:\mathbb{R}^2 \rightarrow \mathbb{R}$ centred at possibly non-uniformly spaced visual angles $\boldsymbol{x}_i \in \mathcal{S}$. The retinal image $b(\boldsymbol{x}_i, t)$ — where $b: \mathcal{S} \times \mathbb{R} \rightarrow \mathbb{R}$ — is computed by spatially correlating the monocularly viewed stimulus $s(\boldsymbol{x}, t)$ with the low-pass spatial filter kernel $h(\boldsymbol{x})$, and then sampling the resultant image[2] at the retinal coordinates corresponding to the visual angles $\boldsymbol{x}_i \in \mathcal{S}$. The retinal image can therefore be expressed as

$$b(\boldsymbol{x}_i, t) = [h * s](\boldsymbol{x}_i, t) \tag{3.1}$$

where $*$ denotes the spatial correlation operation such that

$$[h * s](\boldsymbol{x}, t) \equiv \int_{\mathcal{V}} h(\tilde{\boldsymbol{x}} - \boldsymbol{x}) s(\tilde{\boldsymbol{x}}, t)\, d\tilde{\boldsymbol{x}} \tag{3.2}$$

If the retinal sampling is assumed to be spatially uniform, the photoreceptor RFP h can be viewed as the kernel of an *anti-aliasing filter* (Geisler & Hamilton, 1986), which serves to limit the destructive effect of aliasing caused by retinal sampling[3]. This view necessitates a qualification of the concept of completeness as it applies to the simple cell representation of the visual image, since only the correspondingly spatially low-pass filtered and possibly aliased version $\check{s}(\boldsymbol{x}, t)$ of the visual stimulus $s(\boldsymbol{x}, t)$ is available for encoding by subsequent processing stages. Except where otherwise stated, this qualification is assumed to be implicit in the following discussions.

The output $r(t)$ of a simple cell having *retino-cortical* spatial weighting function $c: \mathcal{S} \rightarrow \mathbb{R}$ is then given by

$$r(t) = \sum_{\boldsymbol{x}_i \in \mathcal{S}} c(\boldsymbol{x}_i) b(\boldsymbol{x}_i, t)$$

---

[1] As a philosophical aside however, a combination of the top-down and bottom-up approaches is probably necessary to significantly further our understanding of biological vision systems. In practice for example, neither approach can or should avoid the iterative loop linking the development of visual theory with the testing of the predictions of that theory against electrophysiological observation. This loop is useful both in guiding the development of top-down algorithms which have a plausible neural implementation, and in avoiding the bottom-up modelling of details of the RFP which are not crucial or are even irrelevant to the functioning of the complete visual system.

[2] This conceptual division of the imaging process into two distinct stages is merely an artefact of the mathematical formalisation, and is not intended to imply the existence of correspondingly distinct physical processes.

[3] There is evidence to suggest that at least in primates, optical diffraction may impose a more severe anti-aliasing effect on early vision than the photoreceptor RFP (Levick, 1993); however, this observation does not qualitatively affect the ensuing conclusion.

$$= \int_{\mathcal{V}} \left[\sum_{\boldsymbol{x}_i \in \mathcal{S}} c(\boldsymbol{x}_i) h(\tilde{\boldsymbol{x}} - \boldsymbol{x}_i) \right] s(\tilde{\boldsymbol{x}}, t)\, d\tilde{\boldsymbol{x}} \tag{3.3}$$

where the expression in (3.3) is obtained by substituting for $b(\boldsymbol{x}_i, t)$ using (3.1), expanding the correlation using (3.2), and reversing the order of the (finite) sum and the (well-behaved) integral. Comparison of (3.3) with (2.1) reveals that the spatial RFP $w(\boldsymbol{x})$ of the simple cell is given by

$$w(\boldsymbol{x}) = \sum_{\boldsymbol{x}_i \in \mathcal{S}} c(\boldsymbol{x}_i) h(\boldsymbol{x} - \boldsymbol{x}_i) \tag{3.4}$$

The simple cell spatial RFP is therefore given by the convolution of the photoreceptor RFP h with the weighted, possibly irregular "bed-of-nails" function c. For evenly spaced sampling points, the former may be viewed as a *reconstruction filter* (Carlson, 1986) for the latter, band-limiting the resultant simple cell RFP $w(\boldsymbol{x})$ to the same frequency range as the retinal image prior to sampling[4]. The expression in (3.4) is also readily extended to account for variations of retinal sampling density or photoreceptor RFP size, as occur for example with increasing retinal eccentricity.

3.1.3 *Filtering and Decomposition*

Bottom-up theories of the spatial processing performed by the simple cells fall into the two main classes of *filtering* and *decomposition*.

Theories of visual cortical spatial filtering relate the simple cell spatial RFP w to the kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ of a position-dependent linear filter whose output $r(\boldsymbol{x}, t)$ is given by

$$r(\boldsymbol{x}, t) = \int_{\mathcal{V}} h(\tilde{\boldsymbol{x}}, \boldsymbol{x}) s(\tilde{\boldsymbol{x}}, t)\, d\tilde{\boldsymbol{x}}$$

The value $r(\boldsymbol{x}_0, t)$ of the filtered image at any given spatial coordinate $\boldsymbol{x}_0 \in \mathbb{R}^2$ is hypothetically represented by the output of a simple cell located at the position $\boldsymbol{x}_0$ in the simple cell layer and having the RFP $w(\boldsymbol{x}) \triangleq h(\boldsymbol{x}, \boldsymbol{x}_0)$. In general the filter may be position-dependent, with the form of $w(\boldsymbol{x})$ varying as a function of the output coordinate $\boldsymbol{x}_0$ (Gutschow & Hecht-Nielsen, 1991), a feature which is of potential biological interest in modelling for example the increase in mean receptive field size with retinal eccentricity. Position-dependent filtering schemes are however excluded from the present discussion due to the paucity of systematic experimental information regarding the position dependence of simple cell spatial RFPs and the lack of available formal completeness results for such schemes. Given these difficulties and the added computational complexity of simulating position-dependent filtering schemes, many theories of visual filtering make the simplifying assumption of position-independence, in which case

$$r(\boldsymbol{x}, t) = \int_{\mathcal{V}} w(\tilde{\boldsymbol{x}} - \boldsymbol{x}) s(\tilde{\boldsymbol{x}}, t)\, d\tilde{\boldsymbol{x}} \tag{3.5}$$

[4]The term "band-limited" should be interpreted loosely in this context, since in practice the photoreceptor RFP h may not be ideally band-limited.

and $w(\boldsymbol{x})$ is the *correlation kernel* of the linear filter. Implementation of the position-independent filter requires an array of simple cells with identical RFPs whose centres densely populate the visual field $\mathcal{V}$. Since, as shown in Section 3.1.2, w is (approximately) band-limited in the spatial frequency domain, the Sampling Theorem (Bracewell, 1986) can be invoked to show that the cortically filtered image can be sampled at an appropriate rate without further loss of information[5]. This sampling corresponds to a reduction in the necessary density of population of the visual field by the RF centres, and hence to a reduction in the number of simple cells required to implement the filter. If two or more filters, each implemented by a sub-population of simple cells, are assumed to be applied simultaneously to the visual image, the requirement for completeness of the individual filters can be relaxed, provided that collectively the filters continue to transmit all the perceptually relevant information. Theories concerning the spatial filtering possibly performed by the simple cells are presented in sections 3.2.2, 3.3, 3.4.5 and 3.5.2.

Theories of visual cortical spatial decomposition, on the other hand, assume that the spatial RFPs of primary visual cortical cells such as the simple cells "constitute the primitives of the biological image code", the "relative presences" of which are signalled by the firing rates of the corresponding cells (Daugman, 1990). According to this view, the stimulus is decomposed into a set $\{c_i(t): i \in \mathbb{Z}_+\}$ of coefficients $c_i(t) \in \mathbb{R}$ corresponding to the set $\{w_i(\boldsymbol{x}): i \in \mathbb{Z}_+\}$ of spatial RFPs such that the spatially low-pass filtered visual stimulus $\check{s}(\boldsymbol{x}, t)$ can be expressed as the expansion

$$\check{s}(\boldsymbol{x}, t) = \sum_i c_i(t) w_i(\boldsymbol{x}) \tag{3.6}$$

and hence $\check{s}$ could be reconstructed if desired from the set of coefficients. The representation of the visual stimulus by the coefficients $\{c_i(t)\}$ is complete if for each image encountered by the visual system there exists a corresponding set of coefficients satisfying (3.6). Assuming in the absence of further information that these images are drawn from the set $L^2(\mathbb{R}^2)$ of finite energy (square-integrable) 2D functions, and denoting by $\mathcal{B} \subset L^2(\mathbb{R}^2)$ the subset of these functions which are appropriately band-limited to the spatial frequency range passed by the retinal imaging process, it is therefore necessary that the set of simple cell spatial RFPs be complete over the set $\mathcal{B}$. Theories concerning the spatial decomposition possibly performed by the simple cells are presented in Section 3.4.

Whilst requiring that any image $\check{s}$ be *reconstructible* from the corresponding coefficients $\{c_i(t)\}$ using (3.6), the completeness condition should not be interpreted as implying that such a reconstruction actually takes place. Recognising the need for this distinction, Daugman (1990) noted that theories concerning the spatial decomposition of the stimulus

[5]Since the band-limitation of w may not be ideal, sampling may in fact incur a further small loss of information due to aliasing.

> ... walk a kind of epistemological tightrope ... demanding that the representation be informationally complete over its domain, while at the same time avoiding the implication that what is entailed by completeness is a cortical image reconstruction (as if on behalf of cortical movie-viewing homunculi).

However the implicit assumption often accompanying spatial decomposition theories of early vision (as documented later in Section 3.4.3), that the output

$$r_i(t) = \int_{\mathcal{V}} w_i(\boldsymbol{x}) s(\boldsymbol{x}, t)\, d\boldsymbol{x}$$

of each simple cell represents (up to a scalar constant) the coefficient in (3.6) corresponding to its own spatial RFP, is problematic. Except for certain special choices of the set of simple cell RFPs — viz. when these functions form an orthonormal basis or a *tight frame* (see e.g. Heil & Walnut (1989)) for the set $\mathcal{B}$ — these outputs will not in fact be the coefficients required by (3.6) (Daugman, 1988a; Martens, 1990). In an attempt to reconcile the apparently conflicting requirements arising from this assumption, a computational scheme whereby the simple cells might compute these coefficients is developed in Chapters 4–6 and critically examined in Chapter 7. In the mean time, sections 3.4.3, 3.4.4 and 3.5.4 of the present chapter pursue the search for suitable special-case RFP sets. Since for most candidate models discussed in Section 2.3.1 for the simple cell spatial RFP the requirements of completeness and mutual orthogonality are conflicting ones, this search focuses on results provided by the theory of tight frames.

3.1.4 Verification of Bottom-up Theories

Bottom-up theories of the spatial processing performed by the simple cells commonly bear on the population distributions of the various RFP parameters. In particular, in order to simplify attempts to ensure representational completeness, both filtering- and decomposition-based theories usually assume highly regular spatial &/or spectral (spatial-frequency) sampling of the retinal image by the simple cell RFPs. Such regular sampling is difficult to refute by means of single- or even current multi-electrode recording techniques, which are hampered by neural sampling biasses introduced among other things by the choice and possibly laminar placement of a particular microelectrode (Robson, 1983; Anderson et al., 1990, p. 215). Electrophysiological evidence for and against such models is therefore frequently confined to a comparison of the correlations between various RFP parameters predicted by the model, and those observed in the recorded population.

3.1.5 Hierarchical Processing

The terms "stage" and "level" used here in connection with the simple cells are strictly speaking a legacy of the hierarchical scheme of visual processing proposed by Hubel

& Wiesel (1962; 1968), according to which the simple cells constitute the first cortical processing stage in the visual pathway, receiving input from neurons in the LGN and sending their outputs to the complex cells. The term "layer" which is sometimes used interchangeably with these terms in this context should not however be confused with the anatomically defined layers of the visual cortex (see e.g. Lund (1988) and Gilbert (1983)). Simple cells are found in cortical layers 2 to 6 of primary visual cortex, and whilst they predominate in layer 4 in which the majority of afferents from LGN are known to terminate[6], simple cells are also found in other layers in which complex cells are at least equally prevalent (see White (1989) for a review). Critics of the strictly hierarchical model — including Stone et al. (1979) and White (1989) — cite evidence at various stages of the visual pathway for connections which circumvent one or more such processing levels, as well as the existence of feedback between various levels. Of particular relevance to the present consideration of the simple cells is the presence of direct synaptic input to V1 complex cells from LGN (White, 1989; Henry et al., 1983) and the probable existence of feedback connections from complex cells onto simple cells, which has been inferred from a number of independent observations (White, 1989, p. 144,187). These findings suggest that the requirement of completeness of the simple cell representation may not be a necessary one, since if pathways exist through which perceptually relevant, visually derived spatial information may reach higher cortical processing areas without passing through the simple cell "stage", such information need not be independently represented by the simple cell population.

Bottom-up theories regarding the spatial computation performed by the simple cells furthermore frequently make the simplifying assumption that these cells respond ideally only to spatial variations of achromatic contrast in the visual image (see e.g. Sakitt & Barlow (1982) or Field (1987)), and exhibit no selectivity for other aspects of visual input such as motion, colour or stereoscopic depth. More recent physiological models of visual processing (Stone et al., 1979; Hubel & Livingstone, 1987) however posit the existence of parallel visual pathways responsible for different combinations of these various aspects of visual input, and which are either anatomically segregated or mixed only selectively at various levels of the visual pathway. The above assumption that the simple cells collectively constitute an homogeneous population concerned exclusively with the processing of spatial information therefore ignores for example the putative segregation (see e.g. White (1989)) of the *X & Y* (Stone et al., 1979; Gilbert, 1983) or *parvo & magno* (Hubel & Livingstone, 1987) streams in layer 4 of feline and primate striate cortex respectively.

[6]White (1989) however points out that the termination of the thalamocortical afferents in layer 4 does not preclude direct LGN input to cells in other layers having dendrites in this layer.

3.1.6 Summary

The present examination of the computational role of the simple cells will focus on bottom-up theories of the monocular spatial processing performed by these cells. The binocular and temporal characteristics of the simple cell RFP are not addressed by the class of theories considered. Where suitable results exist, the qualified completeness or otherwise of potential simple cell image representations will be examined, as well as the ability of each representation to provide a unifying account of the observed variety of simple cell RFPs. Acknowledged shortcomings of bottom-up theories include the failure to incorporate feedback from higher processing areas to mediate attentional and expectational mechanisms[7], and the assumption of a strictly hierarchical processing scheme. Direct verification of such theories is furthermore hampered by experimental difficulties including electrode sampling bias, necessitating a resort to information concerning correlations between RFP parameters over the simple cell population.

3.2 Feature "Detectors"

3.2.1 Nonlinear Detectors

The observation by Hubel & Wiesel (1962) that for simple cells

> The most effective stimulus configurations ... were long narrow rectangles of light (slits), straight-line borders between areas of different brightness (edges), and dark rectangular bars against a light background.

led Barlow (1969a) to postulate that the simple cells were in fact feature detectors whose *trigger features* were luminance bars and edges at a particular binocular disparity (Barlow et al., 1967; Barlow, 1969b). According to this feature detection hypothesis in its crudest form (Barlow, 1969b, p. 220), the output of a simple cell is interpreted as a binary variable signalling the presence or absence of the trigger feature in its receptive field. A more sophisticated form of the hypothesis (Barlow, 1969a; Daugman, 1990) has the firing rate of a feature-detecting simple cell signalling the cell's degree of *certainty* of the presence of the trigger feature. Nevertheless, the firing of a cell is to be interpreted in both cases as making a "symbolic assertion" (Marr & Hildreth, 1980) regarding the presence of the trigger feature.

Edge extraction as a basis for early vision has strong intuitive appeal, given the apparent abundance of information available from such edges (Marr, 1976; Marr & Hildreth, 1980; Marr, 1982; Ullman, 1986) and the lower redundancy (Barlow, 1972) of an edge-based representation, and was for example central to the formation of the *primal*

[7]It has been estimated that only 1% of the neural fibres entering V1 are visual afferents (von der Malsburg, 1990).

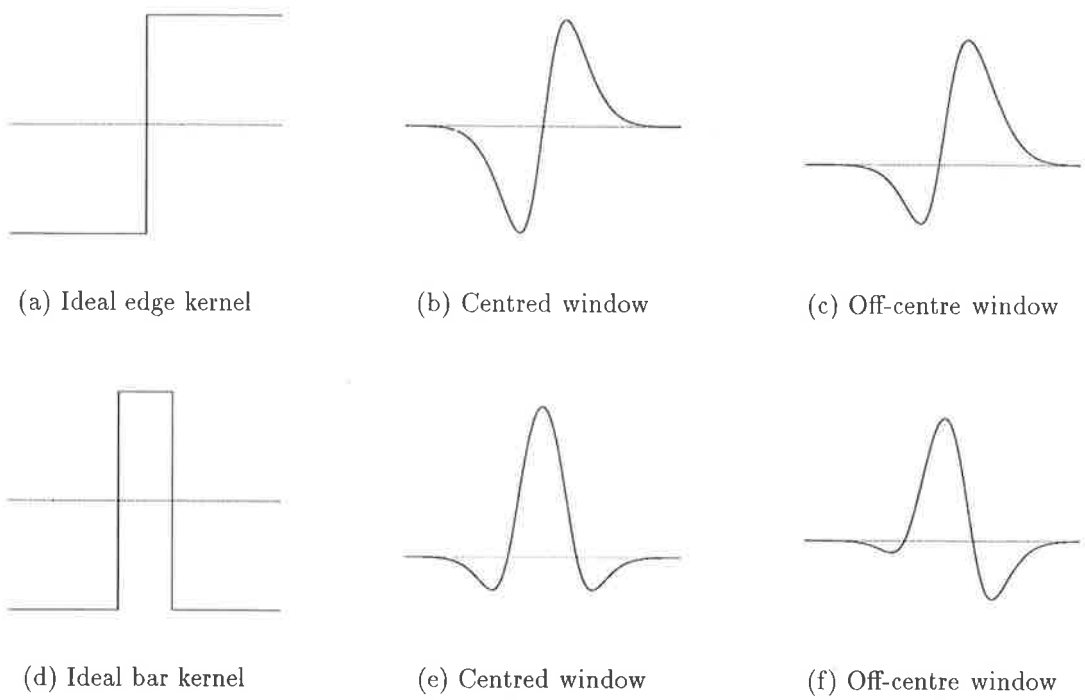
sketch representation used by Marr (1976; 1977; 1982). Early theoretical support for this approach was drawn from results guaranteeing under certain circumstances the complete reconstruction of a multiresolution representation[8] of a signal from its zero crossings (Logan, 1977; Marr et al., 1979), since for the particular multiresolution representation proposed by Marr & Hildreth (1980), collocation and linear alignment of these zero crossings across several spatial scales can be taken as evidence for a luminance edge (Marr & Hildreth, 1980; Marr, 1982). More recently, zero crossing representations have also been proposed in the field of pattern recognition to overcome the lack of translation invariance exhibited by straightforward multiresolution schemes (Mallat, 1989a).

However, Logan's theorem (1977) has been shown to be inapplicable to Marr & Hildreth's (1980) multiresolution Laplacian of Gaussian decomposition (Marr, 1982; Daugman, 1983; Daugman, 1988b), necessitating the incorporation of additional information to ensure complete reconstruction (Marr et al., 1979; Rotem & Zeevi, 1986; Curtis et al., 1989; Mallat, 1989a). Furthermore both the feature detection hypothesis of Barlow (1969a; 1969b) and Marr & Hildreth's (1980) Laplacian zero crossing detector model predict simple cell behaviour which is strongly spatially nonlinear (Marr, 1982), in contrast with later evidence — reviewed earlier in Section 2.2.3 — showing that the spatial behaviour of the simple cells is to a first approximation linear. The response of a simple cell to the third harmonic of a drifting square-wave grating is for example difficult to explain in terms of the detection of the passage of luminance edges over the RF, and militates against a view of the simple cells as edge or line (bar) detectors (Pollen & Ronner, 1982; Henry, 1985).

3.2.2 Linear Matched Filters

The notion of a feature "detector" is severely restricted by the characterisation of the simple cell as a linear device, since in attempting to suppress reponses to non-optimal features falling within the cell's RF, a linear device is denied recourse to the nonlinearities required by models such as those of Barlow (1969a; 1969b) and Marr & Hildreth (1980). The spatial RFP w might however be treated as the kernel of a linear *matched filter* for the desired feature, whose output $r(\boldsymbol{x}, t)$ is given by (3.5). A matched filter is used to improve the detection of a known signal — in this case for example an edge or a line — in the presence of noise (see e.g. Papoulis (1984)). Assuming that such a filter is indeed implemented by an array of simple cells, as described earlier in Section 3.1.3, the simple cell outputs would be subjected by subsequent cortical layers to some form of nonlinearity such as thresholding to detect the presence or absence, as well as position(s),

[8] Multiresolution image representations are discussed briefly in Section 3.5; see also e.g. Mallat (1989b).



(a) Ideal edge kernel (b) Centred window (c) Off-centre window

(d) Ideal bar kernel (e) Centred window (f) Off-centre window

Figure 3.1: 1D cross-sections of Cartesian-separable 2D matched filter kernels for edges and bars. The dotted line indicates the zero weighting level. (a),(d) Ideal edge and bar kernels respectively. (b),(e) Ideal kernels after localisation using a Gaussian window centred on the feature to be matched, and low-pass filtering representing the transformation from visual scene to retinal image. These line weighting functions compare qualitatively with those in Figures 2.1(c) and (f) which are derived from GRGFs. (c),(f) As for (b),(e) except with Gaussian windows centred to the right of the feature to be matched.

of the desired feature in the visual field[9].

One dimensional cross-sections of the Cartesian-separable kernels of ideal matched filters for bars and edges in the visual image are shown in Figures 3.1(a) and (d) respectively. The sections are perpendicular to the direction of elongation of the edge or bar to be detected. Since the sharp transitions in the corresponding stimuli are removed by the optical transformation mapping the visual scene onto the retinal image and are not therefore available to be matched by the filter kernel, these kernels have been similarly low-pass filtered to facilitate a better match to simple cell LWFs[10]. Spatial localisation of the smoothed kernels to reduce the number of retino-geniculo-cortical connections required by each simple cell implementing them can be achieved by multiplying by a

[9]This deferral of the actual detection process until after the simple cell stage avoids the "intellectually criminal" treatment of the simple cells as both linear convolvers and feature detectors, which was condemned by Marr (1982).

[10]The comparison of 1D cross-sections and LWFs — obtained by integrating along the perpendicular direction — of a hypothetical 2D RFP is strictly speaking only valid if the RFP is Cartesian separable. However it is assumed here for simplicity that approximate Cartesian separability applies.

windowing function such as a truncated Gaussian. The resultant localised kernels are illustrated in Figures 3.1(b) and (e) for windows centred on the feature to be matched, and in Figures 3.1(c) and (f) for windows which have been placed off-centre to account for asymmetric RFPs. The edge and bar filters exhibit characteristically bipartite and tripartite LWFs respectively. However, despite the prevalence (Glezer et al., 1989) of simple cell RFPs exhibiting the 2 and 3 subfields required respectively by such kernels, the matched filtering of bars and edges is inconsistent with RFPs exhibiting up to 8 subfields (Glezer et al., 1989), which in the present context should be viewed as kernels for "grating filters". Furthermore whilst the line weighting functions produced by these edge and bar filters compare qualitatively with those resulting from GRGF models, the assumption of Cartesian separability on which they rely is not however justified by the results of Jones & Palmer (1987b).

An undesirable consequence of the localisation achieved by windowing the ideal matched filter kernels is a marked decrease in stimulus selectivity, so that for example a bar filter now responds more vigorously to a high-contrast edge than to a low-contrast bar (Marr & Hildreth, 1980). This problem is exacerbated by offsetting the windowing function from the centre of the bar kernel, which increasingly suppresses the extreme flank in the direction opposite to that of the offset[11], and thereby increases the resemblance between this kernel and the (negative of the) centred *edge* kernel in Figure 3.1(b). Thus the ability of the resultant filters to signal the presence or absence of the stimulus feature to which they are supposedly matched is greatly diminished by localisation of the filter kernel. Notwithstanding these objections, the requirement for the repetition of an identical kernel for each point in the visual field in order to implement a given filter can be relaxed to allow RFPs of differing (a)symmetries at each point, provided each such composite matched filter continues to employ only bipartite or tripartite RFPs.

Under the linear feature detection hypothesis, *at least* one such filter is required for each feature which is to be reliably detected anywhere on the retinal surface; the reliable detection and localisation of edges has for example been found to require the local agreement of edge detection mechanisms at several spatial scales (Marr & Hildreth, 1980). Whilst a position-dependent matched filter could be designed to detect different features at different positions, the detection of different features at a single position still requires multiple matched filters. The enumeration of all possible lines and edges at a given position in the visual field — by independently varying attributes such as scale and orientation — would require a large number of simple cells to implement such a matched filtering scheme. However, estimates of the order of $2-3 \times 10^3$ cells in striate cortex per foveal cone photoreceptor (Wilson et al., 1990)[12], of which perhaps 20% are located in

[11]The left-hand flank in Figure 3.1(f).

[12]This figure reflects for the foveal representation an approximately 1:1 ratio of cones to retinal ganglion cells, a 2–3:1 ratio of LGN cells to retinal ganglion cells, and a 1000:1 ratio of striate cortical

layer 4 (Barlow, 1981, Fig 7) where the simple cells predominate (White, 1989), do not exclude the implementation of at least several hundred matched filters, which may be quite sufficient for most purposes.

The matched filtering hypothesis admits natural extensions to the temporal and binocular domains, which are necessary to account properly for the binocular spatiotemporal behaviour of real simple cells. However this extension requires for each spatial feature the additional enumeration of all relevant binocular disparities and temporal patterns[13]. The consequent explosion in the number of simple cells required, and the relatively poor stimulus selectivity of the localised filter kernels in the first place, render the matched filtering scheme untenable as a theory of simple cell processing.

3.2.3 Summary

In this section it has been argued that linear and nonlinear feature detection, and in particular the detection of contrast edges or lines in the monocular visual field, are unlikely to provide a complete description of the spatial computational role of the simple cells. Despite the improbability of such feature-detecting roles however, some researchers continue to refer to bipartite and tripartite RFPs — and especially those with approximate odd and even symmetries respectively — by the prejudicial terms "edge and bar detectors". Such terms are seen by Koenderink & van Doorn (1990a) as *dangerous* in that they reinforce — by enshrining in the very nomenclature — "speculative interpretations" of the computational role of these cells.

3.3 Directional Spatial Derivatives

Since sharp, spatially coincident transitions in the retinal image at several neighbouring spatial scales may be taken as evidence of the presence of a luminance edge in the monocular visual field (Marr & Hildreth, 1980), an alternative to matched filtering in the linear "detection" of oriented luminance edges is the calculation of *directional spatial derivatives* of the retinal image.

3.3.1 Retino-Cortical Derivative Operators

According to the directional derivative hypothesis of simple cell processing, the retino-cortical weighting function[14] c of each simple cell constitutes a discrete-space approximation to the kernel of a *directional derivative operator* D_θ^i of some order $i \in \mathbb{Z}_+$ (Koenderink

cells to LGN cells (Wilson et al., 1990).

[13]For simplicity, this conservative scheme assumes spatiotemporal separability of the simple cell RFP; inseparability would only serve to further increase the degrees of freedom to be "covered" by the simple cell population.

[14]described earlier in Section 3.1.2.

& van Doorn, 1990a; Koenderink & van Doorn, 1990b) or $i \in \mathbb{R}_+$ (Hungenahally et al., 1992; Hungenahally et al., 1993) on the sampled retinal image $b(\boldsymbol{x}_j, t)$ in the direction $\theta \in [0, 2\pi)$. An approximation to the directional derivative $D^i_\theta b$ of the retinal image is computed by discrete correlation of the image with a retino-cortical spatial weighting function c^i_θ approximating the kernel D^i_θ such that

$$[c^i_\theta \star b](\boldsymbol{x}_j, t) \approx [D^i_\theta b](\boldsymbol{x}_j, t)$$

where $\star$ denotes discrete spatial correlation such that

$$[c^i_\theta \star b](\boldsymbol{x}_j, t) \equiv \sum_{\boldsymbol{x}_k \in \mathcal{S}} c^i_\theta(\boldsymbol{x}_k - \boldsymbol{x}_j) b(\boldsymbol{x}_k, t) \tag{3.7}$$

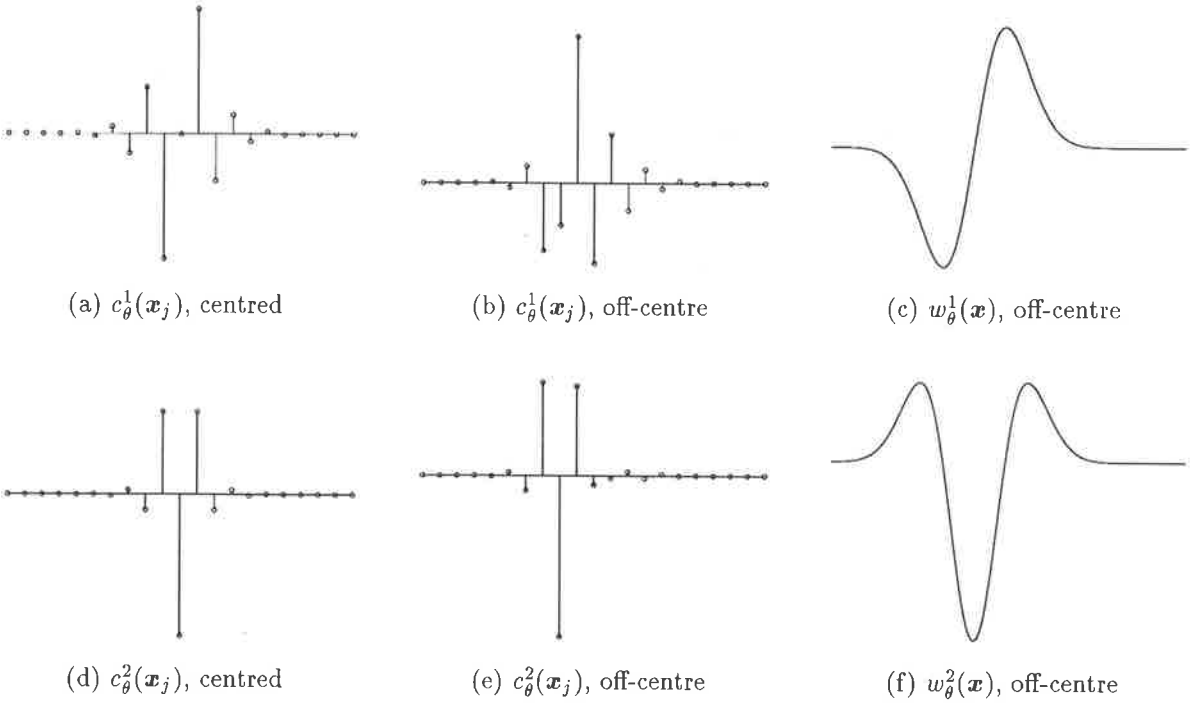
Examples of one-dimensional first- and second-order derivative approximations to be discussed shortly are presented in Figure 3.2. The spatial *scale* of the derivative approximation is determined by the spacing in the direction θ of the photoreceptors from which the simple cell receives input. Multiple scales are achieved in the same retinal location by appropriately undersampling the same photoreceptor array. Whilst thresholding of the first derivative $[c^1_\theta \star b](\boldsymbol{x}_j, t)$ in the direction θ is sufficient — at least in the absence of noise — to detect orthogonally oriented transitions at the chosen spatial scale, the directional derivative hypothesis encompasses the computation of derivatives of arbitrary order.

Like the linear matched filters in Section 3.2.2, each directional derivative operator D^i_θ is ideally approximated by an array of simple cells having identical retino-cortical weighting functions c^i_θ — and hence spatial RFPs $w^i_\theta : \mathbb{R}^2 \to \mathbb{R}$ given by (3.4) — whose centres densely populate the retinal surface. The requirement of identical RFPs for the implementation of a given operator can however be relaxed slightly, since it is possible to derive many different approximations to the kernel of a given derivative operator. For the special case of uniformly spaced sampling points with sampling interval T, one family of derivative approximations results from the differentiation of the 1D interpolation formula

$$f(x) = \sum_{j=-\infty}^{\infty} f(jT) \frac{sin\left(\pi[\frac{x-jT}{T}]\right)}{\pi[\frac{x-jT}{T}]}$$

followed by truncation of the resultant series at an appropriate value of j. However this approximation has non-compact support and a slow rate of decay with increasing $|j|$, requiring an unacceptable number of sampling points in order to obtain a suitable approximation to the required derivative. A second family of derivative approximations, which have strictly compact support and permit irregular sampling, results (see e.g. Kreyszig (1983, Sect. 19.6)) from the differentiation of a 2D *Lagrange interpolating polynomial* (LIP)

$$L(\boldsymbol{x}, t) \triangleq \sum_{\boldsymbol{x}_j \in \mathcal{Q}} \frac{\prod_{i \neq j} \prod_{p=1}^{2} \boldsymbol{x}^{[p]} - \boldsymbol{x}_i^{[p]}}{\prod_{i \neq j} \prod_{p=1}^{2} \boldsymbol{x}_j^{[p]} - \boldsymbol{x}_i^{[p]}} b(\boldsymbol{x}_j, t)$$



(a) $c^1_\theta(\boldsymbol{x}_j)$, centred (b) $c^1_\theta(\boldsymbol{x}_j)$, off-centre (c) $w^1_\theta(\boldsymbol{x})$, off-centre

(d) $c^2_\theta(\boldsymbol{x}_j)$, centred (e) $c^2_\theta(\boldsymbol{x}_j)$, off-centre (f) $w^2_\theta(\boldsymbol{x})$, off-centre

Figure 3.2: "Directional" derivative masks c^i_θ derived from the 21-point 1D LIP for uniformly spaced sampling points. (a),(d) First and second derivative masks respectively for the derivative evaluated at the midpoint (#11) of the sampling array. (b),(e) First and second derivative masks respectively for the derivative evaluated at sampling point #8, 3 samples to the left of the midpoint. (c),(f) The masks in (b),(e) applied to a Gaussian of standard deviation 2.5 samples, representing the photoreceptor RFP h, to produce corresponding simple cell RFPs w^i_θ. As expected, the resultant first and second Gaussian derivatives remain respectively odd and even symmetric about the point of derivative evaluation, despite the apparent asymmetry of the masks in (b),(e) used to produce them.

— where the superscript $^{[p]}$ denotes selection of the pth element of a vector — fitted to the retinal image $b(\boldsymbol{x}_j, t)$ over a local subset $\mathcal{Q} \subset \mathcal{S}$ of the sampling points. Directional differentiation can be achieved by choosing a new Cartesian coordinate system such that the $x^{[1]}$-axis is oriented in the θ direction and then simply evaluating the partial derivative $\frac{\partial^i L}{\partial (x^{[1]})^i}$. For a given point $\boldsymbol{x} \in \mathbb{R}^2$ at which the derivative $D^i_\theta L$ is to be evaluated, the set $\mathcal{Q}$ of points through which the polynomial is fitted can be chosen almost arbitrarily, subject only to the condition that $\boldsymbol{x}$ interpolates this set. Figures 3.2(a) and (b) illustrate two 21-point 1D LIP approximations c^1_θ to the first order derivative operator, where the derivative has been evaluated at and to the left of the centre of the sampling array respectively. Figures 3.2(d) and (e) show the corresponding approximations to the second derivative c^2_θ.

3.3.2 *Fractional Derivatives*

The generalisation of derivatives to *fractional* orders arises in applications including the conduction of heat in solids and electrical signals in cables (Bracewell, 1986). Fractional directional derivatives can perhaps be most easily explained via the Fourier transform domain, through which the identity

$$\frac{\partial^i f(\boldsymbol{x})}{\partial (x^{[k]})^i} \stackrel{\mathcal{F}}{\longleftrightarrow} (j2\pi u^{[k]})^i F(\boldsymbol{u}) \tag{3.8}$$

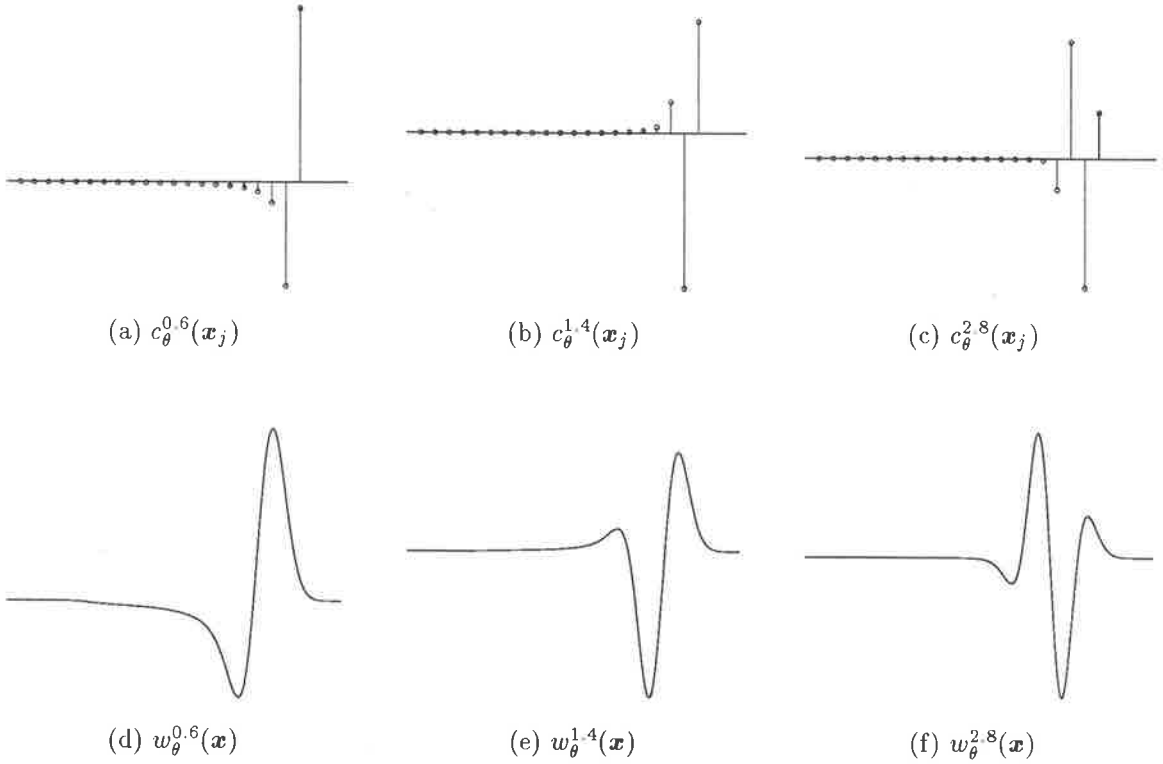
which holds for integer orders $i \in \mathbb{Z}_+$ — where $\stackrel{\mathcal{F}}{\longleftrightarrow}$ indicates a Fourier transform pair and $f,F{:}\mathbb{R}^n \rightarrow \mathbb{C}$ satisfy $f(\boldsymbol{x}) \stackrel{\mathcal{F}}{\longleftrightarrow} F(\boldsymbol{u})$ — can be generalised to fractional orders[15] $i \in \mathbb{R}_+$ (Bracewell, 1986). Thus the fractional partial differentiation $\partial^i / \partial (x^{[k]})^i$ of a function in the spatial domain is equivalent to multiplication of its Fourier transform by a fractional power of $j2\pi u^{[k]}$ in the Fourier domain. The half-order derivatives of an edge and a bar in the direction perpendicular to their orientation are illustrated in Table 8.2 of Bracewell (1986).

Fractional differentiation can be applied to LIPs in the same manner as outlined above for integer order derivatives, to produce fractional discrete-space derivative approximations c_θ^i. An alternative derivation of these fractional derivative operators was provided by Hungenahally et al. (1993), who appears to have used the observation that if the derivative is to be evaluated at the centre of a uniformly spaced 1D sampling array, and the ith order derivative mask is derived from the $(i+1)$-point LIP, the mask values $c_\theta^i(\boldsymbol{x}_j)$ are given by the coefficients of the binomial expansion of $(1-t)^i$. This derivation however suffers from the limitation that for any given sampling grid, even spacing of the sampling points along the $x^{[1]}$ direction of the coordinate axes used to derive the directional derivative cannot simultaneously apply for arbitrary orientations θ. For non-integer values of i, the coefficients resulting from the latter derivation exhibit approximately factorial decay with sample number as illustrated in Figures 3.3(a)-(c), permitting truncation of the series after relatively few terms, and unlike those for $i \in \mathbb{Z}_+$ do not sum to zero. The number n of alternations of sign with increasing sample index is related to the order i of the derivative such that $n = \lceil i \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function.

3.3.3 *Discriminant Functions*

A continuous-space *discriminant function* (Hungenahally, 1991; Hungenahally et al., 1992) w_θ^i can be generated from a discrete-space function c_θ^i using (3.4) with an *aggregation function* h of strictly compact support. Care should however be taken with the

[15]The multiple roots which exist on the right-hand side for non-integer rational orders i produce functions which differ only by a phase shift.



(a) $c_\theta^{0.6}(\boldsymbol{x}_j)$ (b) $c_\theta^{1.4}(\boldsymbol{x}_j)$ (c) $c_\theta^{2.8}(\boldsymbol{x}_j)$

(d) $w_\theta^{0.6}(\boldsymbol{x})$ (e) $w_\theta^{1.4}(\boldsymbol{x})$ (f) $w_\theta^{2.8}(\boldsymbol{x})$

Figure 3.3: (a)-(c) 1D fractional derivative masks $c_\theta^i(\boldsymbol{x}_j)$ with $i \in \mathbb{R}_+ - \mathbb{Z}_+$. The sample index is shown increasing to the left, so that for consistency with (2.1), the mask can be used as a *correlation* kernel. (d)-(f) Corresponding discriminant functions obtained using a Gaussian aggregation function h with standard deviation 2.5 samples.

selection of the finite support aggregation function in order to limit the aliasing resulting from its failure to band-limit the image falling on the retina (cf. Hungenahally et al. (1993, Fig. 7)). Fractional discriminant functions corresponding to the discrete masks shown in Figures 3.3(a)-(c) are illustrated in Figures 3.3(d)-(f), where the requirement of compact support has been relaxed to admit a Gaussian aggregation function. These functions possess no axes of symmetry and exhibit non-zero response to DC illumination (Hungenahally et al., 1993), as is the case for the GRGF model with $\boldsymbol{\theta} = \arg\{\boldsymbol{u}_0\}$ and $\phi \neq 0$.

The salient oscillatory property of the discriminant functions w_θ^i can — for h both symmetric and not too large relative to the sampling period, and for $i \in \mathbb{Z}_+$ — be imitated by multiplying a periodic function having 2 evenly-spaced zero-crossings per period $T \in \mathbb{R}_+$, such as a sinewave, by a compactly-supported window of width $W = iT/2$ to produce a *generalised discriminant function* (Hungenahally, 1991). The width of the window is used to control the number of RFP subfields and hence the order of the generalised derivative approximation. However the extension of this model to the fractional case requires some care, since the periodic function must be truncated to avoid one or

more spurious additional zero-crossings, which would increase the number of subfields and hence the derivative order. At least for the lower order derivatives, this truncation permits only a poor approximation to the factorial tail — evident for example in Figures 3.3(a) and (d) — of the fractional derivative function being generalised. Despite the success of the GRGF model of the simple cell spatial RFP (Jones et al., 1987), to which suitable approximations can be generated using generalised discriminant functions, the construction of the original discriminant functions reflects more directly the physical feedforward mechanisms — formalised in Section 3.1.2 — which contribute to the simple cell RFP.

3.3.4 *Gaussian Derivatives*

Koenderink & van Doorn (1990a; 1990b) proposed a family of kernels w_θ^i known as the *Gaussian-windowed Hermite polynomials* given by the directional derivatives of an isotropic Gaussian. In the present notation, each such kernel results from the application of an unspecified approximation c_θ^i to the corresponding directional derivative operator D_θ^i to an isotropic Gaussian kernel representing the photoreceptor RFP h. Examples of first and second order Gaussian derivatives obtained using the derivative approximations illustrated in Figures 3.2(b) and (e) are shown in Figures 3.2(c) and (f) respectively. To appropriately approximate true Gaussian derivatives, the masks c_θ^i must be derived using sampling points spaced at intervals which are suitably small compared with the standard deviation of the Gaussian h. If this condition is not observed, the θ cross-section of the corresponding simple cell RFP resembles a collection of isolated Gaussian subfields, contrary to experimental observations. The number $i + 1$ of RF subfields reflects the order of the differential operator applied to the retinal image. An attractive property of these kernels is that it is possible to construct a set containing up to nth order derivatives — with $i + 1$ distinct orientations θ for the ith derivative order — which can completely encode the mixed spatial partial derivatives of the blurred image up to order n (Koenderink & van Doorn, 1990a). Young (1985) estimated that for the human visual system the highest derivative order is 4. Despite the attractiveness of this completeness property however, the RFP model relies on an isotropic Gaussian window of the same dimensions as the photoreceptor RF, and odd and even RF symmetries, neither of which seem to be preferred properties amongst the simple cell population (Jones & Palmer, 1987b).

3.3.5 *Summary*

The directional derivative hypothesis of simple cell processing has been shown to account for many features of simple cell spatial RFPs, including subfield alternation and — in the case of fractional derivatives — asymmetry. The theory of fractional derivatives and

discriminant functions, which awaits a full account of the physical significance of the information they extract from an image, provides considerable scope for future research in this area. The mathematical completeness of a directional derivative representation has so far been shown only for the unrealistic case of RFPs given by Gaussian derivatives of integer order. Finally the smallest spatial scale or resolution available to derivative approximations at the simple cell level for any given retinal eccentricity is limited by the photoreceptor sampling interval, while the largest scale is determined by the size of the photoreceptor RFP h, since simple cell spatial RFPs exhibiting spatially isolated blobs of this size are yet to be reported. These two observations place strong limits on the range of spatial scales over which derivative approximations could be evaluated by the simple cells, in potential contrast with observations of simple cell RFPs in the foveal projection with spatial dimensions spanning a range in excess of 30:1 (Daugman, 1985).

3.4 Spatial Frequency Analysis

3.4.1 Introduction

Outlined briefly in Section 2.2.2, the psychophysical results of Campbell & Robson (1968) and Blakemore & Campbell (1969), from which the existence of psychophysical channels with bandpass tuning for 2D spatial frequency magnitude and orientation was inferred, led Pollen et al. (1971) to postulate that

> ... the brain has at its disposal the two-dimensional Fourier transform of the presented brightness distribution.

Noting that sinusoids are the *principal components* (the eigenfunctions of the auto-covariance function) of an image ensemble exhibiting position-independent second order spatial statistics (see e.g. Gaskill (1978)), Bossomaier & Snyder (1986) suggested that a possible advantage of such an analysis is that it removes second-order statistical redundancy from the image (Attneave, 1954; Barlow, 1959; Barlow, 1961). The limited spatial extent of the receptive fields of the spatial-frequency selective cells in primary visual cortex prompted Pollen et al. (1971) to further speculate that these cells participate in "two-dimensional spatial frequency decompositions of subsections of visual space" — henceforth referred to collectively as a *localised Fourier analysis* — a proposition which under the assumption of approximate spatial linearity is more consistent with the relatively broad spatial frequency tuning curves of cells in V1 (Robson, 1983). According to Bossomaier & Snyder (1986), this localisation of the proposed analysis allows it to take advantage of *local* statistical stationarity in reducing the redundancy of an image ensemble whose spatial statistics are globally non-stationary[16]. The interested reader

[16]A more detailed exposition of the underlying statistical theory may be found in Sections 3.4.6 and 3.5.5.

is referred to the following works for further information on the history of the localised Fourier analysis hypothesis of primary visual cortical function: Pollen & Gaska (1987), Henry (1985), Pollen et al. (1985), Shapley & Lennie (1985), Pollen & Ronner (1983; 1982; 1981), Robson (1983), and Maffei & Fiorentini (1973).

3.4.2 The Gabor Expansion

Contrary to the suggestion by Pollen & Taylor (1974) that the complex cells were well-placed to perform the proposed localised Fourier or principal-components analysis, Maffei & Fiorentini (1973) argued that the simple cells were to be preferred for this role because of their more linear spatiotemporal behaviour. The particular mathematical form adopted by Bossomaier & Snyder (1986) for this localised analysis, known as the *Gabor representation* (Marčelja, 1980; Porat & Zeevi, 1988) or *Gabor expansion*[17] (Wexler & Raz, 1990; Hlawatsch & Boudreaux-Bartels, 1992), was first applied by Marčelja (1980) to the understanding of the representation of the visual scene by the simple cells. Named in honour of Gabor (1946), who first proposed the representation of 1D communication signals as a sum of Gabor functions, the Gabor expansion (GE) has since been generalised to encompass possibly non-Gaussian localisation functions (see e.g. Wexler & Raz (1990)). In view of the success of the GRGF model in describing the simple cell RFP however, the present exposition is largely restricted to a consideration of the case of Gaussian localisation functions.

The 1D Gabor expansion using a Gaussian localisation function involves the *decomposition* or *analysis* of a 1D function $s(x)$ into a set $\{c_{ip}\colon\ (i,p) \in \mathbb{Z}^2\}$ of coefficients $c_{ip} \in \mathbb{C}$ corresponding to the set

$$\mathcal{G}_1 \triangleq \{g_{ip}(x)\colon\ (i,p) \in \mathbb{Z}^2\}$$

of functions $g_{ip} : \mathbb{R} \rightarrow \mathbb{C}$ such that

$$s(x) = \sum_{i=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} c_{ip} g_{ip}(x) \tag{3.9}$$

where

$$g_{ip}(x) \triangleq \frac{\sqrt{2}}{S} \cdot \exp\left\{-\pi \left[\frac{x - i\Delta_x}{S}\right]^2\right\} \cdot \exp\{j2\pi p \Delta_u x\} \tag{3.10}$$

[17]The term "Gabor expansion" is used here in preference to "discrete Gabor transform" which appears to have two conflicting definitions, according to which the same set of windowed sinusoids is used either for the synthesis (expansion) or the analysis (filtering) of the signal or image. The first, decomposition-based definition used by Daugman (1988a) involves the calculation of the coefficients in the Gabor expansion to be defined below; the second, filter-based definition used by Heil & Walnut (1989) treats the (complex conjugates of the) same functions as for the Gabor expansion as correlation kernels rather than basis functions. The former definition is equivalent to that used by Hlawatsch & Boudreaux-Bartels (1992) for the *discrete short-time Fourier transform*; the latter is equivalent to that used by Mallat (1989a) for the *discrete window Fourier transform*.

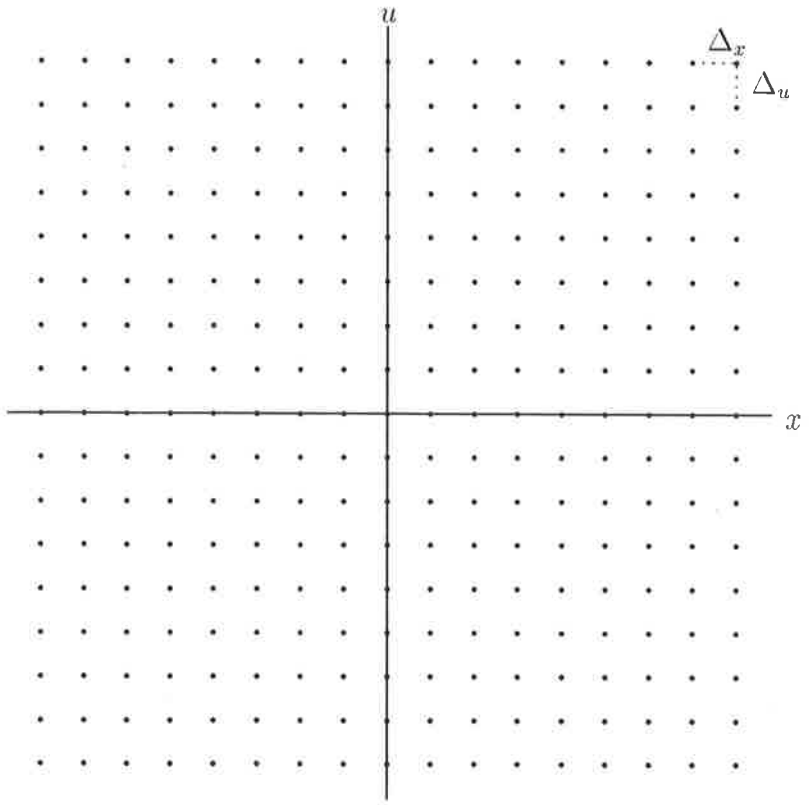


Figure 3.4: Lattice of Gabor function centres $(x_i, u_p) \triangleq (i\Delta_x, p\Delta_u)$ in phase space for 1D Gabor expansion set $\mathcal{G}_1$.

Δ_x, Δ_u specify the separations between neighbouring Gaussian centres along the spatial and spatial-frequency axes respectively[18] as shown in Figure 3.4, and S is as defined for (2.4) with $n = 1$. The comparison of (3.10) with (2.6) reveals that the functions g_{ip} are simply 1D generalised Gabor functions with $a = \frac{\sqrt{2}}{S}$, $x_0 = i\Delta_x$, $u_0 = p\Delta_u$ and

$$\phi = 2\pi i \Delta_x p \Delta_u \tag{3.11}$$

The constant a serves to normalise the Gaussian function to unit L^2 norm, as required by both the GE (Porat & Zeevi, 1988) and the *discrete window Fourier transform*[19] (DWFT) (Mallat, 1989a). Exploitation of the close relationship between $\mathcal{G}_1$ and the corresponding set of *canonical Weyl-Heisenberg coherent states* (Daubechies et al., 1986) — for which $a = \pi^{-\frac{1}{4}}$, $\phi = \pi i \Delta_x p \Delta_u$ and $S = 1$ — allows the adaptation of a classical result in mathematical physics concerning the completeness of these coherent states (see e.g. Bargmann et al. (1971) or Higgins (1977)) to yield the following.

Theorem 3.1 *A necessary and sufficient condition for the infinite set* $\mathcal{G}_1$ *of* translates

[18]The reader should be careful not to confuse these spatial and spectral *intervals* between Gabor functions with the spatial and spectral *spreads* defined in Section 2.3.3 for a single function.

[19]to be discussed in Section 3.4.5

and modulates *of the 1D Gaussian function to be complete for the set* $L^2(\mathbb{R})$ *of finite-energy (square-integrable) 1D functions is that*

$$\Delta_x \Delta_u \leq 1 \tag{3.12}$$

Thus provided this condition is satisfied, any function $s \in L^2(\mathbb{R})$ can be exactly reconstructed from the corresponding set of coefficients. If the equality holds in (3.12), the number of GRGFs required to completely represent a band-limited signal is minimised, and is the same as that required by the Sampling Theorem (Hlawatsch & Boudreaux-Bartels, 1992); the signal is then said to be *critically sampled.* If on the other hand the inequality holds in (3.12), the chosen GRGFs are linearly dependent and are said to *oversample* the signal. Whilst oversampling can be useful in the presence of noise to avoid numerical instability in the computation of the expansion coefficients (Hlawatsch & Boudreaux-Bartels, 1992), the coefficients are in general not unique; this problem is addressed in Chapter 4. For critical sampling, the phase term ϕ given by (3.11) is 0 (mod 2π) for all generalised Gabor functions in $\mathcal{G}_1$, while for oversampling, ϕ is more generally non-zero and different for each function.

Adopting the explicit extension by Porat & Zeevi (1985; 1988) to the more realistic case for early vision of 2D generalised Gabor functions, and making explicit in the notation the temporal dependence of the low-pass filtered visual image $\check{s}(\boldsymbol{x}, t)$, the 2D Gabor expansion — as it may or may not apply to early vision — involves the decomposition of this image into a set $\{c_{ikpq}(t)\colon (i,k,p,q) \in \mathbb{Z}^4\}$ of coefficients $c_{ikpq}(t) \in \mathbb{C}$ corresponding to the set

$$\mathcal{G}_2 \triangleq \{g_{ikpq}(\boldsymbol{x})\colon (i,k,p,q) \in \mathbb{Z}^4\}$$

of functions $g_{ikpq} : \mathbb{R}^2 \rightarrow \mathbb{C}$ such that

$$\check{s}(\boldsymbol{x}, t) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} c_{ikpq}(t) g_{ikpq}(\boldsymbol{x}) \tag{3.13}$$

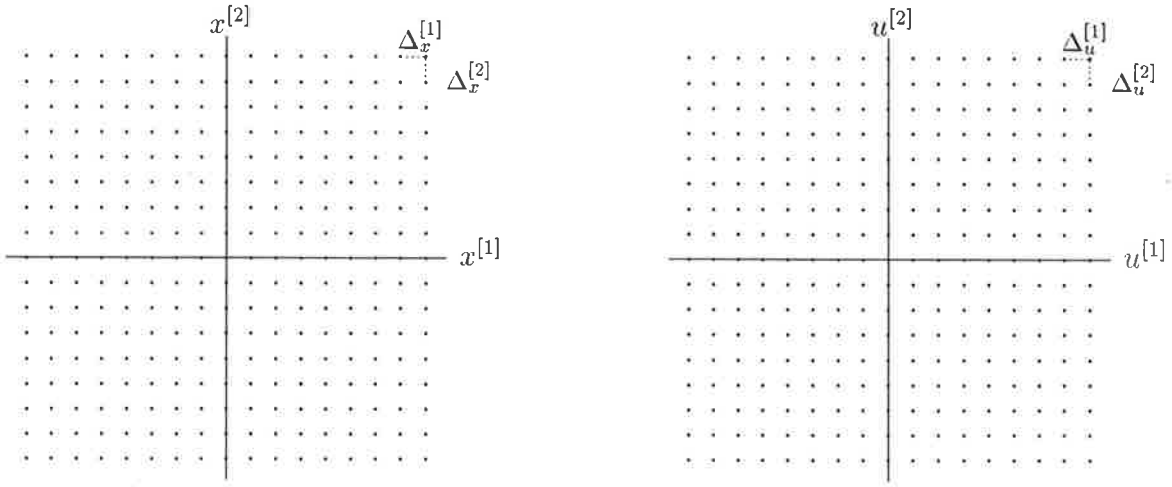
where

$$g_{ikpq}(\boldsymbol{x}) \triangleq \frac{1}{\sqrt{\pi}\det(S)} \cdot \exp\{-\pi \|S^{-1}(\boldsymbol{x} - \boldsymbol{x}_{ik}^T)\|_2^2\} \cdot \exp\{j2\pi \boldsymbol{u}_{pq}^T \boldsymbol{x}\} \tag{3.14}$$

Here S is the diagonal matrix of Gaussian space constants defined earlier in connection with (2.4). The spatial and spectral offset terms

$$\begin{aligned} \boldsymbol{x}_{ik} &\triangleq [i\Delta_x^{[1]}, k\Delta_x^{[2]}]^T \\ \boldsymbol{u}_{pq} &\triangleq [p\Delta_u^{[1]}, q\Delta_u^{[2]}]^T \end{aligned}$$

are defined for notational convenience, with $\Delta_x^{[1]}, \Delta_x^{[2]}$ specifying the separation between neighbouring Gaussian centres along the x_1 and x_2 axes respectively as shown in Figure 3.5(a), and $\Delta_u^{[1]}, \Delta_u^{[2]}$ specifying analogous separations in the spectral domain as shown in Figure 3.5(b). Once again, comparison of (3.14) with (2.6) reveals that the functions



(a) Spatial lattice

(b) Spectral lattice

Figure 3.5: (a) Lattice of generalised Gabor function centres $(x_{ik}^{[1]}, x_{ik}^{[2]})$ in the spatial domain for 2D Gabor expansion set $\mathcal{G}_2$. (b) Corresponding lattice of centres $(u_{pq}^{[1]}, u_{pq}^{[2]})$ in the spatial-frequency domain.

g_{ikpq} are simply 2D generalised Gabor functions with $a = \frac{1}{\sqrt{\pi \det(S)}}$, $\theta_z = 0$, $\boldsymbol{x}_0 = \boldsymbol{x}_{ik}$, $\boldsymbol{u}_0 = \boldsymbol{u}_{pq}$ and

$$\phi = 2\pi \boldsymbol{u}_{pq}^T \boldsymbol{x}_{ik}$$

Theorem 3.2 (Porat & Zeevi (1985)) [20] *A necessary and sufficient condition for the infinite set $\mathcal{G}_2$ of translates and modulates of the 2D Gaussian function to be complete for the set $L^2(\mathbb{R}^2)$ of finite-energy (square-integrable) 2D functions is that*

$$\Delta_x^{[1]} \Delta_u^{[1]} \leq 1 \qquad (3.15a)$$

$$\Delta_x^{[2]} \Delta_u^{[2]} \leq 1 \qquad (3.15b)$$

The variety of phases which would be observed for these 2D generalised Gabor functions whenever the inequalities in (3.15) hold is qualitatively consistent with that observed for simple cell spatial RFPs, and militates against the preference by Pollen & Ronner (1983) for precisely odd- and even-symmetric RFPs.

3.4.3 Do Simple Cells Perform a Gabor Decomposition?

In the previous section it was shown that the spatial RFPs of the simple cells could potentially be described by the expansion functions of a discrete 2D Gabor expansion. The question therefore arises as to whether the simple cells in fact compute the coefficients

[20] Although the ensuing result is not directly stated in their work (Porat & Zeevi, 1985), it can be easily inferred from it.

of such an expansion. Whilst the following investigation of this issue focuses for notational convenience on the 1D case, suitable extensions of the 1D results to the 2D case can be assumed to hold unless stated otherwise.

Marčelja (1980) hypothetically identified each coefficient c_{ip} with the activation level of a corresponding simple cell having 1D spatial RFP $g_{ip}^*(x)$, so that each simple cell contributes to a decomposition of the (1D) visual stimulus $s(x)$ by calculating the coefficient of (the complex conjugate of) its own RFP function in the expansion in (3.9). The hypothesis that the set of Gabor expansion coefficients "characterizes the [primary visual] cortical representation of an image" (Zeevi & Porat, 1984) has been widely and often implicitly adopted by other researchers: Pollen & Ronner (1983) for example refer to the simple cell outputs as encoding "local Fourier coefficients"[21]; Daugman (1989b, p. 243) used observations regarding the coefficients of a 2D Gabor expansion to draw conclusions about the activation levels of simple cells[22]; while MacLennan (1991) hypothesised that the cells in V1 compute the coefficients of a 3D spatiotemporal Gabor expansion of the stimulus[23].

However, Marčelja (1980) noted that since the chosen set of Gabor functions is not orthornomal, the simple cell outputs

$$r_{ip} = \int_{\mathcal{V}} s(x) g_{ip}^*(x)\, dx$$

would only approximate the true coefficients. In fact rather than giving the required coefficients in the Gabor expansion, these hypothetical simple cell outputs are those required for a discrete window Fourier transform (Mallat, 1989a) — to be described shortly in Section 3.4.5 — using the chosen Gaussian window, which is more in line with a view of the simple cell spatial RFP as the correlation kernel of a spatial frequency *filter*, a term used enigmatically by Pollen & Ronner (1983) in the same breath as the term "local Fourier coefficient". Generalising an approach proposed by Bastiaans (1980) both to the case of oversampling and to 2D — with the latter generalisation being omitted from the following formalisation for notational convenience — Porat & Zeevi (1988) showed that the set of coefficients given by

$$c_{ip} = \int_{-\infty}^{\infty} s(x) \cdot \gamma(x - i\Delta_x) \cdot \exp\{-j2\pi p \Delta_u x\}\, dx \tag{3.16}$$

— where the *biorthogonal function* $\gamma : \mathbb{R} \to \mathbb{R}$ is given by

$$\gamma(x) \triangleq \left(\frac{\sqrt{2}}{\Delta_x}\right)^{\frac{1}{2}} \cdot \left(\frac{K_0}{\pi}\right)^{-\frac{3}{2}} \cdot \exp\left\{\pi\left(\frac{x}{\Delta_x}\right)^2\right\} \cdot \sum_{n+\frac{1}{2} \geq \frac{x}{\Delta_x}} (-1)^n \exp\left\{-\pi\left(n+\frac{1}{2}\right)^2\right\}$$

[21] although Pollen et al. (1985) stop short of advocating Marčelja's (1980) Gabor expansion hypothesis.

[22] Elsewhere however, Daugman (1990) conceded that there exist problems with this view.

[23] MacLennan (1991) acknowledged however that a relaxation process — such as that described in Chapter 4 — would be required to calculate the correct coefficients (presumably from the simple cell outputs given by (2.2)).

and $K_0 \in \mathbb{R}$ is a normalisation constant — satisfy the expansion in (3.9) up to a scalar constant. However, the 1D RFP required by a simple cell in order for it to calculate the Gabor expansion coefficient in (3.16) corresponding to a given GRGF is given by

$$w(x) = \gamma(x - i\Delta_x) \cdot \exp\{-j2\pi p \Delta_u x\}$$

which bears little resemblance to either the chosen GRGF (see e.g. Porat & Zeevi (1988, Fig. 17)) or to simple cell 1D RFPs. Nevertheless, since the GRGFs are linearly dependent for the case of oversampling, the choice of coefficients satisfying the Gabor expansion is not unique, and it therefore remains to rule out the possibility that the discrete window Fourier transform might produce an equally valid set of coefficients for the Gabor expansion. In the following section the possibility of a reconciliation of Marčelja's (1980) GE hypothesis with the computation actually performed by the simple cells is therefore investigated.

3.4.4 *Weyl-Heisenberg Frames*

Setting aside the objections discussed earlier in Section 2.3.3 to the implementation by the simple cells of complex-valued RFPs and outputs, as well as their conjugates, and noting that the chosen Gabor functions do not form an orthonormal basis, hopes of validating Marčelja's (1980) hypothesis appear to rest on the generalisation of the notion of a basis provided by the theory of *frames* (Heil & Walnut, 1989).

Definition 3.1 *A set* $\{\psi_i : i \in \mathbb{Z}\} \subset L^2(\mathbb{R})$ *of functions* $\psi_i : \mathbb{R} \to \mathbb{C}$ *is said to form a* frame *for* $L^2(\mathbb{R})$ *if and only if there exists a bounded invertible linear operator* P*, called the* frame operator*, such that for each function* $s \in L^2(\mathbb{R})$

$$[Ps](x) = \sum_i \left[\int s(\tilde{x}) \psi_i^*(\tilde{x})\, d\tilde{x} \right] \psi_i(x) \qquad (3.17)$$

The term in square brackets on the right-hand side of (3.17) can be identified with the output of a simple cell having 1D RFP $\psi_i^*(x)$, while the summation can be identified with the attempted reconstruction of the spatially low-pass filtered original image $\check{s}(x)$ from the frame elements ψ_i using the coefficients given by the simple cell outputs. Note however that since the low-pass filtering operation performed on the image by the retinal imaging process is not invertible, the corresponding operator is not a candidate for the frame operator P. Alternatively, (3.17) can be rewritten as

$$s(x) = \sum_i \left[\int s(\tilde{x}) \psi_i^*(\tilde{x})\, d\tilde{x} \right] P^{-1}\psi_i(x)$$

so that the simple cell output — again in square brackets — can be seen as an expansion coefficient corresponding to the element $P^{-1}\psi_i$ of the *dual frame* $\{P^{-1}\psi_i\}$ (MacLennan, 1991).

The scalar constants κ_1, κ_2 such that

$$\kappa_1 s(x) \leq Ps(x) \leq \kappa_2 s(x) \qquad \forall s \in L^2(\mathbb{R}) \tag{3.18}$$

are called the frame *bounds*. A frame for which $\kappa_1 = \kappa_2$ is said to be *tight*, and in this case P is simply the scalar $\kappa_1 = \kappa_2$, so that the original image can be reconstructed, up to this constant factor, from the simple cell outputs as required by Marčelja's (1980) hypothesis. For a given choice of Δ_x, the set $\mathcal{G}_1$ is known to form a *Weyl-Heisenberg frame* for all sufficiently small values of Δ_u (Daubechies et al., 1986; Heil & Walnut, 1989). However, although this frame becomes increasingly tight ($|\kappa_1 - \kappa_2| \to 0$) as the product $\Delta_x \Delta_u \to 0$ (Daubechies et al., 1986), it is not in fact a tight frame except in the limiting case. This observation therefore rules out the possibility that the output of a simple cell having 1D RFP $g_{ip} \in \mathcal{G}_1$ represents the coefficient of that RFP in the expansion in (3.9), and hence invalidates Marčelja's (1980) hypothesis. Thus as Marčelja (1980) conceded, the coefficient estimate provided by the simple cell is only approximate; bounds on the error of this approximation are provided by (3.18). However, if indeed the set $\mathcal{G}_2$ of 2D Weyl-Heisenberg coherent states does provide an adequate model of the set of simple cell spatial RFPs, a more natural account of the processing performed by simple cells is provided in the following section.

3.4.5 *Discrete Window Fourier Transform*

The 2D *discrete window Fourier transform* (DWFT) of a stimulus $s(\boldsymbol{x}, t)$ employing the window function $v{:}\mathbb{R}^2 \to \mathbb{R}$ — as it may or may not relate to the early visual system — is given by

$$[Ws](\boldsymbol{x}_{ik}, \boldsymbol{u}_{pq}, t) \triangleq \int_{\mathcal{V}} s(\boldsymbol{x}, t) \cdot v(\boldsymbol{x} - \boldsymbol{x}_{ik}) \cdot \exp\{-j2\pi \boldsymbol{u}_{pq}^T \boldsymbol{x}\}\, d\boldsymbol{x}$$

where W is the DWFT operator. Although the DWFT encompasses the use of more general window functions, the choice of a Gaussian for the present purposes leads to a set of 2D spatial RFPs given by the 2D Weyl-Heisenberg coherent states $g_{ikpq}(x)$ as defined in (3.14). For each point $\boldsymbol{u}_{pq}$ on the spatial frequency lattice, $[Ws](\boldsymbol{x}_{ik}, \boldsymbol{u}_{pq}, t)$ can be viewed as the output of a filter having correlation kernel $v(\boldsymbol{x} - \boldsymbol{x}_{ik}) \cdot \exp\{-j2\pi \boldsymbol{u}_{pq}^T \boldsymbol{x}\}$ and sampled at the points $\boldsymbol{x}_{ik}$ on a Cartesian lattice.

Unlike the Gabor expansion, the DWFT is *defined* in terms of the spatial inner product of the stimulus with each of the GRGF elements of this set, and is therefore more plausibly implemented by the simple cells[24]. Like the Gabor expansion however, the DWFT requires a large set of GRGFs having the same Gaussian window (up to a translation) with a varying number of RF subfields (oscillations) "visible" under this window, which despite reports of up to 8 RF subfields (Glezer et al., 1989), does not seem

[24]although see the discussion in Chapter 7

to be the case in general amongst the feline or primate simple cells (Daugman, 1988a), where RFs with 2 and 3 subfields predominate and Gaussian windows of various sizes have been reported (Jones & Palmer, 1987b). This relatively small variation in the number of subfields exhibited by most simple cells is reflected in their approximately constant bandwidths — approximately 1.5 octaves (Pollen & Gaska, 1987) — on a *logarithmic* frequency scale (Daugman, 1988a) over the full range of centre frequencies exhibited by the population, an observation which conflicts with the prediction by the DWFT scheme, due to the use of a single Gaussian window, of constant bandwidth on a linear scale. Furthermore, the 2D DWFT employs a Gaussian window with a single orientation, and therefore fails to account for the variety of orientations observed amongst the simple cell RFPs. It is not however possible on the basis of the above observations to conclusively rule out the parallel implementation of multiple 2D DWFTs employing varying Gaussian window sizes and orientations, for which a sufficient population of simple cells may indeed exist. Nevertheless, the recent development of an alternative class of image representation schemes — to be reviewed in Section 3.5 — which provide a more natural account of the variety of simple cell RFPs has probably sounded the death knell for the DWFT hypothesis of simple cell processing.

3.4.6 *Efficient Coding Through Gabor Expansion*

For discrete-space (e.g. pixellated) images, the phase-space sampling lattice employed by the 2D Gabor expansion is usually truncated both in space, to account for the finite spatial extent of the image, and in spatial frequency, to account for the spectral limitations imposed by spatial sampling[25]. Applying one such truncated 2D Gabor expansion to an image, Daugman (1988a; 1989a) noted, after quantisation of the resultant coefficients to $n = 256$ grey levels, that the first-order entropy estimate

$$\hat{H}_1 \triangleq -\sum_{i=1}^{n} f_i \log_2 f_i \tag{3.19}$$

of the representation, where f_i is the relative frequency of the ith quantisation level, was significantly less than the same measure applied to the pixels of the original image. The symbol $\hat{H}_1$ is used to indicate the fact that the relative frequencies of the various quantisation levels provide only an estimate of the true probability density function for the pixel/coefficient random variables, which are assumed to be identically distributed (see below). When applied to the quantised image, the measure H_1, of which $\hat{H}_1$ is an estimate, is bounded below by the true entropy H of the quantised image source, calculated using the second- and higher-order joint probability distributions of the image pixels in

[25] Although this involves a practical approximation, since due to the strictly infinite spatial and spectral extent of the Gaussian window employed in the true Gabor expansion, the coefficients of the Gabor functions excluded by the truncation are not necessarily exactly zero.

addition to the marginal distributions. H represents a measure of the information content of the image, with 2^H giving the minimum number of discriminable signalling levels required to completely encode each pixel/coefficient of the image in the absence of noise. Contrary to the implication of the term "entropy reduction" used by Daugman (1989a), a lossless transformation of the image involves by definition no loss of information and hence entropy, but may convert the higher-order entropy of joint pixel intensity distributions into the first-order entropy of individual pixel distributions (Field, 1987). Such a conversion might for example be achieved by exploiting the statistical dependence of each pixel on its neighbours to predict the value of that pixel using a linear combination of the neighbouring pixels, and transmitting only the difference between the true value and the prediction (Atick & Redlich, 1991). Since there exist lossless coding algorithms such as the Shannon-Fano algorithm (Carlson, 1986) which can achieve coding entropies close to the first-order entropy, Daugman (1989a) concluded that the Gabor expansion is a more efficient representation of the visual image than the original pixel-based representation.

However, the following formalisation of the above analysis reveals that despite its promulgation in the literature (see e.g. Teuner & Hosticka (1993)), the comparison of the first-order entropy estimates obtained using (3.19) before and after Gabor decomposition of the image is invalid. Let an image be a single outcome of a 2D (spatial) *stochastic process*[26] (Papoulis, 1984). In order to determine the statistics of this process from a single image, we require that the process be *ergodic*, which in turn requires that it also be *stationary* — i.e. its statistics are position-independent. The relevance of the following argument to biological vision is consequently dependent on the extent to which the ensemble of visual images can be characterised by a single ergodic 2D stochastic process[27], a characterisation which is not only unlikely to encompass the statistical diversity of natural scenes, but also ignores the piecewise temporal continuity of the visual environment. In particular we wish to estimate the pixel probability density function on which the first-order entropy measure H_1 is based, and require therefore that the stochastic process which characterises the image is *distribution-ergodic*. The former condition can be met if in addition to being identically distributed the pixel *random variables* are also independent for large pixel separations (Papoulis, 1984)[28], which in the absence of further information is not an unreasonable assumption in the context of natural vision.

These same considerations can be applied to the "image" resulting from Gabor decomposition of the original image, whose "pixels" are now the coefficients of the expansion. However, the requirement that the coefficients also be identically distributed

[26]The reader unfamiliar with the theory of stochastic processes is referred to Papoulis (1984) for definitions and explanations of the terms used in the present exposition.

[27]although more sophisticated schemes for image representation should be adaptive to the more probably changing statistics of the visual environment.

[28]This claim is an extrapolation from the equivalent result proven by Papoulis (1984) for a 1D stochastic process.

places enormous restrictions on the class of allowable transformations between image and coefficients[29]. In the absence of further information regarding the joint probability distributions of the input pixels, the satisfaction of these restrictions can only be guaranteed *apriori* if the transformation is position-independent[30]. The Gabor decomposition is not however position-independent, since it employs Gabor functions with a range of spatial modulation frequencies. Since the coefficients are therefore almost certainly not identically distributed, these random variables cannot be characterised by a single probability density function, in violation of the assumption used to derive the first-order entropy measure. Thus the measure obtained by Daugman (1989a) from the coefficient "image" using (3.19) is not in fact an entropy measure, and its comparison with that obtained for the original image is therefore not justified. In particular, it cannot be concluded on the basis of this comparison that the Gabor expansion coefficients constitute an efficient encoding of the image.

Nevertheless, this objection can be overcome if consideration is restricted to the coefficients of the Gabor expansion corresponding to each spatial frequency channel in turn. Each such channel consists of an array of sensors having spatial RFPs given by the modulated biorthogonal functions in (3.16) centred at points given by the Cartesian spatial sampling grid, and therefore effects a position-independent transformation of the input image. A first-order entropy estimate can therefore be prepared for each channel individually using (3.19), and the resultant estimates averaged over all channels — each of which contributes the same number of coefficients — to produce a per-coefficient first-order entropy. Since in comparison with the pixel distribution of the original image the composite coefficient "distribution" presented by Daugman (1989a, Fig. 5) is strongly concentrated around zero, it is apparent that this corrected first-order entropy estimate will still be considerably lower than the corresponding measure for the input image. However, a full exposition of the significance of this result for biological vision awaits the following: a justification of the assumption that the visual environment can be characterised by a single ergodic 2D stochastic process; a demonstration of the performance of this decomposition in the visual cortex[31]; and a comparison with equivalent results — such as those outlined in Section 3.5.5 — for competing simple cell image representation schemes, including the DWFT and the discrete wavelet transform (described in Section 3.5) to show that the

[29]To see this, assume a linear transformation such as that given by (3.16) and write down the requirements that the distributions of each of the coefficients have identical ith order moments for all $i \in \mathbb{Z}_+$.

[30]Transformations which are position-independent up to any symmetries exhibited by the joint probability distributions of the input pixels/coefficients are sufficient, although these require apriori knowledge of these symmetries. For example if the third- and higher-order joint distributions are zero, and the second-order joint distribution is rotationally symmetric, the transformation need only be position-independent up to rotations of the kernel.

[31]presumably, for reasons outlined in previous sections, by cells other than the simple cells, although see also Chapter 7.

same benefit is not conferred by these other schemes.

3.4.7 Summary

The localised Fourier analysis hypothesis of simple cell spatial computation, that the "simple cells can be said to extract spatially localised spatial frequency components" (Pollen & Gaska, 1987), has been critically examined using two alternative mathematical formalisations: the Gabor expansion (GE), and the discrete window Fourier transform (DWFT). The hypothesis by Marčelja (1980) that the simple cell outputs represent the coefficients of a 2D Gabor expansion of the visual image has been found to be inconsistent with the GRGF model of the simple cell 2D spatial RFP. In contrast the alternative hypothesis, that the simple cells collectively compute a 2D discrete window Fourier transform of the visual image, is consistent with the GRGF RFP model, and therefore provides the more plausible formalisation of the above assertion by Pollen & Gaska (1987). However the DWFT hypothesis has been found to provide at best an incomplete explanation for the variation of the RFP model parameters observed in the simple cell populations of feline and monkey primary visual cortex. The "preservation of both spatial frequency and phase information at a given spatial position" also purported by Pollen & Gaska (1987) furthermore entails the assumption — implicit in both the GE and DWFT hypotheses — that *complex-valued* generalised Gabor functions are implemented amongst the simple cells, an assumption whose criticism by Stork & Wilson (1990) was examined earlier in Section 2.3.3. Finally, it has been argued that with a simple but important modification, the empirical analysis provided by Daugman (1988a), in support of his assertion that the Gabor expansion represents a more efficient coding for the visual image than a pixel-based scheme, can be adapted to demonstrate — at least for the hypothetical image source of which the single image analysed by Daugman (1988a) represents a single outcome — what it purports to.

3.5 Wavelet-like Analysis

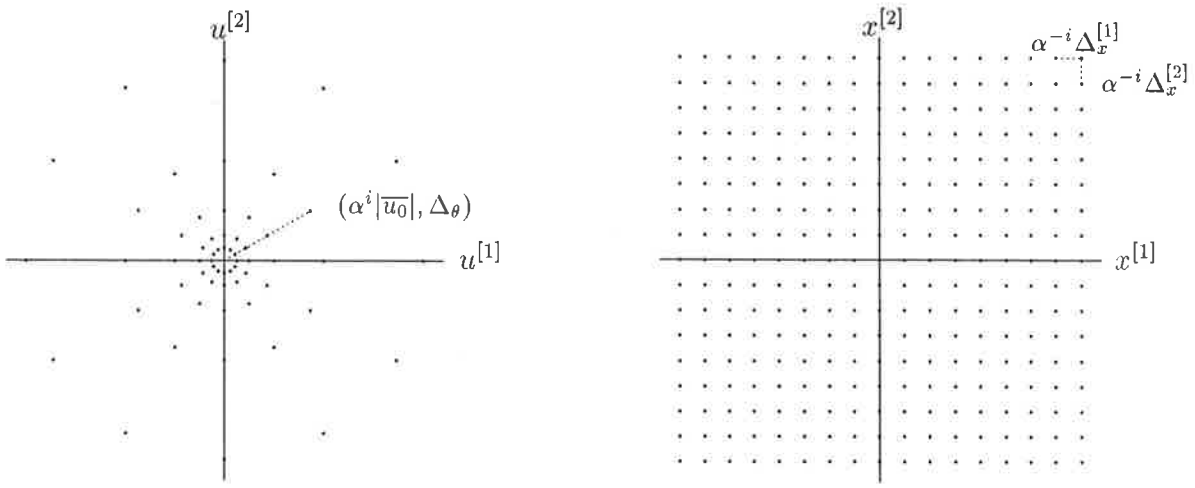
3.5.1 Introduction

According to the DWFT hypothesis of simple cell processing, simple cell RFPs can be generated by the regular *modulation* and *translation* of a common Gaussian window. However it was noted in Section 3.4.5 that the consequent prediction of spatial frequency bandwidths which are approximately constant on a linear scale as the modulation frequency is varied is at odds with the available electrophysiological evidence showing approximately constant bandwidths on a logarithmic scale. Such evidence led Sakitt & Barlow (1982) to propose an alternative scheme for the representation of the visual image by the simple cells, according to which the RFPs of the simple cell population were

assumed to be characterised by the *dilation*, *rotation* and *translation* of a pair of even- and odd-symmetric GRGF templates. Such a scheme is suggestive of an analysis of the visual image at different spatial scales or resolutions, and with the exception of the additional degree of freedom offered by the rotation operation, is consistent with the general class of *multiresolution decomposition* schemes (see e.g. Mallat (1989b)). However, the multi-stage *pyramidal* architectures usually used for the efficient implementation of multiresolution schemes are incompatible with the limited number of cortical layers in V1 in which the simple cells are located. In an effort to both fit the available anatomical and electrophysiological data and to ensure the completeness of the representation, Sakitt & Barlow (1982) chose appropriate dilations, rotations and translations on a somewhat *ad hoc* basis, permitting only empirical arguments in favour of the completeness of the representation. These arguments were based on the coverage of the spatial-frequency domain afforded by the choice of bandwidths and centre frequencies and on the Sampling Theorem, which is not strictly applicable to such a representation scheme since the GRGFs are not strictly band-limited (Mallat, 1989a). A related scheme proposed independently by Watson & Ahumada (1983) suffered from a similar lack of formal completeness results, despite exhibiting greater regularity in the spatial and spectral sampling lattices.

Since the chosen Gabor functions are not mutually orthogonal, a proper treatment of the issue of completeness must once again involve the theory of frames. However, since completeness results are not yet available for such sophisticated sampling schemes, some simplifications are necessary. In particular, the eccentricity-dependent decline in the cutoff frequency exhibited by primate and feline visual systems and included in the models of both Sakitt & Barlow (1982) and Watson & Ahumada (1983), the scale-dependent rotation increments used in the former scheme, and the biological likelihood of somewhat irregular sampling lattices, will be ignored in the present analysis. With these simplifications, one is left with a scheme such as that proposed by Field (1987)[32], which employs Gabor functions distributed on a *log polar* sampling lattice in the spatial frequency domain as shown in Figure 3.6(a), and a frequency-dependent Cartesian lattice in the spatial domain as illustrated in Figure 3.6(b); an explanation of the notation used in Figure 3.6 is provided later in Section 3.5.2. Noting that the generation of a set of functions through the regular dilation, rotation and translation of a single function, along with these resultant spatial and spectral sampling lattices, is highly reminiscent of a representation scheme known as the *discrete wavelet transform* (Daubechies et al., 1986; Mallat, 1989a) or DWT, Daugman (1988a) referred to the resultant functions as a set of *Gabor wavelets*. Since recent interest in wavelets has produced a number

[32]I refer here to the scheme by Field (1987) which uses the true Gabor functions, as opposed to his second scheme — proposed in the same paper — involving sinusoidal functions having windows which are Gaussian on a logarithmic scale. The latter scheme is related to several proposed by Porat & Zeevi (1988) involving distortions of the Gabor functions, which are beyond the scope of the present consideration.



(a) 2D spectral lattice

(b) 2D spatial lattice

Figure 3.6: (a) Lattice of centres $(u_{ik}^{[1]}, u_{ik}^{[2]})$ in the spatial-frequency domain for a wavelet-like 2D representation with $\alpha = 2$. (b) Lattice of centres $(x_{ipq}^{[1]}, x_{ipq}^{[2]})$ in the spatial domain for one value of the scale parameter i. The notation is explained in Section 3.5.2.

of useful results concerning the completeness of wavelet sets, an important step in the search for a completeness proof for the type of representation proposed by Field (1987) is an exploration of the link between this wavelet-like representation and Gabor function wavelet schemes. To simplify the following development, we concentrate initially on the 1D case, and then extend consideration to two dimensions.

3.5.2 Discrete Wavelet Transform

The discrete wavelet transform of a function $f \in L^2(\mathbb{R})$ is given by

$$[Wf](i, p; \psi, \alpha) = \int_{-\infty}^{\infty} f(x)\psi_{ip}(x)\, dx \qquad (i, p) \in \mathbb{Z}^2 \tag{3.20}$$

where the 1D wavelets

$$\psi_{ip}(x) \triangleq \sqrt{\alpha^i} \cdot \psi(\alpha^i x - p\Delta_x)$$

are generated from a single *mother* wavelet $\psi : \mathbb{R} \to \mathbb{C}$ by dilating it by the factor α^{-i}, where $1 < \alpha \in \mathbb{R}_+$, and translating it by $p\alpha^{-i}\Delta_x$, where $\Delta_x \in \mathbb{R}_+$. The set $\{\psi_{ip} : (i, p) \in \mathbb{Z}^2\}$ for a given mother wavelet function ψ is known as a set of 1D *discrete affine coherent states* (Daubechies et al., 1986), since they are generated from the mother wavelet (up to the normalisation factor $\sqrt{\alpha^i}$) by an affine transformation of the spatial variable x. A bandpass function ψ is *admissible* as a mother wavelet if it satisfies the condition (Daubechies et al., 1986)

$$\int_0^{\infty} \frac{|\Psi(u)|^2}{|u|}\, du < \infty \tag{3.21}$$

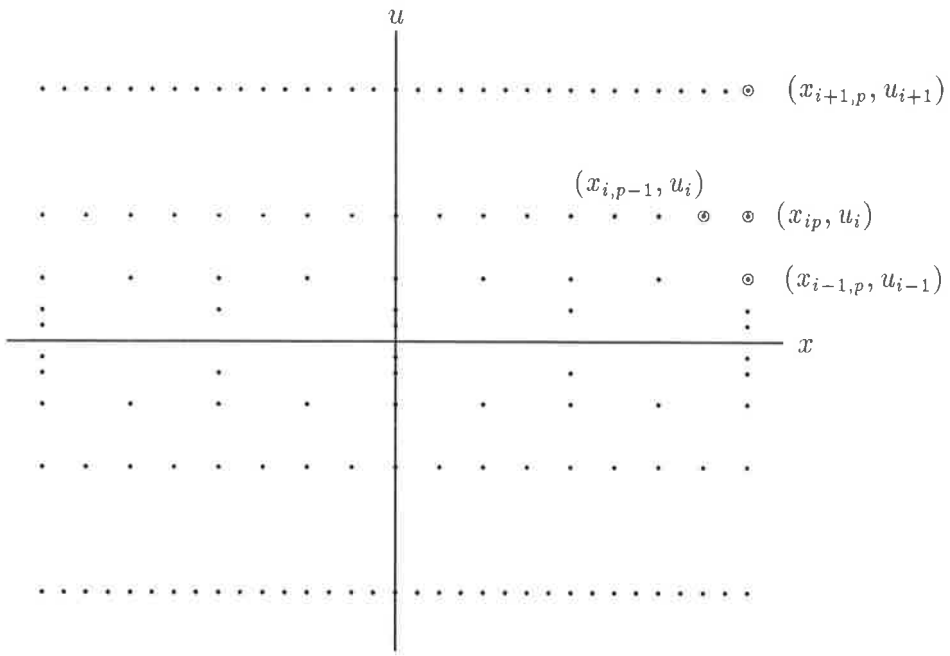


Figure 3.7: Phase-space lattice of centres (x_{ip}, u_i) — shown as dots — for the functions ψ_{ip} used by the 1D wavelet transform with $\alpha = 2$.

where $\Psi \triangleq \mathcal{F}\{\psi\}$ is the Fourier transform of ψ, which in particular requires that $|\Psi(0)| = 0$ (Mallat, 1989a). If ψ is admissible, then the set $\{\psi_i\}$ of discrete affine coherent states which uses ψ as its mother wavelet may, under certain additional conditions on ψ, Δ_x and α, be shown to form a frame for $L^2(\mathbb{R})$, and hence to be complete over this set of square-integrable functions (Daubechies et al., 1986; Heil & Walnut, 1989). Denoting by $\overline{u_0}$ the centre frequency of ψ such that

$$\int_0^\infty (u - \overline{u_0})|\Psi(u)|^2 \, du = 0$$

the distribution in phase space of the spatial and spectral centres

$$\begin{aligned} x_{ip} &= p\alpha^{-i}\Delta_x \\ u_i &= \alpha^i \overline{u_0} \end{aligned}$$

of the functions ψ_{ip} is illustrated in Figure 3.7 for $\alpha = 2$. The centre spatial frequencies u_i are evenly spaced on a logarithmic scale, while as can be seen in Figure 3.7, the spatial sampling interval varies inversely with the centre spatial frequency. Ignoring for the moment the problems associated with implementing possibly complex-valued RFPs, each coefficient $[Wf](i, p; \psi, \alpha)$ of the 1D discrete wavelet transform could according to (3.20) be calculated by a simple cell having corresponding 1D spatial RFP ψ_{ip}, related to all other simple cell RFPs by dilation and translation. This possibility will be investigated further in Section 3.5.3 following an exposition of the extension of wavelets to 2D.

The 2D discrete wavelet transform of a function $f \in L^2(\mathbb{R}^2)$ is given by

$$[Wf](i,p,q;\psi,\alpha) = \int_{-\infty}^{\infty} f(\boldsymbol{x})\psi_{ipq}(\boldsymbol{x})\, d\boldsymbol{x} \qquad (i,p,q) \in \mathbb{Z}^3 \tag{3.22}$$

where the 2D wavelets

$$\psi_{ipq}(\boldsymbol{x}) \triangleq \sqrt{\alpha^i} \cdot \psi(\alpha^i \boldsymbol{x} - [p\Delta_x^{[1]}, q\Delta_x^{[2]}]^T)$$

are generated from the mother wavelet $\psi : \mathbb{R}^2 \rightarrow \mathbb{C}$ by dilating it by the factor α^{-i}, where $1 < \alpha \in \mathbb{R}_+$, and translating it by $\alpha^{-i}[p\Delta_x^{[1]}, q\Delta_x^{[2]}]^T$, where $\Delta_x^{[1]}, \Delta_x^{[2]} \in \mathbb{R}_+$. The admissibility condition

$$\int_0^{\infty} \frac{|\Psi(s\boldsymbol{u})|^2}{s}\, ds < \infty \qquad \forall \boldsymbol{u} \in \mathbb{R}^2 \tag{3.23}$$

is analogous to that in (3.21), where the integral is now performed along a radial line in the frequency domain, and must be finite for all possible radial directions. In particular, admissibility once again requires that the Fourier transform Ψ of the mother wavelet satisfy $|\Psi(\mathbf{0})| = 0$ (Mallat, 1989a). Since the set $\{\psi_{ipq} : (i,p,q) \in \mathbb{Z}^3\}$ does not explicitly incorporate rotations of the mother wavelet, orientation dependence can be incorporated by choosing a rotationally symmetric mother wavelet ψ, dividing its Fourier transform into m sectors, and choosing corresponding wavelets $\{\psi_k : k \in [1,m] \subset \mathbb{Z}_+\}$ whose Fourier transforms Ψ_k are respectively concentrated on each of these sectors and satisfy

$$\sum_{k=1}^{m} |\Psi_k(\boldsymbol{u})|^2 = |\Psi(\boldsymbol{u})|^2 \tag{3.24}$$

(Mallat, 1989a). This construction yields the 2D spatial and spectral sampling lattices illustrated in Figure 3.6, where the sub-wavelets ψ_{ikpq} at a given scale are related to each other by rotation through multiples of $\Delta_\theta = \frac{2\pi}{m}$ and translation. As indicated earlier, these lattices are also employed by the representation scheme proposed by Field (1987).

3.5.3 Do Simple Cells Perform a Discrete Wavelet Transform?

In attempting to determine whether or not the simple cells implement a discrete wavelet transform, the present discussion focuses initially on the Gabor function model of the simple cell RFP; the conclusions drawn from this analysis are then shown to apply to a broad class of RFP models. The question arises as to whether the Gabor functions are suitable candidates for the mother wavelet function ψ in 1D, or for the mother sub-wavelet functions ψ_k in 2D.

The 1D Gabor functions are not admissible as mother wavelets, since all complex-valued Gabor functions have, by virtue of the Gaussian tail, non-zero magnitude at the origin of the spatial frequency plane, and hence fail to satify the admissibility condition (3.21). Since of the real- and complex-valued Gabor functions only the sine-phase RGF

satisfies this condition, the theory of wavelets in its present form cannot, under the assumption of a Gabor function model for the simple cell spatial RFP, provide an adequate description of the variety of simple cell RFPs observed. To rationalise the continued use of the completeness results offered by wavelet theory (Daubechies et al., 1986) in spite of the above objections, the pragmatist may however be tempted to exploit the fact that a complex-valued Gabor function has *approximately* zero DC response for all sufficiently large centre frequencies — given by the frequency of the sinusoid — relative to the size of the Gaussian. However, the justification of this approach would seem to require more RFP subfields than are exhibited by most simple cells.

Nevertheless it will prove instructive to indulge this approximation for the moment, and to examine whether the *2D* Gabor functions are suitable candidates for the sub-wavelet functions ψ_k. Owing to the finite response of these functions at zero frequency, the function Ψ will also have non-zero response at DC, which strictly speaking renders the rotationally symmetric function ψ inadmissible as a mother wavelet according to condition (3.23). However, the satisfaction of (3.24) furthermore requires that Ψ_k be polar separable, which is neither satisfied by the Gabor functions nor characteristic of the spectral RFP of the simple cells[33] (Daugman, 1980; Jones et al., 1987). Thus, in addition to their failure to satisfy the admissibility condition, Gabor functions are further excluded by the multi-orientation two-dimensional wavelet transform on the basis of their inability to satisfy this isometry-preserving condition.

Despite the marked similarities between the 2D discrete wavelet transform and the representation scheme of Field (1987), the latter scheme is therefore not a wavelet scheme, and is consequently not covered by the frame-based completeness results concerning wavelet representations. The required completeness results therefore await future developments in the theory of such wavelet-like schemes. It should further be noted that the two objections raised above — viz. non-zero DC response and lack of polar separability in the spatial frequency domain — to the use of Gabor functions in a discrete wavelet transform extend to most if not all plausible models of the simple cell RFP, and argue against the DWT as a description of the processing performed by the simple cells.

3.5.4 Gabor "Wavelet" Expansion

Using the same wavelet-like set of dilated, rotated and translated Gabor functions g_{ikpq} as the representation scheme of Field (1987), Daugman (1988a; 1989a; 1989c; 1990)

[33]In violation of this simple cell property, polar-separable wavelet-like representation schemes have however been proposed. In the somewhat inaptly-named *cortex transform*, for example, Watson (1987) used sub-wavelets which occupy annular sectors of the spatial frequency plane, with the edges rounded off using a Gaussian profile to improve spatial localisation.

proposed an alternative scheme according to which the coefficients c_{ikpq} in the expansion

$$\check{s}(\boldsymbol{x},t) = \sum_{i=-\infty}^{\infty} \sum_{k=1}^{m} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} c_{ikpq}(t) g_{ikpq}(\boldsymbol{x}) \tag{3.25}$$

are required, rather than simply the outputs of simple cells having the functions $\psi_{ikpq} = g_{ikpq}$ as spatial RFPs, as suggested by (3.22). This scheme is therefore similar to the Gabor expansion, except that the functions used in the expansion are now generated by the dilation, rotation, and frequency-dependent translation, rather than modulation and frequency-independent translation. Not surprisingly it suffers from the same shortcoming: the simple cell outputs do not give the required coefficients, the calculation of which must therefore be deferred until later processing stages. Furthermore, for the reasons outlined in the previous section, it is not possible to make use of the frame-based completeness results for wavelet representations (Daubechies et al., 1986) to demonstrate the completeness or otherwise of this scheme.

3.5.5 *Efficient Coding Through Wavelet-Like Analysis*

Each channel of a wavelet-like image analysis consists of an array of sensors or "cells" whose RFPs ψ_{ikpq} are characterised by a common scale i and orientation k. Field (1987) analysed the Fourier amplitude spectra of six natural images, and demonstrated that these spectra are approximated by a $1/f$ decay in amplitude with spatial frequency. Representation of such images using a scheme whose channels exhibit constant bandwidths on a logarithmic scale and logarithmically spaced centre frequencies therefore results in an approximately equal division of image power between these channels (Field, 1987). The choice of a spatial-frequency bandwidth of approximately 1 octave was found to be optimal in the sense that

> ... a small proportion of cells represents a large proportion of the information with a high signal-to-noise ratio...

so that *for any given image*, relatively few individual sensors had outputs which differed significantly from the mean sensor output (presumed zero), and these accounted for most of the output activity or variance. Field (1987) interpreted this observation to mean that second- and higher-order correlations between the pixels of the stochastic process characterising the original image had been largely converted into first order redundancy in the sensor output distributions for each channel. Having thus been made more explicit, the redundancy in the image could be more easily reduced by subsequent cortical processing stages. This interpretation could equally be applied to the related entropy-based observations by Daugman (1988a; 1989a) concerning the Gabor expansion. Furthermore, the difference between the "entropy" measures obtained after applying to the same image

the Gabor "wavelet" expansion and its inner-product equivalent was observed by Pece (1992) to be slight.

However the analysis from which this conclusion was drawn lumped together the non-identically distributed outputs of all channels, rendering it susceptible to the same criticism — detailed in Section 3.4.6 — as that performed by Daugman (1988a; 1989a) for the Gabor expansion. Nevertheless, Field's (1987) demonstration of approximately equal energy — and hence variance of the output "distribution" if the outputs of each channel are assumed to have zero mean — across channels provides partial reassurance that similar results would have been obtained if the channel-wise analysis had been performed. Furthermore, despite having relatively few sensors in a given channel active for any given image produced by the characteristic 2D stochastic process, the present wavelet-like analysis, by virtue of the identical output distributions of these sensors, distributes the representational effort evenly over all sensors in a given channel for an ensemble of such images (Field, 1987). This property can be contrasted with the tendency of principal component analysis — such as that discussed briefly in Section 3.4.1 — to concentrate as much of the representational effort into as few sensors as possible over the image ensemble (Field, 1987). This latter property would be convenient for reducing the *number of sensors* required to represent the image with a specified sufficiently small loss, as opposed to the *number of distinguishable signalling levels* required to transmit the output of each sensor. Since there appears to be an abundance of simple cells in V1 relative to the number of photoreceptors, and each such cell appears to be relatively noisy (Barlow et al., 1987) and subject to severe temporal limitations on the accuracy with which its output can be signalled (MacLennan, 1992b)[34], the latter scheme would appear to be the more advantageous for the visual cortex.

3.6 Applications of Gabor Functions

Gabor functions have proven popular over the past decade for the preprocessing of images for applications including texture analysis, image compression, and pattern recognition. Since texture is often characterised by its spectral signature, of particular interest to the texture analysis community have been the optimal joint spatial and spectral localisation property of the complex-valued Gabor functions with $\boldsymbol{\theta} = \mathbf{0}$, and the flexibility with which resolution in one domain can be traded for resolution in the other. These properties permit the localisation of boundaries between areas of differing texture, as required for segmentation, through the detection of discontinuities in *local* magnitude &/or phase estimates derived from the outputs of Gabor filters (Turner, 1986; Clark & Bovik, 1989; Fogel & Sagi, 1989; Bovik et al., 1990; du Buf, 1990; Bovik, 1991). The potential for image compression through Gabor function preprocessing derives from the

[34]although see the discussion entitled "Neural Iteration" in Section 7.2.3.

work on entropy reduction and efficient coding by Daugman (1988a; 1989a) and Field (1987), as discussed in Sections 3.4.6 and 3.5.5 respectively. Practical compression has been demonstrated by Daugman (1988a; 1989a) for both the Gabor expansion and the Gabor "wavelet" expansion. The primary motivation for the use of Gabor functions in the majority of pattern recognition applications appears to have been their bandpass tuning for both spatial frequency magnitude and orientation, despite the fact that many other functions exhibit similar 2D spatial frequency tuning properties. A non-exhaustive survey of applications of Gabor functions to image coding and analysis is presented in Appendix C.

Motivated more commonly by engineering objectives than any desire for biological fidelity, such applications have nevertheless drawn inspiration from the results reviewed in Chapter 2 showing some justification for the 2D real-valued Gabor function as a model of simple cell spatial RFPs, and at the same time have contributed to our knowledge concerning the benefits *potentially* conferred on the visual system by such RFP functions. However, applications using the complex-valued Gabor functions in the estimation of local amplitude or phase, or relying strongly on the optimal joint localisation property, depend crucially on the use of complex-valued Gabor functions, the biological implementation of which by the simple cells was argued in Section 2.3.3 to be as yet largely unsubstantiated. The fact that the optimal joint localisation property applies only to the complex-valued Gabor functions is sometimes overlooked, with Beck et al. (1990), for example, erroneously invoking the optimal joint localisation property in justification of an application using only real-valued Gabor functions. Furthermore, the truncation of the Gaussian window necessitated by any computational or physical implementation of a Gabor function renders strict reliance on the optimal joint localisation property of the complex-valued Gabor function untenable, since with the imposition of the constraint of finite spatial support on the search for optimally jointly localised functions, a solution other than the truncated Gabor function will almost certainly emerge. Also of concern to those wishing to relate the results of these applications to biological vision is the fact that little effort has been made to exploit even a significant subset of the degrees of freedom exhibited by spatial RFPs in the simple cell population. As indicated in Appendix C, no systematic variation of the aspect ratio λ or the phase parameter ϕ has for example been attempted.

3.7 Conclusion

A number of theories concerning the spatial computational role of the simple cells in primary visual cortex have been investigated. Special attention has been paid to their ability to account for the variety and distribution of spatial RFPs observed in the simple cell population and to the completeness of the visual representation they postulate. In

addition to the limitations outlined in Section 3.1, the validity of this approach is dependent on the extent to which a linear spatial RFP can be considered or even defined independently of the temporal and binocular properties of the true simple cell RFP, a view which in the light of the analysis in Chapter 2 is largely indefensible. Nevertheless on these grounds alone, strong objections have been raised against most existing theories.

The feature-detection hypothesis was discarded on the basis of its incompatibility with the predominant spatial linearity of simple cell processing. The linear matched filtering hypothesis for bars and edges is rendered untenable by: the poor selectivity of simple cell RFPs for edges and bars (Marr & Hildreth, 1980); the existence of simple cells having more subfields than permitted for bar- and edge-selectivity; and the excessive number of simple cells required by the extension to the binocular spatiotemporal case. With the exception of the fractional discriminant functions, derivative-based theories of simple cell spatial computation were commonly found to admit only purely even- and odd-symmetric RFPs. The Gabor expansion and discrete window Fourier transform were found to use an unrealistic distribution of Gabor function parameters, with an unrotated Gaussian of constant size, and an even distribution of the number of subfields visible under the Gaussian. The biorthogonal functional form of the RFPs required for the Gabor expansion was also shown to be inconsistent with those exhibited by the simple cells. The discrete wavelet transform admits neither the Gabor function nor similarly realistic spatial RFP models as mother sub-wavelets, while the Gabor "wavelet" expansion is once again inconsistent with the spatial inner product performed by the simple cells.

A number of image representation schemes which have been tentatively identified with the simple cells show greater biological realism and (perhaps inevitably) mathematical complexity than those examined in depth in this chapter, including for example a retinal-eccentricity dependent variation in the peak spatial frequency (Sakitt & Barlow, 1982; Watson & Ahumada, 1983). Since the mathematical tools required to establish or refute the completeness of these generalised wavelet-like schemes do not yet exist, the resolution of this issue awaits the future generalisation of the existing tools of wavelet theory.

Chapter IV

NEURAL NETWORKS FOR NON-ORTHOGONAL IMAGE DECOMPOSITION

4.1 Introduction

As discussed in the previous chapter, decomposition-based theories of simple cell processing postulate that each simple cell computes the coefficient corresponding to its own spatial RFP in an expansion of the visual image which uses these RFPs as expansion functions. Implicit in the definition of a coefficient employed by these theories is the notion that each cell signals the relative presence of its spatial RFP in the image, a notion which is reminiscent of feature detection hypotheses of simple cell processing. The concept of the "relative presence" of an expansion function can be quantified by choosing a reconstruction error criterion, the minimisation of which, subject to suitable constraints, uniquely defines the set of expansion coefficients.

In this chapter the decomposition of a sampled multidimensional function using a set of non-orthogonal expansion functions is formulated as a *least squared error* (LSE) quadratic optimisation problem, the solution of which corresponds to the *best linear unbiassed estimate* (BLUE) of the original image when the image is subject to zero-mean spherical Gaussian noise. To standardise the formulation of the problem, the sampled function is first reduced to a vector, and the decomposition is thereafter expressed in terms of matrix-vector algebra which is independent of the function dimension. Thus although the discussion focuses almost solely on the decomposition of an image using 2D expansion functions, it should be remembered that the formulation is not limited to the two-dimensional case. Furthermore, since both the image impinging on the retina and the receptive fields of sensory neurons are naturally represented by real-valued functions of the retinal coordinates, the trivial extension to complex-valued functions and expansion functions is not explicitly addressed. However, this extension is achieved simply by replacing the matrix and vector transpose operators with their Hermetian equivalents in the following formulation.

Since degeneracy of the solutions to the resultant LSE problem might be expected to pose problems for any natural or synthetic perceptual system implementing the required decomposition, the problem is regularised to ensure uniqueness of the coefficients in cases where the expansion functions are linearly dependant. The regularisation has the

additional benefit that coefficients are prevented from becoming arbitrarily large in near-degenerate cases. Following a brief review of conventional methods in linear algebra and optimisation theory for the solution of the LSE and regularised LSE problems, the remainder of the chapter is devoted to a review of *linear time-invariant* (LTI) recurrent artificial neural networks (RANNs) whose dynamics converge to the desired LSE or regularised LSE coefficients.

4.2 LSE Image Decomposition Problem

The reader is referred to any standard text on matrix algebra such as Golub & Van Loan (1989) for general information on many of the concepts and algorithms discussed in this and subsequent chapters.

4.2.1 Matrix Formulation

It is required to represent a real-valued image $i(\boldsymbol{x})$ — where $\boldsymbol{x} = (x, y)$ denotes position within the image — using a non-orthogonal set $\mathcal{G}$ of real two dimensional (2D) expansion functions $\{g_j(\boldsymbol{x})\}$. The reconstruction of $i(\boldsymbol{x})$ from the coefficients $\{a_j\}$ is defined to be

$$\hat{i}(\boldsymbol{x}) \triangleq \sum_j a_j g_j(\boldsymbol{x})$$

For a finite and discretely sampled image, such as that represented by the outputs of the photoreceptors of the retina, the expansion functions are sampled on the same sampling grid to produce a two-dimensional array of samples. This array can then be converted into a sample vector by scanning the array in some regular order. For example, if the sampling grid is rectangular, the array can be represented as a (two-dimensional) matrix, and can be scanned in transpose raster order — using the operator $\boldsymbol{vec}:\mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{mn}$ — to produce the required sample vector. Each such vector sampled from a given expansion function $g_j(\boldsymbol{x})$ of the chosen set can be represented as a row of the expansion function matrix $G \in \mathbb{R}^{n\times m}$, where n is the number of expansion functions and m is the number of points on the sampling grid. The coefficients $\{a_j\}$ are arranged in a column vector $\boldsymbol{a} \in \mathbb{R}^n$ in the same row order as their corresponding expansion functions appear in G. The image $i(\boldsymbol{x})$ after discrete sampling can be converted in the same manner as each of the expansion functions into a column vector $\boldsymbol{i} \in \mathbb{R}^m$, leading to the following matrix-vector formulation of the reconstruction

$$\hat{\boldsymbol{i}} \triangleq G^T \boldsymbol{a} \tag{4.1}$$

where ^{T} denotes the transpose operator. This formulation is generalised to functions of arbitrary dimension simply by adopting an appropriate and consistent convention for the arrangement of image and expansion function samples into a column vector.

4.2.2 *Squared Reconstruction Error (SRE) Minimisation*

The problem of image decomposition using the set $\mathcal{G}$ of expansion functions involves the solution of the set of linear equations

$$G^T \boldsymbol{a} = \boldsymbol{i} \tag{4.2}$$

to identify the coefficient vector $\boldsymbol{a}$ for which the reconstructed image $\hat{\boldsymbol{i}}$ most closely resembles the original observed image $\boldsymbol{i}$, in terms of some reconstruction error metric. If the image is subject to additive zero-mean spherical Gaussian noise, then the *best linear unbiassed estimate* (Mohanty, 1986) of the true image is obtained by minimising the *squared reconstruction error* (SRE)

$$E(\boldsymbol{a}) \triangleq \|\boldsymbol{i} - \hat{\boldsymbol{i}}\|_2^2 = (\boldsymbol{i} - G^T\boldsymbol{a})^T (\boldsymbol{i} - G^T\boldsymbol{a}) \tag{4.3}$$

for each image, where $\|.\|_2$ denotes the Euclidean (l_2) norm. The expected value of E over an ensemble of noisy images is referred to as the *mean* squared reconstruction error, and is zero only if the set $\mathcal{G}$ of expansion functions is complete for the image ensemble and the noise has zero variance.

The SRE E in (4.3) is quadratic in $\boldsymbol{a}$, forming a parabolic error surface defined on the domain of expansion function coefficients. Equating the gradient ∇E to zero and noting that the Hessian $H_E(\boldsymbol{a}) = 2GG^T$ is positive semidefinite reveals that the solutions to the SRE optimisation problem

$$\min_{\boldsymbol{a}} \{E(\boldsymbol{a})\}$$

are the solutions to the *normal equations* (Golub & Van Loan, 1989)

$$GG^T\boldsymbol{a} = G\boldsymbol{i} \tag{4.4}$$

for the set of equations in (4.2). If the expansion functions are linearly independent[1] then G has full row rank and $H_E(\boldsymbol{a})$ is positive definite. The minimum of E is then unique and is given by the (Moore-Penrose) *left pseudoinverse solution* (Ogata, 1987)

$$\boldsymbol{a}^*_{LP} = (GG^T)^{-1}G\boldsymbol{i} \tag{4.5}$$

of the over-determined set of equations in (4.2). For this choice of the coefficients, the reconstructed image $\hat{\boldsymbol{i}}$ is the orthogonal projection of the noisy image $\boldsymbol{i}$ onto the subspace spanned by the rows of G, and the corresponding SRE (in the absence of noise) is the square of the Euclidean distance between the image and its projection. If the expansion functions are orthogonal, then the matrix GG^T is diagonal with diagonal entries given by the square of the Euclidean norm of the corresponding expansion vector, and the inversion is trivial.

---

[1] for which a necessary condition is that $n \leq m$

If the expansion functions are linearly *dependent*[2] then G is row rank deficient and $H_E(\boldsymbol{a})$ is positive *semidefinite* but singular, so that the minimum of E is no longer unique. This degeneracy is usually resolved by seeking among the coefficient vectors corresponding to minima of the cost function that which has the minimum Euclidean norm, giving rise to the following solution (Golub & Van Loan, 1989, Thm 5.5.1)

$$\boldsymbol{a}_+^* = (G^T)^+ \boldsymbol{i} \tag{4.6}$$

where $^+$ denotes the *Moore-Penrose pseudoinverse*[3]. An advantage of this additional minimum Euclidean norm criterion which will become significant during later discussions of implementation of SRE minimisation schemes in analog systems is that the output power required to represent the coefficients is minimised. In the special case where G has full row rank, (4.6) reduces to (4.5) since $\boldsymbol{a}_{LP}^*$ uniquely minimises E, and hence (trivially) also minimises $\|\boldsymbol{a}\|_2$ amongst the minimisers of E (Golub & Van Loan, 1989). If on the other hand there are at least m linearly independent expansion functions (G has full column rank), (4.6) reduces (Hager, 1988) to the (Moore-Penrose) *right pseudoinverse solution*

$$\boldsymbol{a}_{RP}^* = G(G^T G)^{-1} \boldsymbol{i} \tag{4.7}$$

of the under-determined set of equations in (4.2), which gives perfect reconstruction of the image in the absence of noise. Since at most m linearly independent expansion functions are required to completely represent the sampled image, the excess of expansion functions makes this an unlikely scheme for image compression applications; however, it will be seen in Chapter 7 that it warrants closer examination in connection with mammalian visual systems.

4.2.3 *Practical SRE Minimisation*

If the matrix G is known to have full row rank, the SRE E is a positive definite quadratic form, which can be minimised explicitly using optimisation techniques such as *conjugate gradient* algorithms, *quasi-Newton* methods (Press et al., 1988) or iterative matrix techniques, frequently referred to as *relaxation* algorithms. Relaxation algorithms can be characterised as explicit or hybrid explicit-implicit Euler method approximations to the gradient descent equation

$$\dot{\boldsymbol{a}} = -\Gamma \nabla E$$

where $\Gamma \in \mathbb{R}_+^{n \times n}$ is either I_n or $[\mathrm{diag}(GG^T)]^{-1}$, and $diag{:}\mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ returns the matrix obtained by replacing the offdiagonal entries of its matrix argument with zeros. The

[2] which occurs for example if there are more expansion functions n than image sample points m

[3] henceforth referred to simply as the *pseudoinverse*, since there is no potential for confusion with other pseudoinverses in the present context.

application of relaxation algorithms — which include the *Jacobi*, *Gauss-Seidel* and *Successive Overrelaxation* (SOR) algorithms — to the solution of the normal equations in (4.4) has been reviewed by Yan & Gore (1990) in the context of non-orthogonal image decomposition. Alternatively, the normal equations in (4.4) can be solved by direct methods such as matrix inversion or Gaussian elimination. Ill-posedness of the problem — which is exacerbated by the squaring of the condition number during conversion of the equations in (4.2) to normal form — can be to some extent overcome by *scaling* (Golub & Van Loan, 1989) the normal equations — a technique which is examined further in Section 5.4.

If on the other hand the matrix G is known or suspected to be row rank deficient, explicit computation of the pseudoinverse solution in (4.6) is required. The calculation of the pseudoinverse usually involves the *singular value decomposition*[4] (SVD) of G, which produces errors of the order of $\delta\sigma_{max}(G)$ in the singular values, where $\sigma_{max}(G)$ denotes the maximum singular value of G and δ denotes the machine floating-point precision (Golub & Van Loan, 1989). Denoting by $\overline{\sigma}_{min}(G)$ the smallest *non-zero* singular value of G, the *generalised spectral condition number* of G is defined as

$$\overline{\kappa}_2(G) \triangleq \frac{\sigma_{max}(G)}{\overline{\sigma}_{min}(G)} \geq 1 \tag{4.8}$$

For the present purposes, G will be said to be *ill-conditioned* if $\overline{\kappa}_2(G) \gg 1$, and *well-conditioned* if $\overline{\kappa}_2(G) \approx 1$. If G is particularly ill-conditioned, $rank(G)$ — which is required for the calculation of the pseudoinverse — can be difficult to ascertain from the SVD due to the relatively large errors in the estimation of the zero and near zero singular values. Thresholding of the singular values does little to alleviate this problem, since an appropriate threshold is not known *a priori*, and estimates of non-zero singular values can lie below the threshold, while those of zero singular values can lie above it. In the calculation of the pseudoinverse, the estimated singular values which are nonzero after thresholding are inverted to form singular values of the pseudoinverse matrix, while the remaining singular values of the pseudoinverse are set to zero. An overestimate of $rank(G)$ will therefore result in the estimate of the pseudoinverse having one or more singular values which should be zero but are in fact of the order of the inverse of the threshold. Thus the estimated pseudoinverse is a strongly discontinuous function of the estimate of $rank(G)$.

An alternative approximate solution method which does not suffer from the problem of discontinuity with estimated rank is suggested by the observation of den Broeder & Charnes (1957) that the Moore-Penrose pseudoinverse solution given in (4.6) can be expressed as

$$\boldsymbol{a}_+^* = \lim_{\varepsilon \to 0} \left\{ (GG^T + \varepsilon I_n)^{-1} G \right\} \boldsymbol{i} \tag{4.9a}$$

[4]although *Greville's method* (Greville, 1960) is an alternative.

$$= \lim_{\varepsilon \to 0} \left\{ G(G^T G + \varepsilon I_m)^{-1} \right\} \boldsymbol{i} \tag{4.9b}$$

where I_n and I_m are the $n \times n$ and $m \times m$ identity matrices respectively. The equality of the two terms in limits can be established by adding εG to $GG^T G$ and pre- and post-factorising by G. As expected, as $\varepsilon \to 0$ these terms tend respectively to the left pseudoinverse if G has full row rank, and to the right pseudoinverse if G has full column rank. In the more general case, the calculation of either term for suitably small $\varepsilon > 0$ produces an approximation to the true pseudoinverse.

For a given $\varepsilon > 0$

$$\boldsymbol{a}_\varepsilon \triangleq (GG^T + \varepsilon I_n)^{-1} G\boldsymbol{i} \tag{4.10}$$

is the unique stationary point of the error function

$$E_\varepsilon(\boldsymbol{a}; \varepsilon) \triangleq \|\boldsymbol{i} - \hat{\boldsymbol{i}}\|_2^2 + \varepsilon \|\boldsymbol{a}\|_2^2 \tag{4.11}$$

and since the eigenvalues of the Hessian

$$H_{E_\varepsilon}(\boldsymbol{a}) = 2\,(GG^T + \varepsilon I_n)$$

are simply twice those of the positive semidefinite matrix GG^T plus 2ε — from which it follows that $H_{E_\varepsilon}(\boldsymbol{a})$ is positive definite — this stationary point is also the point at which the global minimum of E_ε is attained (Ben-Israel & Greville, 1974). The optimisation problem

$$\min_{\boldsymbol{a}} \{E_\varepsilon(\boldsymbol{a}; \varepsilon)\} \tag{4.12}$$

is referred to as the *ridge regression* problem (Golub & Van Loan, 1989), and from (4.11) has as its solution an *approximate* minimiser of the SRE which has a smaller norm. The *regularisation* (Hager, 1988) or *ridge* (Golub & Van Loan, 1989) parameter $\varepsilon > 0$ controls the trade-off between minimisation of the SRE and minimisation of $\|\boldsymbol{a}\|_2^2$ in the optimisation of E_ε, and can be chosen to avoid undue sensitivity of the computed solution to errors or noise in any one element of the observed image vector $\boldsymbol{i}$ (Golub & Van Loan, 1989). However, for the present purposes it will be assumed that $\varepsilon > 0$ is chosen to be small enough to achieve a suitable approximation to the true pseudoinverse solution.

Since E_ε is a positive definite quadratic form, the application of gradient-based algorithms to the augmented normal equations

$$(GG^T + \varepsilon I_n)\,\boldsymbol{a} = G\boldsymbol{i} \tag{4.13}$$

arising from the ridge regression problem is guaranteed to find the unique global minimum of E_ε. Alternatively, these augmented normal equations can be solved in the same ways — outlined in the beginning of this section — as the original normal equations for the full-rank case. Although as $\varepsilon \to 0$ the condition number of the augmented normal equations approaches the square of that of G, the resultant errors are no longer subject

to the strong discontinuity with estimated rank associated with the explicit calculation of the pseudoinverse. Furthermore the larger the value of ε, the better the conditioning of the augmented equations, although of course the worse the resultant approximation of the true pseudoinverse.

4.3 Neural Networks for SRE Minimisation

For realistic image processing applications, the number m of pixels is frequently of the order of 10^5 (e.g. 512×512), and if accurate reconstruction is required, the number n of expansion functions is of the same order. The computation of the pseudoinverse solution or an approximation thereto by any of the means discussed in Section 4.2.3 is therefore highly computationally intensive, and the use of parallel computation is a necessity for real-time applications. The reader is referred to Golub & Van Loan (1989, Chap. 6) for an overview of parallel matrix computations on generic conventional parallel computer architectures.

In order to investigate the feasibility of the implementation of an SRE minimisation scheme in the mammalian early visual system, the solution of the SRE problem using a class of fine-grained parallel computing architectures known as *neural networks* is examined. Attention is focused on recurrent analog neural networks, formed by the recurrent weighted interconnection of a large number of analog computing elements, each of which consists of a summing integrator with cascaded non-linearity. These networks — of which possibly the best known example is the *Hopfield* network (Hopfield, 1984)— constitute a subset of the class of *additive* neural networks (Grossberg, 1969) which are commonly used to describe biological neural networks at moderate levels of abstraction. Although the identification of suitable analog *very large scale integrated* (VLSI) circuit implementations of such networks is an on-going research issue (see e.g. Schach (1992)), their structural regularity has intuitive appeal in VLSI circuit design. The possible implementation of neural networks discussed in this and the following chapter in the real neural *wetware* (hardware) of the mammalian early visual system is examined in Chapter 7.

The implementation on conventional fine-grained parallel architectures — such as *single instruction multiple data* (SIMD) machines — of direct methods for the computation of the pseudoinverse solution or its approximation — including singular value decomposition, or inversion of (4.13) — requires global coordination of the actions of the individual processing elements. However such coordination, which is also required by optimisation algorithms classified as *conjugate gradient* or *quasi-Newton* methods (Press et al., 1988) to implement the necessary one-dimensional sub-minimisations, is inconsistent with the autonomy of neurons in a recurrent neural network. In contrast, relaxation algorithms require no such coordination and are therefore — with minor modification in some cases to permit simultaneous updating of all coefficients (Yan & Gore, 1990) — well

suited to implementation in recurrent neural architectures. Furthermore, when formulated in continuous-time, relaxation algorithms lead naturally to analog neural network implementations.

In this section, neural network approaches to the least-squares solution of a full-rank over-determined set of linear equations are reviewed. In the following section, the application of neural network implementations of the ridge regression problem to the solution of rank-deficient and under-determined sets of linear equations is examined. In keeping with the current context of image decomposition using non-orthogonal expansion functions, the terminology and much of the discussion in this and the following section centres on SRE optimisation; however it should be born in mind that this is but one application of the least-squares solution of sets of linear equations, to which these neural networks are more generally applicable. The order of presentation of the neural network models has been chosen to facilitate the ordered development of the relevant concepts, and in places does not follow chronological order of publication. Finally, in order to avoid distracting considerations of the capabilities and peculiarities of particular VLSI technologies, architectural comparisons of the various models presented in this and the following section have been performed at the relatively abstract level of counts of connections and ideal components such as summers and integrators. More detailed technological comparisons are beyond the scope of this review.

4.3.1 Daugman (1988a)

An artificial neural network for MSRE image decomposition was first proposed by Daugman (1988a). Illustrated in Figure 4.1, Daugman's network was intended to implement the coefficient update equation (Daugman, 1988a, Eqn 8)

$$\boldsymbol{a}(t+1) = \boldsymbol{a}(t) + [G\boldsymbol{i} - G\,(G^T\boldsymbol{a}(t))] \tag{4.14}$$

the continuous-time equivalent of which performs steepest descent ($\dot{\boldsymbol{a}} = -\nabla E$) on E (Daugman, 1990), and hence is guaranteed to find the unique global optimum. However, (4.14) constitutes an explicit Euler method approximation to the continuous-time case with unit step size, and is known to be unstable whenever an eigenvalue of GG^T exceeds 2. In such cases, stability of Daugman's algorithm can be achieved by decreasing the step size. As will be seen in Section 5.4, the rate of convergence to the optimal solution can be accelerated for a given step size $\alpha \in (0, 1)$ by employing the *diagonal preconditioning* strategy

$$\boldsymbol{a}(t+\alpha) = \boldsymbol{a}(t) + \alpha\Gamma\,[G\boldsymbol{i} - (GG^T)\,\boldsymbol{a}(t)] \tag{4.15}$$

inherent in the *Jacobi iteration* for the system of equations in (4.4), where

$$\Gamma \triangleq [\mathrm{diag}(GG^T)]^{-1}$$

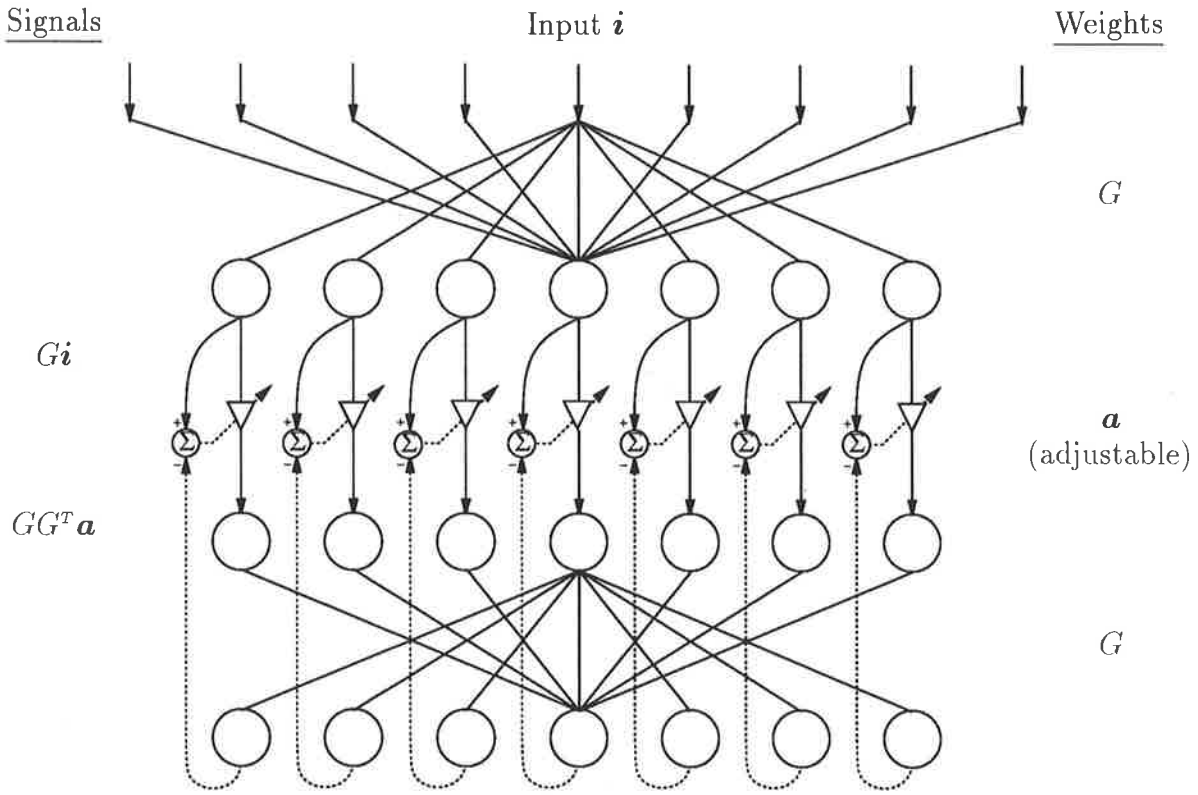


Figure 4.1: Network proposed by Daugman (1988a) for Gabor image decomposition. Reproduced from figure 3 of Daugman (1988a) with matrix-vector notation. The following conventions have been adopted in this and subsequent neural network diagrams (except where otherwise indicated): to avoid clutter, only a subset of the links is shown for each connection matrix; links carrying data in the top-to-bottom and bottom-to-top directions are denoted by solid and dotted lines respectively; open circles represent summers; triangles represent adjustable weights, and the arrowed lines which intersect them represent weight update signals.

is the diagonal *preconditioner* and $diag{:}\mathbb{R}^{n\times n} \rightarrow \mathbb{R}^{n\times n}$ returns the diagonal matrix obtained by setting the offdiagonal entries of its matrix argument to zero. Equation (4.15) reduces to (4.14) in the case of normalised expansion functions ($\mathrm{diag}(GG^T) = I_n$) and unit step size ($\alpha = 1$) (Yan & Gore, 1990). The required coefficients are represented as "weights" on the feedforward connections between the first and second neural layers, and are adjusted according to the difference between the outputs of the first and third layers. The feedforward weight vector of each neuron in the first layer represents a corresponding row of the weight matrix G. Note that the expression of the expansion function corresponding to this row as a weight vector is purely for notational convenience, and that the spatial implementation of these weights is more naturally viewed in the original dimensions of the expansion function.

According to the unwavering description given by Daugman (1988a; 1989c; 1989b; 1990) of the operation of the network, the first and third neural layers are "identical",

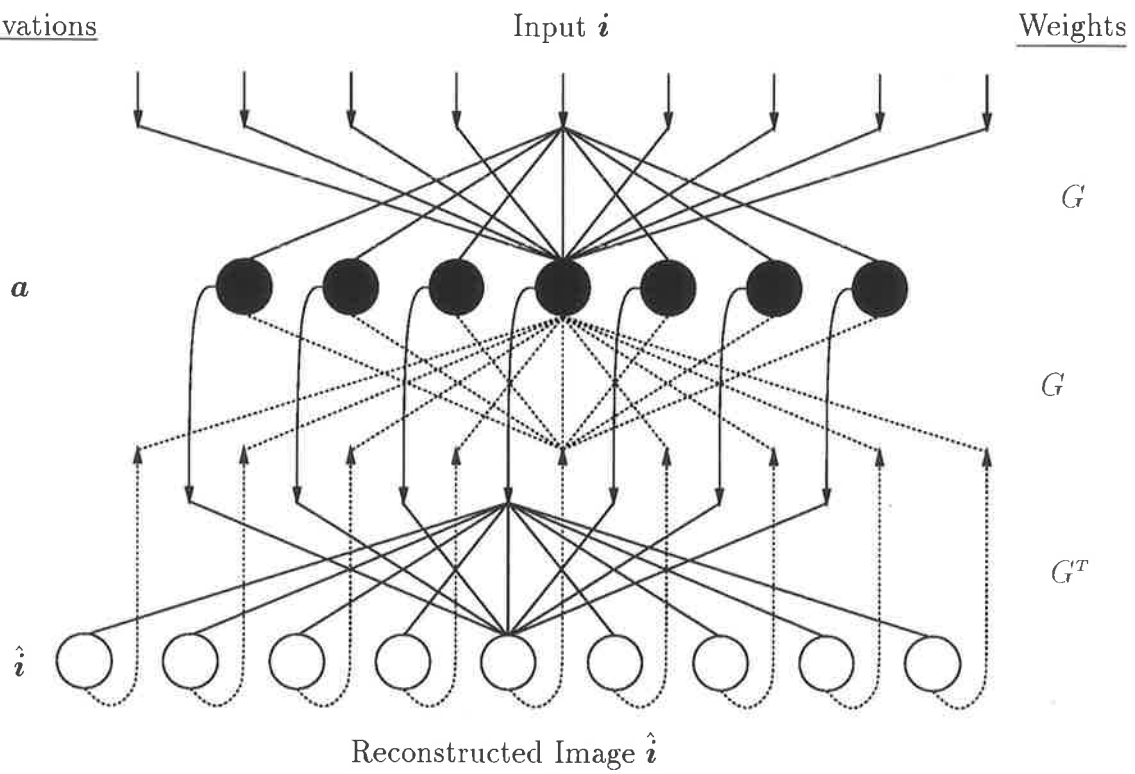


Figure 4.2: Corrected and condensed version of network proposed by Daugman (1988a). Black circles represent (explicit Euler method approximations to) ideal inverting summing integrators.

and in particular they have the same input weight matrix G (as shown in Figure 4.1 and indicated by the explicit parentheses in (4.14)). However, this description is inconsistent with the labelling of the dotted feedback pathway serving as input to the weight adjustment controller, since the second neural layer as drawn is not in a position to perform the necessary prior multiplication of the coefficient vector by the matrix G^T. Furthermore, the description of the coefficient representation as "weights" is also inconsistent with this labelling, since if the description were true, the expression in the label should include the image vector $\boldsymbol{i}$. There are a number of ways in which the network may be redesigned to be consistent with the equations it was intended to implement. Arguably the simplest — although see Section 4.3.2 — is to represent the coefficient vector as the *activations* of a surrogate neural layer in place of the "weight layer" — thereby disposing of the feedforward connections from the first neural layer to the weight "layer" — and to add cross connections representing the matrix G^T between the outputs of the new neural layer and the inputs of the (formerly) second neural layer.

For the purposes of fair comparison with other neural network models, the number of nodes and connections required by the resultant network can be reduced by folding the third layer back onto the first, as illustrated in Figure 4.2. The coefficients are

represented by the *activations* of the resultant composite first-layer nodes, which act as (explicit Euler method approximations to) ideal summing integrators, with the update of their activation governed by their summed weighted inputs and current activation level such that

$$\begin{aligned} \boldsymbol{a}(t+1) &= \boldsymbol{a}(t) + [G\boldsymbol{i} - G\hat{\boldsymbol{i}}] \\ &= \boldsymbol{a}(t) + [G\boldsymbol{i} - G(G^T\boldsymbol{a}(t))] \end{aligned} \tag{4.16}$$

In recognition of the distinction between these ideal integrating nodes and normal summing nodes, the former are henceforth denoted by black circles and the latter as previously by white circles.

The feedforward and feedback weight vectors of each neuron in the first layer are identical (with the exception of a change of sign) and represent a corresponding row of G, while the feedforward weights of each neuron in the second layer represent a corresponding row of G^T. The output of the first layer is the coefficient vector $\boldsymbol{a}$, whilst that of second layer is the vector $G^T\boldsymbol{a}$, which is by definition the reconstructed image. The three separate implementations of the matrix G (or its transpose) — as indicated by the explicit parentheses in (4.16) and illustrated in Figure 4.2 — require a total of $3nm$ connections. In addition, the network requires m summing nodes and n integrators.

4.3.2 *Wang & Yan (1992)*

Yan & Gore (1990, Eqn 13) showed how the SOR algorithm can be modified to allow parallel weight updates by removing the dependence of the update of each coefficient a_i on the most recent updates $\{a_j : \quad j < i\}$. When the rows of G have unit Euclidean norm, this modification yields the update equation

$$\boldsymbol{a}(t+1) = \boldsymbol{a}(t) + \rho\,[G\boldsymbol{i} - GG^T\boldsymbol{a}(t)]$$

where $\rho \in [1, 2)$ is the *relaxation parameter*. However, removal of this dependence is achieved at the expense of the enhanced stability conferred by the implicit portion of the hybrid implicit-explicit Euler method used by the SOR algorithm, rendering the modified algorithm less stable than Daugman's iteration scheme in (4.14). As with the latter scheme, a reduction in the step size is frequently necessary to achieve stability, while as will be seen in Section 5.4 if the rows of G have not been normalised, the rate of convergence may be improved significantly by using the diagonal preconditioning strategy

$$\boldsymbol{a}(t+1) = \boldsymbol{a}(t) + \rho\Gamma\,[G\boldsymbol{i} - (GG^T)\,\boldsymbol{a}(t)] \tag{4.17}$$

inherent in the true SOR algorithm, where $\Gamma \triangleq [\mathrm{diag}(GG^T)]^{-1}$ is the diagonal preconditioner.

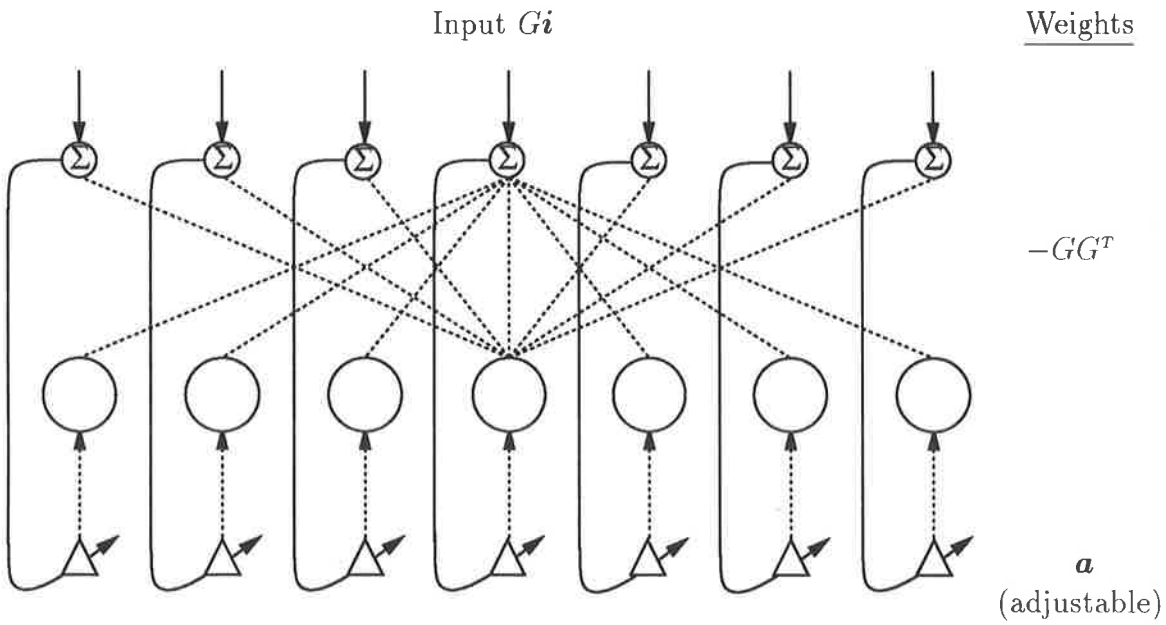


Figure 4.3: Network proposed by Wang & Yan (1992) for Gabor image decomposition. Reproduced from their figure 1 with matrix-vector notation. Open circles *do not* represent summers — see text.

Wang & Yan (1992) modified Daugman's discrete-time network for MSRE image decomposition to implement the modified Successive Overrelaxation (SOR) algorithm in (4.17)[5]; the resultant network is depicted in Figure 4.3. The principal modifications consist of the combination of Daugman's second and third neural layers in Figure 4.1 into a single layer having *output* weight matrix GG^T, the precalculation of the term $G\boldsymbol{i}$ — avoiding the question of how this is to be performed in hardware — the elimination of the erroneous feedforward connections from the signal $G\boldsymbol{i}$ to the adjustable weight layer, and the incorporation of the relaxation parameter ρ and the elements of the preconditioner in the weight update controller. However as with Daugman's network, the use of the term "weights" to describe the representation of the coefficients remains misleading, since the "weights" do not weight a signal input, performing instead the role of adjustable neural activations. Furthermore, the middle layer (open circles) serves no computational purpose, since the weight matrix on its output could equally well be considered to constitute *input* weights to the first (summing) layer, permitting the omission of the middle layer.

[5]Although their equation 13 — which is said to govern the operation of the network — does not incorporate the modification proposed by Yan & Gore (1990) to allow parallel weight updates, it is assumed that it was intended, since otherwise — to use the words of Yan & Gore (1990) in reference to the closely related Gauss-Seidel iteration method —

> ... the weights cannot be updated simultaneously, and thus it is not appropriate for parallel implementation.

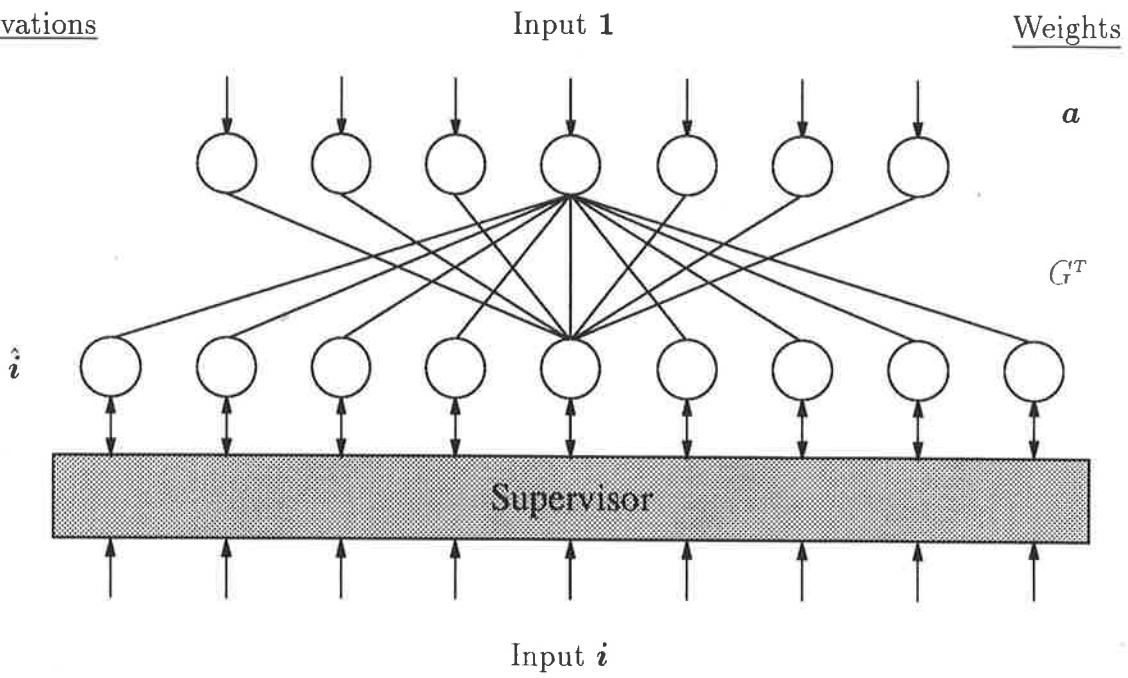


Figure 4.4: Generative Back-Propagation network proposed for MSRE image decomposition by Cohen & Shawe-Taylor (1990). Open circles represent *bidirectional* summers (see text); the direction of signal propagation is controlled by the supervisor.

If these redundant nodes are omitted, the component count is n summing nodes, n^2 weighted connections and n "weight" nodes.

4.3.3 *Cohen & Shawe-Taylor (1990)*

The generative back-propagation (GBP) approach to SRE minimisation (Cohen & Shawe-Taylor, 1990) involves the training by error back-propagation of the input weights to the first layer of a two-layer linear neural network. The network illustrated in Figure 4.4 is presented with a constant input given by the unit vector, which is then weighted by the current estimate $\boldsymbol{a}(t)$ of the MSRE coefficients to produce the output of the first layer nodes. The reconstructed image is then formed as the activations of the second layer nodes by passing the first-layer activations through the weight matrix G^T. The reconstructed image is compared by a supervisor with the original image and the error fed back through the network to adjust the input weights to the first layer.

Although the generative back-propagation network itself is not strictly a relaxation network, its "training" can be seen to implement the following form of the Jacobi-like iteration in (4.14)

$$\boldsymbol{a}(t+1) = \boldsymbol{a}(t) - \alpha G[G^T\boldsymbol{a}(t) - \boldsymbol{i}] \tag{4.18}$$

where $\alpha \in \mathbb{R}_+$. The difference term in brackets — the negative of the *residual* – is evaluated by the supervisor and weighted by G during back-propagation through the

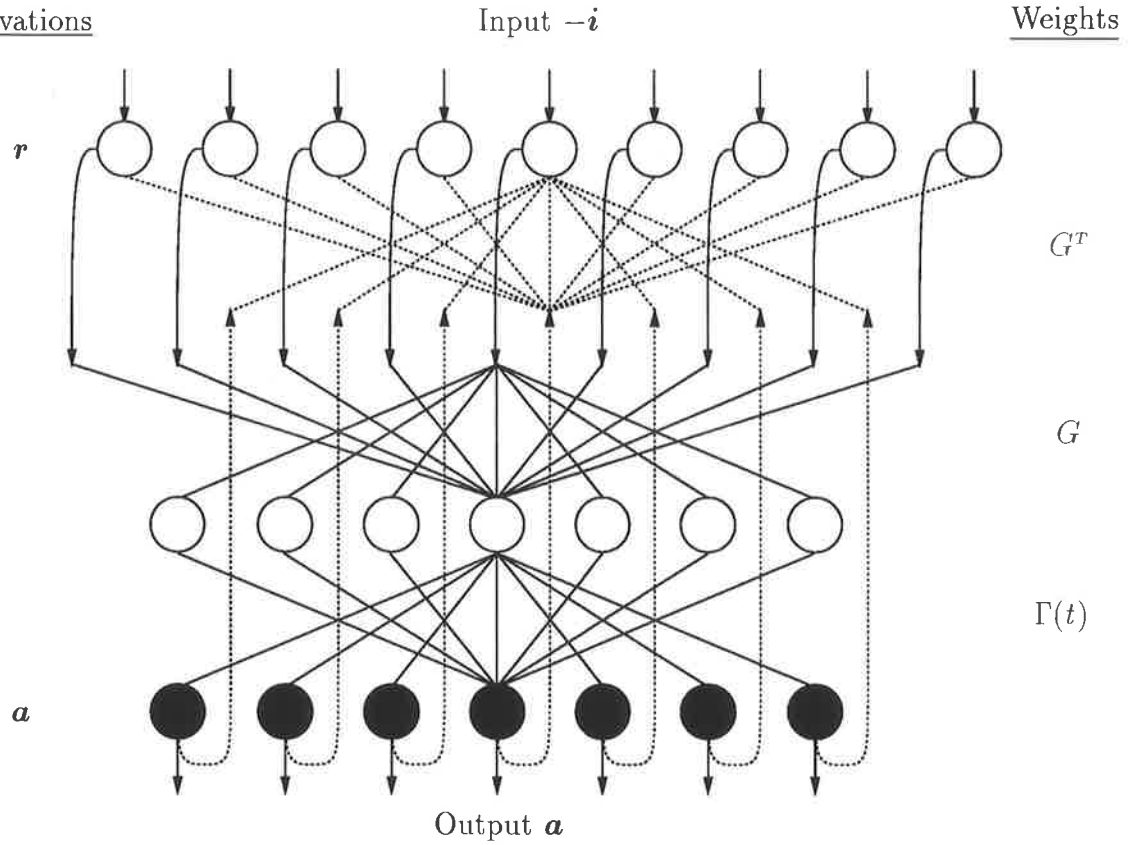


Figure 4.5: Network proposed by Cichocki & Unbehauen (1992) for solving full-rank overdetermined systems of linear equations. Black circles represent ideal inverting summing integrators.

interlaminar weights. The model at first appears to represent a considerable saving in hardware over the modified network in Section 4.3.1, requiring only nm connections and n nodes. However, this saving is achieved at the cost of the additional complexity required by bidirectional weighted connections, and the non-local control by the supervisor of the current direction of signal propagation. These issues are addressed by the following neural network architecture.

4.3.4 *Cichocki & Unbehauen (1992)*

Cichocki & Unbehauen (1992) recently proposed the neural architecture illustrated in Figure 4.5 for the solution of full-rank overdetermined systems of linear equations. The network implements the following set of equations

$$\dot{\boldsymbol{a}} = -\Gamma(t)G\boldsymbol{r} \tag{4.19a}$$

$$\boldsymbol{r} = G^T\boldsymbol{a} - \boldsymbol{i} \tag{4.19b}$$

where $\dot{\boldsymbol{a}}$ denotes the time derivative of $\boldsymbol{a}$ and $\Gamma(t) \in \mathbb{R}^{n\times n}$ is a positive definite preconditioner, for which — provided $\Gamma(t)$ remains positive definite — the SRE defined in (4.3)

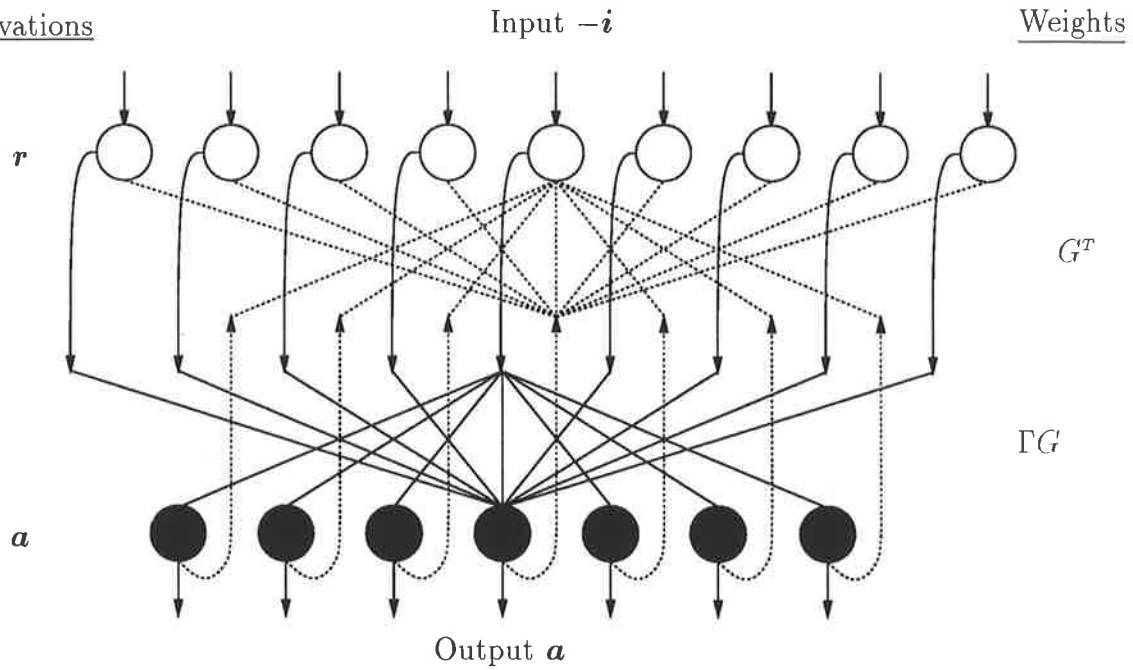


Figure 4.6: Condensed form of network proposed by Cichocki & Unbehauen (1992) for solving overdetermined systems of linear equations. Γ is assumed to be time invariant. Black nodes represent ideal inverting summing integrators.

is a Liapunov function which is monotonically decreasing except at the unique equilibrium point (Cichocki & Unbehauen, 1992). The residual $\boldsymbol{r}$ is evaluated explicitly in the input layer, and then projected onto the Gabor functions followed by $\Gamma(t)$ to modify the coefficient vector $\boldsymbol{a}$.

Cichocki & Unbehauen (1992) showed that the case of non-Gaussian noise distributions on the inputs — such as a Gaussian distribution with outliers — can be dealt with through the use of appropriate non-linear activation functions in the first-layer nodes. In order to achieve linear convergence of the neural network to the optimal solution, the preconditioner $\Gamma(t)$ was adapted according to the length of the residual error vector, albeit at the cost of relatively expensive additional hardware. These two extensions are however beyond the scope of the present analysis.

To facilitate fair and simple comparison with other networks, if Γ is time-invariant and the input nodes have linear activation functions, the second and third neural layers can be condensed into a single layer of ideal summing integrators with input weight matrix ΓG, which can be precalculated. The relative overhead cost involved in this precalculation is negligible provided the system of equations is to be solved repeatedly with different inputs, which is the only case in which one would consider hardware implementation of the neural network. Illustrated in Figure 4.6, the condensed network requires $2nm$ connections, n ideal summing integrators and m summing nodes, which constitutes a saving of nm connections over the modified form of Daugman's network depicted in

Figure 4.2.

The additional cost in connections over the GBP network is due to the replacement of the nm bidirectional connections with two sets of nm unidirectional connections. Furthermore, the computational role of the supervisor has been effectively assumed by the nodes in the first layer, while its role as network scheduler is no longer required, since the coefficients represented by the output of the nodes in the second layer evolve under the system dynamics, rather than by scheduled updates of the network weights.

4.4 Neural Networks for Minimum Norm SRE Minimisation

When applied to an under-determined or rank-deficient set of equations, the foregoing networks find a non-unique minimum of the SRE. This degeneracy of the solutions obtained by the neural network is likely to be problematic in the context of biological or machine vision, since the already complex task of perception is made more difficult by the need to recognise the infinite set of degenerate solutions as representing the same stimulus. Furthermore, which of the degenerate set of solutions is reached by the network is dependent on the state of the network prior to presentation of the current stimulus. This history-dependence is probably undesirable in machine vision applications, although history-dependence is for example an automatic consequence of *automatic gain control* (AGC) mechanisms in biological vision. In this section, neural network approaches to the ridge regression problem — whose unique solution approximates the pseudoinverse solution of a set of linear equations — will be examined.

4.4.1 Culhane et al. (1989)

Culhane et al. (1989) proposed two versions of an analog neural network for computing discrete Fourier and Hartley transforms, for which the matrix G is orthogonal. Shown generalised in Figure 4.7 to arbitrary rectangular G, their "ideal case" model — so called because the summers in the first layer are assumed to have no delay and the two implementations of G are precisely matched — amounts to a special case of the linear programming network of Tank & Hopfield (1986), in which the neurons have linear activation functions and the coefficients of the linear expression to be minimised are zero. Since much of their analysis does not rely on their assumption of the orthogonality of G, it is largely unaffected by the generalisation. The generalised network is similar to the condensed form — shown in Figure 4.6 — of the network proposed by Cichocki & Unbehauen (1992); however in acknowledgement of the difficulty of practical approximation of an ideal integrator, the integrators have been assigned an additional leakage term.

The network is governed by the following set of equations

$$\tau_a \dot{\boldsymbol{a}} = -\boldsymbol{a} - k_a G \boldsymbol{r} \tag{4.20a}$$

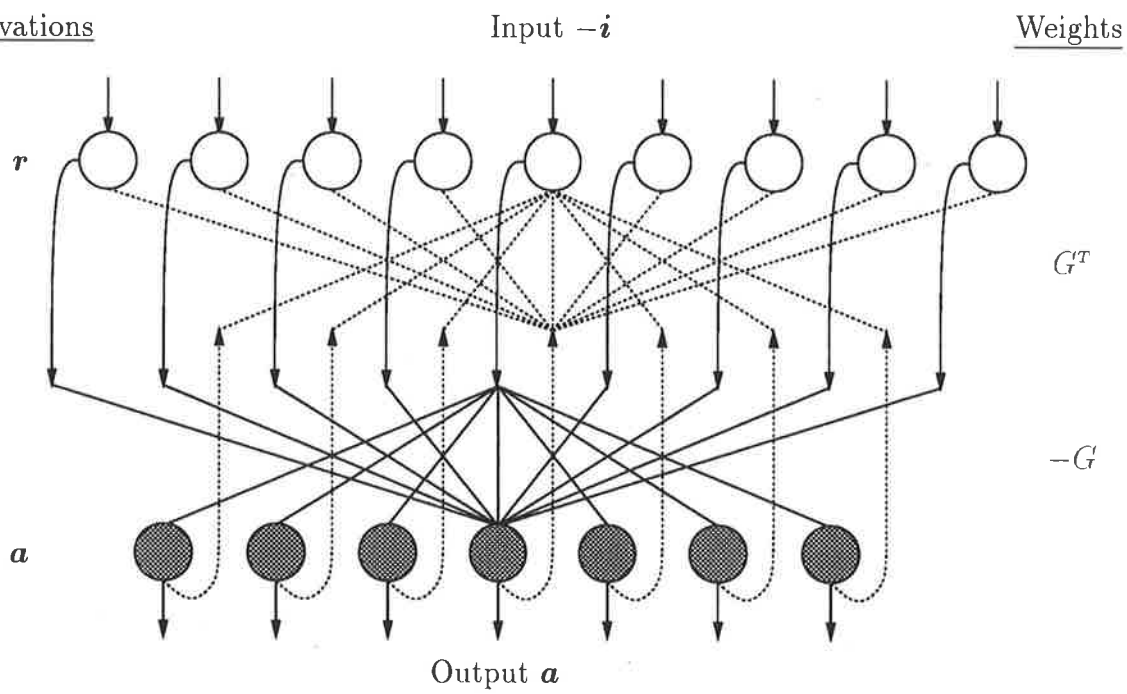


Figure 4.7: Network proposed by Culhane et al. (1989) for computing discrete Fourier and Hartley transforms, as generalised to permit arbitrary rectangular G; see text.

$$\boldsymbol{r} \;=\; k_r \left[G^T \boldsymbol{a} - \boldsymbol{i}\right] \tag{4.20b}$$

where $\boldsymbol{r} \in \mathbb{R}^m$ — which is proportional to the residual — represents the activations of the nodes in the first layer, $\tau_a \in \mathbb{R}_+$ is the integrator time constant, and $k_r, k_a \in \mathbb{R}_+$ are the gains of the summing and summing integrator nodes in the first and second layers respectively. These equations can be combined and rearranged to give

$$\varepsilon \tau_a \dot{\boldsymbol{a}} = -\left(GG^T + \varepsilon I_n\right) \boldsymbol{a} + G\boldsymbol{i} \tag{4.21}$$

where $\varepsilon = (k_a k_r)^{-1}$; (4.21) reveals that the system performs steepest descent on the regularised energy function E_ε in (4.11) and, since the Hessian is positive definite, is globally convergent to the unique equilibrium point given by (4.10). The former observation is consistent with that of Culhane et al. (1989) that

$$E_{k_a k_r} \triangleq \|\boldsymbol{a}\|_2^2 + k_a k_r \|\boldsymbol{i} - G^T \boldsymbol{a}\|_2^2 \tag{4.22}$$

is a Liapunov function for the system, although the authors did not note the approximation of the pseudoinverse solution for large $k_a k_r$ or the consequent extension to the rank-deficient and under-determined cases.

4.4.2 *Yan (1991b)*

Acknowledging the inevitable finite delay (rise-time) associated with any practical implementation of a summer and the difficulty of precisely matching component values,

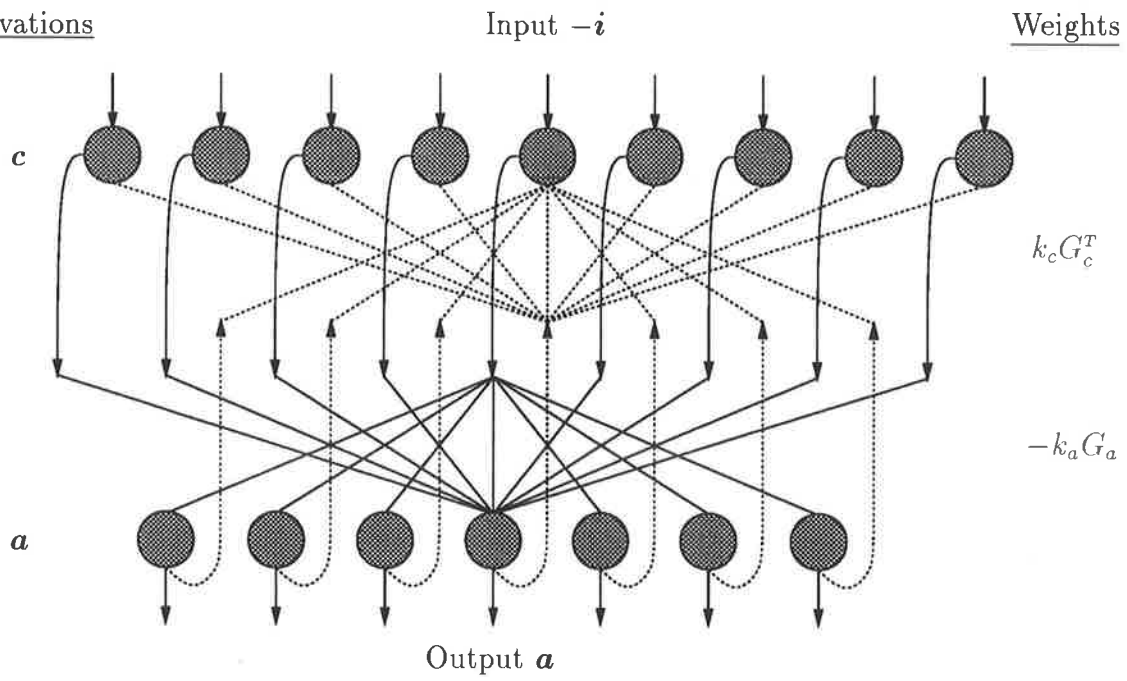


Figure 4.8: Network proposed by Yan (1991a) for Gabor image decomposition.

Culhane et al. (1989, Sect III.B(2)) improved on their "ideal case" network by replacing the summers in the first layer with leaky summing integrators, and allowing the two implementations of the matrix G to differ slightly from their true value. Yan (1991b; 1991a) extended this improved neural network to the case of an incomplete non-orthogonal expansion function set by removing the requirement that the matrix G be symmetric and have orthogonal rows. Shown schematically in Figure 4.8, the network is governed by the following simplified coupled ODEs

$$\tau_a \dot{\boldsymbol{a}} = -\boldsymbol{a} - k_a G_a \boldsymbol{c} \tag{4.23a}$$

$$\tau_c \dot{\boldsymbol{c}} = -\boldsymbol{c} + k_c \left[G_c^T \boldsymbol{a} - \boldsymbol{i} \right] \tag{4.23b}$$

where $\boldsymbol{c}$ and $\boldsymbol{a}$ are the activations of the input and output layers respectively, τ_a and τ_c are the neural time constants in the corresponding layers, G_a and G_c are approximations to the ideal weight matrix G, and k_a and k_c are positive scalar constants. Setting $\varepsilon = (k_a k_c)^{-1}$, the equilibrium point of the system is given as

$$\boldsymbol{a} = (G_a G_c^T + \varepsilon I_n)^{-1} G_a \boldsymbol{i} \tag{4.24}$$

by Yan (1991b) and as

$$\boldsymbol{a} = G_a (G_c^T G_a + \varepsilon I_m)^{-1} \boldsymbol{i} \tag{4.25}$$

by a simple extrapolation of the result of Culhane et al. (1989, p.700). These two claims can be reconciled by noting as in (4.9) the equality of the two right-hand sides for $\varepsilon > 0$,

provided that the minimum eigenvalue of the matrix $G_a G_c^T$ satisfies

$$\lambda_{min}(G_a G_c^T) > -\varepsilon$$

This latter requirement ensures that the inverses exist, and is precisely the network stability criterion (Yan, 1991b). Thus like the "ideal case" network of Culhane et al. (1989), the network proposed by Yan gives an approximation to the pseudoinverse solution of (4.2); however this point and the consequent applicability of the network to the rank-deficient and under-determined cases were not noted by Yan. The accuracy of the approximation is dependent on the choice of the gain parameters k_a, k_c and the accuracy of the two approximations G_a, G_c to G. For good accuracy the product $k_a k_c$ should ideally be large; however, in the presence of errors in G_a, G_c or potential rank-deficiency of G, this product should not be so large as to cause instability of the network.

The application of G and its transpose once each in (4.23a) and (4.23b) requires a total of $2nm$ connections. The replacement of the m summers in Figure 4.6 with leaky summing integrators and of the n ideal summing integrators with their leaky counterparts entails the use of $n + m$ leaky summing integrators.

4.4.3 Pece (1992)

Pece (1992) independently proposed a relaxation model which is functionally identical to that of Yan (1991a), except for a change of sign in the cross-terms and input term in (4.23) and the choice of $k_c = 1$ to yield

$$\tau_a \dot{\boldsymbol{a}} = -\boldsymbol{a} + k_a G \boldsymbol{c} \tag{4.26a}$$

$$\tau_c \dot{\boldsymbol{c}} = -\boldsymbol{c} + [\boldsymbol{i} - G^T \boldsymbol{a}] \tag{4.26b}$$

as depicted in Figure 4.9. The equilibrium point and stability analysis presented by Culhane et al. (1989) can be easily extended to demonstrate the exponential stability[6] of the network, and the equivalence of its fixed point with that presented in (4.24) and (4.25).

4.5 Conclusion

In this chapter the decomposition of an image using a set of non-orthogonal expansion functions was formulated as a *least squared error* (LSE) quadratic optimisation problem, the solution of which corresponds to the *best linear unbiassed estimate* (BLUE) of the original image when the image is subject to zero-mean spherical Gaussian noise. The LSE problem was then regularised to ensure uniqueness of the coefficients in cases where the expansion functions are linearly dependant; the regularised solution tends to the

[6] Exponential stability is defined later in Definition 6.1.

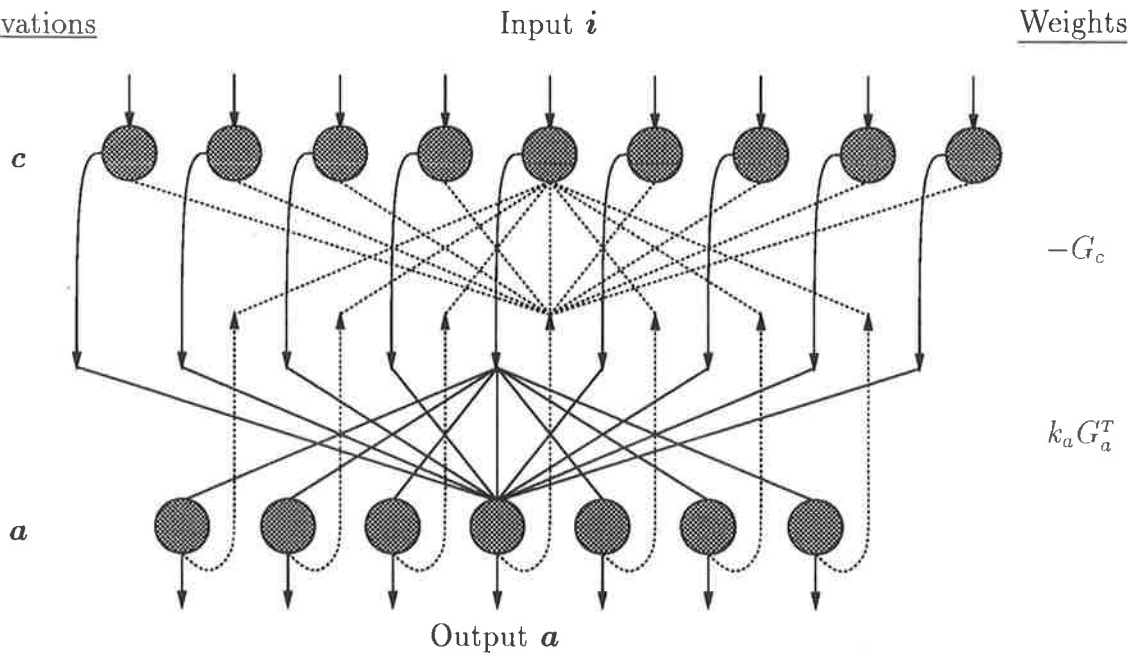


Figure 4.9: Network proposed by Pece (1992) for Gabor image decomposition.

Moore-Penrose pseudoinverse solution of the corresponding set of linear equations in the limit as the regularisation parameter tends to zero. Conventional methods in linear algebra and optimisation theory for the solution of the LSE and regularised LSE problems were briefly reviewed, and it was argued that the underlying algorithms — with the exception of relaxation algorithms — are incompatible with the autonomous fine-grained parallelism offered by analog recurrent artificial neural networks (RANNs). Discrete- and continuous-time RANNs which solve the LSE and regularised LSE problems were then reviewed, and it was shown that several networks proposed as approximate solutions to the LSE problem in fact solve the regularised LSE problem.

Chapter V

SINGLE-LAYERED NEURAL NETWORKS FOR DECOMPOSITION

5.1 Introduction

In the latter part of Chapter 4, linear RANNs were reviewed which solve the LSE or regularised LSE problems. In this chapter, single layered RANNs are proposed to solve these two problems, and are shown in the general case to require less connections or neurons than comparable multi-layered models. These models are shown to bear considerable resemblance to existing resistive grid architectures for solving sets of linear equations arising from the discrete formulation of partial differential equations (PDEs) governing classical problems in machine vision.

5.2 Pattison (1992)

Upon abandoning the discrete-time (SOR) formulation of the SRE optimisation problem, the somewhat problematic architecture of Wang & Yan (1992) can be condensed into a single neural layer to produce a network which had already been proposed independently by Pattison (1992). Depicted in Figure 5.1, the resultant network implements

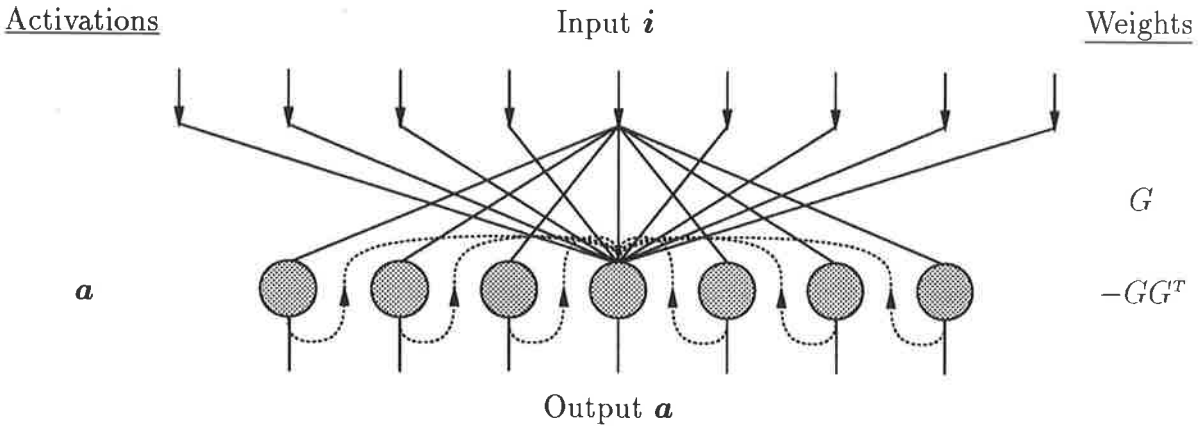


Figure 5.1: Relaxation network proposed by Pattison (1992). Complete input connections shown for centre node only. Grey nodes represent leaky summing integrators.

the following ordinary differential equation (ODE)

$$\dot{\boldsymbol{a}} = \alpha \left[G\boldsymbol{i} - (GG^T)\,\boldsymbol{a} \right] \tag{5.1}$$

where $\alpha \in \mathbb{R}_+$, which performs steepest descent on the SRE (Pattison, 1992), and is therefore exponentially stable provided G has rank $n \leq m$. The matrix $L \triangleq -GG^T$ can be precalculated, and each of its off-diagonal elements represented by a weighted lateral connection between corresponding neurons in a single layered linear recurrent neural network. The value of the weight l_{ik} on the lateral connection from cell k to cell i is given by

$$l_{ik} = -\sum_{\boldsymbol{x}} g_i(\boldsymbol{x}) g_k(\boldsymbol{x}) = l_{ki} \tag{5.2}$$

where $g_i(\boldsymbol{x})$ and $g_k(\boldsymbol{x})$ are the expansion functions on the feedforward weights to the ith and kth neurons respectively. This quantity can be seen to be a discrete approximation to the overlap integral between these two expansion functions, and may be either positive (excitatory) or negative (inhibitory). The diagonal elements of L can be implemented either as external self-inhibitory connections on ideal summing integrator nodes or as internal leakage terms in leaky summing integrator nodes. The latter scheme is preferable, since it obviates the need for the active analog components required to approximate an ideal integrator. The feedforward weight vector of the jth neuron is given by the jth row of G.

This network is substantially similar to resistive networks proposed for the solution of partial differential equations (PDEs) associated with problems in vision, including depth-from-stereo (Chhabra & Grogan, 1989), shape-from-depth (Grimson, 1981) and contour-based optical flow (Poggio & Koch, 1985; Chhabra & Grogan, 1990). The principal difference is that whereas in vision applications the connections are local — reflecting the spatial localization of both discrete approximations to spatial derivatives and penalty terms which enforce constraints such as spatial smoothness — the network depicted in Figure 5.1 may have or approach complete lateral connectivity depending on the chosen expansion functions. Furthermore, unlike these resistive networks, the network of Pattison (1992) may require both positive and negative weights.

The network requires n leaky summing integrators, while the number of weights (connections) required by this implementation — including the self-connection (leakage) term — is $n(n+m) \leq 2nm$ for $n \leq m$, with equality if and only if $n = m$. Thus for general (full-rank) overdetermined sets of linear equations this network improves on the connection economy of the condensed network depicted in Figure 4.6. Furthermore, for expansion functions sets in which each function has non-zero overlap integral with relatively few others, the feedback matrix L is relatively sparse, and can be implemented with considerably less connections than would in general be required. If in addition each of the expansion functions overlaps only with others whose centres are located near its own, then the required spatial extent of the lateral connections can be restricted

by implementing the neural layer as a two-dimensional array of neurons, and assigning the expansion functions to the neurons topologically, according to the locations of their receptive field centres.

For orthogonal expansion functions — in which case the SRE optimisation scheme is of course redundant — the single-layered network requires up to $n(m-1)$ fewer (non-zero) connections than the network shown in Figure 4.6, which in the worst case would still require $2nm$ connections. In general however, even small perturbations of the expansion functions from an orthogonal set will cause the overlap (lateral weight) matrix to become densely populated. An exception is if the expansion functions are localised in image space (of strictly compact support and small compared with the size of the image), in a manner similar to the following simplistic example.

Let each expansion function be circularly symmetric, positive over the range $r \in [0, r_0]$ where r is the radial distance from its centre, and zero for $r > r_0$. Then each expansion function has non-zero overlap integral with all others whose centres lie within $2r_0$ of its centre, and hence the neuron whose feedforward weights represent that expansion function requires a lateral connection to each of the corresponding nearby neurons. A necessary condition for the complete representation of arbitrary images in this neural layer is $n \geq m$; assuming for example that $n = m$, this connection pattern results in (almost) 4 times as many non-zero lateral connections per neuron as non-zero feedforward connections. Thus since the feedforward weight matrix of this model is four times sparser than the lateral weight matrix and the condensed model in Figure 4.6 effectively implements this feedforward weight matrix twice — once in the feedforward path and once in the feedback path — the latter will require only 0.4 times as many connections.

Furthermore, whilst in cases where the overlap of any two expansion functions decays asymptotically with distance between their centres it might seem tempting to reduce the required number of connections by thresholding the lateral weight matrix, it will be seen in Section 5.4 that the errors involved in this truncation are magnified by the condition number of the lateral weight matrix, and in extreme cases the truncation can cause the system to become unstable. This problem will be addressed further in that section. A more suitable procedure for eliminating many of the smaller entries in the lateral weight matrix when the expansion functions are localised is to truncate the expansion functions themselves at some finite distance — for example after 3 standard deviations if the expansion functions have Gaussian weighting functions — and recalculate the lateral weight matrix.

As will be seen in Chapter 6, the network proposed by Pattison (1992) unlike the other neural networks presented in this section — with the exception of that of Wang & Yan (1992) — is readily adapted to the solution of more general quadratic optimisation problems. Furthermore it will be argued in Chapter 7 that this network admits a more feasible mapping onto the wetware of the early visual system than these other networks.

5.3 Additional Leakage Term

The use of a small additional leakage term $-\varepsilon \boldsymbol{a}$ with $\varepsilon \in \mathbb{R}_+$ in (5.1) leads to the network equation

$$\alpha^{-1}\dot{\boldsymbol{a}} = -\left(GG^T + \varepsilon I_n\right)\boldsymbol{a} + G\boldsymbol{i} \tag{5.3}$$

revealing that like the network of Culhane et al. (1989) the modified system performs steepest descent on E_ε and hence is exponentially stable. This modification therefore allows the network to approximate the pseudoinverse solution of the set of equations in (4.2). With the exceptions of non-local lateral connectivity and the need for both positive and negative weights, the resultant network once again bears strong resemblance to resistive grid networks used in machine vision, where the additional leakage has been used to impose a minimum norm penalty term in applications such as area-based optical flow (Lee et al., 1988), and contour-based optical flow, shape-from-depth and depth-from-stereo (Chhabra & Grogan, 1990).

The linear dynamical system

$$\dot{\boldsymbol{a}} = -Q\boldsymbol{a} + G\boldsymbol{i} \tag{5.4}$$

with

$$Q = GG^T \tag{5.5a}$$

$$Q = GG^T + \varepsilon I_n \tag{5.5b}$$

for the neural networks of Section 5.2 and equation (5.3) respectively, is exponentially stable — and hence exponentially convergent to the equilibrium point $\boldsymbol{a}^* = Q^{-1}G\boldsymbol{i}$ for constant $\boldsymbol{i}$ — if and only if the state-feedback matrix Q is positive definite (Lancaster & Tismenetsky, 1985). The use of (5.5b) rather than (5.5a) whenever G is known or suspected to be rank deficient is sufficient to ensure that Q is always positive definite, and hence that the system is stable. For notational convenience in the remainder of this chapter, the matrix Q is therefore assumed to be positive definite and to result from an appropriate choice in (5.5).

This single-layered architecture is readily extended to more general regularised LSE problems. The restoration of a blurred image for example can be formulated (Galatsanos & Katsaggelos, 1992) as

$$\min_{\boldsymbol{a}} \left\{ J_\varepsilon(\boldsymbol{a}) \triangleq \|\boldsymbol{i} - G^T\boldsymbol{a}\|_2 + \varepsilon \|P^T\boldsymbol{a}\|_2 \right\}$$

where $\boldsymbol{i} \in \mathbb{R}^m$ and $\boldsymbol{a} \in \mathbb{R}^m$ are the noisy blurred image and the restored image respectively, and $G^T \in \mathbb{R}^{m\times m}$ and $P^T \in \mathbb{R}^{m\times m}$ are matrices representing the image blurring function and *regularisation operator* respectively. To solve this problem, the network requires G as the feedforward weight matrix and $Q = GG^T + \varepsilon PP^T$ as the lateral weight

matrix. The wide range of LSE and regularised LSE applications to which the single-layered neural network can be readily adapted — as well as the range of problems in machine vision described above — provides strong motivation for addressing the challenging issues in VLSI implementation of non-local lateral connectivity and the need for both positive and negative weights.

5.4 Preconditioning

In Sections 4.3 and 4.4, recurrent neural networks were presented — along with the linear first-order differential or difference equations which describe their dynamics — whose common unique equilibrium point is by design the solution to the normal or augmented normal equations. Each of these neural networks makes use of one or more of the matrices G, GG^T, and $GG^T + \varepsilon I_n$, which it has been tacitly assumed can be implemented without error. However whilst G may be specified exactly through the parametric specification of the expansion functions which when sampled constitute its rows, the hardware implementation of these matrices as weights in a neural network may incur considerable error. In this section, the implications of such implementation errors for the neural networks of Sections 5.2 and 5.3 are examined, and the technique of *preconditioning* is used to mitigate the consequent displacement of the equilibrium coefficient vector $\boldsymbol{a}^*$ and the potential for network instability. The same preconditioning strategy used to reduce sensitivity to weight errors is also shown to accelerate convergence in both continuous- and discrete-time implementations.

In the following analysis, weight implementation rather than computation is assumed to constitute the predominant source of error; thus for example it is assumed that the errors involved in the computation of the matrix GG^T required by the network in Figure 5.1 are negligible in comparison with those incurred by its subsequent implementation. This assumption is clearly reasonable in the case of analog hardware, where the accuracy of components such as conductances and analog multipliers is limited (see e.g. (Schach, 1992)). It may also be significant in floating point implementations where, for reasons of economy, weight precision is low in comparison with that available during the external evaluation of the matrix Q. For the case where errors in the original determination of G constitute the sole or principal source of inaccuracy, a detailed error analysis of the least-squares problem assuming ideal arithmetic in all subsequent computations has been presented by Golub & Van Loan (1989).

5.4.1 Sensitivity to Weight and Derivative Round-Off Errors

The *spectral condition number* $\kappa_2(M) \geq 1$ of a matrix $M \in \mathbb{R}^{m \times n}$ is given by the ratio of its maximum to minimum singular values. Unlike condition numbers based on other matrix norms, the spectral condition number is convenient in the analysis of least-squares

problems since it is derived from the spectral matrix norm, which is induced by the Euclidean vector norm (Horn & Johnson, 1988) with the aid of which such problems are formulated. The matrix Q, and by extension the system of linear equations

$$Q\boldsymbol{a} = G\boldsymbol{i} \tag{5.6}$$

the solution of which is the equilibrium point of (5.4) — is said to be *well-conditioned* if $\kappa_2(Q) \approx 1$ and *ill-conditioned* when $\kappa_2(Q) \gg 1$.

The condition numbers $\kappa_2(Q)$ and $\kappa_2(G)$ serve as a measure of the sensitivity of the equilibrium point of (5.4) to weight implementation errors. If Q is invertible and G is full rank, the relative error

$$e_{\boldsymbol{a}} \triangleq \|\Delta\boldsymbol{a}^*\|_2 / \|\boldsymbol{a}^*\|_2$$

in the equilibrium point $\boldsymbol{a}^* \triangleq Q^{-1}G\boldsymbol{i}$ of the dynamical system in (5.4) caused by the relative errors

$$\begin{aligned} e_Q &\triangleq \|\Delta Q\|_2 / \|Q\|_2 < \kappa_2^{-1}(Q) \\ e_G &\triangleq \|\Delta G\|_2 / \|G\|_2 \\ e_{\boldsymbol{i}} &\triangleq \|\Delta\boldsymbol{i}\|_2 / \|\boldsymbol{i}\|_2 \end{aligned}$$

in Q, G and $\boldsymbol{i}$ respectively, can be shown to satisfy

$$e_{\boldsymbol{a}} \leq \kappa_2(Q)\left\{e_Q + \kappa_2(G)\left[e_G + e_{\boldsymbol{i}}\right]\right\} + O^2(\|\Delta Q\|_2, \|\Delta G\|_2, \|\Delta\boldsymbol{i}\|_2) \tag{5.7}$$

where $O^2(\cdot,\cdot,\cdot)$ denotes second and higher order terms and cross-terms of its arguments. Thus for small perturbations, the condition number gives an indication of the sensitivity of the solution to errors in Q, G and $\boldsymbol{i}$. Thus in the case of the full rank normal equations for example, substituting (5.5a) into (5.7) and noting that $\kappa_2(GG^T) = \kappa^2(G)$ reveals that if G is even mildly ill-conditioned, the solution of the set of equations in (5.6) obtained at the equilibrium point of the system can be highly sensitive to implementation errors in the matrices GG^T and G. If G is rank deficient and $\boldsymbol{i}$ lies in the row-space of G, (5.7) is valid provided $\kappa_2(G)$ is replaced by $\overline{\kappa}_2(G)$ as defined in (4.8); extension to the more general case is beyond the scope of the present discussion.

The condition number $\kappa_2(Q)$ also provides a measure of the sensitivity of the equilibrium point to floating point round-off errors in the evaluation of the temporal derivative in (5.4). This sensitivity arises from the fact that an absolute residual error $\|G\boldsymbol{i} - Q\boldsymbol{a}\|_2$ in the normal or augmented normal equations which is less than the unit roundoff δ results from any approximation $\hat{\boldsymbol{a}}$ to the true coefficient vector $\boldsymbol{a}^*$ satisfying

$$\|\hat{\boldsymbol{a}} - \boldsymbol{a}^*\|_2 \leq \sqrt{n}\kappa_2(Q)\delta$$

(Golub & Van Loan, 1989). Thus the presence of round-off error results in a region about the true equilibrium point — the largest dimension of which is proportional to

$\kappa_2(Q)$ — in which all points yield $\dot{\boldsymbol{a}} = 0$ to floating point precision, which will result in inaccurate location of the equilibrium point for most discrete-time simulation algorithms. Whilst the resultant sensitivity to round-off may not represent a significant problem under computer-based double-precision arithmetic, in floating point hardware implementations it may necessitate the use of excessive numerical precision, thereby increasing the cost of implementation. This particular argument is of course predicated on the assumption that it is not sufficient that the coefficients obtained by the network correspond to a small residual in the normal equations — and hence produce a good approximation of the true image — but that it is also necessary that the coefficients closely approximate the optimal coefficients. The validity of this assumption will usually depend on the sensitivity of any subsequent processing of the coefficients to errors in those coefficients.

Finally, since the minimum singular value of Q is the spectral-norm distance from the state-feedback matrix to the set of rank-deficient matrices (Golub & Van Loan, 1989, Thm 2.5.2), the minimum relative perturbation in Q required to make the state-feedback matrix Q in (5.4) singular is given by the inverse $\kappa^{-1}(Q)$ of the condition number. Thus since any further perturbation could make the state-feedback matrix indefinite — and hence make the linear dynamical system unstable — the larger the value of $\kappa_2(Q)$ the more susceptible the system is to instability resulting from inaccurate implementation of the state-feedback matrix.

In summary, if the matrix G — and hence Q — is ill-conditioned, the equilibrium point of the network can be excessively sensitive to implementation errors in these matrices, and to floating-point round-off errors in the evaluation of the time derivative. Furthermore, ill-conditioning of the state-feedback matrix Q renders the network susceptible to instability resulting from errors in its implementation. Thus for both analog and digital implementations, it is desirable to address potential ill-conditioning of the dynamical system in (5.4) in order to decrease its sensitivity to weight implementation and floating-point round-off errors.

5.4.2 *Diagonal Preconditioning*

Preconditioning the linear dynamical system (5.4) involves choosing non-singular matrices $\Gamma, B \in \mathbb{R}^{n\times n}$ called *preconditioners* such that the state-feedback matrix $S \triangleq \Gamma Q B$ of the preconditioned dynamical system

$$\dot{\boldsymbol{u}} = -\Gamma Q B \boldsymbol{u} + \Gamma G \boldsymbol{i} \tag{5.8a}$$

$$\boldsymbol{a} = B\boldsymbol{u} \tag{5.8b}$$

is better conditioned than Q. Careful selection of the preconditioners can ensure that

$$\frac{\kappa_2(Q)}{\kappa_2(\Gamma)\kappa_2(B)} \leq \kappa_2(S) \leq \kappa_2(Q)$$

In this section, preconditioning is used to address the problems associated with ill-conditioning of the dynamical system in (5.4), and guidance is given on a suitable choice of preconditioners.

The equilibrium point $\boldsymbol{u}^*$ of the preconditioned system satisfies

$$\Gamma Q B \boldsymbol{u} = \Gamma G \boldsymbol{i} \tag{5.9}$$

which upon substitution of (5.8b) and factorisation by Γ yields the same equilibrium point $\boldsymbol{a}^*$ as (5.4). In the special case of diagonal preconditioners, preconditioning of the system in (5.4) simply produces *row-* and *column-scaling* (Golub & Van Loan, 1989) of the normal or augmented normal equations which describe the equilibrium point of the system.

A sufficient condition for preserving the positive definiteness of the state-feedback matrix and hence the stability of the system is that the preconditioners Γ, B be symmetric positive definite (Horn & Johnson, 1988, Thm 7.6.3). This condition also ensures that the matrix $B\Gamma$ is positive definite, so that under the dynamics of the preconditioned system, for a given input image the error function V which denotes E under the choice of unpreconditioned state-feedback matrix in (5.5a) and E_ε under (5.5b) satisfies

$$\dot{V} = (\nabla V)^T \dot{\boldsymbol{a}} = \frac{1}{2} \nabla^T V B \Gamma \nabla V \leq 0 \tag{5.10}$$

with equality only at the unique stationary point of V. Thus for symmetric positive definite preconditioners, V is a global Liapunov function for the preconditioned dynamical system in (5.8) and decreases monotonically with time.

However except for some special classes of matrices (see e.g. Ku & Kuo (1992)), strategies for choosing non-diagonal preconditioners are not well developed. Furthermore the generalisation from diagonal to non-diagonal preconditioners considerably increases the amount of potentially error-prone computation required during preconditioning, in violation of the assumption usually made in perturbation analyses that the preconditioning step can be performed without error (see e.g. Golub & Van Loan (1989)). Consideration is therefore restricted to diagonal preconditioners $\Gamma, B \in \mathbb{R}_+^{n \times n}$, a restriction which in Chapter 6 — where bound constraints are imposed on the quadratic optimisation problem — is not only convenient but necessary.

For ideal preconditioning, Γ and B should be proportional to the inverse of corresponding factors of GG^T, giving $\kappa_2(S) = \kappa_2(I) = 1$. However, since they are constrained to be diagonal, Γ and B are at best a reasonable approximation to scalar multiples of such inverses. Now since Q is Hermitian, Schur's theorem on the strong majorisation of eigenvalues by the diagonal entries (see e.g. Horn & Johnson (1988, p.193)) can be used to show that

$$\kappa_2(Q) \geq \frac{\max_i \{q_{ii}\}}{\min_i \{q_{ii}\}} \tag{5.11}$$

where q_{ii} is the ith diagonal entry of Q. Substitution of (5.11) into a standard condition number inequality yields

$$\kappa_2(S) \geq \frac{\kappa_2(Q)}{\kappa_2(\Gamma)\kappa_2(B)} \geq \frac{\min_i\{\gamma_i\}}{\max_i\{\gamma_i\}} \frac{\max_i\{q_{ii}\}}{\min_i\{q_{ii}\}} \frac{\min_i\{\beta_i\}}{\max_i\{\beta_i\}} \tag{5.12}$$

where γ_i and β_i are the ith diagonal entries of Γ and B respectively. According to (5.12), an obvious strategy is to assign the diagonal entries of Γ and B in the reverse order of corresponding diagonal elements of Q and with a smaller overall spread on a log scale, in order to move the diagonal entries s_{ii} of S closer together than the corresponding elements of Q. If these diagonal entries of S are further required to be in the same order as those of Q, (5.12) becomes simply

$$\kappa_2(S) \geq \frac{\max_i\{s_{ii}\}}{\min_i\{s_{ii}\}}$$

which constitutes a limited extension of the above corollary of Schur's theorem to the case where S is not necessarily Hermitian (i.e. $\Gamma \neq B$). One simple choice satisfying the reverse-ordering condition is

$$\gamma_i = \frac{k}{\sqrt{q_{ii}}} = \beta_i \qquad k > 0 \tag{5.13}$$

which under certain conditions on Q can be shown to yield the optimal or near-optimal diagonal preconditioners (Greenbaum & Rodrigue, 1989). In addition to ensuring that S is by virtue of a lower condition number less susceptible than Q to weight implementation errors, in the case of the unaugmented normal equations this choice also improves the condition of the matrix G with the same benefit, since for $\Gamma = B$

$$\kappa_2(\Gamma G G^T B) = \kappa_2^2(\Gamma G)$$

This result extends approximately to the case of the augmented normal equations for sufficiently small ε. Furthermore, since under this choice of preconditioners the right-hand side in (5.12) becomes unity, no additional restriction is imposed on $\kappa_2(S)$, since by definition $\kappa_2(S) \geq 1$.

The preconditioned dynamical system in (5.8) can be re-written as

$$\dot{\boldsymbol{a}} = -B\Gamma Q\boldsymbol{a} + B\Gamma G\boldsymbol{i} \tag{5.14}$$

revealing that under ideal conditions the same trajectory $\boldsymbol{a}(t; \boldsymbol{a}_0, t_0)$ could have been obtained through the use of a single composite preconditioner $\Gamma' \triangleq B\Gamma$. Under the choice of preconditioners in (5.13), this composite preconditioner is simply

$$\Gamma' = [\mathrm{diag}(Q)]^{-1}$$

as is indeed used in discrete-time relaxation algorithms such as the Jacobi iteration and the SOR method (discussed briefly in Section 4.3.2). However, it has been observed

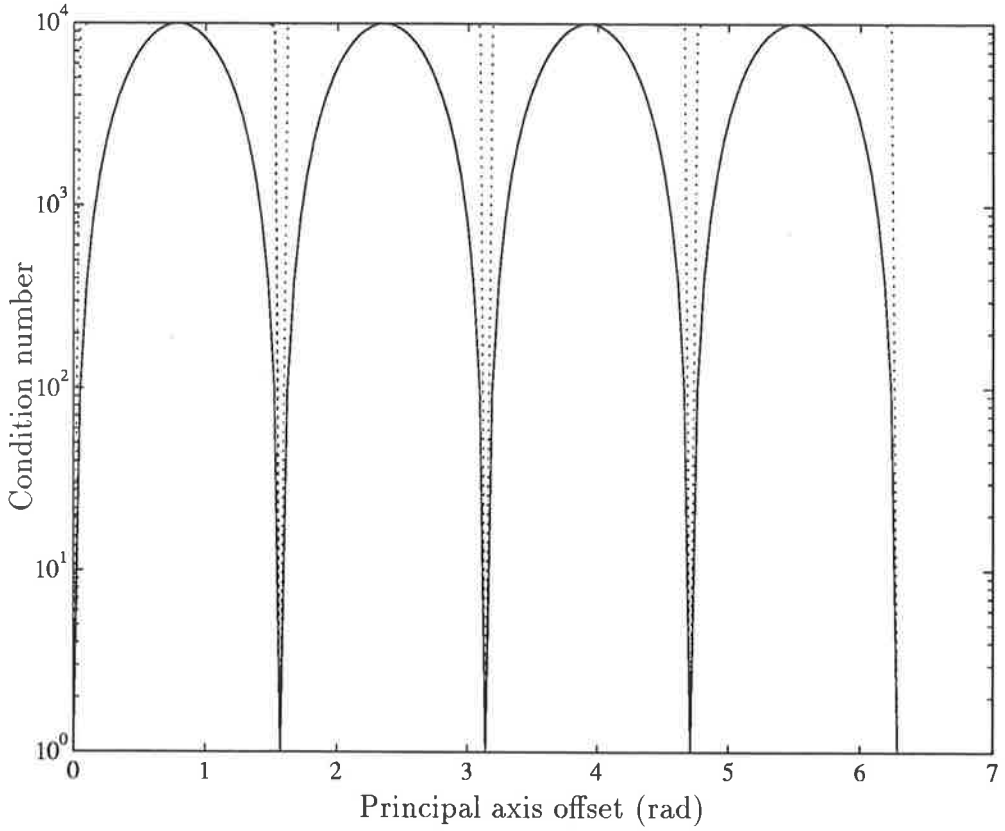


Figure 5.2: Condition numbers of ΓQB (solid) and $B\Gamma Q$ (dotted) for the choice of preconditioners in (5.13) with $k = 1$ vs the angle between the x-axis and the principal eigenvector of Q.

heuristically that the use of dual preconditioners B, Γ produces in many cases a considerably better conditioned state-feedback matrix, thereby reducing the potential of the system to become unstable as a result of weight implementation errors. This observation is illustrated in Figure 5.2 for the 2×2 case, in which with the exception of near-diagonal Q, the symmetrically preconditioned matrix is considerably better conditioned. For $\kappa_2(Q) = 10^4$, an improvement in the condition number of two orders of magnitude is obtained just outside the near-diagonal range.

On recovery of the desired variable $\boldsymbol{a}$ from $\boldsymbol{u}$ using (5.8b), the reduction in the bound on relative error $e_{\boldsymbol{a}}$ of the equilibrium point of the system obtained through preconditioning is at least partially undone, since

$$e_{\boldsymbol{a}} \leq \kappa_2(B) e_{\boldsymbol{u}}$$

Thus the relative error $e_{\boldsymbol{u}}$ in $\boldsymbol{u}$ is amplified by a factor which by (5.13) is bounded above by $\kappa_2(B) = \sqrt{\kappa_2(\mathrm{diag}(Q))}$. In some cases, this amplification factor may outweigh any reduction in sensitivity to weight errors achieved by preconditioning, and Golub & Van

Loan (1989) advise that the corresponding scaling of the normal equations should be used with caution and on a case-by-case basis.

However, since in the case of the unaugmented normal equations the ith diagonal entry q_{ii} of Q is simply given by the square $\|g_i(\boldsymbol{x})\|_2^2$ of the norm of the ith sampled expansion function, if freedom is allowed in choosing the amplitudes of the expansion functions, the above preconditioning can be achieved implicitly by normalising the rows of G, thereby avoiding the pitfall of error amplification on recovery of the true coefficients. If the normalisation option is available, the new state-feedback matrix becomes

$$S = \Gamma G G^T B \tag{5.15a}$$

for the normal equations and

$$S = \Gamma G G^T B + \varepsilon I_n \tag{5.15b}$$

for the augmented normal equations with Γ, B chosen according to (5.13). In the latter case this strategy is only approximately equivalent to preconditioning the matrix Q in (5.5b); however normalisation is easily shown to improve the conditioning of this choice of Q whenever it improves the conditioning of that in (5.5a).

5.4.3 *Accelerating Convergence*

In this section, it is shown that diagonal preconditioning of the networks in Sections 5.2 and 5.3 can improve their rates of convergence when subject to a limitation on the minimum permissible integrator time-constant in the case of analog implementation and on the maximum eigenvalue of the system feedback matrix in the case of discrete-time simulation or implementation.

The solution $\boldsymbol{u}(t; \boldsymbol{u}_0, t_0)$ to the initial value problem (IVP) associated with the dynamical system in (5.8a) for time-invariant input $\boldsymbol{i}$ presented at time t_0 is bounded, continuous and unique, and is given by

$$\boldsymbol{u}(t; \boldsymbol{u}_0, t_0) = S^{-1}\left[I - e^{-S(t-t_0)}\right]\Gamma G \boldsymbol{i} + e^{-S(t-t_0)}\boldsymbol{u}_0 \tag{5.16}$$

which converges exponentially to the unique equilibrium point $\boldsymbol{u}^* = B^{-1}Q^{-1}G\boldsymbol{i}$. The slowest mode of convergence to the equilibrium point is governed by the eigenvalue of S having the smallest real part. For example if S is normal (e.g. Hermitian) and diagonalisable, it can be shown that Golub & Van Loan (1989)

$$\|\boldsymbol{a} - \boldsymbol{a}^*\|_2 \leq e^{-\min_i\{\Re(\lambda_i(S))\}t}\|\boldsymbol{a} - \boldsymbol{a}_0\|_2$$

where $\Re(\cdot)$ returns the real part of its complex argument. In theory at least, the convergence of the system in (5.4) can be arbitrarily accelerated by multiplying the right hand side by some $1 < \alpha \in \mathbb{R}_+$. However, this strategy is limited in real implementations by an upper limit on the diagonal entries s_{ii} of S imposed by a technology-dependent

lower limit τ_{min} on the integrator time constants. Furthermore in discrete-time simulations, some algorithms such as the explicit Euler method require a trade-off between the maximum eigenvalue and the temporal step size, so that increasing α which increases the convergence rate necessitates a proportionate decrease in the temporal step size, so that the same number of steps and hence computation is required for the network state to approach within a specified distance of the true equilibrium point for a given starting point.

If the state feedback matrix S of the dynamical system in (5.4) is particularly ill-conditioned, convergence of the slowest mode can be orders of magnitude slower than would in general be dictated by the smallest permissible time-constant in analog implementations or by the largest eigenvalue in discrete-time simulations. For example, in the special case where S is symmetric — in which case the eigenvalues are real and since S is also positive definite are equal to the singular values — this situation is guaranteed by widely spread diagonal entries — which by Schur's theorem on the strong majorisation of diagonal entries by the eigenvalues is a sufficient condition for ill-conditioning — and ill-conditioning respectively.

Assuming symmetrical diagonal preconditioning ($\Gamma = B$) to preserve the symmetry of Q in S, this observation appears to suggest the following respective remedies

Heuristic 5.1 *Choose $\alpha = \max_i\{q_{ii}\}\tau_{min}$ to take advantage of the available computational speed, and then precondition αQ in such a way as to increase the minimum diagonal entry whilst holding constant the maximum diagonal entry, thereby compressing the range of the diagonal entries of S compared with that of Q.*

Heuristic 5.2 *Precondition Q whilst holding constant the (real) maximum eigenvalue, in order to increase the minimum singular — and hence eigen- — value and hence accelerate convergence for the same simulation step-size.*

Unfortunately for the former scheme however, a small spread of the diagonal entries does not guarantee that the above problem of sub-optimal convergence rate will not arise. Furthermore since it may not always be convenient to calculate or estimate the largest eigenvalue of Q — using for example the *power method* (see e.g. Golub & Van Loan (1989)) — a more practical albeit heuristic strategy in the latter remedy is to hold constant the largest diagonal entry. Since in extreme cases either strategy has the potential to worsten the convergence rate — although in practice such cases do not appear to be common — the above remedies should only be afforded the status of heuristics.

A comparison of Heuristics 5.1 and 5.2 for the 2×2 case is illustrated in Figure 5.3. For the chosen $\kappa_2(Q) = 10^4$, a speed-up in excess of an order of magnitude is obtained for near-diagonal matrices. Since the modified version of Heuristic 5.2 is equivalent to Heuristic 5.1 with $\alpha = 1$, the difference in convergence rate between the original and

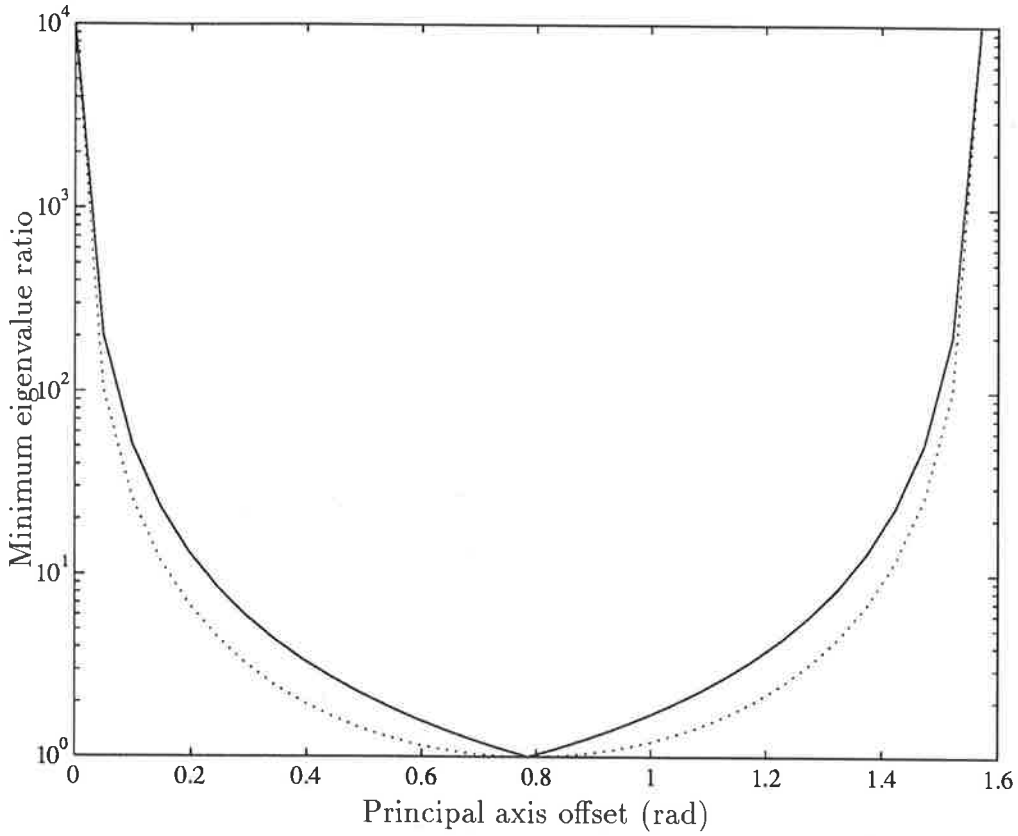


Figure 5.3: Minimum eigenvalues of $S = \Gamma QB$ for the choice of preconditioners in (5.13) vs the angle between the x-axis and the principal eigenvector of Q, with $\kappa_2(Q) = 10^4$. k has been chosen to equalise the maximum diagonal entry (solid) and eigenvalue (dotted) of S and Q, in accordance with heuristics 5.1 (with $\alpha = 1$) and 5.2 respectively. The minimum eigenvalues have been normalised by the minimum eigenvalue of Q. Greatest improvement in convergence rate is observed when Q is near-diagonal.

modified strategies — which is at most a factor of 2 — does not appear to justify the extra computational effort involved in calculating the maximum eigenvalue, especially for higher dimensional problems.

If S is not symmetric — as is indeed the case when the system preconditioned according to (5.13) is mapped back into the original state space as in (5.14) — the relationship between the diagonal entries and the eigenvalues provided by Schur's theorem no longer holds in general. However since

$$\Gamma QB\boldsymbol{x} = \lambda\boldsymbol{x} \quad \Longrightarrow \quad B\Gamma Q\,(B\boldsymbol{x}) = \lambda\,(B\boldsymbol{x})$$

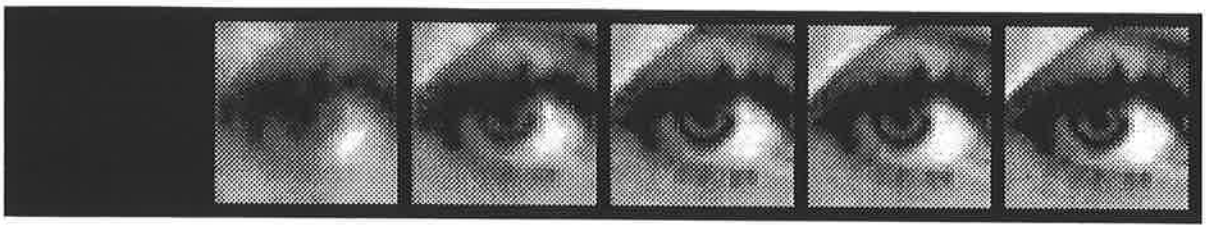
the eigenvalues of the matrices ΓQB and $B\Gamma Q$ are identical. In addition, since the preconditioners are chosen to be diagonal, these two matrices have the same diagonal entries. Thus a given choice of the preconditioners has the same effect on the diagonal entries and eigenvalues of the dynamical system expressed in either state space.

5.5 Examples

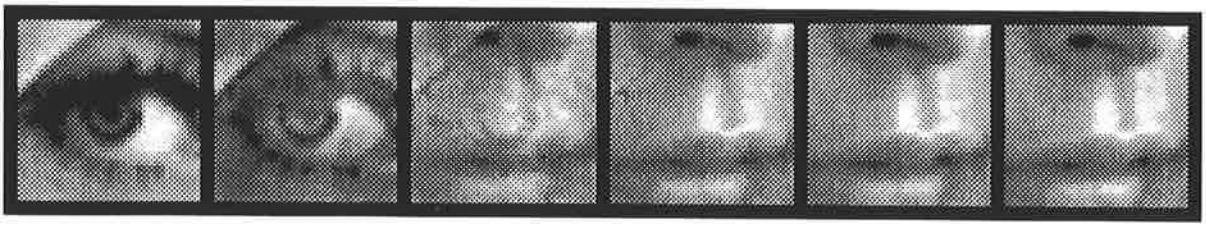
Example 5.1 The following example illustrates the application of the neural network in (5.3) to the minimum SRE decomposition of an image using non-orthogonal expansion functions. The expansion functions used in this example are two-dimensional Gabor functions as described in Chapter 2, with the parameters distributed uniformly over ranges which approximate those typically observed in cats (Jones & Palmer, 1987b; Glezer et al., 1989). The centre coordinates (x_i, y_i) were chosen from a uniform 2D distribution across the $m = 40 \times 40 = 1600$ pixel input image, and the phases ϕ and orientations $\arg(\boldsymbol{\omega})$ distributed uniformly over the range $0-360°$. The magnitude $|\boldsymbol{\omega}|$ of the spatial frequency vector for each Gabor function was chosen from a uniform distribution over the range 0.5 cycles/pixel – 1.0 cycles/picture covering the full range of available frequencies up to the Nyquist frequency for the rectangular image sampling grid. The major axis of each 2D Gaussian was constrained to be parallel to the spatial frequency vector $\boldsymbol{\omega}$, and the variance σ_a on this axis chosen so that the number of cycles of the sinusoid within $\pm 3\sigma_a$ of the centre of the Gaussian was distributed uniformly over the range 1.0–4.0 cycles. The variance σ_b on the minor axis of each 2D Gaussian was then chosen to give an aspect ratio σ_a/σ_b uniformly distributed over the range 1.0–2.0.

A better fit to the non-uniform distributions of some of these parameters found in nature was not attempted since the object of the exercise was not to investigate suitable sets of Gabor functions but to demonstrate the effectivenes of the proposed neural network for any given set of expansion functions. Furthermore, any more systematic scheme for choosing the set of expansion functions which relied on a precise relationship between the parameters of the different expansion functions would seem to be dubious for error-prone implementations, and unrealistic in the biological context. Since the resultant set of $n = 1600$ expansion functions was therefore not guaranteed to be linearly independent, and the determination of the rank of the matrix $G \in \mathbb{R}^{1600\times1600}$ was potentially ill-conditioned, the decomposition was formulated as the minimisation of the regularised SRE in (4.11) with $\varepsilon = 0.001$. The neural network in (5.3) was chosen to perform this minimisation, and since the freedom was available to choose the expansion functions to be (Euclidean) normalised, no explicit preconditioning was required. The effects of preconditioning on a related neural network are illustrated in Chapter 6.

The neural network with $\boldsymbol{a} = \mathbf{0}$ was first presented with the image of an eye, and the neural activations allowed to attain their equilibrium values. The output coefficients were recorded at several stages during the transition from $\boldsymbol{a} = \mathbf{0}$ to the equilibrium point, and were later used to produce the corresponding sequence of reconstructed images depicted in Figure 5.4(a). The network was then presented with the image of a mouth and nose, and the neural activations once again allowed to attain their equilibrium values. The output coefficients were again recorded at several stages during the transition from the



(a)



(b)

Figure 5.4: (a) Sequence of images reconstructed from the outputs of the network in (5.3) at times $t = 0, 1, 10, 100$ and 1000ms after step presentation of the image of an eye (shown at right) to the network input. The network activation vector at $t = 0$ was $\boldsymbol{a}_0 = \mathbf{0}$. For ease of comparison, each reconstructed image was scaled (i.e. its intensity linearly transformed) to utilise the full available grey-scale range. (b) Sequence of reconstructed images at times $t = 0, 1, 10, 100$ and 1000ms after step presentation of the image of a mouth (shown at right) to the same network following equilibration on the eye in (a). Simulations were performed in double-precision arithmetic using the Runge-Kutta method of order 4 with adaptive step-size selection.

first to the second equilibrium point, and were later used to produce the corresponding sequence of reconstructed images depicted in Figure 5.4(b). The real time scale shown in the caption was fixed by assuming a time constant for real cortical pyramidal cells of 20ms (Stratford et al., 1989). The network is known to be exponentially stable, and the expected near-complete convergence of the coefficients within 3 time constants (60ms) was indeed observed.

5.6 Conclusion

In this chapter, a single layered RANN has been presented which solves the LSE problem and — through the simple addition of an extra nodal leakage term — the regularised LSE problem associated with non-orthogonal image decomposition. The network has been shown in the general case to require less connections or neurons than comparable multi-layered networks. However for expansion functions of local and strictly compact support — which in biological visual systems are most likely to be of interest — the number of non-zero weighted connections may in fact exceed that for these other networks unless

a pixel to coefficient ratio (compression factor) of at least 2.5 is applied. The network was found to bear a resemblance to existing resistive grid architectures for solving sets of linear equations arising from the discrete formulation of partial differential equations (PDEs) governing classical problems in machine vision. The two major distinctions — (potentially) complete lateral connectivity and both positive and negative weights — pose a considerable challenge for analog VLSI implementation.

Symmetric diagonal preconditioning of the network dynamics has been shown to result in a potential reduction of the condition number of the state feedback matrix, thereby reducing the susceptibility of the equilibrium point and network stability to weight implementation errors, and in the case of floating point implementations, of the equilibrium point to derivative evaluation errors. Although it has not been theoretically guaranteed that the amplification of the perturbation of the equilibrium point on conversion of the neural activations back to the required coefficient vector will not outweigh the original benefits of preconditioning, it has been observed empirically that this is usually the case. It has also been argued that diagonal preconditioning can be used to accelerate convergence of the network for a given maximum eigenvalue or minimum neural time constant, although once again cases exist where the opposite effect may result. However, in cases where the freedom exists to choose the amplitude of (i.e. scale) each expansion function, Euclidean normalisation of the expansion functions has the same effect as the proposed diagonal preconditioning scheme, whilst avoiding these potential though apparently uncommon pitfalls.

Chapter VI

NEURAL NETWORK FOR BOUND-CONSTRAINED QUADRATIC OPTIMISATION

6.1 Introduction

In Chapter 4, the decomposition using the basis function matrix $G \in \mathbb{R}^{n \times m}$ of a signal or image vector $\boldsymbol{i} \in \mathbb{R}^m$ into a coefficient vector $\boldsymbol{a} \in \mathbb{R}^n$ was formulated as the regularised but unconstrained SRE optimisation problem

$$\min_{\boldsymbol{a}} \left\{ E_\varepsilon(\boldsymbol{a}; \varepsilon) \triangleq \|\boldsymbol{i} - G^T \boldsymbol{a}\|_2^2 + \varepsilon \|\boldsymbol{a}\|_2^2 \right\}$$

— with $\varepsilon = 0$ for $rank(G) = n \leq m$ — and various linear recurrent neural networks which perform the required quadratic optimisation for $\varepsilon = 0$ and $\varepsilon > 0$ were reviewed in Sections 4.3 and 4.4 respectively, and in Chapter 5. In Chapter 7, the potential implementation of such networks in the feline early visual system for the purposes of nonorthogonal decomposition of the retinal image is to be investigated. However, in order to facilitate a comparison between the various networks on the basis of neurological plausibility, it is first necessary to address the following problem.

If the coefficients of such a decomposition are assumed to be signalled by the mean or instantaneous firing frequencies of spiking neurons in the feline visual cortex, then these coefficients are necessarily constrained to be non-negative and have a finite upper limit. In the primary visual cortex, such neurons — and the simple cells in particular — are furthermore known to exhibit remarkably low spontaneous firing rates — see e.g. (Ferster, 1988) — the suppression of which might otherwise be used to signal negative values to neurons capable of measuring such departures from the spontaneous rate. An engineering solution to the analog electronic implementation of a neural network whose task is to minimise the (regularised) SRE using integrators exhibiting analogous output range limitations would be to simply scale and shift the original problem to accommodate these limitations. Since the appropriate coordinate transformation would have to be performed for each new image and is not readily automated, a more practical approach to this problem is to impose range or *bound* constraints on the (regularised) SRE problem and to reformulate the network to seek the corresponding *constrained* minimum.

In this chapter, it is shown that a linear recurrent neural network can be programmed to optimise a general positive semidefinite quadratic form — of which the SRE and regularised SRE are examples — and that the requisite independent bound constraints can

be imposed on the optimisation variables through the use of a piecewise linear saturating activation function in each of the nodes.

6.2 Unconstrained Quadratic Optimisation

The problem of optimising a quadratic functional arises frequently in engineering and science. The unconstrained semidefinite quadratic optimisation problem

$$\min_{\boldsymbol{x}} \left\{ J(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^T W \boldsymbol{x} - \boldsymbol{x}^T \boldsymbol{r} \right\} \tag{6.1}$$

involves the minimisation over the vector $\boldsymbol{x} \in \mathbb{R}^n$ of the quadratic cost function $J(\boldsymbol{x})$ where $W \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix and $\boldsymbol{r} \in \mathbb{R}^n$ is a vector constant.

The quadratic form $J(\boldsymbol{x})$ is quite general. For example, the maximisation over $\boldsymbol{x}$ of the quadratic form

$$G(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T M \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{z}$$

where $M \in \mathbb{R}^{n \times n}$ is negative semidefinite and $\boldsymbol{z} \in \mathbb{R}^n$ is constant can be converted into the above minimisation problem by substituting $W = -M$ and $\boldsymbol{r} = -\boldsymbol{z}$ and minimising the resulting expression over $\boldsymbol{x}$. Furthermore, adding any constant to $J(\boldsymbol{x})$ affects only the value of $J(\boldsymbol{x})$ for any given $\boldsymbol{x}$, and not the location of its global optima. Thus for example the regularised SRE minimisation problem can be reduced to the form of (6.1) by rewriting E_ε as

$$E_\varepsilon(\boldsymbol{a}; \varepsilon) = 2[\frac{1}{2}\boldsymbol{a}^T (GG^T + \varepsilon I_n)\, \boldsymbol{a} - \boldsymbol{a}^T G \boldsymbol{i} + \frac{1}{2}\boldsymbol{i}^T \boldsymbol{i}] \tag{6.2}$$

noting that both the factor 2 and the last term in square brackets are independent of $\boldsymbol{a}$ and can therefore be omitted from the optimisation over $\boldsymbol{a}$, and letting $W = GG^T + \varepsilon I_n$, $\boldsymbol{x} = \boldsymbol{a}$ and $\boldsymbol{r} = G\boldsymbol{i}$.

Although the global unconstrained minimum of $J(\boldsymbol{x})$ in (6.1) clearly occurs where

$$\nabla J = \frac{1}{2}(W + W^T)\boldsymbol{x} - \boldsymbol{r} = \boldsymbol{0}$$

which may be solved by calculating the inverse of

$$Q \triangleq \frac{1}{2}(W + W^T) \tag{6.3}$$

the inversion process becomes computationally intensive for large n, and numerically unstable for ill-posed problems. As has already been demonstrated in Chapter 4, these problems can be in part overcome by harnessing the massive parallelism of recurrent analog neural networks, and through diagonal preconditioning respectively. The single-layer recurrent neural network depicted in Figure 6.1 with lateral weight matrix $-Q$, input vector $\boldsymbol{r}$ and output vector $\boldsymbol{x}$ governed by the following ODE

$$\dot{\boldsymbol{x}} = \boldsymbol{r} - Q\boldsymbol{x}$$

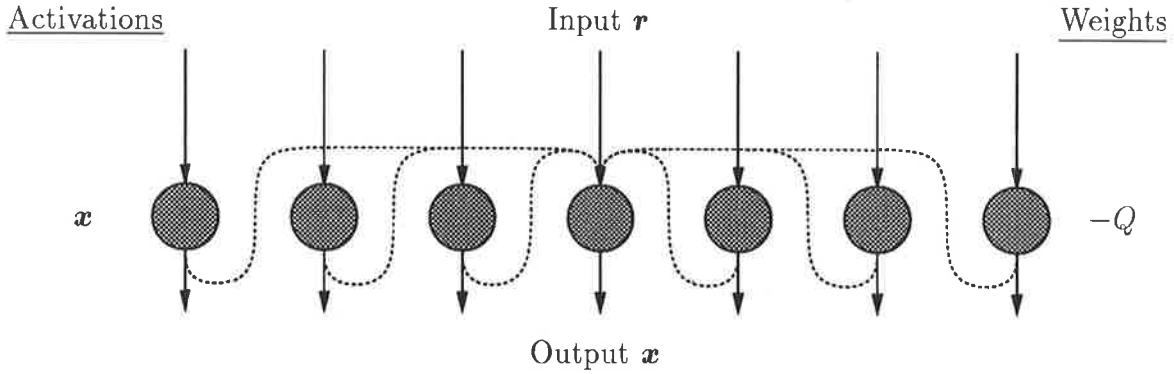


Figure 6.1: Linear recurrent network for the unconstrained minimisation of a positive semidefinite quadratic $J(\boldsymbol{x})$.

performs steepest descent on $J(\boldsymbol{x})$ and thereby implements the required optimisation. The network structure in Figure 6.1 can be modified by the addition of weighted feed-forward connections as used in Chapter 4 to form explicitly terms of the form $\boldsymbol{r} = G\boldsymbol{i}$, such as that required to solve the (regularised) SRE problem; it is henceforth assumed that such connections are invoked wherever appropriate.

6.3 Bound-Constrained Quadratic Optimisation

The optimisation of $J(\boldsymbol{x})$ is complicated by the imposition of constraints on the solution, since a closed-form solution no longer exists. The bound-constrained semidefinite quadratic optimisation (BCSQO) problem

$$\min \{J(\boldsymbol{x}) \,:\, \boldsymbol{\mu} \leq \boldsymbol{x} \leq \boldsymbol{\nu}\} \tag{6.4}$$

with $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^n$ can arise in such diverse topics as rigid body mechanics, fluid dynamics, elastic-plastic torsion (Moré & Toraldo, 1991), and relaxation image labelling. Bound constraints on the optimisation variables may also be imposed by the medium in which a scheme such as the network in Figure 6.1 for the *unconstrained* quadratic optimisation in (6.1) is to be implemented. For example, analog amplifier saturation and the positivity constraint on neural firing rate are two factors which may impose such constraints.

Sudharsanan & Sundareshan (1991) modelled the saturation characteristics of the amplifiers in an analog neural network for unconstrained quadratic optimisation as the piecewise linear neural activation function $g{:}\mathbb{R} \rightarrow \mathbb{R}$ illustrated in Figure 6.2(a), by which the neural outputs $\{x_i\}$ are constrained to lie between upper and lower limits $\mu, \nu \in \mathbb{R}$ placed symmetrically about the origin ($\mu = -\nu$). In order to simplify the definition of the activation function, the assumed output limits μ and ν are mapped through the inverse of the activation function gain (slope) $\beta \in \mathbb{R}$ to produce the corresponding activation

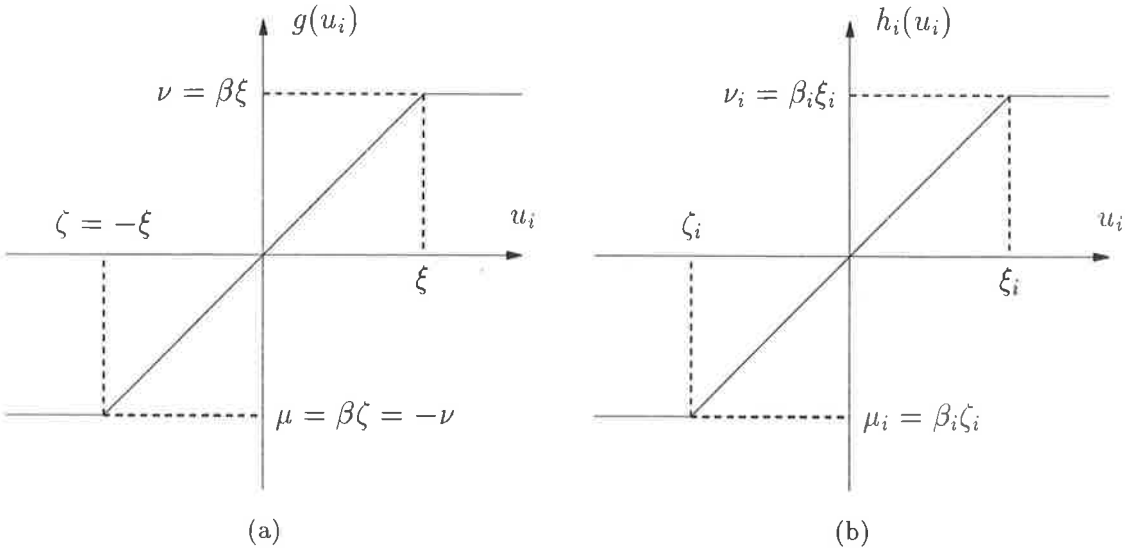


Figure 6.2: (a) One element $g(u_i)$ of the activation function $\boldsymbol{g}(\boldsymbol{u})$ used by Sudharsanan & Sundareshan (1991). (b) One element $h_i(u_i)$ of the generalised constraint enforcement function $\boldsymbol{h}(\boldsymbol{u})$ used by Bouzerdoum & Pattison (1993b).

values ζ and ξ respectively with $\zeta = -\xi$. The vector-valued neural activation function $\boldsymbol{g}:\mathbb{R}^n \rightarrow \mathbb{R}^n$ can then be defined for notational convenience as

$$g_i(\boldsymbol{u}) \equiv g(u_i) = \begin{cases} \beta\zeta & u_i < \zeta \\ \beta u_i & u_i \in [\zeta, \xi] \\ \beta\xi & u_i > \xi \end{cases}$$

such that each of its elements $g(u_i)$ constitutes the activation function of a node in the network.

The neural activation (state) vector $\boldsymbol{u}$ of their network is governed by

$$\dot{\boldsymbol{u}} = \boldsymbol{y} - C\boldsymbol{g}(\boldsymbol{u}) - A\boldsymbol{u} \tag{6.5a}$$

$$\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{u}) \tag{6.5b}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is the external input, $\boldsymbol{x} \in \mathbb{R}^n$ is the network output, $C \in \mathbb{R}^{n\times n}$ is the symmetric lateral feedback matrix with zero diagonal entries, and $A \in \mathbb{R}^{n\times n}$ is a positive diagonal matrix representing the passive decay rate of the activation vector. To map onto their neural network the unconstrained quadratic optimisation problem (6.1) with W symmetric positive definite, Sudharsanan & Sundareshan (1991) set

$$\boldsymbol{y} = \boldsymbol{r} \tag{6.6a}$$

$$A = \beta\text{diag}(Q) \tag{6.6b}$$

$$C = \text{offdiag}(Q) \tag{6.6c}$$

where $diag{:}\mathbb{R}^{n\times n} \rightarrow \mathbb{R}^{n\times n}$ and $offdiag{:}\mathbb{R}^{n\times n} \rightarrow \mathbb{R}^{n\times n}$ select the diagonal and offdiagonal elements respectively of their matrix arguments. One node of the network with these assignments is illustrated in Figure 6.3(a). However, it was left to the user to ensure that

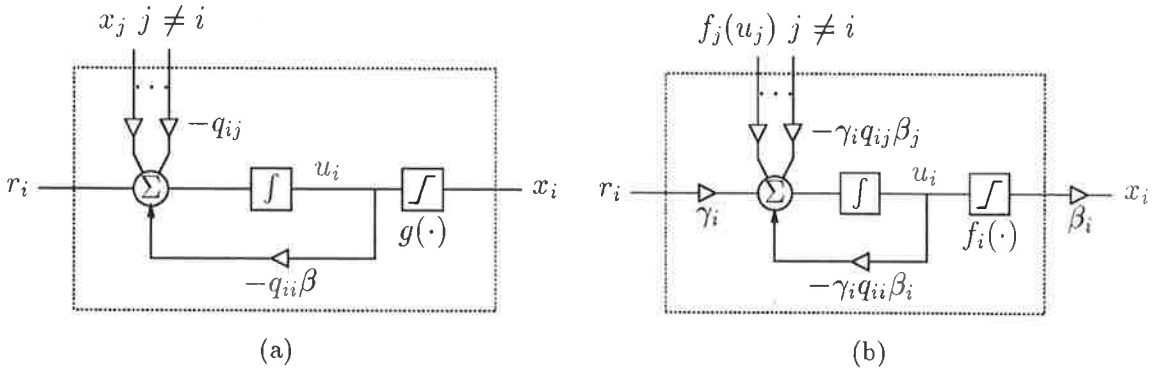


Figure 6.3: Neural models of (a) Sudharsanan & Sundareshan (1991) and (b) Bouzerdoum & Pattison (1993b) as implied by equations (6.5) & (6.6) and equations (6.9) & (6.10) respectively.

the global optimum of the quadratic function lay within the region defined by the output limits; no attempt was made to ensure that if in fact the global optimum lay outside this region, the network would seek the constrained optimum of the quadratic cost function. Furthermore, the proof of exponential convergence offered by Sudharsanan & Sundareshan (1991) for this network, has since been shown to be flawed (Davis & Pattison, 1992). Finally their proof of global convergence of the network relies on the erroneous presumption of the invertibility and (arguably) differentiability of the activation function g.

These problems have been addressed by Bouzerdoum & Pattison (1993b), who generalised the neural output constraints used by Sudharsanan & Sundareshan (1991) by defining the constraint enforcement function $\boldsymbol{h}{:}\mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{u}) \triangleq B\boldsymbol{f}(\boldsymbol{u}) \tag{6.7}$$

where $B \in \mathbb{R}_+^{n\times n}$ is the diagonal matrix of activation function gains $\{\beta_i\}$ and $\boldsymbol{f}{:}\mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined by

$$f_i(\boldsymbol{u}) \equiv f_i(u_i) = \begin{cases} \zeta_i & u_i < \zeta_i \\ u_i & u_i \in [\zeta_i, \xi_i] \\ \xi_i & u_i > \xi_i \end{cases} \tag{6.8}$$

The neural activation vector $\boldsymbol{u} \in \mathbb{R}^n$ is permitted to vary without constraint, and $\boldsymbol{\zeta}, \boldsymbol{\xi} \in \mathbb{R}^n$ are the constraints $\boldsymbol{\mu}, \boldsymbol{\nu}$ on the output $\boldsymbol{x}$ mapped onto corresponding activation values $\boldsymbol{u}$ such that $\boldsymbol{\zeta} = B^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\xi} = B^{-1}\boldsymbol{\nu}$. A typical element $h_i(u_i)$ of the constraint enforcement function $\boldsymbol{h}(\boldsymbol{u})$ is illustrated in Figure 6.2(b), and represents a generalisation of

the neural activation function used by Sudharsanan & Sundareshan (1991) in that each element $h_i(u_i)$ has its own activation limits ζ_i, ξ_i and gain β_i, and that the activation limits need not be symmetric about the origin. The latter generalisation is of interest where, for example, the optimisation variables are constrained to have a specified sign, or prior knowledge places an estimate of the unconstrained optimum at some considerable offset from the origin.

Finally, for the purposes of symmetric preconditioning as discussed in the previous chapter, the right-hand side of (6.5a) can be premultiplied by the positive diagonal preconditioner $\Gamma \in \mathbb{R}^{n \times n}$. An appropriate choice of Γ will be discussed further in Section 6.9. The network resulting from these generalisations will henceforth be referred to as the *bound constraint projection* (BCP) network, in recognition of the fact that the nonlinearity $\boldsymbol{h}$ maps or *projects* an infeasible neural activation vector $\boldsymbol{u}$ (i.e. $B\boldsymbol{u}$ is infeasible) onto a feasible output vector $\boldsymbol{x}$. The activation vector $\boldsymbol{u}$ of the BCP network is governed by

$$\dot{\boldsymbol{u}} = \boldsymbol{y} - C\boldsymbol{f}(\boldsymbol{u}) - A\boldsymbol{u} \tag{6.9a}$$

$$\boldsymbol{x} = B\boldsymbol{f}(\boldsymbol{u}) \tag{6.9b}$$

with

$$\boldsymbol{y} = \Gamma \boldsymbol{r} \tag{6.10a}$$

$$A = \operatorname{diag}(\Gamma Q B) \tag{6.10b}$$

$$C = \operatorname{offdiag}(\Gamma Q B) \tag{6.10c}$$

Since B and Γ are positive definite and Hermitian and Q is positive semidefinite, $\Gamma Q B$ is positive semidefinite (Horn & Johnson, 1988, Thm 7.6.3), and in particular A is non-negative diagonal. By excluding the trivial case where the ith row and column of Q are zero, A is guaranteed to be positive diagonal as required.

The neuron models in Figures 6.3(a) and (b) differ not only in the generalisation of the neural activation function, but also in the explicit factorisation of the activation function into a unity gain saturation function and a gain term which is applied explicitly only to obtain the network output. The lateral connection matrix C is accordingly modified to apply the necessary gain term implicitly during lateral feedback. This factorisation is necessary in order to obtain a diagonally preconditioned lateral weight matrix which as seen in Section 5.4 can be made less susceptible to implementation errors. For the same reason, the gain term γ_i is used outside the inner loop and incorporated into the lateral- and self-feedback matrices, rather than inside this loop which would otherwise be convenient for the local encapsulation of γ_i.

In the following two sections, it is shown that: each equilibrium point of the BCP network corresponds to a solution of the BCSQO problem; for each solution of the BCSQO problem there exists a unique corresponding equilibrium point of the BCP network;

and that under the neural dynamics the network outputs converge to the set of such solutions. The proofs presented here differ considerably in a number of details from those presented by Bouzerdoum & Pattison (1993b); these differences have been in part necessitated by the relaxation of their requirement that W be positive definite to the more general condition that W be positive semidefinite.

6.4 Equilibrium Point Analysis

The first step in the analysis of the proposed network is to establish the location and nature of the equilibrium points. To assist in this endeavour, it will prove convenient to define the new energy function

$$E(\boldsymbol{u}) \triangleq \frac{1}{2}\boldsymbol{h}^T(\boldsymbol{u})Q\boldsymbol{h}(\boldsymbol{u}) - \boldsymbol{h}^T(\boldsymbol{u})\boldsymbol{r} \tag{6.11}$$

which, by identifying $\boldsymbol{h}(\boldsymbol{u})$ with $\boldsymbol{x}$ *confined to the constraint region*, can be seen to be identical to $J(\boldsymbol{x})$ over this region. The BCSQO problem in (6.4) is thereby reduced to the *unconstrained* minimisation of $E(\boldsymbol{u})$ over the network activation vector $\boldsymbol{u}$.

Observation 6.1 *The Kuhn-Tucker optimality conditions (Moré & Toraldo, 1991)*

$$\nabla_i J \quad \begin{cases} \geq 0 & x_i = \mu_i \\ = 0 & x_i \in (\mu_i, \nu_i) \\ \leq 0 & x_i = \nu_i \end{cases} \qquad \forall i \tag{6.12}$$

are necessary and sufficient for a constrained minimum $\boldsymbol{x}^\star$ of J.

Proof. For the optimisation of a convex function over a convex set, the Kuhn-Tucker optimality conditions (6.12) are both necessary and sufficient (Bazaraa & Shetty, 1979, Thm 4.2.11) for the constrained optimum. The required result follows from the observation that J is a convex function — since its Hessian Q is positive semidefinite — and the hyper-rectangular constraint region is a convex set.

This result is now used to verify the following observation, that the output of the network at equilibrium is a constrained minimum of J.

Observation 6.2 *Each equilibrium point $\boldsymbol{u}^*$ of the BCP neural network in (6.9) is mapped by $\boldsymbol{h}$ onto a constrained minimum $\boldsymbol{x}^\star$ of J, and for each constrained minimum there exists a unique corresponding equilibrium point.*

Proof. It will first be shown that any equilibrium point $\boldsymbol{u}^\star$ is mapped by $\boldsymbol{h}$ onto a constrained optimum $\boldsymbol{x}^\star$. Differentiating (6.11) with respect to $\boldsymbol{h}(\boldsymbol{u})$ and substituting equation (6.9) gives

$$\frac{dE}{d\boldsymbol{h}(\boldsymbol{u})} = \Gamma^{-1}\left(-\dot{\boldsymbol{u}} + A[\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{u}]\right) \tag{6.13}$$

Since $\dot{\boldsymbol{u}} = 0$ at all equilibrium points $\boldsymbol{u}^\star$, this reduces to

$$\left.\frac{dE}{d\boldsymbol{h}(\boldsymbol{u})}\right|_{\boldsymbol{u}=\boldsymbol{u}^\star} = \Gamma^{-1}A[\boldsymbol{f}(\boldsymbol{u}^\star) - \boldsymbol{u}^\star]$$

from which it can be deduced that the equilibrium points of the dynamical system presented in (6.9) must satisfy

$$\left.\frac{\partial E}{\partial h_i(u_i)}\right|_{\boldsymbol{u}=\boldsymbol{u}^\star} = \begin{cases} \beta_i q_{ii}(\zeta_i - u_i^\star) > 0 & u_i^\star < \zeta_i \\ 0 & u_i^\star \in [\zeta_i, \xi_i] \\ \beta_i q_{ii}(\xi_i - u_i^\star) < 0 & u_i^\star > \xi_i \end{cases} \quad \forall i \tag{6.14}$$

where β_i and q_{ii} are the ith diagonal elements of B and Q respectively. Rewriting (6.14) in terms of $\boldsymbol{x}^\star = \boldsymbol{h}(\boldsymbol{u}^\star)$ gives

$$\nabla_i J|_{\boldsymbol{x}=\boldsymbol{x}^\star} \quad \begin{cases} \geq 0 & x_i^\star = \mu_i \\ = 0 & x_i^\star \in (\mu_i, \nu_i) \\ \leq 0 & x_i^\star = \nu_i \end{cases} \quad \forall i$$

which by Observation 6.1 implies that $\boldsymbol{x}^\star$ is a constrained optimum of J as required.

The existence and uniqueness of the equilibrium point $\boldsymbol{u}^\star$ corresponding to a given constrained optimum $\boldsymbol{x}^\star$ of J will now be established by showing that under the system dynamics, all trajectories $\boldsymbol{u}(t; \boldsymbol{u}_0, t_0)$ with starting points $\boldsymbol{u}_0$ satisfying $\boldsymbol{h}(\boldsymbol{u}_0) = \boldsymbol{x}^\star$ — including $\boldsymbol{u}_0 = B^{-1}\boldsymbol{x}^\star$ — converge to a unique equilibrium point $\boldsymbol{u}^\star$ satisfying $\boldsymbol{h}(\boldsymbol{u}^\star) = \boldsymbol{x}^\star$.

Since $\boldsymbol{x}^\star$ satisfies the Kuhn-Tucker conditions, any starting point $\boldsymbol{u}_0$ such that $\boldsymbol{h}(\boldsymbol{u}_0) = \boldsymbol{x}^\star$ satisfies

$$\left.\frac{\partial E}{\partial h_i(u_i)}\right|_{\boldsymbol{u}=\boldsymbol{u}_0} \quad \begin{cases} \geq 0 & u_i \leq \zeta_i \\ = 0 & u_i \in [\zeta_i, \xi_i] \\ \leq 0 & u_i \geq \xi_i \end{cases} \tag{6.15}$$

Since the function $h_i : \mathbb{R} \to \mathbb{R}$ is invertible for $u_i \in (\zeta_i, \xi_i)$, all such starting points $\boldsymbol{u}_0$ must have in common those elements u_i which lie in their linear range. These elements are henceforth denoted by $u_i^\star$. Furthermore, substitution of (6.15) into (6.13) reveals that under the system dynamics all such elements must have and continue to have zero time-derivative until such time as $\boldsymbol{h}(\boldsymbol{u})$ changes. Rearrangement of (6.13) followed by substitution of (6.15) also reveals that until this happens, each saturated component $u_i^\dagger \notin (\zeta_i, \xi_i)$ will converge *monotonically* (exponentially) towards the unique value

$$u_i^\star = \begin{cases} \zeta_i - \frac{\gamma_i}{a_{ii}}\nabla_i J \leq \zeta_i & u_i^\dagger \leq \zeta_i \\ \xi_i - \frac{\gamma_i}{a_{ii}}\nabla_i J \geq \xi_i & u_i^\dagger \geq \xi_i \end{cases}$$

which is clearly in the same saturation region, so that at no time will an initially saturated component enter the linear range of its corresponding activation function. Thus

since $\boldsymbol{h}(\boldsymbol{u})$ has remained unchanged, the point $\boldsymbol{u}^*$ is an equilibrium point of the system and satisfies $\boldsymbol{h}(\boldsymbol{u}^*) = \boldsymbol{x}^\star$ as required. Furthermore, since it has been established that all elements $u_i^* \in (\zeta_i, \xi_i)$ and $u_j^* \notin (\zeta_j, \xi_j)$ are uniquely determined for the specified constrained optimum $\boldsymbol{x}^\star$, this equilibrium point is clearly unique.

Observation 6.2 indicates that at any equilibrium point of the system, the network output constitutes a constrained minimum of J, and that for each constrained minimum there exists a unique corresponding equilibrium point. If the given matrix W is positive definite, J is strictly convex and the constrained minimum $\boldsymbol{x}^\star$ is unique (Bazaraa & Shetty, 1979), and Observation 6.2 guarantees that the equilibrium point of the neural network exists and is unique. If on the other hand W has one or more zero eigenvalues, J is convex but not strictly convex, and the constrained optimum is not necessarily unique.

6.4.1 *Multiple Constrained Minima*

Multiple constrained minima clearly do arise in some cases, and in such cases the corresponding neural network has by Observation 6.2 multiple equilibrium points, each of which produces a network output which optimises J over the constraint region. To see that the constrained minimum is not unique in some cases, note firstly that the parabolic cost function $J(\boldsymbol{x})$ has zero curvature in the directions of the eigenvectors with zero eigenvalue. Differentiating $J(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, using the eigenvalue decomposition for Q real, symmetric and diagonalisable, and setting the gradient of J to zero in order to locate the stationary points yields

$$\nabla J = \sum_i \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T \boldsymbol{x} - \boldsymbol{r} = 0 \tag{6.16}$$

where λ_i and $\boldsymbol{e}_i$ are the ith eigenvalue and eigenvector of Q respectively. For convenience the *kernel* or *nullspace* of Q is denoted by $\ker(Q) = \{\boldsymbol{e}_i | \lambda_i = 0\}$. If $\boldsymbol{r}$ has a component in the direction of any eigenvector $\boldsymbol{e}_i \in \ker(Q)$ of Q, $J(\boldsymbol{x})$ has constant non-zero slope in that direction; thus the rightmost equality in (6.16) cannot be satisfied, and $J(\boldsymbol{x})$ has no global minimum. However, with the imposition of the proposed bound constraints a constrained minimum is introduced. Furthermore J has zero slope in the direction of each eigenvector $\boldsymbol{e}_i \in \ker(Q)$ (if any) to which $\boldsymbol{r}$ is orthogonal, and there exists an infinite set of unconstrained minima in this direction. In some cases, the imposition of constraints can resolve this degeneracy; this occurs for example if the set of unconstrained minima lies outside the constraint region, and the corresponding eigenvector with zero eigenvalue is not parallel to any of the boundary surfaces of the constraint region. This example illustrates the potential existence of multiple constrained minima when W — and hence Q — is singular.

Problem	Parameters				Minima	
	Q	$\boldsymbol{r}$	$\boldsymbol{\mu}$	$\boldsymbol{\nu}$	$\boldsymbol{x}^\circ$	$\boldsymbol{x}^\star$
I	$\begin{matrix}2 & 2 & 1\\ 2 & 2 & 1\\ 1 & 1 & 2\end{matrix}$	$\begin{matrix}0.5\\ -0.5\\ 0.5\end{matrix}$	-1	1	—	$\begin{matrix}1.0\\ -1.0\\ 0.25\end{matrix}$
II	"	$\begin{matrix}0.5\\ 0.5\\ -0.5\end{matrix}$	"	"	$\begin{matrix}s\\ 0.5-s\\ -0.5\end{matrix}$	$\begin{matrix}-0.5 \leq s \leq 1.0\\ 0.5-s\\ -0.5\end{matrix}$

Table 6.1: Parameters for the two BCSQO problems in Example 6.1.

Example 6.1 The parameters for two bound constrained positive semidefinite quadratic optimisation problems with W singular are listed in Table 6.1, along with the unconstrained and constrained minima $\boldsymbol{x}^\circ$ and $\boldsymbol{x}^*$ respectively. In problem II, $\boldsymbol{r}$ is orthogonal to the eigenvector of Q having zero eigenvalue, and there exist multiple unconstrained and constrained minima (expressible in parametric form). In problem I on the other hand, this orthogonality condition does not hold, and a single constrained minimum is observed, while the unconstrained minimum exists only at infinity.

6.5 Convergence Analysis

So far it has been established that the equilibrium points of the system are in one-to-one correspondence with the constrained minima of J. All trajectories of the system are now shown to converge to the set of equilibrium points, from which it can be concluded that the output of the network converges to a constrained minimum of the quadratic problem. To achieve this it is necessary to invoke LaSalle's invariance principle (LaSalle, 1968; LaSalle, 1976).

Observation 6.3 *The energy function $E(\boldsymbol{u})$ given in (6.11) is a global Liapunov function for the BCP neural network described by (6.9).*

Before the proof of this observation is presented, it should be remarked that although Sudharsanan & Sundareshan (1991) made the equivalent observation regarding their neural network, their proof relied on both invertibility and differentiability of their activation function g, neither of which hold. The following proof rectifies these errors.

Proof. Since the function $\boldsymbol{f}(\boldsymbol{u})$ (and hence $\boldsymbol{h}(\boldsymbol{u})$) is continuous and bounded, $E(\boldsymbol{u})$ is also continuous and bounded. Denoting by D_+ the right lower derivative operator (LaSalle, 1976), it therefore remains to show that the forward time derivative D_+V of

$V(t) \triangleq E(\boldsymbol{u}(t; u_0, t_0))$ [1] satisfies

$$D_+ V \leq 0$$

and hence — since $E(\boldsymbol{u})$ is continuous — that $E(\boldsymbol{u})$ is non-increasing along trajectories.

Applying the chain rule to the differentiation of $V(t)$, defining $F{:}\mathbb{R}^n \to \mathbb{R}^{n\times n}$ such that

$$F_{ij}(\boldsymbol{u}) = \begin{cases} 0 & j \neq i \\ \begin{cases} D_+ f_i & \dot{u}_i > 0 \\ 0 & \dot{u}_i = 0 \\ D_- f_i & \dot{u}_i < 0 \end{cases} & j = i \end{cases} \tag{6.17}$$

and noting that $D_+ \boldsymbol{u} = \dot{\boldsymbol{u}}$ for continuous input gives

$$D_+ V = \left(\frac{dE}{d\boldsymbol{h}}\right)^T \frac{d\boldsymbol{h}}{d\boldsymbol{f}} F(\boldsymbol{u})\dot{\boldsymbol{u}} \tag{6.18}$$

In particular, F accounts appropriately for the discontinuity in the derivative of f_i at ζ_i and ξ_i by substituting at each point the left or right derivative of f_i depending on the sign of the time-derivative of u_i. Substituting for the first and third terms in (6.18) using (6.13) and (6.17) respectively, and noting that the second term is simply B and that $F(\boldsymbol{u})(\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{u}) = \boldsymbol{0}$ gives

$$D_+ V = -\dot{\boldsymbol{u}}^T \Gamma^{-1} B F(\boldsymbol{u})\dot{\boldsymbol{u}} \leq 0 \tag{6.19}$$

since $\Gamma^{-1} B F(\boldsymbol{u})$ is non-negative diagonal and hence positive semidefinite. Therefore, the energy function $E(\boldsymbol{u})$ is non-increasing along trajectories, and hence is a global Liapunov function for the BCP network.

Theorem 6.1 *Trajectories of the BCP network in (6.9) converge to the set of equilibrium points.*

The network is said to be *quasiconvergent* if the equilibrium points are spatially contiguous, and *globally convergent* if the equilibrium point is unique.

Proof. Because the dynamical system (6.9) is bounded for positive diagonal A (see Appendix D.1), every forward trajectory converges to a non-empty compact and connected set, the *positive limit set*, which is invariant under the dynamics (LaSalle, 1976). Since $E(\boldsymbol{u})$ is a global Liapunov function for the system, all trajectories will by LaSalle's invariance principle converge to $\mathcal{M}$, the largest invariant subset of the set in which the Liapunov function is constant on orbits. Setting $D_+ V = 0$ in (6.19) and noting that $\Gamma^{-1} B F(\boldsymbol{u})$ is non-negative diagonal gives

$$\mathcal{M} \subset \mathcal{E} \triangleq \{\boldsymbol{u} : \dot{u}_i = 0 \quad \text{or} \quad F_{ii}(\boldsymbol{u}) = 0\}$$

[1] Since $E(\boldsymbol{u}(t; u_0, t_0))$ is strictly a function of the spatial variable $\boldsymbol{u}$, $V(t)$ is introduced so that the application of D_+ obtains the forward *temporal* derivative.

Consider first the set $\mathcal{R} \subset \mathcal{E}$ which consists of all points $\boldsymbol{u} \in \mathcal{E}$ which are mapped by $\boldsymbol{h}$ onto points $\boldsymbol{x}$ *not* satisfying the Kuhn-Tucker conditions. It will be shown that any trajectory $\boldsymbol{u}(t; \boldsymbol{u}_0, t_0)$ with $\boldsymbol{u}_0 \in \mathcal{R}$ passes outside of $\mathcal{E}$, and hence that the set $\mathcal{R}$ is excluded by the invariant set $\mathcal{M} \subset \mathcal{E}$.

Let the Kuhn-Tucker optimality conditions (6.12) be violated by the current network output state $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{u})$. Then

$$\exists i \mid \quad \nabla_i J \begin{cases} < 0 & x_i = \mu_i \\ \neq 0 & x_i \in (\mu_i, \nu_i) \\ > 0 & x_i = \nu_i \end{cases} \tag{6.20}$$

Rewriting this in terms of the network activation vector $\boldsymbol{u}$ and the unconstrained cost function E and using (6.13) yields

$$\exists i \mid \quad \dot{u}_i \begin{cases} > 0 & u_i \leq \zeta_i \\ = -\gamma_i \beta_i \nabla_i E \neq 0 & u_i \in (\zeta_i, \xi_i) \\ < 0 & u_i \geq \xi_i \end{cases} \tag{6.21}$$

In particular an offending component $u_i \in (\zeta_i, \xi_i)$ must have a non-zero time derivative, and therefore cannot belong to a point in the set $\mathcal{E}$. On the other hand, an offending component $u_i \notin (\zeta_i, \xi_i)$ has by (6.21) a time derivative which will return it to the linear region of its activation function, and the first moment at which one or more such components satisfies $u_i \in \{\zeta_i, \xi_i\}$ will now be considered. Since the aim is to identify the invariant set $\mathcal{M} \subset \mathcal{E}$, the trajectory $\boldsymbol{u}(t)$ is assumed to remain in $\mathcal{E}$ until this time, so that no other component $u_j \in (\zeta_j, \xi_j)$ — and hence $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{u})$ — has changed before this happens. Then for each offending component $u_i \in \{\zeta_i, \xi_i\}$, $F_{ii}(\boldsymbol{u}) = 1$ and $\dot{u}_i \neq 0$, so that at this stage the trajectory passes outside of $\mathcal{E}$. Thus it is concluded that $\mathcal{M} \subset \mathcal{E} - R$.

Now since the Kuhn-Tucker conditions are sufficient for a constrained minimum of J and hence a global minimum of E, the set $\mathcal{E} - R$ must be invariant under the system dynamics, since otherwise E must increase at some stage, contrary to Observation 6.3. Thus $\mathcal{M} = \mathcal{E} - R$, and it remains only to prove convergence of all trajectories which start in $\mathcal{E} - R$ to the set of equilibrium points. Since this result has already been established in the last part of the proof of Observation 6.2, it can be concluded that all trajectories of the system converge to the set of equilibrium points, as required.

6.6 Related Neural Networks

6.6.1 The Hopfield Network

The networks described by Equations (6.5) and (6.9) fall into the general class of *additive* networks (Grossberg, 1969), and are similar to the associative memory model studied by Hopfield (1984) and used by Hopfield & Tank (1985) for combinatorial optimisation.

The Hopfield model has a sigmoidal activation function $\boldsymbol{s}:\mathbb{R}^n \to \mathbb{R}^n$ such that $s_i(\boldsymbol{u}) \equiv \nu_i s(u_i) \in [0, \nu_i]$ where ν_i is the maximum activation of the ith node, and the nodal activation vector $\boldsymbol{u}$ is governed by

$$\begin{aligned} C\dot{\boldsymbol{u}} &= \boldsymbol{r} + T\boldsymbol{s}(\boldsymbol{u}) - R^{-1}\boldsymbol{u} \\ \boldsymbol{x} &= \boldsymbol{s}(\boldsymbol{u}) \end{aligned}$$

where $\boldsymbol{r}, \boldsymbol{x} \in \mathbb{R}^n$ are the network input and output respectively, $T \in \mathbb{R}^{n\times n}$ is the symmetric lateral interconnection matrix, and $C, R \in \mathbb{R}_+^{n\times n}$ are diagonal matrices representing the nodal capacitance and resistance respectively. In the *high gain limit* as the activation function tends to the unit step function, the quadratic function

$$V_H(\boldsymbol{x}) \triangleq -\frac{1}{2}\boldsymbol{x}^T T\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{r}$$

is a Liapunov function for the Hopfield network (Hopfield & Tank, 1986). Thus in the high gain limit for T negative semidefinite the network solves the optimisation problem

$$\min_{\boldsymbol{x}} \{V_H(\boldsymbol{x}) : x_i \in \{0, \nu_i\} \quad \forall i\}$$

which differs from the BCSQO problem in that solutions are additionally constrained to lie only on vertices of the constraint region.

The mapping of the overdetermined full-rank SRE optimisation problem onto the BCP network bears superficial resemblance to the mapping by Tank & Hopfield (1986) of the non-orthogonal decomposition decision problem — the object of which is to produce a binary *decision* vector $\boldsymbol{x}$ such that x_i indicates the presence or absence of the ith basis function in the input signal — onto the Hopfield network. The latter mapping involved setting

$$\begin{aligned} \boldsymbol{r} &= G\boldsymbol{i} + \frac{1}{2}\boldsymbol{vdiag}\,(GG^T) \\ T &= -\text{offdiag}\,(GG^T) \\ \boldsymbol{\nu} &= \mathbf{1} \end{aligned}$$

where $\boldsymbol{vdiag}:\mathbb{R}^{n\times n} \to \mathbb{R}^n$ returns the diagonal entries of its matrix argument as a column vector. However, the mappings differ in that the diagonal entries of GG^T are used in the Hopfield model as constant nodal input ("bias") terms, and in the BCP model as nodal self-inhibitory or decay terms.

6.6.2 *Generalised Brain-State-In-A-Box (GBSB) Network*

Golden (1992) recently generalised the discrete-time *Brain-State-in-a-Box* (BSB) model proposed by Anderson et al. (1977) to be governed by the update equations

$$\boldsymbol{x}(t+1) = \boldsymbol{f}\,(\boldsymbol{x}(t) + \alpha\Gamma\boldsymbol{\delta}(t)) \tag{6.22a}$$

$$\boldsymbol{\delta}(t) \triangleq (W + W^T)\boldsymbol{x} + \boldsymbol{r} \tag{6.22b}$$

where $\boldsymbol{r}, \boldsymbol{x} \in \mathbb{R}^n$ are the network input and output respectively, $\boldsymbol{\delta}(t) \in \mathbb{R}^n$ is the output update term, $W \in \mathbb{R}^{n \times n}$ is not necessarily positive semidefinite, $\alpha \in \mathbb{R}_+$ is the temporal step size, and $\Gamma \in \mathbb{R}_+^{n \times n}$ is a diagonal preconditioner[2] with diagonal entries $\gamma_i \in (0, 1)$. The function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^n$ is equivalent to that defined in (6.8) with $\boldsymbol{\zeta} = \boldsymbol{\mu}$ and $\boldsymbol{\xi} = \boldsymbol{\nu}$. Golden (1992) showed that for sufficiently small step size α the quadratic cost function

$$V_G(\boldsymbol{x}) \triangleq -\boldsymbol{x}^T W \boldsymbol{x} - \boldsymbol{x}^T \boldsymbol{r}$$

is a Liapunov function for the generalised BSB (GBSB) network. In the case where W is negative semidefinite, this network can therefore be shown to solve the BCSQO problem in (6.1). However, in the more general case where W is not necessarily positive semidefinite the gradient descent strategy employed by the GBSB and BCP networks which will be discussed in Section 6.7 is not guaranteed to find the constrained minimum.

6.6.3 *Continuous-time GBSB Network*

A continuous-time analogue of the GBSB network with preconditioning — henceforth referred to as the *continuous Generalised Brain-State-in-a-Box* (CGBSB) network — can be formulated as

$$\dot{u}_i = \begin{cases} \max\{0, \eta_i\} & u_i \leq \zeta_i \\ \eta_i & u_i \in (\zeta_i, \xi_i) \\ \min\{0, \eta_i\} & u_i \geq \xi_i \end{cases} \tag{6.23a}$$

$$\boldsymbol{\eta}(t) \triangleq \boldsymbol{y} - C\boldsymbol{u} \tag{6.23b}$$

$$\boldsymbol{x} = B\boldsymbol{u} \tag{6.23c}$$

where

$$\begin{aligned} \boldsymbol{y} &= \Gamma \boldsymbol{r} \\ C &= \frac{1}{2} \Gamma \left(W + W^T \right) B \end{aligned}$$

and $\boldsymbol{\zeta}, \boldsymbol{\xi}$ are as defined previously. For the case where W is positive semidefinite this network is similar to the BCP network except that the nodal activation $\boldsymbol{u}$ is itself implicitly hard-limited by the network dynamics whenever it attempts to exceed the range $[\boldsymbol{\zeta}, \boldsymbol{\xi}]$, and consequently $\boldsymbol{f}(\boldsymbol{u})$ may be replaced by $\boldsymbol{u}$ provided the network is initialised such that $\boldsymbol{u}_0 \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$. This change addresses the fact that in active analog implementations of the BCP network, the nodal activation vector $\boldsymbol{u}$ is subject to qualitatively the same saturation which limits the range of outputs, and hence cannot be allowed to assume arbitrarily large values, as would be required in extreme cases in order for the network to attain equilibrium. In Appendix D.2, it is shown that for W positive semidefinite and

[2] Although Golden (1992) suggested neither this specific purpose nor any suitable choice for Γ.

$\boldsymbol{u}_0 \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$, the CGBSB network converges to the set of constrained minima of $J(\boldsymbol{x})$ and hence solves the BCSQO problem.

However, if a trajectory starts or is caused by noise to stray outside of the feasible region, it will not necessarily return to this region under the network dynamics. If $\boldsymbol{u}$ is physically constrained to the feasible region by for example amplifier saturation, this problem will not arise. However, to ensure for robustness that in the absence of such physical constraints the network produces only feasible outputs, (6.23a) should be reformulated as

$$\dot{u}_i = \begin{cases} \max\{\zeta_i - u_i, \eta_i\} & u_i \leq \zeta_i \\ \eta_i & u_i \in (\zeta_i, \xi_i) \\ \min\{\xi_i - u_i, \eta_i\} & u_i \geq \xi_i \end{cases} \tag{6.23d}$$

This has the effect of collapsing each set of adjacent equilibrium points lying outside this region onto the single equilibrium point given by their projection onto the feasible region, and causing trajectories starting from infeasible solutions to return to the feasible region. This latter function is performed in the GBSB network by the saturation function $\boldsymbol{f}:\mathbb{R}^n \to \mathbb{R}^n$. However, a full stability analysis of this network will be deferred until a later publication.

6.7 Optimisation Strategy

Conventional approaches to bound-constrained quadratic optimisation reviewed by Moré & Toraldo (1991) share two essential elements: an algorithm — such as the conjugate gradient (CG) method — for searching a selected boundary face of the constraint region, and a rule or algorithm for deciding when and how to select a new boundary face for such a search. Although the necessary matrix computations are parallelisable, the core of such algorithms — including the series of 1D searches required by the CG algorithm — is inherently sequential. In contrast, the BCP neural network is formulated in continuous time and admits a direct analog implementation, rendering direct comparisons on the basis of computational efficiency impossible. Instead, the optimisation *strategy* employed by the BCP network is compared in this section with those of conventional and neural network approaches to BCSQO. It is noted in passing that since any (stable) discrete-time simulation of the BCP network automatically constitutes a numerical method for solution of the BCSQO problem, efficiency comparisons between the above algorithmic methods and such simulations *are* possible for any suitable discrete-time simulation algorithm. Such comparisons are however beyond the scope of this thesis.

6.7.1 Description & Comparisons

Noting that for $\dot{\boldsymbol{u}}$ continuous, $D_+\boldsymbol{x} = BF(\boldsymbol{u})\dot{\boldsymbol{u}}$ and substituting for $\dot{\boldsymbol{u}}$ using (6.13) yields

$$D_+\boldsymbol{x} = -\Gamma BF(\boldsymbol{u})\nabla J \tag{6.24}$$

where $F(\boldsymbol{u})$ defined in (6.17) is diagonal. For the special case $\Gamma = B = F(\boldsymbol{u}) = I_n$, (6.24) performs steepest descent on J, as illustrated below in Example 6.2. As will be seen from Example 6.3, the effect of the preconditioners Γ, B is to modify the search direction from that of steepest descent — viz. the negative of the true gradient; however, since Γ and B are positive diagonal, the network still performs gradient descent — i.e. $(D_+\boldsymbol{x})^T \nabla J < 0$ — and E strictly decreases with time.

The effect of premultiplication of the gradient ∇J by $F(\boldsymbol{u})$ in (6.24) is twofold. Firstly, it sets to zero the forward time derivative of any output component x_i which under gradient descent would be forced out of the constraint region. Illustrated in Example 6.2 below, this effect results from the fact that since the output trajectory $\boldsymbol{x}(t)$ is obtained by mapping $\boldsymbol{u}(t)$ through $\boldsymbol{h}$, it is necessarily confined to the constraint region, and hence is constrained to move only in feasible directions. In contrast, conventional *penalty function* methods — upon which previous neural network approaches to constrained linear and nonlinear optimisation have been based (Tank & Hopfield, 1986; Kennedy & Chua, 1988; Rodríguez-Vázquez et al., 1990; Chen et al., 1992; Maa & Shanblatt, 1992a) — can generate infeasible solutions (Bazaraa & Shetty, 1979; Maa & Shanblatt, 1992b). The two networks proposed respectively by Rodríguez-Vázquez et al. (1990) and Maa & Shanblatt (1992b) in their attempts to rectify this deficiency have yet to be proven stable[3], and both ensure feasibility at the expense of considerable additional hardware and network complexity.

The optimisation method described so far and henceforth referred to as *constrained* gradient descent — or in the special case where $\Gamma = B = I_n$ *constrained* steepest descent — falls into the general class of *gradient methods* for constrained nonlinear optimisation (Gellert et al., 1989), in which the search direction used for minimisation is any feasible downhill direction. However, the second effect of premultiplication of the gradient vector by $F(\boldsymbol{u})$ is to set to zero the forward time derivative of any output component x_i for which u_i is strictly in the saturation region of f_i, even if $\nabla_i J$ is such that under gradient descent x_i would move in a feasible direction away from the corresponding boundary face. Although this has the potentially undesirable effect of confining the search process to the current boundary face, the following argument shows that this confinement is only temporary, as expected given the convergence result in Section 6.5.

For such points $\boldsymbol{x}$ on the boundary of the constraint region, at least one of the components x_i lies on one of its corresponding constraint surfaces $x_i = \mu_i$ and $x_i = \nu_i$, and the relevant constraint is said to be *active*. If there exist feasible directions pointing away from an active constraint surface and in which J is decreasing, then the Kuhn-Tucker optimality conditions (6.12) must be violated, and by (6.21) the time derivative of the element u_i corresponding to this constraint is such that u_i will eventually be brought

[3]The stability of the network of Rodríguez-Vázquez et al. (1990) has in fact been questioned by Maa & Shanblatt (1992a).

into the linear region of its activation function f_i. In the mean time, all other output components x_j corresponding to activations $u_j \in [\zeta_j, \xi_j]$ are free to pursue constrained gradient descent on J with the search confined at least to the ith boundary face of the constraint region. Once u_i has returned to the linear region, x_i can become unstuck from the relevant constraint boundary, and the search is then no longer confined to this boundary face.

In the case where the system is not immediately free to pursue search directions away from a boundary face, the return of u_i into the linear region of its corresponding activation function is governed by

$$\dot{u}_i = \begin{cases} a_{ii}(\zeta_i - u_i) - \gamma_i \nabla_i J & x_i = \mu_i \\ a_{ii}(\xi_i - u_i) - \gamma_i \nabla_i J & x_i = \nu_i \end{cases} \tag{6.25}$$

where a_{ii} is the ith diagonal entry of A. In the special case where the desired search direction is perpendicularly away from the boundary face — ie $\boldsymbol{x}$ minimises J over this active boundary face — u_i converges exponentially towards a point inside the linear region of its activation function. However, as soon as u_i hits the linear region, $\dot{u}_i = -\gamma_i \nabla_i J$ and the search is free to pursue the component of the gradient in the ith direction. This behaviour is evident in Example 6.4 below.

It is therefore concluded that the optimisation method employed by the BCP network is gradient-based, with the search constrained to feasible directions and at worst temporarily to one or more boundary faces on which the current point $\boldsymbol{x}$ is located. Although in Example 6.4 this latter effect does not cause an appreciable delay in obtaining the optimal solution, it has the potential to do so in more extreme cases. This problem is overcome by the GBSB and CGBSB models which prevent saturation of the neural activation, leaving them free on presentation of a new network input to pursue constrained gradient descent on the new cost function J. The optimisation strategy of the GBSB network for $\Gamma = I_n$ is closely related to the *gradient projection* (GP) method used by Moré & Toraldo (1991) to move between boundary faces, with the principal difference being in the use of a fixed rather than adaptive Euler step size α. This method involves taking a step in the direction of the gradient, and then projecting the resultant point back onto the constraint region — a function performed by the nonlinearity in the GBSB model. The continuous-time equivalent of this strategy employed by the CGBSB model is to constrain the search direction itself to feasible directions by modifying the gradient accordingly. Although the GP method is itself sufficient to locate the bound-constrained optimum of a convex cost function, more computationally efficient though inherently serial methods — such as the CG algorithm — are often employed to search any new boundary face located by an abbreviated GP search (Moré & Toraldo, 1991).

6.7.2 *Examples*

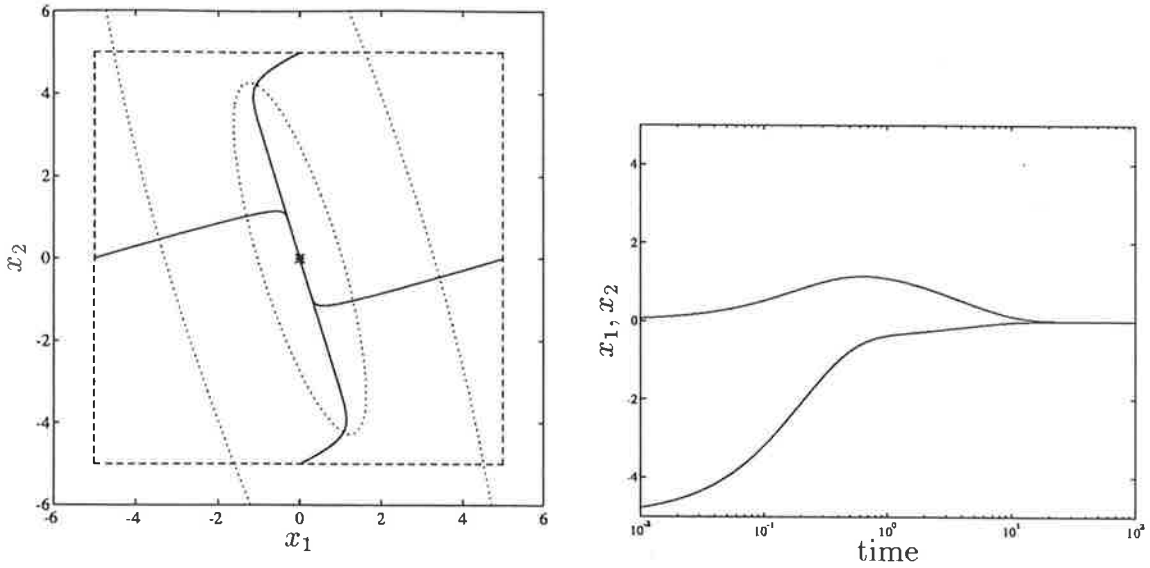
Problem	Parameters				Minima	
	Q	$\boldsymbol{r}$	$\boldsymbol{\mu}$	$\boldsymbol{\nu}$	$\boldsymbol{x}^\circ$	$\boldsymbol{x}^\star$
I	4.585 1.341 1.341 0.6648	0 0	−5 −5	5 5	0 0	0 0
II	"	4.368 −0.6307	"	"	3 −7	2.415 −5.000

Table 6.2: Parameters for the two BCSQO problems in Example 6.2; $\boldsymbol{x}^\circ$ and $\boldsymbol{x}^\star$ denote the unconstrained and constrained minima respectively.

Initialisation		Equilibria			
		Problem I		Problem II	
$\boldsymbol{u}_0$	$\boldsymbol{x}_0$	$\boldsymbol{u}^\star$	$\boldsymbol{x}^\star$	$\boldsymbol{u}^\star$	$\boldsymbol{x}^\star$
−5 0	−5 0	0 0	0 0	2.415 −5.820	2.415 −5.000
0 5	0 5	"	"	"	"
5 0	5 0	"	"	"	"
0 −5	0 −5	"	"	"	"

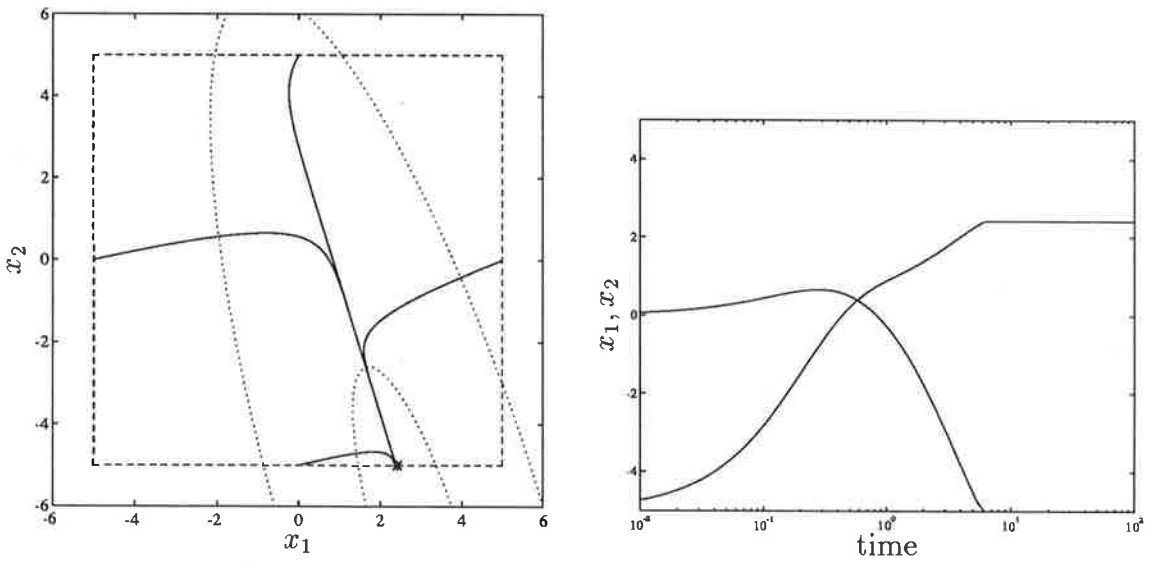
Table 6.3: Network equilibria for the application of the BCP network with $\Gamma = I_n = B$ to the BCSQO problems in Table 6.2. The network output trajectory is illustrated in Figure 6.4 for problem I and in Figure 6.5 for problem II; $\boldsymbol{u}^\star$ and $\boldsymbol{x}^\star$ denote the neural activation and output vectors respectively at equilibrium.

Example 6.2 The BCP network with $\Gamma = I_n = B$ was applied to the two BCSQO problems whose parameters are listed in Table 6.2, along with the corresponding constrained and unconstrained minima — denoted $\boldsymbol{x}^*$ and $\boldsymbol{x}^\circ$ respectively — of J, which are coincident in problem I but distinct for problem II. Spatial and temporal representations of the trajectories $\boldsymbol{x}(t; \boldsymbol{x}_0, 0)$ of the network output for four different starting points $\boldsymbol{x}_0$ for each of problems I and II are presented in Figures 6.4 and 6.5 respectively, and the network equilibria listed in Table 6.3. In the interior of the constraint region as expected each trajectory performs steepest descent on J — and hence E — moving perpendicular to the contours and converging to the constrained optimum. However, when the trajectories in Figure 6.5(a) reach the constraint boundary $x_2 = -5$, they are prevented by the nodal activation function f_2 from leaving the constraint region. Characteristic of the steepest descent strategy is the rapid descent on steep inclines followed by slow convergence along the gently sloping valley floor, as can be seen from the temporal plots in Figures 6.4 and 6.5 of the trajectory starting from $\boldsymbol{x}_0 = [-5, 0]^T$ for each problem.



(a) Spatial (b) Temporal

Figure 6.4: (a) Spatial representation of the trajectories $\boldsymbol{x}(t; \boldsymbol{x}_0, 0)$ (solid) of the BCP network without preconditioning ($\Gamma = I_n = B$) applied to problem I in Table 6.2 for the four starting points listed in Table 6.3. Logarithmically spaced contours of the cost function J (dotted) along with the four boundary faces of the constraint region (dashed) are shown superimposed. The constrained minimum is marked with an asterisk. (b) Temporal representation of the trajectory shown in (a) starting from $\boldsymbol{x}_0 = [-5, 0]^T$. The Runge-Kutta method of order 4 with adaptive step size was used to simulate the network dynamics over temporal intervals which were equal on a logarithmic scale.



(a) Spatial (b) Temporal

Figure 6.5: Description as for Figure 6.4 but with reference to problem II in Table 6.2.

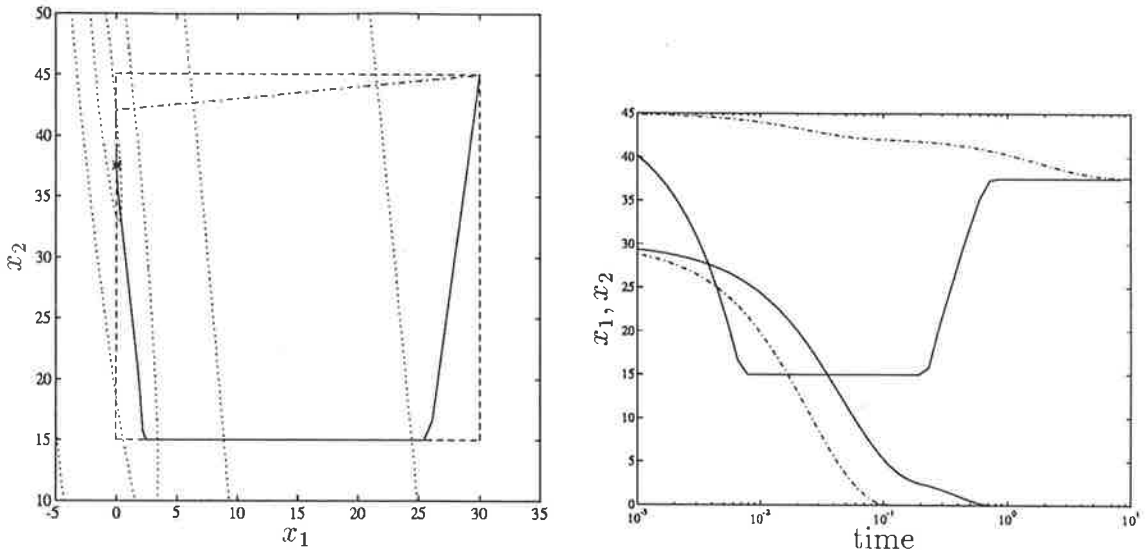
Parameters					Minima	
Q		$\boldsymbol{r}$	$\boldsymbol{\mu}$	$\boldsymbol{\nu}$	$\boldsymbol{x}^\circ$	$\boldsymbol{x}^\star$
39.60	3.960	138.6	0	30	-1	0
3.960	0.5307	19.92	15	45	45	37.56

Table 6.4: Parameters for the BCSQO problem in Example 6.3.

Preconditioning		Initialisation		Equilibria	
$\Gamma = B$		$\boldsymbol{u}_0$	$\boldsymbol{x}_0$	$\boldsymbol{u}^\star$	$\boldsymbol{x}^\star$
I_n		30 45	30 45	-0.2562 37.56	0.000 37.56
0.7361	0	40.75	"	-0.3448	"
0	6.359	7.077		5.903	

Table 6.5: Network equilibria for the application of the BCP network with two different settings of the preconditioners to the BCSQO problems in Table 6.4. The network output trajectory is illustrated in Figure 6.6.

Example 6.3 The modification of the search direction by preconditioning is illustrated by the following example. In anticipation of the discussion in Section 6.9 on the appropriate choice of preconditioners, the preconditioning strategy described in Section 5.4 is used. The BCP network was applied with and without preconditioning to the BCSQO problem whose parameters are listed along with the corresponding constrained and unconstrained minima of J in Table 6.4. The unconstrained optimum lies outside the constraint region. Note in passing that in keeping with the generalisation of the activation limits used in the formulation of the BCP network, the bound constraints are different for the two optimisation variables, and neither pair of constraints is placed symmetrically about the origin. Spatial and temporal representations of the trajectories $\boldsymbol{x}(t; \boldsymbol{x}_0, 0)$ of the network output both with and without preconditioning are presented in Figure 6.6 and the network equilibria listed in Table 6.3. For the preconditioned network, the preconditioners are chosen to satisfy (5.13), with the scaling factor $k = 4.633$ ensuring that the preconditioned matrix ΓQB has the same maximum eigenvalue as Q, so that a fair comparison of rates of convergence can be made. As can be seen from both the spatial and temporal views, both networks converge to the optimal solution. In the interior of the constraint region, the unpreconditioned system progresses as expected in the direction of steepest descent (perpendicular to the contours). Once the trajectory of the unpreconditioned system hits the constraint boundary $x_1 = 0$, it is constrained to move along that constraint in a direction as near as possible to that of steepest descent. The path taken by the preconditioned system, on the other hand, deviates strongly from steepest descent, spending some time in saturation on the boundary $x_2 = 15$.



(a) Spatial (b) Temporal

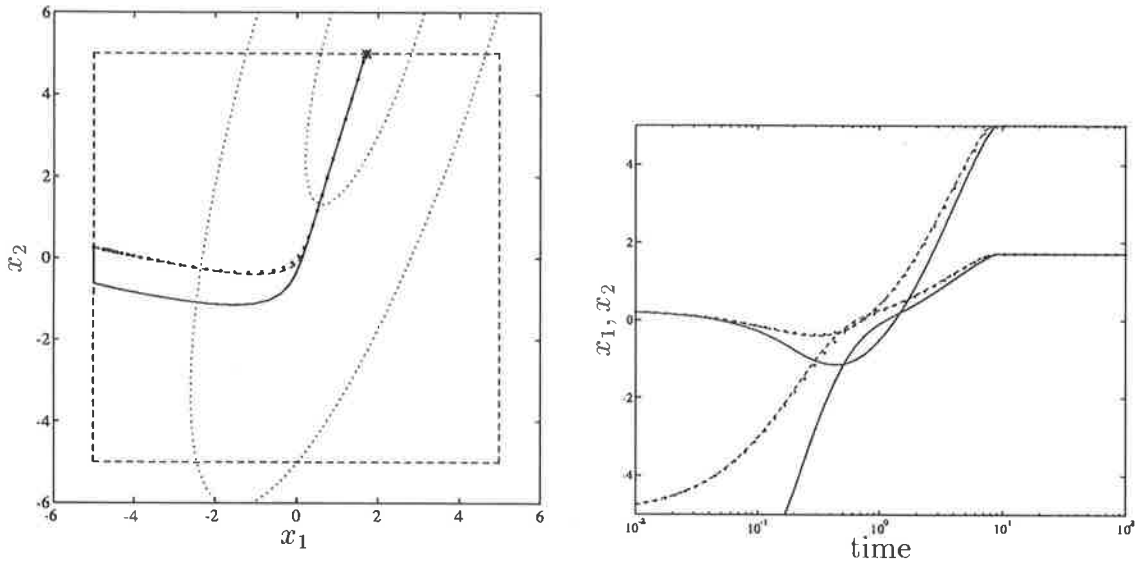
Figure 6.6: Spatial and temporal representations of the trajectories $\boldsymbol{x}(t;\boldsymbol{x}_0,0)$ of the unpreconditioned (dot-dashed) and preconditioned (solid) systems detailed in Tables 6.4 and 6.5 applied to the BCSQO problem whose parameters are listed in Table 6.4. Logarithmically spaced contours of the cost function J (dotted) along with the four boundary faces of the constraint region (dashed) are superimposed on the spatial representation. The constrained minimum is marked with an asterisk.

Problem	Parameters				Minima	
	Q	$\boldsymbol{r}$	$\boldsymbol{\mu}$	$\boldsymbol{\nu}$	$\boldsymbol{x}^\circ$	$\boldsymbol{x}^\star$
I	4.585 −1.341	−51.47	−5	5	−20	−5.000
	−1.341 0.6648	6.876	−5	5	−30	0.2565
II	"	1.124	"	"	2	1.708
		1.307			6	5.000

Table 6.6: Parameters for the two BCSQO problems in Example 6.4.

Example 6.4 In this example the optimisation strategies of the GBSB, CGBSB and BCP networks are compared. Each network was first applied without preconditioning to the solution of problem I in Table 6.6. For the BCP network, this had the effect of driving u_1 well into the saturation region of its corresponding activation function, whereas the activations of the GBSB and CGBSB networks were restricted to their linear ranges by their respective difference or differential equations. Each network was then presented with the new input $\boldsymbol{y} = \boldsymbol{r}$ in problem II of Table 6.6. Spatial and temporal representations of the trajectories $\boldsymbol{x}(t;\boldsymbol{x}_0,0)$ of the output of each network subsequent to the presentation of the new input are presented in Figure 6.7 and the network equilibria listed in Table 6.7.

The BCP network takes some time to bring u_1 back into its linear region, during which



(a) Spatial (b) Temporal

Figure 6.7: Spatial and temporal representations of the trajectories of the GBSB (dotted), CGBSB (dashed) and BCP (solid) networks without preconditioning. Logarithmically spaced contours of the cost function J (dotted) along with the four boundary faces of the constraint region (dashed) are superimposed on the spatial representation. The constrained minimum is marked with an asterisk.

Network	Initialisation		Equilibria	
	$\boldsymbol{u}_0$	$\boldsymbol{x}_0$	$\boldsymbol{u}^\star$	$\boldsymbol{x}^\star$
GBSB	−5.000 0.2565	−5.000 0.2565	1.708 5.000	1.708 5.000
CGBSB	"	"	"	"
BCP	−11.15 0.2565	"	1.708 5.410	"

Table 6.7: Network equilibria for the application of the GBSB, CGBSB and BCP networks to problem II in Table 6.6. The output trajectory of each network is illustrated in Figure 6.7.

$\boldsymbol{x}(t)$ is constrained to move along the constraint boundary $x_1 = -5$ in a direction as near as possible to that of steepest descent. In contrast, both the GBSB and CGBSB networks are immediately free to pursue constrained steepest descent on the new error function J. As expected, the trajectories of these latter two networks are almost identical, with the slight discrepancies being attributable to the use by the GBSB network of an explicit Euler method approximation to the ODE governing the CGBSB network.

6.7.3 Constraint Generalisation

The constraints on the optimisation variables considered in this chapter are quite specific: either the constraint region is hyper-rectangular and its principal axes are aligned with the optimisation variable axes, or it can be described by

$$\boldsymbol{\mu} \leq P\boldsymbol{x} \leq \boldsymbol{\nu} \tag{6.26}$$

where $P \in \mathbb{R}^{n \times n}$ is non-singular, which can be linearly — although not necessarily efficiently — transformed into this form through substitution of $\boldsymbol{x} = P^{-1}\boldsymbol{z}$ followed by a change of variables. However, the technique by which the outputs of the BCP network are limited to strictly feasible solutions should be amenable to extension to general convex constraint sets. The constraints which constitute the current minimum upper and maximum lower permissible limits on the range of a given optimisation variable x_i should be used to set upper and lower thresholds ζ_i and ξ_i on the corresponding activation function f_i. In general these thresholds will be a function of $\boldsymbol{x}(t)$, making stability analysis and practical implementation of the resulting time-varying dynamical system significantly more challenging.

An alternative strategy suggested by the CGBSB network for generalising the constraints is to subtract from the negative of the cost function gradient vector those components which are in the direction of the outward facing normal to each active constraint, thereby limiting the search to feasible directions. This strategy can be expressed mathematically as

$$\dot{\boldsymbol{x}} = -\nabla J - \sum_{i \in \mathcal{A}} r(-\nabla J^T \boldsymbol{n}_i)\boldsymbol{n}_i$$

where $\mathcal{A}$ is the set of active constraints, $\boldsymbol{n}_i$ is the outward facing normal to the ith constraint, and $r{:}\mathbb{R} \to \mathbb{R}$ with $r(\cdot) \triangleq \max(0, \cdot)$ is the half-wave rectification function. The convergence of such a network to the set of minima of the positive semidefinite cost function over a convex constraint set is readily argued through consideration of the underlying strategy of gradient descent constrained to feasible directions. Although for linear constraints the normals are independent of the current point $\boldsymbol{x}$ so that the requisite inner product can be implemented in a linear (non-integrating) node having predetermined weights and activation function r, in the more general case of nonlinear constraints the "weights" would need to be continuously recomputed. Furthermore the

output of each such node would have to be gated by the output of a second node which is responsible for determining whether the corresponding constraint is currently active. Thus, although this scheme for the more general case is clearly inherently parallel and amenable at least theoretically to fine-grained parallel implementation, the engineering and biological plausibility of such implementations awaits future investigation.

6.8 Exponential Stability

In this section it is shown that under appropriate conditions on the matrices A and C, the BCP network is exponentially stable. Sufficient conditions on the given matrix Q and the preconditioners Γ and B are then presented to ensure that in cases where the constrained minimum of the corresponding quadratic cost function is unique, the neural network produced by mapping the optimisation problem onto the BCP architecture converges exponentially to the optimal solution. For convenience $\lambda_{min}(A)$ is defined to be the minimum eigenvalue of A and $\sigma_{max}(C)$ to be the maximum singular value of C.

6.8.1 Theoretical Results

Definition 6.1 *A dynamical system is said to be* exponentially stable *in the sense of Liapunov if* $\exists \beta, \eta \in \mathbb{R}_+$ *such that for all pairs* $\{\boldsymbol{x}(t; t_0, \boldsymbol{x}_0), \boldsymbol{x}(t; t_0, \boldsymbol{x}_1)\}$ *of solutions*

$$\|\boldsymbol{x}(t; t_0, \boldsymbol{x}_1) - \boldsymbol{x}(t; t_0, \boldsymbol{x}_0)\| \leq \beta e^{-\eta(t-t_0)} \tag{6.27}$$

Definition 6.2 *The* degree *of an exponentially stable dynamical system is the largest non-negative value of* η *for which (6.27) holds.*

Theorem 6.2 *If the self-feedback and lateral connection matrices* A *and* C *respectively satisfy*

$$\eta \triangleq \lambda_{min}(A) - \sigma_{max}(C) > 0 \tag{6.28}$$

then the BCP neural network (6.9) is exponentially stable with a lower bound on the degree of exponential stability given by η.

Proof. Since $\boldsymbol{f}(\boldsymbol{u})$ is continuous and Lipschitzian, the solution $\boldsymbol{u}(t) = \boldsymbol{u}(t; t_0, \boldsymbol{u}_0)$ of the initial value problem associated with (6.9) is continuous and unique. The solution is also bounded for A positive diagonal (see Appendix D.1), and is given by

$$\boldsymbol{u}(t) = e^{-A(t-t_0)}\boldsymbol{u}_0 + \int_{t_0}^{t} e^{-A(t-s)}\boldsymbol{y}(s)\, ds - \int_{t_0}^{t} e^{-A(t-s)} C \boldsymbol{f}(\boldsymbol{u}(s))\, ds \tag{6.29}$$

Subtracting $\boldsymbol{u}(t)$ from a second trajectory $\boldsymbol{v}(t)$ starting from $\boldsymbol{v}_0$ at time t_0 and taking the norm yields

$$\|\boldsymbol{v}(t) - \boldsymbol{u}(t)\| \leq \|e^{-A(t-t_0)}\| \|\boldsymbol{v}_0 - \boldsymbol{u}_0\| + \int_{t_0}^{t} \|e^{-A(t-s)}\| \|C\| \|\boldsymbol{f}(\boldsymbol{v}(s)) - \boldsymbol{f}(\boldsymbol{u}(s))\|\, ds \tag{6.30}$$

Using the *spectral norm*, the matrix norm induced by the Euclidean vector norm (Horn & Johnson, 1988), gives

$$\begin{aligned} \|\boldsymbol{f}(\boldsymbol{v}(s)) - \boldsymbol{f}(\boldsymbol{u}(s))\| &\leq \|\boldsymbol{v}(s) - \boldsymbol{u}(s)\| \\ \|e^{-A(t-s)}\| &= e^{-\lambda_{min}(A)(t-s)} \\ \|C\| &= \sigma_{max}(C) \end{aligned}$$

Substituting these results into (6.30) and multiplying both sides by $e^{\lambda_{min}(A)t}$ gives

$$\|\boldsymbol{v}(t) - \boldsymbol{u}(t)\|e^{\lambda_{min}(A)t} \leq \|\boldsymbol{v}_0 - \boldsymbol{u}_0\|e^{\lambda_{min}(A)t_0} + \sigma_{max}(C)\int_{t_0}^{t} \|\boldsymbol{v}(s) - \boldsymbol{u}(s)\|e^{\lambda_{min}(A)s}\, ds$$

Applying Gronwall's Inequality (Reinhard, 1986) and evaluating the integral in the resulting expression produces the following inequality

$$\|\boldsymbol{v}(t) - \boldsymbol{u}(t)\| \leq \|\boldsymbol{v}_0 - \boldsymbol{u}_0\|e^{-\eta(t-t_0)} \tag{6.31}$$

Therefore a sufficient condition for exponential stability of the BCP network is that the exponent in (6.31) is strictly negative, from which condition (6.28) follows. Furthermore if (6.28) is satisfied, η provides a lower bound on the degree of exponential stability, which completes the proof.

Theorem 6.2 shows that provided the network weight matrices A and C satisfy condition (6.28), the Euclidean distance between any two trajectories $\boldsymbol{u}(t; \boldsymbol{u}_0, t_0)$ and $\boldsymbol{v}(t; \boldsymbol{v}_0, t_0)$ will decrease at least exponentially with time. In particular, if $\boldsymbol{v}_0$ is chosen to be an equilibrium point of (6.9), and the given input vector $\boldsymbol{y}$ does not vary with time, all trajectories $\boldsymbol{u}(t; \boldsymbol{u}_0, t_0)$ must converge exponentially to that equilibrium point. However, since this is not possible if the system has multiple equilibrium points[4], (6.28) clearly cannot be satisfied by any choice of the preconditioners Γ and B if the matrix Q has one or more zero eigenvalues and the constrained minimum is not unique. Recall however that according to Theorem 6.1 the network will still converge to the set of constrained minima, even though it is not exponentially convergent.

6.8.2 *Practical Considerations*

The required singular value for the evaluation of (6.28) is the square root of the largest eigenvalue of CC^T, and can be efficiently computed using the *power method* (Golub & Van Loan, 1989). However, whilst Theorem 6.2 provides a lower bound on the rate of convergence of the network, it gives little indication of how best to choose the elements of the preconditioners Γ and B when mapping the constrained optimisation problem onto the proposed network, or whether or not there exist *any* such choices which will ensure exponential convergence or improve the convergence properties of the network.

[4]Try choosing $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ to be distinct equilibrium points.

Two sets of sufficient conditions for exponential convergence are now presented which make appropriate choices of these preconditioners more apparent, albeit at the expense of sacrificing some of the generality of Theorem 6.2.

For convenience, the ith diagonal entries of Γ and B are denoted by γ_i and β_i respectively, and γ_{max} and β_{max} are defined to be the largest of these, $q_{min} = \min_i\{q_{ii}\}$ to be the minimum diagonal entry of Q, and $\lambda_{max}(Q_o) = \max_i\{|\lambda_i(Q_o)|\}$ to be the maximum magnitude eigenvalue of the matrix $Q_o \triangleq \text{offdiag}(Q)$.

Lemma 6.1 *If the matrix Q defined in (6.3) satisfies*

$$q_{min} > \lambda_{max}(Q_o) \tag{6.32}$$

then choosing Γ and B such that

$$min_i\{\gamma_i q_{ii} \beta_i\} = \gamma_{max} q_{min} \beta_{max} \tag{6.33}$$

ensures that the BCP network is exponentially stable, with a lower bound on the degree of exponential stability given by $\gamma_{max}(q_{min} - \lambda_{max}(Q_o))\beta_{max}$.

Proof. Noting that since Γ and B are diagonal, $\Gamma Q_o B = \text{offdiag}(\Gamma Q B)$ yields

$$\begin{aligned} \gamma_{max}\lambda_{max}(Q_o)\beta_{max} &= \|\Gamma\|\|Q_o\|\|B\| \\ &\geq \|\Gamma Q_o B\| \\ &= \sigma_{max}(C) \end{aligned} \tag{6.34}$$

Starting with (6.33) and using (6.32) followed by (6.34) gives

$$\begin{aligned} \min_i\{\gamma_i q_{ii}\beta_i\} &= \gamma_{max} q_{min} \beta_{max} \\ &> \gamma_{max}\lambda_{max}(Q_o)\beta_{max} \\ &\geq \sigma_{max}(C) \end{aligned} \tag{6.35}$$

which is simply (6.28). Thus conditions (6.32) and (6.33) are sufficient to ensure exponential stability of the network.

Furthermore, by (6.35) the lower bound η on the degree of exponential convergence satisfies

$$\begin{aligned} \eta &\geq \min_i\{\gamma_i q_{ii}\beta_i\} - \gamma_{max}\lambda_{max}(Q_o)\beta_{max} \\ &= \gamma_{max}(q_{min} - \lambda_{max}(Q_o))\beta_{max} \end{aligned}$$

as required.

Equation (6.33) is satisfied by the choice of preconditioners in (5.13). Note that even if Q does not satisfy (6.32), there may still exist choices of Γ and B which ensure the satisfaction of (6.28), and hence guarantee exponential convergence of the system. The

power method can be used to efficiently compute $\lambda_{max}(Q_o)$ in order to verify whether or not Q satisfies condition (6.32). However, the following corollary replaces condition (6.32) in Lemma 6.1 with a condition which is computationally less expensive to evaluate whilst still sufficient to satisfy (6.32).

Corollary 6.1 *If the matrix Q defined in (6.3) satisfies*

$$min_i\{q_{ii}\} > max_i\{\sum_{j\neq i}|q_{ij}|\} \tag{6.36}$$

then choosing Γ and B according to (6.33) is sufficient to ensure exponential stability of the BCP network described by (6.9).

Proof. It can be shown that any matrix $M \in \mathbb{R}^{n\times n}$ satisfies $\|M\| \leq \sqrt{\|M\|_1\|M\|_\infty}$, where $\|M\|_\infty \triangleq \max_i\{\sum_j |p_{ij}|\}$, $\|M\|_1 \triangleq \max_j\{\sum_i |p_{ij}|\}$, and $\|M\|$ is the spectral norm of M (Golub & Van Loan, 1989, Cor. 2.3.2). Furthermore, if M is symmetric this reduces to $\|M\| \leq \|M\|_\infty$. Thus

$$\max_i\{\sum_{j\neq i}|q_{ij}|\} = \|Q_o\|_\infty \geq \|Q_o\| = \lambda_{max}(Q_o)$$

which indicates that if (6.36) is satisfied, then so is (6.32), as required.

6.9 Preconditioning

If ΓQB satisfies (6.28), multiplying either Γ or B by a scalar constant $\alpha > 1$ increases the rate of convergence guaranteed by Theorem 6.2 by a factor of α, and hence — at least in theory — convergence of the neural network can be made arbitrarily fast by making α arbitrarily large. However, as seen in Section 5.4.3, practical considerations limit the utility of this strategy, and in such cases preconditioning should be used instead to improve the convergence rate subject to these limitations.

In Section 5.4 it was shown that preconditioning of the linear networks in Sections 5.2 and 5.3 for SRE and minimum-norm SRE optimisation can both mitigate the sensitivity of the equilibrium point to weight implementation and floating-point derivative evaluation errors and accelerate convergence to the equilibrium point. In this section the implications of these results for the nonlinear network for constrained quadratic optimisation are discussed.

The nonlinear network can be viewed as an essentially linear network in which the offdiagonal entries in the ith column of the state-feedback matrix are set to zero whenever the activation of the ith neuron saturates, and a constant term $c_{ji}\mu_i$ or $c_{ji}\nu_i$ — where c_{ji} is the (j,i)th element of C — added to the input of the jth neuron $(\forall j \neq i)$ in compensation. These "changes" are reversed when the ith neuron comes out of saturation. Thus for as long as the trajectory remains in the linear region of each of the neural

activation functions for example, the network can be treated as a linear network with state-feedback matrix $C + A$. On the other hand, while all neurons are operating strictly in their saturation regions, the network can be considered to be linear with effective feedback matrix A which is diagonal, indicating the temporary uncoupling of the neural activations.

The condition number of the effective state-feedback matrix varies[5] between those of $C + A = \Gamma QB$ and A. The relevant condition number from the point of view of sensitivity of the equilibrium point to weight implementation and derivative evaluation errors is that which applies in the region in which the true equilibrium point is located, which is not generally known *apriori*. Furthermore, for robust stability in the presence of weight implementation errors the perturbed system matrix $\widehat{\Gamma QB}$ must remain non-singular after the above modification for each of the regions of state-space through which the trajectory passes, which again is not usually known in advance. Thus in the absence of such prior knowledge, a general preconditioning strategy which improves the conditioning of both ΓQB and A — and hence all other intermediate effective state-feedback matrices — should be employed. This is achieved by the symmetric diagonal preconditioning strategy detailed in Section 5.4.

Using a singular value inequality for the matrix sum (Horn & Johnson, 1991, Thm 3.3.16) yields

$$\eta = \lambda_{min}(A) - \sigma_{max}(C) \leq \sigma_{min}(\Gamma QB) \tag{6.37}$$

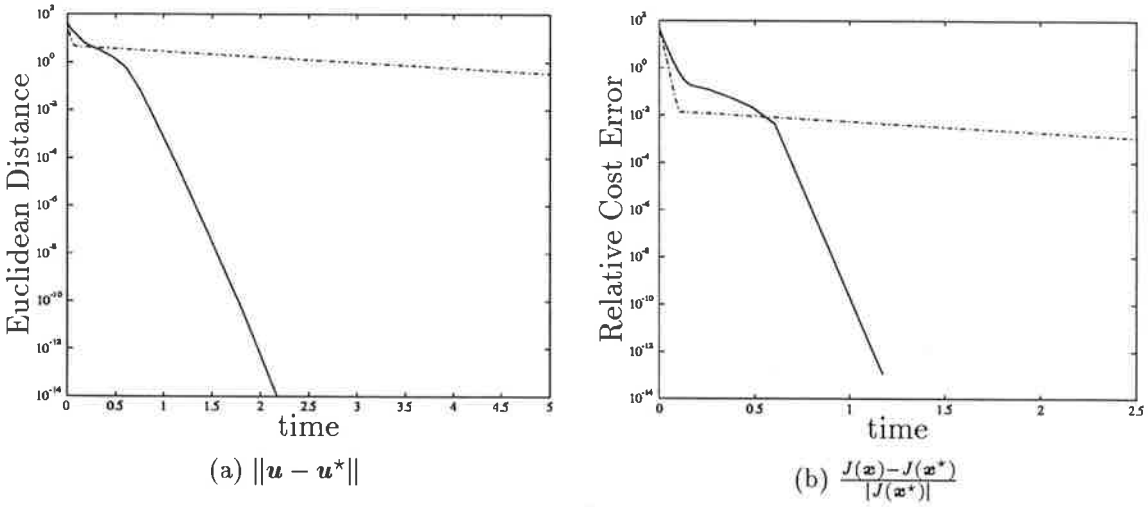
which provides an upper bound on the rate of exponential convergence guaranteed by Theorem 6.2. For the choice of diagonal preconditioners detailed in (5.13) with k chosen to preserve the maximum eigenvalue — and hence in view of symmetry the maximum singular value — as in Heuristic 5.2, this bound is guaranteed to exceed that of the unpreconditioned system whenever the preconditioning is effective in reducing the condition number of Q. However, except in the case of a 2×2 matrix with symmetric preconditioning — for which the equality in (6.37) can be shown to hold — increasing this upper bound on η does not ensure that either η or the degree of exponential stability will be increased.

Example 6.5 The effects of preconditioning on the condition number of the extreme effective state-feedback matrices $\Gamma QB = C + A$ and A for the matrix Q given in Table 6.4 of Example 6.3 are illustrated in Table 6.8. The preconditioners were chosen to satisfy (5.13), with the scaling factor $k = 4.633$ chosen in accordance with Heuristic 5.2 to ensure that the preconditioned matrix ΓQB has the same maximum eigenvalue as Q. The improvement in the condition number of both matrices indicates a reduced susceptibility to weight implementation and floating-point derivative evaluation errors. Since the eigenvalues of a symmetric positive semidefinite matrix are also its singular

[5] Unproven observation.

Preconditioning	Condition #		Eigenvalues		Degree $\geq$
$\Gamma = B$	$\kappa_2(\Gamma QB)$	$\kappa_2(A)$	$\lambda_{\max}(\Gamma QB)$	$\lambda_{\min}(\Gamma QB)$	η
I_n	300.0	74.62	40.00	0.1333	0.1333
$\begin{matrix} 0.7361 & 0 \\ 0 & 6.359 \end{matrix}$	13.69	1.000	40.00	2.922	2.922

Table 6.8: Effects of preconditioning on the condition number of the extreme effective state-feedback matrices $\Gamma QB = C + A$ and A for Example 6.3. The extreme eigenvalues of ΓQB and the lower bound η on the degree of convergence are also shown.



(a) $\|\boldsymbol{u} - \boldsymbol{u}^\star\|$ (b) $\frac{J(\boldsymbol{x}) - J(\boldsymbol{x}^\star)}{|J(\boldsymbol{x}^\star)|}$

Figure 6.8: (a) The Euclidean distance of the activation vector $\boldsymbol{u}$ from the equilibrium point $\boldsymbol{u}^\star$ as a function of time for the network in (6.9) applied to the problem in Example 6.3 both with (solid) and without (dot-dashed) preconditioning. (b) The corresponding relative cost error as a function of time.

values, the improvement by a factor in excess of 20 in the lower bound η on the degree of convergence which is also evident from Table 6.8 can be seen to be consistent with (6.37). As indicated by the plot of distance from the equilibrium point against time in Figure 6.8(a), the actual rate of convergence for the chosen starting points is also markedly accelerated by preconditioning, with an estimated increase by a factor of 40 in the slope on a log scale of the final segments of each trajectory. This same speed-up is also observed in Figure 6.8(b) for the monotonic decrease of the cost $J(\boldsymbol{x})$ to its optimal value. In contrast, the constrained steepest descent strategy employed by the network without preconditioning produces rapid initial reduction of the cost and the distance from the equilibrium point, but soon gives way to painfully slow convergence.

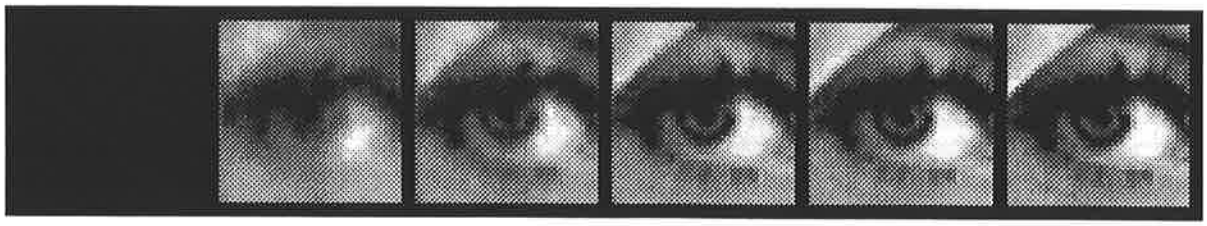
For a given application or implementation of the proposed recurrent neural network, there will of course be other considerations involved in the choice of the preconditioners.

For example, Γ and B can be used to control the dynamic range of internal network signals for a given externally imposed range of the variables $\boldsymbol{r}$ and $\boldsymbol{x}$. In such cases care should be taken not to worsten the conditioning of the system, and a trade-off between these possibly conflicting requirements may need to be investigated.

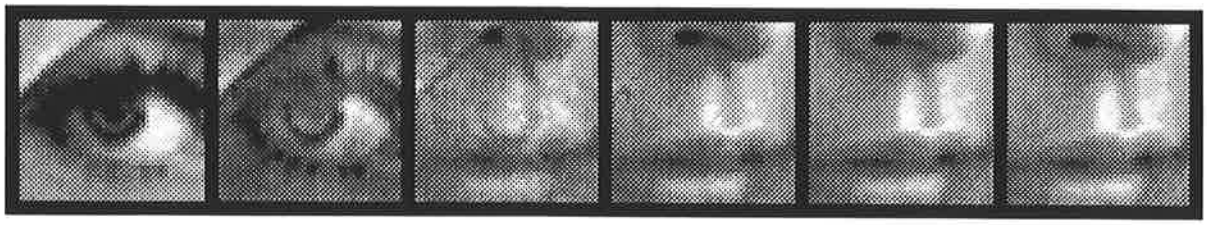
6.10 Bound Constrained SRE Minimisation

The goal of this chapter — to develop a neural network capable of solving the quadratic SRE minimisation problem arising from non-orthogonal image decomposition subject to bound constraints on the basis function coefficients imposed by the biological or electronic neural implementation medium — has now been achieved. The following example illustrates the effect on the reconstructed image of such constraints.

Example 6.6 The BCP neural network was applied to the regularised SRE image decomposition problem described in Example 5.1 upon which the three sets of bound constraints $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{-\mathbf{2}, \mathbf{2}\}, \{-\mathbf{0.1}, \mathbf{0.1}\}$ and $\{\mathbf{0}, \mathbf{10}\}$ were independently imposed. As in Example 5.1, due to the prior normalisation of the basis functions, no explicit preconditioning was required. The procedure used to obtain each of the three corresponding pairs of reconstructed image sequences depicted in figures 6.9, 6.10 and 6.11 respectively was the same as that used in Example 5.1. The near-complete convergence of the coefficients observed for each of the image sequences within 75ms — 3 time constants — of presentation of the image is consistent with exponential or near-exponential convergence. For both input images, the minimum regularised SRE coefficients for the unconstrained case in Example 5.1 exhibited a zero mean, approximately Gaussian distribution, with a standard deviation in the vicinity of 0.1 and all coefficients lying inside the range $(-2, 2)$. Consequently, the final image in each sequence reconstructed from the coefficients attained by the network having $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{-\mathbf{2}, \mathbf{2}\}$ shows no appreciable difference from the corresponding solution reached by the unconstrained network in Example 5.1; this observation is confirmed by the fact that for both image sequences, the two networks produce the same final SRE value. Constraint of the coefficients to the range $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{-\mathbf{0.1}, \mathbf{0.1}\}$ — which was exceeded by approximately 30% of the unconstrained coefficients — produced a barely noticeable effect on the image sequence, despite an almost 50% increase in the SRE for the image at $t = 1000$ms. The effect of the constraint of the coefficients to the range $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{\mathbf{0}, \mathbf{10}\}$ — which excludes approximately 50% of the unconstrained coefficients — is similarly difficult to discern from the corresponding image sequence, despite an almost 300% increase in the SRE for the image at $t = 1000$ms over the unconstrained case. Whilst these examples by no means constitute an exhaustive analysis of the effects of bound constraints on the coefficients, they do at least suggest that a great deal of the visual information in an image is retained despite such potentially restrictive constraints.

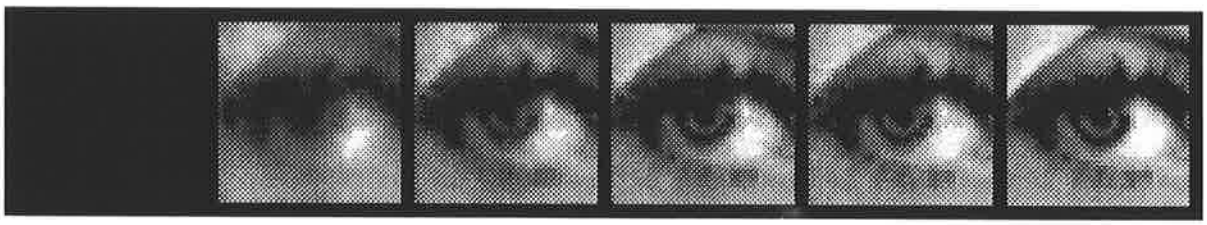


(a)

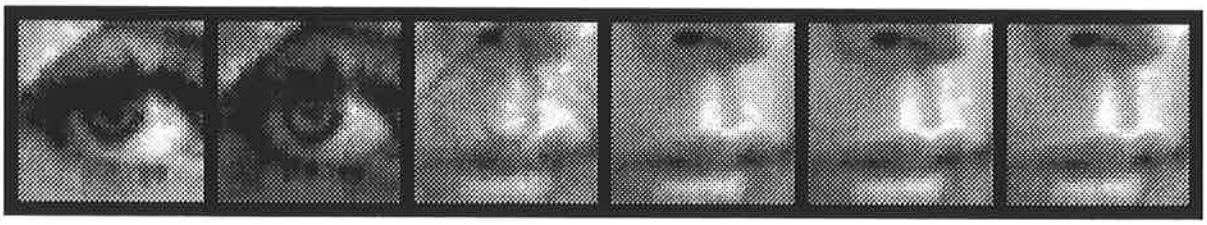


(b)

Figure 6.9: (a) Sequence of images reconstructed from the outputs of the BCP network with $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{-\mathbf{2}, \mathbf{2}\}$ at times $t = 0, 1, 10, 100$ and 1000ms after step presentation of the image of an eye (shown at right) to the network input. The network activation vector at $t = 0$ was $\boldsymbol{u}_0 = \mathbf{0}$. For ease of comparison, each reconstructed image was normalised to utilise the full available grey-scale range. (b) Sequence of reconstructed images at times $t = 0, 1, 10, 100$ and 1000ms after step presentation of the image of a mouth (shown at right) to the same network following equilibration on the eye in (a).



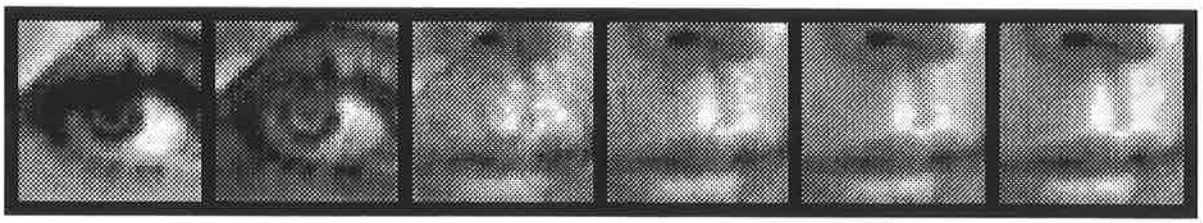
(a)



(b)

Figure 6.10: Sequence of reconstructed images for the BCP network with $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{-\mathbf{0.1}, \mathbf{0.1}\}$. All other details are the same as those for Figure 6.9.

(a)



(b)

Figure 6.11: Sequence of reconstructed images for the BCP network with $\{\boldsymbol{\mu}, \boldsymbol{\nu}\} = \{\mathbf{0}, \mathbf{10}\}$. All other details are the same as those for Figure 6.9.

6.11 Conclusion

In this chapter, the Bound Constraint Projection (BCP) network proposed by Bouzerdoum & Pattison (1993b) has been presented and shown to converge to the set of optima of the bound-constrained semidefinite quadratic optimisation (BCSQO) problem. The BCP network is a generalisation of the network of Sudharsanan & Sundareshan (1991), who considered the application of their network only to the unconstrained optimisation of a definite quadratic function. The extension of the requisite proofs presented by Bouzerdoum & Pattison (1993b) to cover the case where the matrix W is singular — and the network potentially quasiconvergent — necessitated a considerable strategic departure from the former treatment. The proof of exponential convergence — under certain conditions on the Hessian Q of the quadratic function and the preconditioners Γ, B — for the case where W is positive definite is however substantially unchanged from their paper. These proofs have superceded those of Sudharsanan & Sundareshan (1991) for global convergence and exponential stability, which have been shown here and by Davis & Pattison (1992) respectively to be flawed.

The optimisation strategy of the BCP network has been shown to be one of gradient descent with the search constrained to feasible directions and at times temporarily to one or more boundary faces of the constraint region. In contrast with penalty function methods — on which most previous neural network approaches to constrained linear and non-linear optimisation are based — which are guaranteed to produce feasible solutions only in the limit as the weighting of the constraint-violation penalty function tends to

infinity, the bound constraints are strictly enforced in the BCP network by the piecewise linear nodal activation function. The algorithm underlying this optimisation strategy differs from that of conventional serial methods for solution of the BCSQO problem — which require the sequential application of parallelisable matrix computations — in that it is inherently fully parallel. Ironically, those portions of such conventional algorithms which necessitate a sequential approach are those such as the Conjugate Gradient algorithm which were introduced to accelerate the optimisation process on serial architectures.

Comparison of the optimisation strategy of the BCP network with that of the Generalised Brain-State-in-a-Box (GBSB) network proposed by Golden (1992) revealed that the latter did not suffer from the unproductive temporary constraint of the search to one or more boundary faces of the constraint region exhibited by the former. In order to answer the potential objection to the BCP network that the nodal activations can become excessive, the Continuous-time Generalised Brain-State-in-a-Box (CGBSB) network was proposed as a continuous-time analogue of the GBSB, and trajectories starting in the feasible region were shown to converge to the set of solutions of the BCSQO problem. An analysis of the stability of the network after the proposed modification to force the return of the trajectory to the feasible region following a "glitsch" or infeasible initialisation awaits future investigation, as does a suitable analog implementation of the modified CGBSB network.

The BCP network was applied to the regularised squared reconstruction error (SRE) minimisation problem arising from the decomposition of an image using a set of Gabor basis functions. For the two example images, the quality of the image reconstructed from the coefficients obtained by the network was found to be robust to the imposition of restrictive bound constraints on the coefficients, suggesting that the nonlinear mapping from image to constrained coefficients implemented by the network at equilibrium may still transmit the vast majority of the visual information inherent in the input image. This hypothesis however awaits more thorough future investigation.

Chapter VII

DO THE SIMPLE CELLS PERFORM IMAGE DECOMPOSITION?

7.1 Introduction

Decomposition-based theories of simple cell processing postulate that each simple cell computes the coefficient corresponding to its own spatial RFP in an expansion of the visual image which uses these RFPs as expansion functions. Implicit in the definition of a coefficient employed by these theories is the notion that each cell signals the relative presence of its spatial RFP in the image. As indicated earlier in Section 4.1, the concept of the "relative presence" (Daugman, 1990) of an expansion function is effectively quantified by the choice of a reconstruction error measure, since the minimisation of this error measure subject to appropriate constraints uniquely defines a set of expansion coefficients. In Chapters 4–6, recurrent artificial neural networks (RANNs) were examined which solve, subject to these constraints, the least squared error (LSE) problem resulting from the use of the squared reconstruction error (SRE) measure. Minimisation of this measure is known to produce the best linear unbiassed estimator (BLUE) of the original image in the presence of zero-mean spherical Gaussian noise.

MacLennan (1993b), Pattison (1992) and Pece (1992) have argued that several of these RANNs could potentially be mapped onto the neural architecture of the early visual system. The potential for the early visual implementation of these networks raises the possibility that the simple cells may have at their disposal the computational machinery to compute, by relaxation, the coefficients of non-orthogonal image expansions such as those discussed in Chapter 3. In an investigation of this hypothesis, the proposed early visual implementation of each of these SRE-minimising neural networks is examined in Section 7.2 for its consistency with the neuroanatomy of the retino-geniculo-cortical pathway and with general principles of neural modelling and computation. Those implementations which are, according to these criteria, biologically plausible are examined further in Section 7.3 for consistency with the results of the RFP identification experiments of Jones & Palmer (1987b) and Palmer et al. (1991). A discussion and summary of the findings is then presented in Section 7.4.

The following treatment does not pretend to provide an exhaustive analysis of the plausibility of early visual implementations of SRE-minimising networks. However it will

be argued, on the basis of inconsistencies between the predicted and actual results of the reverse-correlation identification experiments of Jones & Palmer (1987b), that sufficient doubt exists to cause one to question the motivation behind the promulgation of these networks as models of early visual processing. In particular, why should the output of a simple cell be assumed to signal the relative presence of its spatial RFP in the visual image, rather than the spatial inner product of that RFP with the image, *in terms of which the spatial RFP is defined*? It is concluded that this decomposition hypothesis of simple cell processing, for which no apriori justification is evident, can be discounted as a vestige of the earlier feature-detection hypothesis, which was discussed briefly in Section 3.2.

7.2 Biological Implementation of Decomposition Networks

In this section, the early visual implementation of the recurrent neural networks for SRE or regularised SRE minimisation presented in Chapters 4–6 is investigated. Although most of these networks were not intended by their authors as models of early visual processing, this investigation is occasioned by the observations of MacLennan (1993b), Pattison (1992) and Pece (1992) that several of these RANNs might be amenable to early visual implementation. In order to assess the relative biological plausibility of the various models, it is first necessary to examine some principles of, and neurophysiological and neuroanatomical constraints on, neural and neural network modelling[1] (MacLennan, 1993b; Shepherd, 1990; Crick & Asanuma, 1986; Sejnowski, 1986), with particular emphasis on the functional architecture of the feline and primate early visual systems (Henry, 1991; Taylor, 1990; Thorpe & Imbert, 1989; White, 1989; Sereno, 1988; Lund, 1988; Gilbert, 1983; Poggio, 1980). Since this section combines disparate lines of reasoning to establish or refute the biological plausibility or otherwise of the various neural networks under consideration, it may appear somewhat disjointed on a first reading; in order to assist the reader, the principal messages of each sub-section are highlighted, where appropriate, in bold face type.

7.2.1 Plausibility Considerations

Orthodromic Transmission

Conventional models of neural information processing have been based on the *orthodromic* (feedforward) neural transmission of signals, with passive *electrotonic* spread of axo-dendritic synaptic input through the dendritic tree to the axon hillock via the soma,

[1]The term "neural modelling" is used here to refer to the modelling of the signal- and information-processing role of individual neurons, as opposed to the overall processing performed by the neural network of which it may be a part.

and active propagation along the axon to synapses located at each axon terminal. However cortical microcircuits are known to be considerably more complex than is suggested by this simplified model (see e.g. Shepherd (1990; 1988)). For example, the generation of an action potential at the axon hillock produces a wave of membrane depolarisation which spreads *antidromically* (i.e. backwards) into the dendritic tree (Shepherd, 1988), raising the possibility of antidromic as well as orthodromic signal transmission in neural circuits, such as would be required by the GBP network (Cohen & Shawe-Taylor, 1990) depicted in Figure 4.4. Nevertheless, the effective retrograde transmission of a signal across a synapse and antidromically along an axon, as is also required by this network, has yet to be documented in the cerebral cortex. **The GBP network is therefore not supported by the available electrophysiological evidence.**

Synaptic Model

The "strength" of a synaptic connection between two neurons is dependent on the physical properties of the neural membrane on either side of the synaptic cleft (which are in turn affected by the history of activity at that synapse). Many artificial neural networks employ a crude linear and usually short-term time-invariant model of synaptic transmission, in which synaptic strength is represented by a weight[2]. Since this weight therefore reflects membrane properties which are localised to the synaptic cleft, it cannot be directly signalled to other neurons. **The neural network models depicted in Figures 4.1, 4.3 and 4.4 are clearly implausible on this count.**

Neural Model

The linear model of synaptic transmission postulates that the pre-synaptic signal is transformed linearly into a post-synaptic trans-membrane current. This current adds to those established at nearby synaptic sites to produce a nett current which spreads electrotonically along the dendrite. Currents flowing in converging branches of the dendritic tree are progressively summed as they flow towards the soma. Each patch of neural membrane can to a first approximation be represented by an equivalent linear electrical circuit (Hodgkin & Huxley, 1952; Shepherd, 1988), according to which, in the absence of synaptic input, the dendritic membrane shunts some of this current, while the somatic membrane behaves as a leaky integrator for the remainder reaching the soma. Thus to the extent to which the linear model of synaptic transmission and the membrane equivalent electrical circuit can be said to reflect the processing actually performed by

[2]More realistic models of the dynamics of the post-synaptic membrane, in which the membrane conductance is controlled by synaptic input, lead under certain circumstances to multiplicative or shunting-inhibitory synaptic transmission (see e.g. Bouzerdoum (1991, Ch. 2) or Nabet & Pinter (1991, Ch. 3)). However there is evidence to suggest that shunting inhibition may not play a major role in shaping the responses of visual cortical cells (Douglas et al., 1988; Berman et al., 1991; Dehay et al., 1991).

the neuron, and to which the transmission losses along the dendrites can be modelled linearly and hence lumped into a modified synaptic weight, the dendro-somatic neural subsystem may be treated as a leaky summing integrator[3]. In particular, an artificial neural model consisting only of a summer is excluded by the above considerations on the basis of its (theoretically) infinitely short rise-time. Similarly, a neural model consisting only of a summing integrator lacks the leakage term arising from the membrane conductance required to prevent saturation of the somatic potential. **The networks depicted in Figures 4.1–4.7 are therefore implausible due to their use of summer or summing integrator neurons.**

Activation Function

In spiking neurons such as the simple cells, the somatic membrane potential is converted to a train of impulses, the frequency of which is usually assumed to encode the output of the dendro-somatic subsystem[4]. The relationship between the somatic potential — or more correctly the potential at the axon hillock — and the output firing frequency exhibits a generally monotonic increase in firing rate with input level, with a small amount of spontaneous activity in the absence of input, and saturation at high firing frequencies. The observation that firing frequency cannot be negative necessitates, in the case of the idealised linear neurons employed by the networks in Chapters 4 and 5, the postulation of cell pairs whose outputs signal the positive and negative halves of the linear activation function respectively. This scheme could be achieved by cell pairs having identical inputs but weights of opposite sign, as have been proposed and investigated by Pollen & Ronner (1981; 1982; 1983). However, **the neural activation function employed by the Bound Constraint Projection (BCP) network presented in Chapter 6 can be used to impose more realistically the constraints of non-negativity and saturation exhibited by the firing frequencies of real neurons.**

Although the linearity of this activation function between the available extremes may at first appear to be simply a convenient idealisation, there is some empirical evidence in favour of such a model. Bialek et al. (1991) showed that the horizontal velocity of a randomly-moving visual stimulus could be near-optimally (in the least-squares sense) reconstructed from the spike train of the blowfly H1 neuron using a linear filter, the impulse response of which resembles a typical synaptic impulse response function. In a realistic nonlinear model of spike generation, a linear filter was also shown to effect a reconstruction of the waveform of injected currents used to produce the simulated spike

[3]A number of other important factors have been ignored here for simplicity, including for example: the matter of relative synaptic arrival times of, and dendritic delays experienced by, different synaptic inputs; the presence of voltage-dependent membrane conductances; and the possibility of dendritic action potentials.

[4]although see the discussion in Section 7.2.3 for possible alternative neural signalling schemes.

train. Thus

> **In this view of computation with spike trains, the combination of nonlinearities in spike generation and the filter characteristics of the synapse results in an essentially linear transmission of analog signals from pre-synaptic cell bodies to post-synaptic dendrites.**

(Bialek et al., 1991). The lower and upper activation limits are, however, still necessary to represent the limits on the sign (Bialek et al., 1991) and maximum magnitude of the transmitted signal.

Dale's Law & Interneurons

An important question in assessing the biological validity of artificial neural network models concerns the likelihood of a single neuron exhibiting both excitatory and inhibitory influences on different post-synaptic target neurons. A much misunderstood (Shepherd, 1988) hypothesis in neurobiology, known as *Dale's Law* (1935), states that a given neuron will use the same neurotransmitter at all of its axon terminals. However, since at least some neurotransmitters have been found to exert excitatory and inhibitory influences on the post-synaptic membrane at different synapses depending on which receptors are present, Dale's hypothesis (which has come under fire in recent years) *cannot* be interpreted as meaning that a neuron must either excite all or inhibit all of its post-synaptic targets (Shepherd, 1988).

Nevertheless, **experimental evidence of cortical neurons which excite some cells and directly (mono-synaptically) inhibit others appears to be lacking** (Crick & Asanuma, 1986) or at least rare. Whilst the concept of disynaptic pathways mediated by abundant, locally connected *interneurons* can still be invoked to satisfy the need for mixed post-synaptic influences in biological implementations of the recurrent networks under consideration, Crick & Asanuma (1986) argue against neural connection schemes in which these interneurons simply change the sign of the signals they transmit. Additional delays introduced by the use of interneurons in the feedback pathways may furthermore have a deleterious effect on the stability of each network. In Appendix D.3 for example, it is demonstrated that the sufficient condition for exponential stability derived for the BCP network in Section 6.8 becomes (exponentially) more difficult to satisfy as a uniform delay on all lateral feedback connections is introduced and increased. The latter observation is indicative of the more general need to minimise transmission delays in the feedback paths of such networks, and suggests that **the use of interneurons to mediate the mixed post-synaptic influences exerted by neurons in the recurrent neural networks under consideration is likely to be strongly constrained by network stability requirements.**

Reciprocal Connections

In accordance with the Principle of Sloppy Workmanship (Huggins & Licklider, 1951; Grzywacz & Yuille, 1990), **networks employing reciprocal connections between neurons should not rely on a high degree of precision in the matching of either the weights or delays on these connections for their solution accuracy or stability**. The network models by Culhane et al. (1989) and Yan (1991b), for example, allow for the possibility of an imprecise match of reciprocal weights, and the stability analyses which have been presented for these networks can be readily extended to cover the single-layered networks presented in Chapter 5. The effects of weight errors on the precision of the solutions achieved by the latter networks, and the potential of preconditioning[5] to mitigate these effects, were examined in Chapter 5. Although symmetric preconditioning was employed for mathematical tractability throughout the preconditioning analysis, the network could be preconditioned asymmetrically, removing the requirement for symmetrically weighted reciprocal connections.

Image Decomposition

The results of Jones & Palmer (1987a) demonstrate that the GRGF provides a good fit to the spatial RFP of the simple cell[6], reflecting the relationship between the contrast of the visual stimulus and the response of the simple cell. Similarly, decomposition-based theories of simple cell processing postulate that the response of a given simple cell signals the relative presence of the spatial RFP of that cell in the visual stimulus. However, **the discrete-space formulation of the networks presented in Chapters 4–6 corresponds more naturally to the decomposition of the image formed by the outputs of the photoreceptor array**, which constitute the "pixels" of the photo-transduced image. In particular, since these networks clearly could not have direct access to the visual image prior to transduction, and since any attempted reconstruction of that image by a neural array would necessarily be performed in discrete-space, those governed by equations in which the image vector $\boldsymbol{i}$ or the image reconstruction $G^T\boldsymbol{a}$ appear explicitly must be viewed as attempting a *discrete-space* decomposition of the *photo-transduced* retinal image.

In contrast however, each element of the vector $G\boldsymbol{i}$ is the result of the discrete-space equivalent of the continuous-space inner product

$$\int_{\mathcal{V}} g_j(\boldsymbol{x}) i(\boldsymbol{x}, t)\, d\boldsymbol{x}$$

[5] combined with the simple scaling of the outputs of the preconditioned system as described in (5.8b).

[6] As was discussed in Chapter 2, this claim is dependent on the extent to which the spatial RFP of a cell having an inseparable spatiotemporal RFP can even be defined. Refer to Sections 2.2.4 and 7.3 for further discussion of this issue.

between the *feedforward* spatial weighting function[7] (FSWF) $g_j : \mathbb{R}^2 \rightarrow \mathbb{R}$ of an idealised simple cell and the spatiotemporal stimulus $i(\boldsymbol{x}, t)$. Similarly the elements of the matrix GG^T are the equivalents of the pairwise inner products

$$\int_{\mathcal{V}} g_j(\boldsymbol{x}) g_k(\boldsymbol{x})\, d\boldsymbol{x}$$

between the simple cell FSWFs. Networks governed by equations in which the input and reconstructed images appear only in expressions of the explicit form $G\boldsymbol{i}$ and $(GG^T + \varepsilon I_n)\boldsymbol{a}$ with $\varepsilon \geq 0$ can therefore be seen to have direct continuous-space analogues which are capable of decomposing the stimulus i using continuous-space expansion functions $\{g_j\}$. The linear recurrent network in (5.4), for example, can be reformulated by defining a vector $\boldsymbol{g} : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ whose elements are the FSWFs $\{g_j : j \in [1, n]\}$ of the individual neurons such that $\boldsymbol{g}(\boldsymbol{x}) \triangleq [g_1(\boldsymbol{x}), \ldots, g_n(\boldsymbol{x})]^T$. The elements Q_{jk} of the corresponding feedback matrix $Q \in \mathbb{R}^{n \times n}$ are then given by

$$Q_{jk} = \int_{\mathcal{V}} g_j(\boldsymbol{x}) g_k(\boldsymbol{x})\, d\boldsymbol{x} + \delta_{jk}\varepsilon \tag{7.1}$$

where δ_{jk} is the Kronecker delta function, and the resultant network is governed by

$$\dot{\boldsymbol{a}}(t) = -Q\boldsymbol{a}(t) + \int_{\mathcal{V}} \boldsymbol{g}(\boldsymbol{x}) i(\boldsymbol{x}, t)\, d\boldsymbol{x} \tag{7.2}$$

which, not surprisingly, bears a close resemblance to the original discrete-space formulation in (5.4). The BCP network in Chapter 6 is clearly amenable to continuous-space reformulation in the same manner. Thus **each of the single-layered recurrent networks in Chapters 5 and 6 admits a continuous-space formulation, whose neural implementation in the early visual system would permit the decomposition of the visual stimulus using the *continuous-space* simple cell FSWFs.**

7.2.2 Corticofugal Feedback

Cortico-Retinal Feedback

Since retinal ganglion cells have spatial RFPs which consist of concentric antagonistic "centre and surround" subfields, as opposed to the single-subfield RFPs of the photoreceptors, the outputs of the photoreceptors in the feline and primate visual systems have already been substantially modified by the time they are encoded as the outputs of the retinal ganglion cells, whose axons constitute the optic nerve. The vector $\boldsymbol{i}$ of unmodified photoreceptor outputs is therefore clearly only available in the retina. Thus if the expansion function coefficients for the image formed by the photoreceptor outputs are

[7]The FSWF of a cell is a weighting function which characterises the nett spatial processing performed by the feedforward pathway to that cell. If the feedback pathway were absent, the RFP would simply correspond to the FSWF of the cell. In the presence of feedback however, the RFP of a cell differs in general from the FSWF. This issue will be examined further in Section 7.3.

to be given by the simple cells, the implementation of the terms of the form $[G^T\boldsymbol{a} - \boldsymbol{i}]$ in Equations (4.19b), (4.20b) and (4.23b) within the visual system — for which they were not intended by their respective proponents — would require massive corticofugal feedback to the retina, which has not been observed. **The networks depicted in Figures 4.6–4.8 respectively are therefore incompatible with the neuroanatomy of the early visual system.**

Cortico-Geniculate Feedback

Corticofugal feedback *is* however known to occur in the primate (Lund, 1988) and feline visual systems from layer 6 of primary visual cortex to the LGN, and this observation was used by Pece (1992) as the basis for his mapping of the neural network depicted in Figure 4.9 onto the early visual system. The neurons in the first layer were identified with cells in the LGN, and those in the second with unspecified cells in V1. In contrast with the previous mapping, the network input was consequently identified with the nett feedforward input to the LGN, which, being derived from the outputs of retinal ganglion cells having centre-surround spatial RFPs, represents a significantly transformed version of the photoreceptor image. Pece (1992) argued that the opposite signs on the feedforward (geniculo-cortical) and feedback (cortico-geniculate) pathways were consistent with the observation by Murphy & Sillito (1987) that the "activity of LGN neurons is suppressed by stimuli which are optimal for excitation of V1 neurons". However this argument is difficult to sustain in the light of the model's prediction that an excited cortical neuron will inhibit some LGN cells but excite others[8]. Nevertheless, **the geniculo-cortical implementation proposed by Pece (1992) for his two-layered network remains an at least superficially attractive explanation for the massive and as yet somewhat neglected cortico-geniculate projection.**

Given that the GRGF model had earlier been successfully fitted by Jones & Palmer (1987a) to the spatial RFPs of simple cells, the specification of discrete-space GRGF weighting functions on the feedforward geniculo-cortical connections to the unidentified cortical cells suggests that these cells were intended to represent simple cells. However since each spatial RFP is by definition measured relative to the visual input, while each geniculo-cortical weighting function weights the outputs of the LGN, it remains to reconcile Pece's (1992) specification of the latter with the verification by Jones & Palmer (1987a) of the GRGF model for the former. Ignoring for the moment the effect of the cortico-geniculate feedback pathway, so that the predicted spatial RFP of the simple cell is given simply by its FSWF, a loose reconciliation is suggested by the following argument. If the LGN cell feedforward spatial weighting functions (FSWFs) are assumed

[8]More generally, the feedback influence suggested by this model on the network layer identified with the LGN takes the form of the reconstructed image, and consequently has approximately the same structure as the feedforward input to these cells from the retinal ganglion cells.

for simplicity to be identical up to a translation, the geniculo-cortical spatial weighting function is related to the simple cell FSWF by a discrete convolution with the LGN cell FSWF. The spatially low-pass afferent retino-geniculate processing might therefore be viewed as acting, at least approximately, as a spatial reconstruction filter (Carlson, 1986) for the discrete-space geniculo-cortical weighting functions; conversely, the geniculo-cortical weighting functions would, according to this approximation, be viewed as sampled, anti-aliased versions of the corresponding simple cell FSWFs.

This reconstruction filter approximation should however be treated with some caution, since the LGN cell FSWFs exhibit relatively poor low-pass behaviour in comparison with the ideal "brick wall" low-pass reconstruction filter, and, like the retinal ganglion cell spatial RFPs, are likely to vary quite considerably in size over the LGN cell population. More importantly, however, the above open-loop analysis also ignores the very real possibility that the cortico-geniculate feedback pathway might alter the simple cell spatial RFP. Thus even if the reconstruction filter approximation were to hold, **it remains to show that the simple cell spatial RFP is described by its corresponding GRGF FSWF**. This question is investigated in Section 7.3.

Since the geniculo-cortical weighting functions are given by the chosen discrete-space expansion functions, the fact that the network proposed by Pece (1992) solves the *augmented* normal equations for the decomposition problem is consistent with the observation that the simple cells significantly outnumber cells in the LGN (i.e. $n \gg m$). However, **an open question in the verification of the proposed implementation is the extent to which the inter-areal delay in the feedback path would jeopardise the network stability**. Although these connections can be reasonably presumed to mediate some form of visual feedback system which *is* stable, the critical question here is whether or not the *proposed* network could be stable in the presence of the associated inter-areal delays.

7.2.3 Intracortical Feedback

Daugman's Network

Illustrated in Figure 4.2, the corrected version of Daugman's (1988a) network shown in Figure 4.1 computes the difference $G\boldsymbol{i} - G[G^T\boldsymbol{a}]$ between the spatial inner products of the expansion functions with the original and reconstructed images, a computation which could potentially be performed entirely within V1. The small intra-areal delay associated with a cortical implementation offers a substantial reduction in feedback delay, and hence improvement in network stability, over the inter-areal pathways required by the networks discussed in the previous section. Nevertheless, **as a model of biological visual processing** — for which Daugman did *not* intend it — **Daugman's (1988a) network also predicts an explicit cortical reconstruction of the retinal image,**

for which there is as yet no evidence. Furthermore the regularisation term used in the formulation of the augmented normal equations necessitated by the estimated 1000:1 ratio of simple cells n to photoreceptors m (Wilson et al., 1990)[9] cannot be readily accommodated by the (corrected) Daugman architecture.

Pattison's Network

By identifying the neurons in Pattison's (1992) single-layered network depicted in Figure 5.1 (or the continuous-space equivalent thereof described by (7.2)) with the simple cells, the improved stability offered by an intracortical implementation of the feedback pathway can be exploited without involving an explicit reconstruction of the input image.

The GRGF weights on the feedforward connections to each neuron encapsulate the nett feedforward spatial processing performed by the retino-geniculo-cortical pathway, while the reciprocal lateral connections between a given pair of simple cells are weighted by the negative of the inner product between the GRGFs on the corresponding feedforward connections. The decay or self-feedback term for each neuron is simply increased by ε to implement the necessary regularisation of the normal equations. The grossly topological organisation and finite spatial support of the receptive fields of primary visual cortical cells should limit the number and spatial extent of the lateral connections required by the network model to well below those required for complete lateral interconnection. Although, as explained in Section 5.2, this network requires more connections in general than those of Culhane et al. (1989), Yan (1991b), Cichocki & Unbehauen (1992) and Pece (1992), intrastriate connections are likely to be metabolically cheaper than the inter-areal projections required by these networks. Thus the enhanced stability of the single-layered network is likely to be realised for little or no additional biological expense.

Evidence for the relatively dense lateral connectivity between simple cells predicted by Pattison's (1992) single-layered network will be examined shortly. In the mean time, the linear activation function required by this network will be replaced in the following section by the more realistic piecewise linear saturating activation function of the BCP network.

BCP Network

If the linear activation function of the single-layered networks presented in Chapter 5 is replaced by the piecewise linear saturating function depicted in Figure 6.2(b), the resultant networks can be seen to be special cases of the BCP network of Bouzerdoum

[9]The derivation of this estimate from the information provided by Wilson et al. (1990) is explained in Section 3.2.2.

& Pattison (1993b; 1993a) in which (ignoring preconditioning for the moment)

$$\begin{aligned} \boldsymbol{y} &= G\boldsymbol{i} \\ A &= \mathrm{diag}(GG^T) + \varepsilon I_n \\ C &= \mathrm{offdiag}(GG^T) \end{aligned}$$

with $\varepsilon \geq 0$. The saturation parameter ξ_i can be chosen to reflect the input level at which the output activity of a simple cell saturates, while since the linear portion of the activation function must lie on a line passing through the origin and the neural firing rate is by definition non-negative, ζ_i is ideally zero. Although simple cells exhibit remarkably little spontaneous activity, this model would require a minor correction[10] if a finite membrane hyperpolarisation (i.e. negative input) were required to silence a simple cell, or if the cell could remain silent in the presence of a finite membrane depolarisation (i.e. positive input).

Rather than requiring idealised *linear* nodes, whose implementation necessitates the invocation of paired simple cells with weights differing only by a sign change (Pollen & Ronner, 1981,1982,1983), to solve the *unconstrained* SRE minimisation problem posed by image decomposition, **an early visual implementation of the BCP network would use this more realistic *nonlinear* activation function to solve the decomposition problem *subject to constraints* imposed on the individual neural firing rates by their respective activation functions**. Examples of the solutions attained by this network under increasingly severe constraints were presented in Figures 6.9–6.11, and indicated that the effect of a positivity constraint, such as that imposed by pulse-frequency encoding of neural activation, is surprisingly mild[11]. As mentioned previously, the effect of uniform delays on the lateral connections of the BCP network incurred by the proposed cortical implementation is to make the sufficient condition for exponential stability, derived in Section 6.8 and extended in Appendix D.3, exponentially harder to satisfy, although this condition may not of course be necessary for global convergence.

Hebbian Weight Development

Since the feedforward and feedback terms of the proposed single-layered networks can be calculated using the continuous-space spatial RFP of each simple cell, these networks, when modified as detailed in Section 7.2 to use spatially continuous FSWFs, are capable of performing a decomposition of the visual image prior to transduction. The question of course arises as to how the spatial RFP of each cell, which encapsulates the nett

[10] viz. the subtraction of the vertical offset of the true activation function from the signal transmitted to all other neurons.

[11] A more extensive analysis would of course be required before any firm conclusions could be drawn from these preliminary observations. In particular, the effects of the choice of GRGFs, images, and relative numbers of neurons n and inputs m would need to be investigated.

spatial processing performed by the retino-geniculo-cortical pathway to that cell[12], could be explicitly available at the level of the primary visual cortex for the computation of the required inner products for the lateral connection weights. Since the early visual system undergoes significant developmental changes in the early postnatal period (see e.g. Honavar (1990) for a review), the lateral weights are unlikely to be hard-coded or genetically pre-specified, and must therefore result from some form of self-organisation.

A simple form of self-organisation is that resulting from the family of Hebbian synaptic learning rules, which are reviewed, along with their biological plausibility, by Brown et al. (1990). Drawing inspiration from a postulate by Hebb (1949) on the neural basis of learning, Hebbian learning rules are characterised by the adaptation of a synaptic weight according to an expression resembling the correlation or covariance between the pre- and post-synaptic activity. The potential for the *relationship* between the feedforward and feedback weights required by the network of Pattison (1992) to develop according to the particular form of Hebbian learning law proposed by Földiák (1989) is demonstrated in Appendix E. Whilst the mature network will not necessarily exhibit the required Gabor function weights on the feedforward connections, a number of Hebbian learning algorithms which develop orientation-sensitive and in some cases Gabor-like spatial RFPs in single-layered networks have been reported elsewhere (von der Malsburg, 1973; Linsker, 1986b; Yuille et al., 1989; Rubner & Schulten, 1990; Sanger, 1990). An indication that these RFPs are not simply artefacts of the somewhat unrealistic single-layered and mono-synaptic model of the retino-geniculo-cortical pathway employed by most competing models can be gained from the observation that Linsker's (1986c; 1986b; 1986a) results were obtained using a more realistic series of neural layers. Viewed together, these results are indicative of the possibility that **the necessary feedforward and feedback weights might develop through Hebbian self-organisation as a result of exposure to a statistically realistic approximation to the natural visual environment.**

Evidence for Intracortical Feedback

A literal interpretation of the networks of Pattison (1992) and Bouzerdoum & Pattison (1993a) as models of simple cell processing would require the presence of monosynaptic excitatory and inhibitory connections between simple cells. A looser interpretation would admit the possibility that the necessary interactions may also be mediated by di- or even poly-synaptic pathways involving possibly one or more other cortical layers. In accordance with this looser interpretation, the following analysis examines evidence in favour of any intrastriate feedback pathways connecting these cells.

Based on evidence from various sources (reviewed by e.g. White (1989, p. 141-148)

[12]including, at least notionally, the effect of any feedback interactions; an examination of the predicted effects of feedback interactions on the spatial RFP is presented in Section 7.3.

and Douglas & Martin (1991)), generic models of cortical circuitry incorporating massive intracortical feedback are nowadays commonplace. Specific models of simple cell processing include those postulating mutual inhibition between iso- or cross-oriented simple cells (or both) in order to sharpen their orientation tuning[13] (see e.g. Ferster & Koch (1987) or Wörgötter & Koch (1991) for a review). Nevertheless, *direct* evidence for feedback pathways between cells functionally identified as simple cells is hard to come by, and the following overview is therefore necessarily based on a number of indirect sources[14].

Simple cells in feline V1 receive monosynaptic excitation and disynaptic inhibition from cells in the LGN (Toyama et al., 1974; Ferster & Lindström, 1983; Martin & Whitteridge, 1984). The inhibitory input is presumed to be mediated by inhibitory interneurons, most likely the basket or clutch cells (Ferster & Koch, 1987). The geniculo-cortical afferents providing input to these interneurons appear to be specialised to produce rapid post-synaptic activation, and hence to expedite the resultant inhibitory input to excitatory simple cells (Douglas & Martin, 1991), such as the spiny stellate cells in layer 4 (White, 1989). As might be expected from this arrangement, a blockade of intracortical inhibition using an antagonist of the inhibitory neurotransmitter GABA leads to a disturbance of the receptive field properties of the simple cells (Sillito, 1975; Tsumoto et al., 1979; Sillito et al., 1980). However, the profound nature of this disturbance, including a loss or reduction of the distinction between on- and off-excitatory subfields, of direction sensitivity, and in some cases of orientation sensitivity (Sillito, 1975; Tsumoto et al., 1979; Sillito et al., 1980) is inconsistent with the simple withdrawal of inhibitory input to each subfield, since the excitatory input alone could for example retain both subfield segregation and orientation sensitivity. In particular, these observations suggest the involvement of input from cells whose inputs are themselves either directly or indirectly affected by the local intracortical blockade of inhibition, as proposed for example by Ferster & Koch (1987) in connection with the effects on orientation selectivity.

Having demonstrated the presence of spatially opponent inhibitory input to simple cells in feline V1, Ferster (1988) concluded that

> ... the most obvious candidates for the pre-synaptic inhibitory neurons are other simple cells.

in accordance with which Palmer & Davis (1981b) had earlier identified a number of adjacent simple cells having overlapping receptive fields whose corresponding subfields were of opposite excitatory type. A less direct feedback pathway between cells in layer 4,

[13]This topic is addressed briefly in Section 7.4.

[14]This should not however be interpreted as an indication that the required pathways are either weak or rare — the predictive power of the model by Wörgötter & Koch (1991) involving feedback between simple cells argues against this conclusion — but rather as evidence of both the lack of experiments designed to investigate this specific question, and the difficulty of piecing together a coherent picture from disparate sources such as functional, morphological and anatomical studies of the striate cortex.

in which the simple cells predominate, is via cells in layer 6 whose axons contribute to the cortico-geniculate projection, and collaterals of which produce monosynaptic excitation and disynaptic inhibition of cells in layer 4 (McGuire et al., 1984; Ferster & Lindström, 1985). However despite the existence of feedforward pathways from layer 4 to layer 6 via intervening layers (see e.g. Douglas & Martin (1991)), the presence of direct geniculate input to at least some cells in layer 6 (McGuire et al., 1984; Ferster & Lindström, 1985) makes it difficult to establish this as a purely feedback pathway as required by the proposed neural network models.

The most direct evidence obtained so far for feedback interactions between orientation-sensitive cells in the same cortical layer comes from experiments involving the cross-correlation of simultaneous electrophysiological recordings from pairs of cells in feline primary visual cortex. Using cross-correlation techniques, Hata et al. (1988) identified unidirectional, short latency (about 1.4ms), probably monosynaptic inhibition between 5 of 82 pairs of cells in layer 4, while Ts'o et al. (1986) reported unidirectional lateral excitation between cells in layers 2–3 showing similar orientation preference. The strength of the interaction was found in the latter case to vary approximately with the degree of receptive field overlap, as predicted broadly by the proposed single-layered decomposition networks, although excitation was also found between cell pairs having apparently non-overlapping receptive fields. However, whilst in the former study the recorded cells were in layer 4, in which simple cells predominate, neither group functionally identified the recorded cells as simple cells. Furthermore the lack of reciprocity in the lateral interactions inferred from these experiments is inconsistent with the proposed network models, although mutual inhibition was observed by Toyama et al. (1981) between cells in layer 4 exhibiting a single subfield[15]. **The relative infrequency of unequivocal instances of lateral connectivity (reciprocal or otherwise) between simple cells does little to dispel such doubts about the true cortical plausibility of these network models, which in general require dense lateral interconnections.**

Neural Iteration

The success of any relaxation algorithm relies on the propagation of the state vector with sufficient accuracy to avoid a catastrophic accumulation of errors, such as may occur in the presence of floating-point round-off. In the present investigation of the biological implementation of SRE minimisation networks, it has so far been assumed that each node in the network should be identified with a real neuron. However Marr (1982) and, more recently, MacLennan (1993a) have argued that if the mean firing frequency over a given interval is taken to signal the real-valued neural output, then a neural relaxation network would take an infeasibly long time to perform a small number of iterations with

[15]The investigators chose not to class these cells as simple since the lack of additional subfields hindered conventional distinctions between the functional classes.

even a modest degree of precision. In criticising the neural implementation proposed by Pattison (1992) for the unregularised ($\varepsilon = 0$) network in Chapter 5, MacLennan (1993a) argued in particular that the Weyl-Heisenberg ("Gabor") Uncertainty Principle in the temporal and temporal-frequency domains places a fundamental lower limit on the time taken to signal a coefficient to a specified accuracy using pulse-frequency coding. A review of the temporal constraints on this and related neural signalling mechanisms, and their implications for the feasibility of the neural implementation of relaxation algorithms in early vision has been presented by Thorpe & Imbert (1989).

However, these arguments and associated estimates of relaxation times are predicated on the unwarranted assumption that a neuron receiving the output pulse train must constantly *re-acquire* its estimate of the sending frequency. Under this assumption each acquisition, which delimits a notional "iteration" of the corresponding neural relaxation network (Thorpe & Imbert, 1989), starts *ab initio* and is therefore subject to the above trade-off between acquisition time and accuracy. In contradiction of this assumption, the duration of the impulse response function of the linear filter used by Bialek et al. (1991) to successfully decode a neural spike train in the visual system of the fly, which may be taken as the time required for acquisition of the encoded signal, was of the order of only 5–10 times the minimum interspike interval. According to Marr (1982), only 0.5–1 decimal digits of precision could be signalled in this interval, in apparent contrast with estimates by Bialek et al. (1991) of average information rates in insect sensory systems approaching 3 bits per spike, which would permit the transmission of 5–9 decimal digits (15–30 bits) in this same interval. This conflict is in part[16] resolved by the observation that once the signal has been acquired, it is thereafter only necessary to *track* it. Tracking allows the receiver of the spike train to maintain a relatively accurate estimate of the transmitted signal whilst requiring a considerably lower information rate to update this estimate than to continually re-acquire the signal. This tracking process would be facilitated by the relatively low temporal rate of change of the output imposed by the low-pass filtering effect of the proposed relaxation networks on the input.

The removal of the assumed need to constantly re-acquire the signal conveyed by the neural spike train renders impotent the objections raised by Marr (1982) and MacLennan (1993a) to the neural implementation of relaxation algorithms. In particular, the imposition of temporal constraints on the relaxation process does not entail the severe trade-off with the accuracy of transmission of the neural state vector alleged by MacLennan (1993a). Since similar assumptions concerning temporal constraints on neural signalling mechanisms underly many of the objections raised by Thorpe & Imbert (1989) to the

[16]The relatively high information rates reported by Bialek et al. (1991) probably also indicate the use of other parameters of the spike train, such as the relative timing of the spikes, to convey information in addition to that conveyed by the spiking frequency.

use of relaxation in early visual processing, these may be similarly discounted[17]. **These considerations therefore reopen the possibility that relaxation algorithms may indeed be implemented in the neural wetware of the early visual system**. We note in passing that Sejnowski (1986, p. 378-80) reached a similar conclusion concerning the plausibility of the neural implementation of stochastic relaxation algorithms.

Dendritic Networks

The relaxation algorithm in (5.1) proposed by Pattison (1992) for unregularised non-orthogonal image decomposition was developed independently by MacLennan (1993b). Whereas the former proposed a neural implementation of this algorithm, consisting of classical neurons communicating via axo-dendritically transmitted impulse trains, the latter suggested its implementation in a densely connected *dendritic network*, in which electrotonic communications are mediated by dendro-dendritic synapses. The principal argument presented by MacLennan (1993b) in favour of a dendritic implementation was that the use of electrotonic signalling would overcome the alleged timing problems associated with the acquisition[18] of signals transmitted using axonal spike trains. Since variables such as the displacement of the membrane potential from its resting value may be either positive or negative, the use of electrotonic communication furthermore suggests the possibility of avoiding the positivity constraint imposed on the coefficients of the corresponding neural implementation of the network — and hence on the SRE minimisation — by the presumed pulse-frequency signalling mechanism. According to decomposition-based theories of simple cell processing, however, each of the coefficients produced by the relaxation process must ultimately be translated into the output neural spike train of the corresponding simple cell. The positivity constraints, which are consequently inevitable under the presumed method of output signalling, should therefore be imposed on the SRE minimisation rather than on the coefficients of an unconstrained optimisation performed using electrotonic signalling.

Although the general feasibility of dendritic computation was discussed at length by MacLennan (1993b), neither the functional classification of the neurons whose dendrites

[17]Even assuming an ideal analog signalling mechanism however, the demonstration of short-latency neural and behavioural responses which are dependent on the classification or recognition of the visual stimulus seems to place strong constraints on the implementation of iterative algorithms in early visual processing (Thorpe & Imbert, 1989), since the *initiation* of these responses must be based on short-latency, crude early reponses of such algorithms. The possibility that a given classification or recognition, or the confidence attached thereto, might be dynamically refined as more accurate information becomes available as a result of subsequent "iteration" is not, however, precluded by such observations. The exponential convergence of the proposed networks with neural time constants of the order of 20ms (Stratford et al., 1989), in comparison with initial latencies in excess of 100ms (Thorpe & Imbert, 1989), may be fast enough to render such refinements relatively transparent to the experimental techniques used in these experiments.

[18]As was noted in the previous section however, the important distinction between signal acquisition and tracking was not noted by MacLennan (1993b).

might contribute to this network, nor the specific biophysical parameters to be identified with its state variables, were specifically identified. In lieu of the former, the dendrites of the proposed network are henceforth tentatively assumed to belong to the simple cells. Nevertheless, **a full assessment of the biological plausibility of the proposed dendritic network, and the extension of this implementation to incorporate the regularisation term used in (5.3), await a more detailed specification of the network's neuroanatomical and biophysical substrate.**

7.2.4 Summary

Hypothetical early visual implementations of the networks proposed in Chapters 4–6 for the decomposition of the visual image using non-orthogonal expansion functions have been considered. Some networks were found to be clearly implausible, requiring the communication of synaptic weights and, in the case of the GBP network, bidirectional synaptic and axonal signal transmission. Of the remainder, the network implementation proposed by Pece (1992) as a possible explanation for the prominent corticofugal feedback projection was found to be at odds with one of the electrophysiological observations it was intended to explain. Some (albeit mostly indirect) evidence was found in support of the direct or indirect interconnection of simple cells, as required by the linear networks presented in Chapter 5 and the BCP network of Bouzerdoum & Pattison (1993b). It was furthermore argued that the nonlinear activation function of the BCP network may represent a reasonably accurate characterisation of the nett processing performed by the cascaded mechanisms of spike generation and synaptic transmission. The BCP network is therefore the most plausible candidate among those networks proposed for *neural* implementation in the primary visual cortex. On the other hand, with the simple addition of the necessary regularisation term the *dendritic* implementation suggested by MacLennan (1993b) for the linear unregularised network in Chapter 5, proposed independently by MacLennan (1993b) and Pattison (1992), is potentially the most plausible implementation of the *linear* regularised network also presented in Chapter 5. Nevertheless, large residual uncertainties concerning the existence of the appropriate neural or dendritic connectivity, the suitability of the neural models employed, and the modes of interneural signalling suggest that caution should be exercised in attempting to draw firm conclusions from these tentative findings.

Amongst the networks amenable to more plausible neural or dendritic implementations, different networks were found to effect the decomposition of the "image" formed at different stages of the visual pathway, ranging from the raw image prior to optical blurring, to the output of the retinal ganglion cells. However, the attempts by Pattison (1992), Pece (1992) and MacLennan (1993b) to relate these networks to the early visual system were only necessitated in the first place by their adherence to the decomposition

hypothesis of simple cell processing, according to which the expansion functions used in the decomposition are the spatial RFPs of the simple cells, which are by definition referred to the raw visual image. However attractive one might find the neural implementation of networks, such as that proposed by Pece (1992), which use the simple cell spatial RFP to decompose anything other than the raw image, these networks therefore fail to satisfy their very *raison d'être* as models of early visual processing.

7.3 Predicted and Identified Spatiotemporal RFPs

7.3.1 Introduction

Decomposition-based models of simple cell processing postulate that each simple cell signals the coefficient of its own spatial RFP in an expansion of the visual image which uses these RFPs as expansion functions. However, in Chapter 3 it was shown that except in cases where the simple cell spatial RFPs are mutually orthogonal or collectively form a tight frame, the computation of the required coefficients cannot be achieved using the spatial inner product of the image with the RFPs, in terms of which each spatial RFP is defined. A potential resolution of this dilemma is suggested by the analysis in the previous section, where it was argued that the simple cells could be tentatively identified with the output nodes of one of several different recurrent neural networks, each of which is capable of finding the minimum SRE (MSRE) expansion coefficients and amenable to implementation in the wetware of the early visual system. Among the many questions to be answered in assessing the consistency of this solution with the available electrophysiological evidence is the following:

> Can the spatial RFPs identified by Jones & Palmer (1987b) be taken to be the expansion functions used by an early visual implementation of the proposed SRE minimisation networks?

or conversely

> Would the reverse-correlation identification technique of Jones & Palmer (1987b) correctly identify the expansion functions used by these networks?

The latter statement of this question reflects the approach to be pursued in this section to the assessment of the electrophysiological plausibility of the proposed relaxation networks.

In this section it is shown that contrary to the purely spatial description of the simple cells assumed by the very theories which give rise to their formulation, the postulated relaxation networks would, if implemented in the early visual system, produce *spatiotemporal* simple cell behaviour. Cells in the proposed networks are furthermore shown to have

spatiotemporally inseparable RFPs, which strictly speaking renders the concept of a spatial RFP inapplicable. Thus if these predicted spatiotemporal RFPs are to be reconciled with the "spatial RFPs" identified and modelled by Jones & Palmer (1987b), the results of the reverse-correlation identification techniques used by Jones & Palmer (1987b) and Palmer et al. (1991) must therefore be interpreted in the light of the predictions of the model.

In Appendix B it was shown that provided the stimuli and the output time bins approximate spatiotemporal and temporal impulses respectively, reverse-correlation analysis identifies the simple cell spatiotemporal RFP to within a spatial DC term. Although this analysis assumed the special case of spatiotemporally separable RFPs, it is readily extended to cover the generic form of the simple cell spatiotemporal RFP predicted by the proposed relaxation networks. The effects of the finite-sized rectangular spatiotemporal stimuli and temporal bins used in these experiments on the estimation of the true spatiotemporal RFP are examined in Section 7.3.3. To permit a preliminary comparison of the identified "spatial RFP" with the predicted spatiotemporal RFP in the mean time however, these identification experiments are naively assumed in Section 7.3.2 to yield temporal slices through the spatiotemporal RFP of each recorded simple cell at the end of each post-stimulus time bin. Each such section is henceforth referred to as the *instantaneous* spatial RFP of the cell at the corresponding sampling instant. The unwanted spatial DC term is ignored in order to simplify the ensuing discussion and thereby focus attention on the critical issues.

7.3.2 Theoretical RFPs

In this section, analytic expressions are derived for the spatiotemporal RFPs of the output neurons of the one- and two-layered recurrent neural networks in Section 4.4 and Chapter 5 respectively, whose implementations in the early visual system were deemed in Section 7.2 to be, broadly speaking, biologically plausible. Notwithstanding the differences discussed in Section 7.2, the dendritic implementation proposed by MacLennan (1993b) for the networks in Chapter 5 is, under the idealisations inherent in a purely linear analog model, formally equivalent to the neural implementation proposed by Pattison (1992), and is therefore not addressed separately. Furthemore, since analytic expressions cannot be obtained for RFPs in the BCP network, and the domain of validity of simulation studies would be restricted to the particular distributions of spatial GRGF parameters used in the simulations, analysis is restricted to the case of neurons having linear activation functions. For generic networks of both the one- and two-layered types, the instantaneous spatial component of each predicted spatiotemporal RFP is then compared for all post-stimulus times with the expansion function whose coefficient is purportedly signalled by the corresponding neuron. Speculations concerning the BCP

network are based on the approximate analogy with the corresponding linear network.

One-Layered Networks

In Section 7.2, the output neurons of the single-layered linear recurrent neural networks presented in Chapter 5 were tentatively identified with the simple cells. Their FSWFs, given by the corresponding rows of the matrix G or elements g_j of the function $\boldsymbol{g}$, were accordingly taken to represent linear approximations to the nett spatial processing performed by the afferent retino-geniculo-cortical pathway. The response $\boldsymbol{a}(t)$ of the linear recurrent network in (7.2) to a spatiotemporal stimulus $s(\boldsymbol{x}, t)$ is given by

$$\boldsymbol{a}(t) = \int_{-\infty}^{t} e^{-Q(t-\tau)} \int_{\mathcal{V}} \boldsymbol{g}(\boldsymbol{x}) s(\boldsymbol{x}, \tau)\, d\boldsymbol{x}\, d\tau \tag{7.3}$$

Defining the combined spatiotemporal RFP $\boldsymbol{w}: \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^n$ to be the vector whose elements are the spatiotemporal RFPs $w_j: \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ of the individual simple cells[19], comparison of (7.3) with (2.2) reveals that

$$\boldsymbol{w}(\boldsymbol{x}, t) = e^{-Qt} \boldsymbol{g}(\boldsymbol{x}) \tag{7.4}$$

Assuming for simplicity that Q is diagonalisable with discrete eigenvalues $\lambda_i \in \mathbb{R}$ (since Q is symmetric) and corresponding eigenvectors $\boldsymbol{e}_i$, the spatiotemporal RFP can be expressed as

$$\boldsymbol{w}(\boldsymbol{x}, t) = \sum_{i=1}^{n} e^{-\lambda_i t} \cdot \{[\boldsymbol{e}_i \boldsymbol{e}_i^T]\, \boldsymbol{g}(\boldsymbol{x})\} \tag{7.5}$$

revealing that the RFP of any given cell is a sum of spatiotemporally separable terms, a property it shares for example with the linear quadrature model of motion processing proposed by Watson & Ahumada (1983). The interesting spatiotemporal properties of the linear quadrature model, which were outlined in Section 2.2.4, engender initial confidence in the possibility that the spatiotemporal RFP of a real simple cell might be generated by the proposed single-layered network.

The spatial component of each term in the expansion in (7.5) consists of a linear combination of the FSWFs g_j of each neuron, while the temporal component is a simple exponential decay[20]. Thus whilst the instantaneous spatial RFP of the jth cell at time $t = 0$ is given simply by the jth expansion function g_j, as required, this spatial RFP changes form as the various terms in (7.5) decay at different rates. Furthermore, since

[19]The combined RFP $\boldsymbol{w}$ is closely related to the spatiotemporal impulse response function of the network, the only difference being a change of sign of the spatial variable consistent with the spatial correlation (as opposed to convolution) used in (7.3).

[20]In the case of repeated eigenvalues, temporal components of the form $t^k e^{-\lambda_i t}$, where $k \in [0, p_i]$ and p_i is the multiplicity of λ_i, are also introduced.

the end of the first 50ms post-stimulus time bin[21] — from which Jones & Palmer (1987b) obtained the spatial RFPs to which the GRGF model was successfully fitted — occurred 50ms after stimulus presentation, a period which is long compared with estimated neural time constants of the order of 20ms (Stratford et al., 1989), the instantaneous spatial RFP identified by Jones & Palmer (1987b) could be expected to differ significantly from the predicted instantaneous RFP at $t = 0$.

Nevertheless, if the chosen expansion functions are linearly dependent[22], the possibility arises that at one or more additional post-stimulus times $t > 0$, the expansion functions might be *re-synthesised* from non-trivial (i.e. $e^{-Qt} \not\propto I_n$) linear combinations of the others to give the required instantaneous spatial RFPs. However, the additional constraint that one such time must coincide at least approximately with the end of the time bin used by Jones & Palmer (1987b) makes this prospect similarly unattractive.

Two-Layered Networks

The simple cell spatiotemporal RFPs predicted by the two-layered networks in Section 4.4 are derived in the same way as those for the single-layered networks. The requisite analysis is outlined below for the network proposed by Pece (1992), which is governed by (4.26); extension of this analysis to the network by Yan (1991b) is trivial. In accordance with Pece's (1992) mapping of this network onto the early visual system, as discussed in Section 7.2, (4.26) can be reformulated in continuous space and using matrix notation to yield

$$\begin{bmatrix} \tau_a I_n & 0_{nm} \\ 0_{nm}^T & \tau_c I_m \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{a}} \\ \dot{\boldsymbol{c}} \end{bmatrix} = - \begin{bmatrix} I_n & -k_a G \\ k_c G^T & I_m \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{c} \end{bmatrix} + \int_{\mathcal{V}} \begin{bmatrix} \mathbf{0}_n \\ \boldsymbol{r}(\boldsymbol{x}) \end{bmatrix} s(\boldsymbol{x}, t)\, d\boldsymbol{x}$$

where $\boldsymbol{r}(\boldsymbol{x}) \triangleq [r_1(\boldsymbol{x}), \ldots, r_n(\boldsymbol{x})]^T$ is the vector of LGN cell FSWFs $r_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, 0_{nm} is the $n \times m$ zero matrix, and $\mathbf{0}_n$ is the n-dimensional zero vector. The remaining notation was described earlier in connection with (4.26), with the exception of the gain factor $k_c \in \mathbb{R}_+$, which has been added to make apparent the extension of this analysis to the network of Yan (1991b). The resultant spatiotemporal RFP $\boldsymbol{w}(\boldsymbol{x}, t)$ is then given by

$$\boldsymbol{w}(\boldsymbol{x}, t) = \frac{1}{\tau_c} e^{-Pt} \begin{bmatrix} \mathbf{0}_n \\ \boldsymbol{r}(\boldsymbol{x}) \end{bmatrix} \tag{7.6}$$

[21]Since the approximately 50ms latency of simple cell response to retinal stimulation is not explicitly modelled in the proposed networks, this latency is henceforth subtracted from all post-stimulus time measurements reported by Jones & Palmer (1987b) in order to facilitate comparisons with the temporal predictions of these networks. Accordingly, the numbering of the post-stimulus time bins has been adjusted such that the "first" bin from the point of view of these networks is that spanning the practical post-stimulus interval 50–100ms.

[22]as would for example be the case if these functions form a frame for the set of square-integrable real-valued functions.

where

$$P \triangleq \begin{bmatrix} \frac{1}{\tau_a} I_n & -\frac{k_a}{\tau_a} G \\ \frac{k_c}{\tau_c} G^T & \frac{1}{\tau_c} I_m \end{bmatrix}$$

with the first n elements of $\boldsymbol{w}$ giving the simple cell spatiotemporal RFPs. By analogy with the case of the one-layered network, the instantaneous spatial RFP predicted for each simple cell by the two-layered model is therefore a time-varying linear combination of the LGN cell FSWFs. The coefficients of this linear combination are given by the corresponding row of the instantaneous geniculo-cortical weight matrix $M(t) \in \mathbb{R}^{n \times m}$, defined as the top right-hand $n \times m$ block of the matrix e^{-Pt}. Consistent with observations indicating a latency of simple cell response to retinal stimulation (see e.g. Ikeda & Wright (1975b); Hamilton et al. (1989)), the instantaneous simple cell spatial RFPs predicted by the generic two-layered network are zero at time $t = 0$.

Implicit in Pece's (1992) choice of a discrete-space GRGF for the geniculo-cortical weight vector of each simple cell — given for the jth cell by the jth row of the geniculo-cortical weight matrix G — is the assumption that this weight vector is related to the cell's spatial RFP $g_j(\boldsymbol{x})$ by anti-aliased spatial sampling. It was argued in Section 7.2.2 that if the time dependence of the simple cell spatial RFP could be ignored, this assumption could be reconciled with the observations of Jones & Palmer (1987b; 1987a) provided that the LGN cell FSWFs collectively approximated a spatial reconstruction filter, such that

$$\boldsymbol{g}(\boldsymbol{x}) \approx G\boldsymbol{r}(\boldsymbol{x}) \tag{7.7}$$

However, the hypothesised early visual implementation of the proposed two-layered network would predict that both the instantaneous spatial RFP and the instantaneous geniculo-cortical weighting function of each simple cell should be time-varying, so that their respective temporal behaviours cannot be ignored. The spatial-domain identification techniques used by Jones & Palmer (1987b; 1987a) to test the GRGF model of the simple cell spatial RFP have been assumed, for the present purposes, to identify the instantaneous spatial RFP of the simple cell (at the end of the corresponding time bin). In order to reconcile Pece's (1992) specification of the geniculo-cortical weight matrix with the identified simple cell *instantaneous* spatial RFPs, conditions must therefore be established under which the instantaneous geniculo-cortical weight matrix $M(t)$ satisfies

$$\boldsymbol{g}(\boldsymbol{x}) = M(t)\boldsymbol{r}(\boldsymbol{x}) \tag{7.8}$$

at one or more post-stimulus times t.

In general, the solution of (7.8) for the required instantaneous geniculo-cortical weight matrix would in turn require the identification of the FSWF of each contributing cell in the LGN. However, this requirement is obviated if the instantaneously measured simple cell RFP can be assumed, at least approximately, to be related to the instantaneous

geniculo-cortical spatial weighting function by anti-aliased spatial sampling. Notwithstanding the objections to this approximation presented in Section 7.2.2, this relationship simplifies the necessary reconcilitation, since a comparison of Equations (7.7) and (7.8) reveals that it then only remains to establish conditions under which $M(t) \approx G$ for some time or times t. An examination of (7.6) and the matrix P reveals that a sufficient condition for equality to hold is that

$$e^{-Pt} \propto -P \tag{7.9}$$

with a positive constant of proportionality. The ease or otherwise with which this condition can be satisfied will depend on the particular choice of the basis functions $\{g_j\}$ via a dependence on the matrix G. In general however, the matrix P is positive definite (Yan, 1991b) and approximately anti-symmetric (provided $\frac{k_a}{\tau_a} \approx \frac{k_c}{\tau_c}$); if it is also diagonalisable with distinct eigenvalues, e^{-Pt} is characterised by exponentially damped oscillatory modes. Thus in order to produce the required negative definite result, these modes must all be negative at time t, a condition which may prove difficult to satisfy. This difficulty would once again be compounded by the additional constraint that one such time must coincide at least approximately with the end of the time bin used by Jones & Palmer (1987b).

Summary and Discussion

The weights in the biologically plausible networks proposed by MacLennan (1993b), Pattison (1992) and Pece (1992) were chosen according to the GRGF model of the simple cell spatial RFP. However, this model was validated by Jones & Palmer (1987b; 1987a) using reverse-correlation identification techniques which yield an approximation to the *instantaneous* spatial RFP of each simple cell at the end of the corresponding time bin. For those cells in each network hypothetically identified with the simple cells, conditions have therefore been examined under which the instantaneous spatial RFP is given by the chosen GRGF expansion functions.

It has been shown that the simple cell instantaneous spatial RFP predicted by the generic one-layered network for the instant $t = 0$ is the corresponding expansion function. However it was argued that the instantaneous spatial RFP obtained by Jones & Palmer (1987b) at the end of the first post-stimulus time bin, which occurs approximately 1 neural time constant later, is likely to differ significantly from that which would be measured at $t = 0$. It was furthermore concluded that for both the one- and two-layered networks, the simple cell instantaneous spatial RFPs at the end of the first time bin are unlikely to coincide with the chosen expansion functions. It therefore appears improbable that the instantaneous spatial RFPs identified experimentally by Jones & Palmer (1987b) result from an early visual implementation of either generic network using these functions as expansion functions.

To prove the negative case however — viz. that the instantaneous spatial RFPs identified experimentally by Jones & Palmer (1987b) *do not* result from an early visual implementation of either generic network using these functions as expansion functions — would in general require the exhaustive elimination of all alternatives. Efforts to this end are hampered by the unavailability of important information such as a detailed knowledge of the population distribution of simple cell RFP parameters. The above conclusions are sufficient, however, to suggest the need for a fundamental re-evaluation of the decomposition hypothesis of simple cell processing, a re-evaluation of its very motivation which should be undertaken before any attempt is made to disprove it through exhaustive experimentation. This matter will be discussed further in subsequent sections following an attempt to eliminate non-ideal aspects of the reverse correlation identification technique as potential sources of the noted discrepancies.

7.3.3 Practical Identification

The arguments presented in the previous section are based on a comparison between the spatiotemporal RFPs predicted theoretically by the relaxation networks proposed in Chapter 5, and those actually identified in real simple cells by the experiments of Jones & Palmer (1987b) and Palmer et al. (1991). It remains to address the possibility that approximations inherent in the reverse-correlation identification technique used by these investigators can account for the expected discrepancies between the predicted and identified RFPs.

An analysis of the reverse-correlation technique is presented in Appendix B for the ideal case where the stimuli can be assumed to approximate spatiotemporal impulses and the output of a cell is given by (B.1). Under the additional assumptions that the third and higher odd-order terms of the Taylor series expansion of the nonlinear function f are negligible, which is clearly true for the proposed *linear* relaxation networks, and that the DC offset of the spatial component of the spatiotemporally separable RFP is negligible, it was concluded that the spatiotemporal RFP estimate obtained by this technique would be the ideal RFP. This conclusion is furthermore unaffected by a simple extension of the analysis to the present case, in which each of the spatiotemporal RFPs in Equations (7.4) and (7.6) is given by a sum of spatiotemporally separable terms. As indicated by (B.6), the effect of a non-zero DC offset in the spatial component of the ideal RFP is simply to introduce an additional time-varying DC offset in the identified spatial component.

However the true stimuli used by Jones & Palmer (1987b) and Palmer et al. (1991) are rather poor approximations to spatiotemporal impulses. This is especially true in the temporal domain, where the stimulus duration of 50ms is large relative to reported time constants of the order of 20ms for cortical pyramidal cells (Stratford et al., 1989), and to a lesser extent in the spatial domain, where the stimulus elongation in the direction of

subfield elongation is up to $\frac{3}{16}$ of the width of the receptive field. A revision of the analysis in Appendix B using rectangular stimuli of constant size and assuming a strictly linear system, such as the proposed relaxation network, reveals that the effect of these non-ideal stimuli on the result of the reverse correlation identification technique is to filter the true RFP (plus parasitic terms) with a rectangular filter given by the stimulus[23]. The temporal binning procedure used in the estimation of the neural firing rate results in additional temporal smoothing, with the corresponding smoothing filter having a rectangular ("box-car") impulse response function whose width is that of the temporal bins. To reduce the computational effort, the identification procedure calculates, or in effect samples, the estimate of the smoothed spatiotemporal RFP in the temporal domain at the ends of these bins, and in the spatial domain at the centres of the divisions of the spatial stimulus grid.

In a simplified analysis of the application of the reverse correlation technique used by Jones & Palmer (1987b) to the un-regularised network presented in Chapter 5, Pattison (1992, App. B) approximated the temporal smoothing effect of a rectangular stimulus as a temporal integration. Since the RFP is simply the (spatially reversed) spatiotemporal impulse response function of the neuron, its integration yielded the corresponding temporal step response function. This step response approximation is valid for times t not exceeding the duration of the stimulus pulse, so that Pattison's (1992) analysis is valid only over this interval; for times exceeding this interval, the neural response — and hence spatial RFP — should be derived using the true rectangular stimulus instead of the temporal step function approximation. The step response was then temporally smoothed to account for the 50ms time bin used by Jones & Palmer (1987b), and sampled at the end of the first post-stimulus time bin to obtain an estimate of the spatial RFP[24]. Since the transient components of the step response had time constants of the order of half the width of the time-bin integration period their effect on the predicted spatial RFP was assumed to be largely overshadowed by the steady state component, which is given by the corresponding row of the expression

$$Q^{-1}\boldsymbol{g}(\boldsymbol{x}) \tag{7.10}$$

where $Q \in \mathbb{R}^{n\times n}$ is defined in (7.1). However, this steady-state assumption was based on the neural time constants, whose inverses are given by the diagonal entries of the matrix Q. The time constants of the transient components of the network step response are in fact given by the inverse eigenvalues of Q, which may differ from the neural time constants

[23]The technique used by Jones & Palmer (1987b) and Palmer et al. (1991) differs slightly from the ideal reverse correlation technique, described for example by Victor (1992) — for which the resultant smoothing filter is given by the autocorrelation of the stimulus — in that the response to the actual stimulus is in effect correlated with the idealised (impulsive) stimulus rather than with the actual stimulus.

[24]In practice, smoothing and sampling were effected simultaneously through an integration of the step response over the first post-stimulus time-bin.

by orders of magnitude, so that the steady-state assumption may in some cases involve a rather gross approximation. The unwanted additional time-varying spatial DC term introduced by the reverse-correlation technique was not included in Pattison's (1992) analysis, and is neglected here simply because it makes no useful contribution to the resultant spatial RFP estimate[25]. Pattison (1992) concluded that under the assumption that the neural output is given by the instantaneous firing frequency and that this is (unrealistically) permitted to be negative, the relaxed spatial RFP of each neuron would bear little resemblance to its FSWF $g_j(\boldsymbol{x})$. An interesting exception to this conclusion is provided by the case where the expansion functions $\{g_j\}$ form a tight frame. In this case

$$\sum_i \left[\int_{\mathcal{V}} g_i(\boldsymbol{x}) g_j(\boldsymbol{x}) \, d\boldsymbol{x} \right] g_i(\boldsymbol{x}) = \lambda^{-1} g_j(\boldsymbol{x}) \qquad \forall j, \boldsymbol{x}$$

giving that

$$Q^{-1}\boldsymbol{g}(\boldsymbol{x}) = \lambda \boldsymbol{g}(\boldsymbol{x}) \qquad \forall \boldsymbol{x}$$

and hence the relaxed spatial RFP of each cell is linearly related to the corresponding expansion function. However in this case, the relaxation process is unneccessary in the first place. Whilst the spatial smoothing resulting from the use of rectangular stimuli was not taken into account in Pattison's analysis, such smoothing is not of the required form to undo the premultiplication by the matrix Q^{-1} in (7.10).

The additional temporal resolution achieved by Palmer et al. (1991) through the use of 1ms time bins to minimise the temporal smoothing effect of binning permitted them to obtain estimates of simple cell *spatiotemporal* RFPs, although these estimates were still subject to the temporal smoothing occasioned by the use of rectangular stimuli. Assuming the residual effect of temporal binning to be negligible in comparison with that of the rectangular stimulus, and ignoring the unwanted effects of spatial smoothing and DC offset, the spatiotemporal RFP estimate $\hat{\boldsymbol{w}}_1$ predicted by the proposed single-layered networks over the 50ms stimulus presentation interval is

$$\begin{aligned} \hat{\boldsymbol{w}}_1(\boldsymbol{x}, t) &\approx Q^{-1}\left(I_n - e^{-Qt}\right)\boldsymbol{g}(\boldsymbol{x}) && 0 \leq t \leq 50\text{ms} \\ &\approx Q^{-1}\boldsymbol{g}(\boldsymbol{x}) && 20\text{ms} \ll t \leq 50\text{ms} \end{aligned}$$

the instantaneous spatial component of which, since the exponential term has time constants of the order of 20ms, approximates the RFP presented in (7.10) for $20\text{ms} \ll t \leq 50\text{ms}$. Although the accuracy of this approximation is once again largely dependent on the ratio of the inverse of the smallest eigenvalue of Q to the 20ms neural time constants, it is assumed to be sufficient for the present qualitative purposes. Barring the possibility of linearly dependent expansion functions, this spatiotemporal RFP is inconsistent with the hypothesis that the neurons in the proposed network would at any time during this

[25] Although it may of course significantly affect the best-fit GRGF parameters obtained by Jones & Palmer (1987a).

interval exhibit instantaneous spatial RFPs resembling the expansion functions given by their FSWFs. A similar analysis for the generic two-layered network yields the vector $\hat{\boldsymbol{w}}_2$ of instantaneous spatial RFPs

$$\begin{aligned}\hat{\boldsymbol{w}}_2(\boldsymbol{x},t) &\approx \frac{P^{-1}}{\tau_c}\left(I_{n+m}-e^{-Pt}\right)\begin{bmatrix}\mathbf{0}_n\\ \boldsymbol{r}(\boldsymbol{x})\end{bmatrix} && 0\le t\le 50\text{ms}\\ &\approx \frac{P^{-1}}{\tau_c}\begin{bmatrix}\mathbf{0}_n\\ \boldsymbol{r}(\boldsymbol{x})\end{bmatrix} && 20\text{ms}\ll t\le 50\text{ms}\end{aligned}$$

of which the first n elements belong to the simple cells. Since the top right-hand block of the matrix P^{-1} differs in general from G, the model simple cells again fail to exhibit the required spatial RFPs.

Analysis of the reverse correlation technique with rectangular stimuli is complicated by the addition of a static nonlinearity f as used in (B.1). However as was the case with impulsive stimuli, even-order terms of this nonlinearity result in components of the estimated RFP which cancel out. Furthermore, for the nonlinearity used by the BCP network with saturation limits $\mu_i = 0$ and $\nu_i \gg 0$, and with appropriate smoothing at the origin to ensure the existence of the relevant derivatives, odd-order derivatives are approximately zero. Thus provided the stimuli produce only small deviations about the origin of the nonlinearity, the BCP network appears under reverse correlation identification to be approximately linear regardless of the precise form of the stimulus, and the identified spatiotemporal RFPs for this network should therefore not differ qualitatively from those obtained for the corresponding linear networks.

Qualification and Summary

A full assessment of the implications, for the decomposition hypothesis of simple cell processing, of the use of temporal binning and rectangular spatiotemporal stimuli in the reverse correlation identification techniques employed by Jones & Palmer (1987b) and Palmer et al. (1991) would necessarily include an analysis, similar to the temporal analysis presented above, of the effects of spatial smoothing on the identified instantaneous spatial RFP. Such an analysis would require detailed assumptions regarding the spatial forms of each of the expansion functions, assumptions which, in the absence of sufficient electrophysiological information, have been avoided here so as not to unduly limit the generality of the discussion. Pending a full spatiotemporal analysis of the effects of temporal binning and rectangular spatiotemporal stimuli, it is therefore difficult to draw firm conclusions from the above observations.

The preliminary discussions in this section suggest nevertheless that spatiotemporal smoothing does not account for the predicted discrepancies between the "spatial RFPs" which would be measured by Jones & Palmer (1987b) and Palmer et al. (1991) and the expansion functions whose relative presences in the visual image the simple cells are

hypothesised to calculate by relaxation. Given the as yet complete lack of evidence in favour of the latter hypothesis, the onus should therefore be placed on proponents of this hypothesis to show that such an explanation might be forthcoming.

7.3.4 Summary and Discussion

The hypothetical early visual implementations proposed by MacLennan (1993b), Pattison (1992) and Pece (1992) for their respective SRE minimisation networks have been shown to predict spatiotemporally inseparable simple cell RFPs, the instantaneous spatial components of which differ, at almost all post-stimulus times, from the expansion functions whose coefficients are calculated by these networks. It has furthermore been demonstrated that the reverse-correlation identification techniques used by Jones & Palmer (1987b) would, when applied to a model simple cell in any of the proposed networks, yield a temporal slice, taken at the end of the corresponding time bin, through a spatiotemporally smoothed and spatially DC-shifted version of the spatiotemporal RFP. Neglecting initially the effects of smoothing and DC-shifting, it was therefore tentatively concluded that the "spatial RFPs" identified by Jones & Palmer (1987b) do not correspond to the expansion functions used by an early visual implementation of any of these networks. In particular, it was found that the GRGF model successfully fitted to these "spatial RFPs" by Jones & Palmer (1987a) cannot be interpreted as describing the expansion functions used by such a network. It was furthermore argued that this conclusion extends, at least to a first approximation, to the proposed neural implementation of the more realistic BCP network.

Although it was argued in Section 7.3.3 that the effect of spatiotemporal smoothing is not in fact negligible, temporal smoothing, when considered in isolation, was found to be incapable of effecting the required reconciliation. The effects of spatial smoothing and DC-shifting were not considered in detail.

It is not possible, on the basis of the above observations, to *conclusively* rule out agreement between the expansion functions and measured "spatial RFPs" for special choices of the expansion functions and the stimulus and measurement parameters of the identification technique. There is however sufficient evidence to compel those who might choose to persist with the notion that these "spatial RFPs" are used by the early visual system as non-orthogonal expansion functions to consider alternative neural schemes, and possibly error criteria, for the computation of the required coefficients.

The comparison of the RFPs identified by Jones & Palmer (1987b) and Palmer et al. (1991) with those predicted by the application of the same identification technique to the proposed relaxation network models, is of course predicated on the assumption that the

instantaneous spiking frequency[26] of each real simple cell represents a corresponding coefficient estimate in the theoretical network. However, provided there exists a monotonic — ideally linear — relationship between the firing rate and the value actually being signalled, this comparison retains approximate validity. Nevertheless, the possibility exists that a more sophisticated method is in fact used to encode an analog value in the output spike train, and the above comparisons must therefore be viewed as contingent upon the verification of the neural firing rate signalling hypothesis.

7.4 Discussion

Implicit in the proposed encoding of the chosen expansion functions on the feedforward connections of both the one- and two-layered networks is the assumption that the simple cell spatial RFP is determined solely by its feedforward weighting of the retino-geniculocortical input, an assumption which was disproved in Section 7.3 on the basis of its neglect of the temporal dimension imparted to the RFP by the presence of feedback. This assumption is furthermore at odds with the evidence in favour of models of the early visual system which postulate the presence of mutual inhibition between iso- or cross-oriented simple cells in order to sharpen their orientation tuning (see e.g. Ferster & Koch (1987) or Wörgötter & Koch (1991) for a review). These models commonly assume that an orientation bias is established in the feedforward pathway to a given neuron, through for example the alignment of LGN afferents (Hubel & Wiesel, 1962), and that this bias is enhanced by intracortical inhibition between the simple cells. Wörgötter & Koch (1991) showed that an eclectic model combining several intracortical inhibitory mechanisms to achieve this enhancement of orientation tuning predicts, among other things, the effect on this orientation tuning of the application of GABA lateral to the recording site.

Nevertheless, the negative findings in Section 7.3 do not of course eliminate the possibility that the simple cell spatial RFPs might still be used as expansion functions in a decomposition of the visual image. For example, it would be possible to propose alternative reconstruction error criteria (Cichocki & Unbehauen, 1992), and, presumably, biologically plausible neural networks capable of minimising these errors. Alternatively, in a departure from the hypothesis that the simple cells signal the expansion coefficients, cells such as the complex cells, located in subsequent processing "stages" and receiving input from the simple cells, might be identified with the output nodes of the proposed SRE-minimisation networks. However, before such alternatives can be given serious consideration, it is necessary to re-evaluate the motivation behind the decomposition hypothesis.

[26] An estimate of the spiking frequency is obtained for each post-stimulus time bin by counting the number of spikes occuring in that bin, dividing by the width of the bin, and averaging the result over repeated presentations of the same stimulus.

As Daugman (1988a) pointed out,

> ... it goes without saying that the purpose of vision is *not* to reconstruct the retinal image ...

Linsker (1990) therefore raised the question

> If the original scene is never "reconstructed" by the brain (and why should it be?), what is the meaning of the fidelity criterion?

In particular, why go to the computational and metabolic expense of minimising the reconstruction error when that error is never realised in a cortical reconstruction? For decomposition-based theories of simple cell processing, the answer is that this "fidelity criterion" is, as indicated earlier, simply a means of quantifying the concept of the "relative presence" of a simple cell spatial RFP in the visual image. According to this interpretation, the minimisation of some measure of reconstruction error clearly does not imply that a reconstruction of the image is to be performed.

However it remains to justify the underlying assumption which gave rise to this diversionary debate on reconstruction error minimisation in the first place — viz. that the simple cells spatial RFPs are used in the early visual system as expansion functions. In seeking justification for the possible neural implementation of a reconstruction error minimisation algorithm amongst the simple cells, the more pertinent question is therefore, "Why should the output of a simple cell be assumed to signal the relative presence of its spatial RFP in the visual image, rather than the spatial inner product of that RFP with the image, *in terms of which the spatial RFP is defined*?" Apart from the observation that this assumption, which is central to decomposition-based theories of simple cell processing, is probably a vestige of the earlier feature-detection hypothesis, no clear motivation or *apriori* justification for this assumption is in evidence. Although the sceptic might argue that this quite specific assumption appears in hindsight to have been somewhat ill-conceived, and that it is not possible on this basis to rule out the calculation of the required expansion coefficients elsewhere in the early visual system using the simple cell outputs, he or she should be challenged to justify the need for and the utility of the hypothesised expansion.

7.5 Conclusion

The implementation of the SRE minimisation networks of MacLennan (1993b), Pattison (1992) and Pece (1992) in the feline early visual system has been shown to be at least broadly speaking biologically feasible. However, the identification of their output nodes with the simple cells has been found to be almost certainly inconsistent with the hypothesis that the latter use the "spatial RFPs" obtained by Jones & Palmer (1987b)

as expansion functions in a minimum SRE decomposition of the visual image. Although this finding is by no means fatal for more general decomposition-based theories of early visual processing, it suggests the need for a fundamental re-evaluation of their motivation in the absence of *apriori* justification. In particular, claims — either explicit or implicit — of "biological motivation" for image processing algorithms which use models of the simple cell "spatial RFP" as visual expansion functions should be treated with extreme caution.

Chapter VIII

CONCLUSIONS

8.1 Overview

This thesis has reviewed models of the simple cell RFP and its variation over the simple cell population, and has used artificial neural networks to investigate the multi-dimensional signal processing role of the RFP in the formation of a cortical representation of the visual image.

The degree of linearity of simple cell processing, and the validity of omitting one or more stimulus dimensions from the RFP, were examined in Section 2.2. The GRGF model of the simple cell RFP was then evaluated in Section 2.3. Theories concerning the variation of RFP parameters across the simple cell population were investigated in Chapter 3 with regard to their ability to account for the true variety of simple cell RFPs observed experimentally. These theories were divided into the two categories of filtering and decomposition, depending on their use of the spatial RFP as either the kernel of a visual filter or a visual expansion function. RANNs which solve the SRE minimisation problem associated with the decomposition of an image using non-orthogonal expansion functions were reviewed in Chapter 4 and developed in Chapters 5 and 6. The biological plausibility of these networks was assessed in Chapter 7, and the RFPs which they predicted were compared with the expansion functions used by these networks in order to evaluate the viability of the decomposition hypothesis of simple cell processing.

8.2 Summary and Conclusions

In Chapter 2 it was concluded that the simple cell can be considered, to a first approximation, to be a linear device characterised by its binocular spatiotemporal RFP. The omission of the temporal or the second spatial dimension from the RFP was found to be valid only in the special case where the RFP is Cartesian separable. In particular the concept of a "spatial RFP", which is used extensively in Chapter 3, is strictly speaking undefined except for the Cartesian separable case. To the extent to which it *can* be defined, the identified simple cell spatial RFP is best fitted by the GRGF model presented in Section 2.3. The fact that functions have been found which provide a better fit to the identified spectral RFP is not sufficient to discount the GRGF spatial model, since the observed discrepancy between the spectral RFP and the Fourier transform of the spatial

RFP (Jones & Palmer, 1987a) is most likely attributable to nonlinear components of simple cell processing.

The recent controversy concerning the alleged optimal joint spatial and spectral localisation property of simple cell processing was found to hinge on the credibility assigned to the implementation of complex-valued Gabor functions by pairs of simple cells whose RFPs are in spatial phase quadrature. However even if one accepts that such pairs are widespread, as Daugman (1993) would have us believe, it was also argued that since the particular localisation measure used in (2.10) is sensitive to rotations of the coordinate axes, it cannot be simultaneously minimised by a population of simple cells exhibiting arbitrary relative orientations. This property, which holds for all component-wise localisation measures, casts grave doubts on the relevance of optimising some measure of the joint spatial and spectral localisation to the processing performed by the simple cells.

A number of theories concerning the spatial processing of the visual stimulus performed by the simple cells were examined in Chapter 3. Few if any were found to make any serious attempt to account for all the degrees of freedom exhibited by simple cell spatial RFPs, with many models assuming one or more of the following: odd or even RFP symmetry, constant window size or orientation across the population, mutual orthogonality with respect to spatial integration, phase quadrature pairs, and highly regular spatial and spectral sampling lattices. It should be noted however that for those which did attempt to answer even some of these objections, determination of the completeness of the resultant RFP sets generally exceeds our current mathematical capabilities. Although this by no means constitutes an argument against their suitability as models of simple cell processing, it is difficult, without established completeness results, to rule out the existence of large classes of stimuli which are in effect invisible to the feline or primate visual systems (Daugman, 1988b). This observation limits the utility in machine vision of many of the more realistic models of simple cell RFP variety.

The decomposition hypothesis of simple cell processing was found to be in conflict with the very definition of the spatial RFP. Whilst the RFP model(s) employed by such theories may enjoy electrophysiological support, the RFP definition supports their use as the kernels of (possibly position-dependent) spatial filters. However, this conclusion is later reviewed in Chapter 7 as a result of the development of biologically plausible RANNs which calculate the required expansion coefficients.

The theory of non-orthogonal image decomposition was reviewed in Chapter 4. The squared reconstruction error (SRE) minimisation problem, whose solution yields the best linear unbiassed estimator (BLUE) in the presence of spherical Gaussian noise, was formulated, and a minimum-norm regularisation term added to overcome the problems associated with ill-conditioning of the normal equations. ANNs which solve the SRE and minimum-norm SRE minimisation problems were critically reviewed. Most notably, it was shown that the two-layered networks in Section 4.4 which were previously though

to provide only approximate solutions to the SRE minimisation problem in fact provide exact solutions to the minimum-norm regularised form of the problem.

In Chapter 5, a one-layered analog RANN was presented which solves the SRE and minimum-norm SRE minimisation problems. It was shown to be closely related to resistive grid networks proposed independently for the discrete-space solution of minimum-norm SRE minimisation problems arising in machine vision applications. A strategy for diagonally preconditioning the network equations was presented in order to improve the stability of the network in the presence of weight implementation and state propagation errors.

The neural activation function of the single-layered RANN proposed in Chapter 5 was modified in Chapter 6 to more accurately reflect the constraints on the firing rate of real neurons. The global and exponential stability of the resultant boundary constraint projection (BCP) network was then established, along with its optimisation of the chosen positive (semi)definite quadratic cost function subject to the imposed bound constraints on the optimisation variables. This network was compared with similar RANN models, including that of Sudharsanan & Sundareshan (1991) and the generalised Brain-State-In-A-Box (GBSB) network of Golden (1992). A continuous-time equivalent of the GBSB network, the continuous GBSG (CGBSB) network, was then proposed to overcome a perceived shortcoming of the BCP network for engineering applications, and a partial stability proof for it outlined. Finally, various examples and comparisons of the operation of these networks were presented.

In Chapter 7, the biological plausibility of the SRE minimisation networks reviewed in Chapter 4 and those developed in Chapters 5 and 6 was first examined. The hypothesis that the expansion coefficients purportedly signalled by the simple cells are computed by an early visual implementation of an SRE minimisation network was then investigated through a comparison of the RFPs predicted by these networks with the corresponding expansion functions. It was concluded that if such a network were indeed implemented in the early visual system, the "spatial RFPs" identified by Jones & Palmer (1987b) would not be the expansion functions used by that network. Although this does not conclusively discount the possibility that the simple cells perform a decomposition using other expansion functions, it does at least argue against the postulated Gabor-like expansions presented in Chapter 3.

8.3 Discussion

If the discussion in Chapters 2, 3 and 7 could be said to have a weak point, it is that much of the more detailed analysis relies on data from only a few experiments. In particular, the reverse-correlation experiments of Jones & Palmer (1987b) and their subsequent evaluation of the GRGF model of the simple cell spatial RFP (Jones & Palmer, 1987a)

constitute the principal source of electrophysiological information. This reliance is largely necessitated by the lack of alternative electrophysiological investigations of the 2D spatial RFP which are both quantitative and free of strong *apriori* assumptions such as symmetry or Cartesian separability.

8.4 Contributions

The principal contributions of this thesis are as follows:

- A critical review of the evidence for and against the GRGF model of the simple cell RFP.
- A critical review of bottom-up theories of simple cell spatial processing.
- A brief review of applications of GRGFs in image processing.
- A comparison and critical review of existing neural networks for solving the SRE minimisation problem associated with non-orthogonal image decomposition.
- The recognition of the fact that the two-layered RANNs of Culhane et al. (1989), Yan (1991b) and Pece (1992) solve, exactly, the minimum-norm SRE minimisation problem.
- The development of a single-layered linear network for minimum-norm SRE minimisation, and a comparison with resistive grid networks used in other early vision applications.
- The development of a diagonal preconditioning strategy to mitigate the effects on network stability and solution accuracy of errors in weight implementation or state-vector propagation.
- The development of the BCP network, the proof of its global and exponential stability, and its comparison with similar ANN models. This work was based on earlier work undertaken jointly and published by Bouzerdoum & Pattison (1993b).
- The demonstration that
  - the decomposition theory of simple cell processing contradicts the very definition of the simple cell spatial RFP,
  - its biologically plausible RANN implementation predicts spatiotemporally inseparable simple cell RFPs, in which case the required spatial RFPs are not even defined, and
  - the instantaneous spatial RFPs of the hypothetical simple cells in such implementations are not, in general, the chosen expansion functions.

Appendix A

SPATIAL AND SPECTRAL RF INVESTIGATIONS

A.1 Spatial RF

Table A.1 contains a non-exhaustive summary of single-unit electrophysiological investigations of the spatial receptive field organisation, profile &/or 1D response plane (RP – see eg. Palmer & Davis (1981b)) of the simple cells in primate and feline primary visual cortex. The stimulus consists of a contrast bar, edge or spot, and is either stationary (flashed once on-off or periodically modulated) or moving with constant drift velocity. The receptive field map is based on measurements of response (mean or peak firing frequency or probability) or reverse correlation of the output spike train with the stimulus. A "bar" has for the present purposes been classified as a spot if its length was substantially less than the elongation of a subfield. The specification of 1D in the RFP column includes both the true LWF, obtained using long thin bars, and 1D sections through the RFP in the direction of subfield alternation, obtained using spots or small bars. Stimulus and measurement specifications refer only to those portions of each study concerned with the spatial properties of simple cell RFPs.

A.2 Spectral RF

Table A.2 contains a non-exhaustive summary of single-unit electrophysiological investigations of the spatial frequency magnitude, orientation and phase tuning properties of the simple cells in primate and feline primary visual cortex. The stimulus consists of a contrast bar, edge, or sinewave or squarewave grating, and is either stationary (usually flashed or counter-phase modulated) or drifting with constant velocity. The tuning curves are based on measurements of either response (mean or peak firing frequency) or contrast sensitivity (CS – inverse of minimum stimulus contrast required to elicit a specified response). Several studies do not differentiate between simple, complex and hypercomplex cell types. Stimulus and measurement specifications refer only to those portions of each study concerned with the relevant tuning curves of simple cells.

Investigators	RFP Property			Species		Stimulus							Measurement	
	Org.	RFP	RP	Cat	Monkey	Stat.	Drift.	On-Off	Mod.	Bar	Edge	Spot	Resp.	Corr.
Henry & Bishop (1972)	1D			•		•	•	•		•	•		•	
Schiller et al. (1976a)		1D			•	•	•		•	•	•	•	•	
Movshon et al. (1978b)		1D		•		•			•	•			•	
Glezer et al. (1980)		1D		•			•			•	•		•	
Kulikowski & Bishop (1981a)		1D		•			•			•	•		•	
Kulikowski et al. (1981)		1D		•		•	•	•		•	•		•	
Kulikowski & Bishop (1981b)		1D		•		•	•	•		•			•	
Palmer & Davis (1981b)			•	•		•		•				•	•	
Palmer & Davis (1981a)			•	•		•	•	•			•	•	•	
Glezer et al. (1982)		1D		•			•			•			•	
Mullikin et al. (1984)			•	•		•		•		•			•	
Camarda et al. (1985b)		1D		•		•	•		•		•		•	
Camarda et al. (1985a)		1D		•		•	•		•	•			•	
Maske et al. (1985)		1D		•			•			•			•	
Peterhans et al. (1985)		1D		•		•	•		•	•	•		•	
Yamane et al. (1985)		1D		•		•	•		•	•			•	
Field & Tolhurst (1986)		1D		•		•			•	•			•	
Heggelund (1986b)		1D		•		•		•		•			•	
Jones & Palmer (1987b)		2D		•		•		•				•		•
Glezer et al. (1989)		1D		•		•	•	•		•			•	
McLean & Palmer (1989)		1D		•		•		•				•		•
Palmer et al. (1991)		2D		•		•		•				•		•

Table A.1: A non-exhaustive summary of investigations of the spatial structure of the simple cell RF. See text for an explanation of symbols and abbreviations.

Investigators	Tuning Parameter			Species		Stimulus						Measurement	
	Mag.	Orient.	Phase	Cat	Monkey	Stat.	Drift.	Bar	Edge	Sine	Square	CS	Response
Henry & Bishop (1972)		•		•			•	•					•
Maffei & Fiorentini (1973)	•			•		•				•			•
Henry et al. (1974)		•		•			•	•	•				•
Rose & Blakemore (1974)		•		•			•	•					•
Watkins & Berkley (1974)		•		•			•	•	•				•
Ikeda & Wright (1975b)	•	•		•		•	•			•	•		•
Ikeda & Wright (1975a)	•	•		•		•	•			•	•		•
Finlay et al. (1976)		•			•		•	•	•				•
Schiller et al. (1976b)		•			•		•	•	•	?	?		•
Schiller et al. (1976c)	•				•		•	•		•	•		•
Heggelund & Albus (1978)		•		•			•	•					•
Movshon et al. (1978b)			1D	•		•				•			•
Movshon et al. (1978a)	•			•			•			•		•	•
Albrecht et al. (1980)	•			•	•	•	•	•		•		•	•
Glezer et al. (1980)	•		1D	•			•			•			•
Kulikowski & Bishop (1981a)	•			•		•	•	•	•				•
de Valois et al. (1982)	•				•		•			•		•	
de Valois et al. (1982)		•			•		•	•		•			•
Pollen & Ronner (1982)	•			•			•			•	•		•
Spitzer & Hochstein (1985)	•		1D	•		•				•			•
Webster & de Valois (1985)	•	•		•			•			•			•
Hawken & Parker (1987)	•	•			•		•			•		•	
Jones et al. (1987)	•	•		•			•			•			•
Reid et al. (1987)			1D	•		•				•			•

Table A.2: A non-exhaustive summary of investigations of the spectral structure of the simple cell RF. See text for an explanation of symbols and abbreviations.

Appendix B

RFP IDENTIFICATION USING IMPULSES

Let the response $r(t)$ of a spatiotemporal system to an input $s(\boldsymbol{x},t)$ be given by

$$r(t) = f\left(\int_{-\infty}^{\infty} h(t-\tau)\int_{-\infty}^{\infty} w(\boldsymbol{x})s(\boldsymbol{x},\tau)\,d\boldsymbol{x}\,d\tau\right) \tag{B.1}$$

where $(\boldsymbol{x},t)$ are the spatiotemporal coordinates, $f{:}\mathbb{R}\rightarrow\mathbb{R}$ is a nonlinear function applied to the output of an otherwise linear system, $h(t)$ is the temporal impulse response function, and $w{:}\mathbb{R}^2\rightarrow\mathbb{R}$ is a spatial weighting function. The aim is to estimate the function w using the reverse correlation technique, which for each possible stimulus position $\boldsymbol{x}_0$ involves presenting the system with a stream of spatiotemporal impulses occurring at regular temporal intervals of T seconds and at spatial positions $\boldsymbol{x}_i$ chosen from a uniform distribution over the support[1] of w, and correlating the resultant output with the spatiotemporal impulse presented at position $\boldsymbol{x}_0$. Prior to correlation with the stimulus, the response of the system is averaged over a number of trials, with the time axes of the trials aligned so that the impulse at position $\boldsymbol{x}_0$ occurs at the same time in each trial, which is arbitrarily taken to be time $t=0$.

The stimulus and response are given by

$$\begin{aligned} s(\boldsymbol{x},t) &= \pm\sum_{i=-\infty}^{\infty}\delta(\boldsymbol{x}-\boldsymbol{x}_i)\delta(t-iT) \\ r_{\pm}(t) &= f\left(\pm\sum_i w(\boldsymbol{x}_i)h(t-iT)\right) \end{aligned} \tag{B.2}$$

respectively. Assuming that the system is operating in the vicinity of the origin of the nonlinearity f, and that f is analytic in this region so that its Taylor series expansion about the origin[2] exists, then averaging (B.2) over a series of trials gives

$$r_{\pm}(t;\boldsymbol{x}_0) = \left\langle\sum_{j=0}^{\infty}(\pm1)^j\frac{f^{(j)}(0)}{j!}\left[\sum_i w(\boldsymbol{x}_i)h(t-iT)\right]^j\right\rangle \tag{B.3}$$

$$\begin{aligned} &= f(0)+\sum_{j=0}^{\infty}(\pm1)^j\frac{f^{(j)}(0)}{j!}w^j(\boldsymbol{x}_0)h^j(t) \\ &\quad \pm f^{(1)}(0)\sum_{i\neq0}\langle w(\boldsymbol{x}_i)\rangle\,h(t-iT) \\ &\quad +\frac{1}{2}f^{(2)}(0)\sum_{i\neq0}\sum_k\langle w(\boldsymbol{x}_i)w(\boldsymbol{x}_k)\rangle\,h(t-iT)h(t-kT)+\ldots \end{aligned} \tag{B.4}$$

---

[1] The support of a function is the smallest region outside of which it evaluates to zero everywhere.

[2] viz. its MacLauren series expansion

where $\langle\cdot\rangle$ denotes the average over trials and $f^{(j)}(0)$ denotes the jth derivative of f at 0. The parameterisation of the averaged response $r_\pm$ by the spatial variable $\boldsymbol{x}_0$ serves as a reminder that the average is over trials in which a stimulus impulse occurs in this position at time $t = 0$. Estimates $\hat{w}_+, \hat{w}_-: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ of the spatiotemporal RFP are obtained by repeating the above process for all stimulus positions $\boldsymbol{x}_0$, weighting each trial-averaged response $r_\pm(t; \boldsymbol{x}_0)$ by the corresponding spatial stimulus $\pm\delta(\boldsymbol{x} - \boldsymbol{x}_0)$, and adding the results such that

$$\hat{w}_\pm(\boldsymbol{x}, t) = \pm \int_{-\infty}^{\infty} r_\pm(t; \boldsymbol{x}_0)\delta(\boldsymbol{x} - \boldsymbol{x}_0)\, d\boldsymbol{x}_0 \tag{B.5}$$

Adding the two estimates obtained with positive and negative impulses respectively [3] yields the combined spatiotemporal RFP estimate

$$\hat{w}(\boldsymbol{x}, t) \triangleq \frac{1}{2}\left[\hat{w}_+(\boldsymbol{x}, t) + \hat{w}_-(\boldsymbol{x}, t)\right]$$

Rather than denoting a fixed spatial position, the notation $\boldsymbol{x}_i$ should be interpreted as the position of the stimulus presented iT seconds after the stimulus $\boldsymbol{x}_0$. This position varies randomly over trials, having a uniform distribution over the support of w so that

$$\left\langle w^j(\boldsymbol{x}_i)\right\rangle = \overline{w^j} \triangleq \int_{-\infty}^{\infty} w^j(\boldsymbol{x})\, d\boldsymbol{x}$$

Similarly, the average of the jth order cross terms in $w(\boldsymbol{x}_i)$ gives the area $\overline{R_w^j}$ under the jth order (spatial) autocorrelation R_w^j of w. Thus the RFP estimates obtained with positive and negative impulses are identical except for a difference in the sign of the even order terms, which upon addition (averaging) of the two estimates results in cancellation of these terms to yield

$$\begin{aligned}\hat{w}(\boldsymbol{x}, t) \;=\; & \sum_{j=0}^{\infty} \frac{f^{(2j+1)}(0)}{(2j+1)!} w^{2j+1}(\boldsymbol{x}) h^{2j+1}(t) + \Bigg\{ f^{(1)}(0)\overline{w} \sum_{i\neq 0} h(t - iT) \\ & + \frac{1}{6} f^{(3)}(0)\overline{R_w^3} \sum_{i\neq 0} \sum_{k} \sum_{l} h(t - iT)h(t - kT)h(t - lT) + \ldots \Bigg\}\end{aligned}$$

This cancellation of the even-order terms — which is not dependent on the present assumption of spatiotemporal separability or even linearity of the RFP — can be seen to result from a failure of these terms to preserve the sign of the input stimulus, a property shared by any even-order static nonlinearities occurring prior to the spatial summation (which are not incorporated in the model of (B.1)). This latter observation provides a formal basis for the assertion by McLean & Palmer (1989) that

> ... spontaneous activity and spatial regions excited by both bright and dark stimuli tend to cancel.

[3] This is equivalent to the *subtraction* of the two estimates obtained by Jones & Palmer (1987b) since they did not account for the negative sign of the negative contrast stimuli until this stage, whereas it has already been incorporated in the present analysis in (B.5).

Assuming that the third and higher odd-order derivatives of f are negligible at the origin yields

$$\hat{w}(\boldsymbol{x},t) \approx f^{(1)}(0)\left[w(\boldsymbol{x})h(t) + \overline{w}\sum_{i\neq 0} h(t-iT)\right] \tag{B.6}$$

where $\overline{w}$ denotes the DC component (mean value) of w, so that for any given time t, $\hat{w}$ is not proportional to w unless the latter has no DC component. The effect on $\hat{w}$ of such a DC component is to introduce a DC offset term which is independent of $\boldsymbol{x}$, but varies with the time t after presentation of the stimulus. Assuming that the impulse response function $h(t)$ decays monotonically for $t > T$, the weighting of this DC term will increase relative to that of the desired term as each subsequent stimulus is received.

The use of spatiotemporal impulses in the above analysis immitates the ideal experimental situation which Jones & Palmer (1987b) sought to approximate with finite rectangular spatiotemporal pulses. The model assumed for the simple cell is a deliberately simple one, consisting of a spatiotemporally separable linear system followed by a nonlinear function. Spatiotemporal separability was found to hold for approximately 40% of the simple cells examined by Palmer et al. (1991). However, even for such a simple model and in the absence of the nonlinearity, the proposed reverse correlation technique does not yield the desired RFP except when that RFP integrates to zero over its support. The RFP model fitted by Jones & Palmer (1987a) satisfies these conditions only for odd symmetry ($\phi = n\pi/2, n = \pm 1, \pm 3\ldots$) which was found to be no more common than any other value of spatial phase. This finding is consistent with that of Jones et al. (1987), who noted that the spectral RFP of some simple cells did not appear to decay to zero for small spatial frequencies.

With the introduction of a nonlinearity, such as the halfwave rectification nonlinearity commonly used to model the conversion of somatic membrane potential to firing rate (see eg. Palmer et al. (1991)), additional terms including those of the form $f^{(2j+1)}(0)w^{2j+1}(\boldsymbol{x})h^{2j+1}(t)/(2j+1)!$ are introduced, the second term of which is in general nonzero, making the estimate $\hat{w}$ less reliable for nonlinearities having non-zero higher odd-order derivatives at the origin. However, in the particular case of the half-wave rectification nonlinearity (with appropriate smoothing at the origin) these odd-order terms are zero, thereby eliminating this potential source of error. The use of the Taylor series expansion of f about zero is justified since the cells probed by Jones & Palmer (1987b) operated around the spiking threshold and were never strongly activated. The existence and convergence of this expansion is assumed since the relation between membrane potential and firing rate or probability is likely to be appropriately smooth, even though this is not the case for the idealised halfwave rectification function commonly used to approximate it.

Appendix C

GABOR FUNCTION APPLICATIONS IN IMAGE CODING AND ANALYSIS

Tables C.1 and C.2 present a non-exhaustive summary of applications of Gabor-function representation schemes to computational tasks involving image coding and analysis. Table C.1 indicates whether the chosen Gabor functions are used as filters or expansion functions in the chosen image representation, the type of information extracted from the representation, and the type of application. Table C.2 summarises the systematic variation of the Gabor function parameters over the set chosen for each application.

In the "Application" column of Table C.1, each application is classified into one of the following categories: texture (Text.) including segmentation (S), discrimination (D) or classification (C); image compression (Comp.), including vector quantisation (VQ); pattern recognition (PR), perhaps to achieve some type of invariance (I); stereopsis (Stereo.); and miscellaneous (Misc.). The "Operation" column specifies whether the image is filtered by or expanded using the chosen Gabor functions. Filtering schemes, which include the discrete window Fourier transform and the wavelet-like transforms by Watson & Ahumada (1983) and Field (1987), may involve either convolution or correlation — denoted respectively by $*g_i(\boldsymbol{x})$ and $*g_i(-\boldsymbol{x})$ where $*$ indicates 2D convolution — of the image with each of the chosen Gabor function kernels. Filtering is generally followed by sampling of the result on a grid of Gabor function centres in the input space, the form and spacing of which are specified in the $\boldsymbol{x}_0$ column of Table C.2, which is to be described shortly. Expansion schemes, including the Gabor expansion and Gabor "wavelet" expansion, may be implemented either by correlation — denoted $*\gamma_{ik}(-\boldsymbol{x})$ — with the modulated 2D biorthogonal functions $\gamma_{ik}(\boldsymbol{x}) \triangleq \gamma(\boldsymbol{x}) \cdot \exp\{j2\pi[i\Delta_u^{[1]}, k\Delta_u^{[2]}]^T\boldsymbol{x}\}$ (Porat & Zeevi, 1988) in the case of the GE, or more generally by relaxation (Relax.) as described in Chapter 4. The information used for further processing may consist of the real &/or imaginary parts $(\Re\{\cdot\}, \Im\{\cdot\})$ or the corresponding magnitude and phase $(|\cdot|, \angle\cdot)$ of each pixel/coefficient after filtering or decomposition. Representations employing only cosine- or only sine-phase GRGFs are for convenience classified in the real $(\Re\{\cdot\})$ or imaginary $(\Im\{\cdot\})$ columns respectively, even though they don't in fact implement complex-valued Gabor functions. Applications which adaptively select the outputs of an appropriate subset of the functions used are indicated in the column marked "Adapt.".

Table C.2 describes the systematic variation of the Gabor function parameters used

Investigators	Operation				Information				Adapt.	Application				
	$*g_i(\boldsymbol{x})$	$*g_i(-\boldsymbol{x})$	$*\gamma_{ik}(-\boldsymbol{x})$	Relax.	$\Re\{\cdot\}$	$\Im\{\cdot\}$	$\lvert\cdot\rvert$	$\angle\cdot$		Text.	Comp.	PR	Stereo	Misc.
Daugman & Kammen (1986)		•					•		•		•			
Daugman & Kammen (1986)		•			•	•			•		VQ			
Paler & Bowler (1986)		•					•					•		
Turner (1986)		•					•		•	D				
Daugman (1988a; 1989b; 1990)				•			•			S				
Daugman (1988a; 1989a; 1989b; 1990)				•	•	•					•			
Porat & Zeevi (1988)			•		•	•			•					•
Buhmann et al. (1989)	•						•					I		
Clark & Bovik (1989)	•						•	•		S				
Flaton & Toborg (1989)	•						•		•			•		
Fogel & Sagi (1989)		•					•			D,S				
Lawton (1989)	?				•	•								•
Porat & Zeevi (1989)			•				•			S				
Zetzsche & Caelli (1989)	•				?		?					I		
Beck et al. (1990)				•	•					S				
Bovik et al. (1990)	•						•	•	•	S				
Gopal et al. (1990)	•						•			•				
Gutschow & Hecht-Nielsen (1991)		•			•	•						•		•
du Buf (1990)	•							•		D				
Leung et al. (1990)	?				•	•	•	•				•		
Rubenstein & Sagi (1990)		•					•			D				
Tan & Constantinides (1990)	•						•			C				
Bovik (1991)	•						•		•	S				
Jain & Farrokhnia (1991)		•			•				•	S				
Theimer & Mallot (1992)	•						•	•					•	
Wang & Yan (1992)				•	•	•					•			
Zhou & Chellappa (1992)		•			•	•							•	

Table C.1: Applications of Gabor functions in image analysis and machine vision. Corresponding selections of function parameters (where published) are detailed in Table C.2. Refer to text for an explanation of abbreviations used. Where insufficient information is provided, "?" indicates the inferred or presumed classification.

Investigators	$\boldsymbol{x}_0$				$\boldsymbol{u}_0$					s			
	C	$C_{1/\lvert\boldsymbol{u}_0\rvert}$	LP	$H_{1/\lvert\boldsymbol{u}_0\rvert}$	C	P	LP	$1/\lvert\boldsymbol{x}_0\rvert$	Other	F	$1/\lvert\boldsymbol{u}_0\rvert$	$\lvert\boldsymbol{x}_0\rvert$	Other
Daugman & Kammen (1986)				•			•				•		
Daugman & Kammen (1986)	•						•				•		
Paler & Bowler (1986)	•					?				•			
Turner (1986)	•						•			•			
Daugman (1988a; 1989b; 1990)	•				•					•			
Daugman (1988a; 1989a; 1989b; 1990)		•					•				•		
Porat & Zeevi (1988)	•				•					•			
Buhmann et al. (1989)		•					•				•		
Clark & Bovik (1989)	?										•		
Flaton & Toborg (1989)	•						•				•		
Fogel & Sagi (1989)	•								•	•			
Porat & Zeevi (1989)	•				•					•			
Zetzsche & Caelli (1989)	?						•				•		
Beck et al. (1990)		•							•				
Bovik et al. (1990)	•								?				•
Gopal et al. (1990)	•						•				•		
Gutschow & Hecht-Nielsen (1991)			•					•				•	
du Buf (1990)	•						•				•		
Rubenstein & Sagi (1990)	•					•				•			
Tan & Constantinides (1990)	?						•				•		
Bovik (1991)	•						•				?		
Jain & Farrokhnia (1991)	•						•				•		
Theimer & Mallot (1992)	•						•				•		
Zhou & Chellappa (1992)	•						•				•		

Table C.2: Selection schemes for Gabor function or GEF parameters used in the applications detailed in Table C.1. See text for an explanation of symbols used. Where insufficient information is provided, "?" indicates the inferred or presumed classification.

Representation	$\boldsymbol{x}_0$				$\boldsymbol{u}_0$					s			
	C	$C_{1/\lvert\boldsymbol{u}_0\rvert}$	LP	$H_{1/\lvert\boldsymbol{u}_0\rvert}$	C	P	LP	$1/\lvert\boldsymbol{x}_0\rvert$	Other	F	$1/\lvert\boldsymbol{u}_0\rvert$	$\lvert\boldsymbol{x}_0\rvert$	Other
GE,DWFT	•				•					•			
GWE,DWT		•					•				•		

Table C.3: Illustration of how the image representation schemes discussed in Sections 3.4 and 3.5 would be classified according to the conventions used in Table C.2. See text for an explanation of symbols used.

to construct the set of Gabor functions used in each of the applications for which this information is provided. The notation adopted in Table C.2 is as follows. Independent parameters may be either fixed (F) or distributed evenly on a Cartesian (C), polar (P), log-polar (LP) or hexagonal (H) grid, the spacing of which is a function of the subscripted parameter (if any). Linearly or inverse-linearly dependent parameters have a single value for each value of the independent parameter — indicated in the column heading — on which they are dependent. With the exception of the application by Turner (1986), the Gabor function phase parameter ϕ was chosen to be 0, corresponding to a pair of GRGFs with phases 0 and $\frac{\pi}{2}$ (although in some cases, as indicated, only one or the other was used). The aspect ratio λ of the Gaussian was for most applications 1, and was invariably chosen to be the same for all Gabor functions employed by the application. In the former case, the Gaussian rotation parameter θ_z is irrelevant, and in both cases, since the aspect ratio s_1/s_2 remains constant, the variation of the dimensions of the 2D Gaussian window — detailed in the "s" column of Table C.2 — can be completely characterised by variations of the single parameter $s = s_1$ (say). The Gaussian size parameter s could be: fixed; varied directly with the spatial ($|\boldsymbol{x}_0|$) eccentricity; varied inversely with the spectral ($|\boldsymbol{u}_0|$) eccentricity; or otherwise varied. The spatial lattice over which the Gaussian centres are distributed — or equivalently the filtered image sampled — can be classified as: Cartesian; Cartesian with grid spacing varying inversely with spatial frequency ($C_{1/|\boldsymbol{u}_0|}$); log-polar; or hexagonal with grid spacing varying inversely with spatial frequency ($H_{1/|\boldsymbol{u}_0|}$). The spatial frequency lattice over which the Gaussian centres are distributed is classified as: Cartesian; polar; log-polar; inversely dependent on spatial eccentricity ($1/|\boldsymbol{x}_0|$); or varied in some other way. Table C.3 indicates how the GE, DWFT, GWE and DWT would be classified according to these categories.

Appendix D

STABILITY IN NONLINEAR NETWORKS

D.1 Boundedness of Solutions of the BCP Network

It is required to prove that for bounded inputs $\{\boldsymbol{y}(t) : \|\boldsymbol{y}\| \leq Y\}$ and bounded starting points $\{\boldsymbol{u}_0 : \|\boldsymbol{u}_0\| \leq U\}$, with $Y, U \in \mathbb{R}_+$ finite, the trajectory of the system in (6.9) with A positive diagonal is bounded.

Taking the Euclidean norm in (6.29) and substituting for the spectral norms of the relevant matrices using the results of Section 6.8 gives

$$\|\boldsymbol{u}(t; t_0, \boldsymbol{u}_0)\| \leq e^{-\lambda_{min}(A)(t-t_0)}\|\boldsymbol{u}_0\| + \int_{t_0}^{t} e^{-\lambda_{min}(A)(t-s)} \{\sigma_{max}(C)\|\boldsymbol{f}(\boldsymbol{u}(s))\| + \|\boldsymbol{y}(s)\|\}\, ds$$

Now clearly $\|\boldsymbol{f}(\boldsymbol{u})\| \leq F$ for some finite $F \in \mathbb{R}_+$. Thus

$$\begin{aligned} \|\boldsymbol{u}(t)\| &\leq \|\boldsymbol{u}_0\| e^{-\lambda_{min}(A)(t-t_0)} + [\sigma_{max}(C)F + Y] \int_{t_0}^{t} e^{-\lambda_{min}(A)(t-s)}\, ds \\ &= \|\boldsymbol{u}_0\| e^{-\lambda_{min}(A)(t-t_0)} + \left[\frac{\sigma_{max}(C)F + Y}{\lambda_{min}(A)}\right] \left(1 - e^{-\lambda_{min}(A)(t-t_0)}\right) \end{aligned}$$

which is equal to $\|\boldsymbol{u}_0\|$ at $t = t_0$ and converges exponentially to $(\sigma_{max}(C)F + Y)/\lambda_{min}(A)$. Thus an upper bound on $\|\boldsymbol{u}(t; t_0, \boldsymbol{u}_0)\|$ is given by

$$\max\left\{\|\boldsymbol{u}_0\|, \frac{\sigma_{max}(C)F + Y}{\lambda_{min}(A)}\right\} \leq \max\left\{U, \frac{\sigma_{max}(C)F + Y}{\lambda_{min}(A)}\right\}$$

and solutions of the system will remain bounded, as required. If in addition the system initialization satisfies

$$\|\boldsymbol{u}_0\| \leq \frac{\sigma_{max}(C)F + Y}{\lambda_{min}(A)}$$

an upper bound on $\|\boldsymbol{u}(t; t_0, \boldsymbol{u}_0)\|$ is given by

$$\frac{\sigma_{max}(C)F + Y}{\lambda_{min}(A)}$$

D.2 Convergence Proof for CGBSB Network

The proof of convergence or quasiconvergence of trajectories of the CGBSB network starting in the feasible region proceeds along the same lines as the global convergence proof for the BCP network; the necessary steps and proofs are therefore simply outlined below.

Observation D.1 *Trajectories of the CGBSB network for which* $\boldsymbol{u}_0 \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$ *will under the system dynamics remain in this region.*

This observation is readily verified by inspection of (6.23a).

Observation D.2 *Any equilibrium point* $\boldsymbol{u}^*$ *of the CGBSB network for* W *positive semidefinite satisfying* $\boldsymbol{u}^* \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$ *is mapped by* B *onto a constrained minimum of* $J(\boldsymbol{x})$*, and for each constrained minimum there exists a unique corresponding equilibrium point* $\boldsymbol{u}^* \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$.

Proof. The proof involves establishing that the network output at equilibrium satisfies the Kuhn-Tucker conditions for a constrained optimum. Let $P{:}\mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ be defined such that

$$P_{ij}(\boldsymbol{u}) = \begin{cases} 0 & j \neq i \\ \begin{cases} 1 & u_i \leq \zeta_i \quad \text{and} \quad \eta_i > 0 \\ 1 & u_i \in (\zeta_i, \xi_i) \\ 1 & u_i \geq \xi_i \quad \text{and} \quad \eta_i < 0 \\ 0 & \text{otherwise} \end{cases} & j = i \end{cases} \tag{D.1}$$

permitting the expression of (6.23a) as

$$\dot{\boldsymbol{u}} = P(\boldsymbol{u})\boldsymbol{\eta}(\boldsymbol{u}) = -P(\boldsymbol{u})\Gamma\nabla J \tag{D.2}$$

Setting $\dot{\boldsymbol{u}} = 0$ to locate the equilibrium points yields

$$\nabla_i J|_{\boldsymbol{u}=\boldsymbol{u}^*} \begin{cases} \leq 0 & u_i \leq \zeta_i \\ = 0 & u_i \in (\zeta_i, \xi_i) \\ \geq 0 & u_i \geq \xi_i \end{cases}$$

which upon expression in terms of $\boldsymbol{x}$ yields the required Kuhn-Tucker conditions provided $\boldsymbol{u}_0 \in [\boldsymbol{\zeta}, \boldsymbol{\xi}]$.

The existence and uniqueness of an equilibrium point corresponding to a given constrained optimum are both guaranteed by the invertibility of the mapping B.

Theorem D.1 *Trajectories of the CGBSB network converge to the set of equilibrium points.*

Proof. Trajectories of the CGBSB network are clearly continuous and bounded. Assuming that they are also unique[1] it is sufficient to show that $J(\boldsymbol{x})$ is a Liapunov function for the network, and that it has zero time derivative only at equilibrium points of the system. Since diagonal matrices commute and $P^T(\boldsymbol{u})P(\boldsymbol{u}) = P(\boldsymbol{u})$ it follows from (D.2) that $\dot{\boldsymbol{u}} = P^T(\boldsymbol{u})\dot{\boldsymbol{u}}$. Therefore

$$\frac{dJ}{dt} = \nabla^T J \dot{\boldsymbol{x}} = -\dot{\boldsymbol{u}}^T \Gamma^{-1} B \dot{\boldsymbol{u}}$$

[1] which is not obvious since $\dot{\boldsymbol{u}}$ is no longer Lipschitzian, but can be argued through consideration of the constrained gradient descent strategy implemented by the network, as explained in Section 6.7

so that since Γ, B are both positive diagonal, $J(\boldsymbol{x})$ is a Liapunov function for the network as required, and has zero time derivative only at equilibrium points of the system.

Since by Observation D.1 trajectories starting in the feasible region will remain in this region, such trajectories therefore converge to the set of constrained minima of J.

D.3 Uniform Delay on Lateral Connections of BCP Network

We now examine the effect on the BCP network of the introduction of a uniform delay on all lateral connections.

Observation D.3 *The network described by the time-delayed differential equation*

$$\dot{\boldsymbol{u}} = \boldsymbol{y}(t) - C\boldsymbol{f}(\boldsymbol{u}(t-\tau)) - A\boldsymbol{u}(t) \qquad (D.3)$$

has the same equilibrium points as the BCP network.

Proof. For a given time-invariant input $\boldsymbol{y}$, the network at equilibrium at some time t_1 must by definition satisfy

$$\boldsymbol{u}(t-\tau) = \boldsymbol{u}(t) \quad \forall t > t_1 + \tau$$

However, with this substitution, the differential equation (D.3) reduces to that governing the BCP network with un-delayed lateral connections, and hence has the same equilibrium points, as required.

Theorem D.2 *The network described by (D.3) is exponentially asymptotically stable for A positive diagonal provided*

$$\eta \triangleq \lambda_{min}(A) - e^{\lambda_{min}(A)\tau}\sigma_{max}(C) > 0 \qquad (D.4)$$

Proof. Since $\boldsymbol{f}(\boldsymbol{u})$ is continuous and Lipschitzian, the solution $\boldsymbol{u}(t) = \boldsymbol{u}(t; t_0, \boldsymbol{u}(T))$ for all $T \in [t_0 - \tau, t_0]$ of the initial value problem associated with (D.3) is continuous and unique. The solution is given by

$$\boldsymbol{u}(t) = e^{-A(t-t_0)}\boldsymbol{u}(t_0) + \int_{t_0}^{t} e^{-A(t-s)}\boldsymbol{y}(s)\,ds - \int_{t_0}^{t} e^{-A(t-s)}C\boldsymbol{f}(\boldsymbol{u}(s-\tau))\,ds \qquad (D.5)$$

which can be shown — by a trivial modification of the proof in Appendix D.1 — to be bounded for A positive diagonal. Subtracting $\boldsymbol{u}(t)$ from a second trajectory $\boldsymbol{v}(t)$ starting from $\boldsymbol{v}(t_0)$ at time t_0 and taking the norm yields

$$\begin{aligned} \|\boldsymbol{v}(t) - \boldsymbol{u}(t)\| \;\leq\; & \|e^{-A(t-t_0)}\|\|\boldsymbol{v}(t_0) - \boldsymbol{u}(t_0)\| + \\ & \int_{t_0}^{t} \|e^{-A(t-s)}\|\|C\|\|\boldsymbol{f}(\boldsymbol{v}(s-\tau)) - \boldsymbol{f}(\boldsymbol{u}(s-\tau))\|\,ds \end{aligned} \qquad (D.6)$$

Choosing the Euclidean vector norm, substituting the spectral norms given in the proof of Theorem 6.2 into (D.6) and multiplying both sides by $e^{\lambda_{min}(A)t}$ gives

$$\begin{aligned} \|\boldsymbol{v}(t) - \boldsymbol{u}(t)\| e^{\lambda_{min}(A)t} \;\leq\; & \|\boldsymbol{v}(t_0) - \boldsymbol{u}(t_0)\| e^{\lambda_{min}(A)t_0} + \\ & \sigma_{max}(C) \int_{t_0}^{t} \|\boldsymbol{v}(s-\tau) - \boldsymbol{u}(s-\tau)\| e^{\lambda_{min}(A)s}\, ds \end{aligned} \tag{D.7}$$

Changing the variable in the integral and using the fact that the integrand is non-negative gives

$$\int_{t_0}^{t} \|\boldsymbol{v}(s-\tau) - \boldsymbol{u}(s-\tau)\| e^{\lambda_{min}(A)s}\, ds \leq e^{\lambda_{min}(A)\tau} \int_{t_0-\tau}^{t} \|\boldsymbol{v}(\rho) - \boldsymbol{u}(\rho)\| e^{\lambda_{min}(A)\rho}\, d\rho$$

Substituting this result into (D.7), applying Gronwall's Inequality (Reinhard, 1986) and evaluating the integral in the resulting expression produces the following inequality

$$\|\boldsymbol{v}(t) - \boldsymbol{u}(t)\|_2 \leq e^{\theta\sigma_{max}(C)\tau} \|\boldsymbol{v}(t_0) - \boldsymbol{u}(t_0)\|_2 e^{-\eta(t-t_0)} \tag{D.8}$$

where

$$\theta \;\triangleq\; e^{\lambda_{min}(A)\tau} \tag{D.9a}$$

$$\eta \;\triangleq\; \lambda_{min}(A) - \theta\sigma_{max}(C) \tag{D.9b}$$

A sufficient condition for exponential stability of the solutions is that η is strictly positive. Note that since (6.28) is already a fairly stringent condition on the system matrices A and C, the exponential function θ of τ in (D.9b) means that if at all, exponential convergence is only likely to be guaranteed by Theorem D.2 for small delays on the lateral connections.

The issue of limit cycles in the delayed network however remains to be addressed, awaiting a full convergence analysis along the lines of that performed for the BCP network.

Appendix E

HEBBIAN WEIGHT DEVELOPMENT

Földiák's (1989) weight development equations are

$$W(t+1) = W(t) - \alpha \text{offdiag}\left(\Sigma_Y(t)\right) \tag{E.1a}$$

$$Q(t+1) = Q(t) + \beta\left[T(t)\Sigma_X - \text{diag}(\Sigma_Y(t))Q(t)\right] \tag{E.1b}$$

with associated update equations

$$T(t) \triangleq \left(I - W(t)\right)^{-1} Q(t) \tag{E.2a}$$

$$\Sigma_Y(t) = T(t)\Sigma_X T(t)^T \tag{E.2b}$$

where $W(t)$ is the offdiagonal nxn matrix representing the lateral weights; $Q(t)$ is the nxm feedforward weight matrix; $T(t)$ is the nxm transfer function of the relaxed network; Σ_X and Σ_Y are the mxm input and nxn output (spatial) covariance matrices respectively; I is the nxn identity matrix; α and β are positive scalars $\ll 1$; $W(0)$ is the zero matrix; and Q(0) is randomly initialised. Now $W(t)$ and $Q(t)$ are both time-invariant iff both of the following hold

$$\Sigma_Y = \text{diag}(\Sigma_Y) \tag{E.3a}$$

$$T\Sigma_X = \text{diag}(\Sigma_Y)Q \tag{E.3b}$$

Postmultiplying (E.3b) by T^T and substituting using (E.3a) gives

$$\Sigma_Y(I - QT^T) = 0 \tag{E.4}$$

Now Földiák found that the n rows of T developed to span the subspace spanned by the n most dominant eigenvectors of Σ_X. Thus we can write T as follows

$$T = \left[\begin{array}{cc} A & O \end{array}\right] X$$

where A is an nxn matrix of rank n, O is the nx$(m-n)$ zero matrix, and the rows of X are the n eigenvectors of Σ_X in descending order of corresponding eigenvalues. Thus we can now write

$$\Sigma_Y = T\Sigma_X T^T = \left[\begin{array}{cc} A & O \end{array}\right] X\Sigma_X X^T \left[\begin{array}{c} A^T \\ O^T \end{array}\right] \tag{E.5}$$

Now $X\Sigma_X X^T$ is the diagonal matrix containing the eigenvalues of Σ_X in descending order. Since Σ_X is a covariance matrix, it must be non-negative definite, and hence has

all non-negative eigenvalues. Let us assume that Σ_X is of rank r such that $n \leq r \leq m$, and define the nxn matrix B to be the diagonal matrix containing the first n eigenvalues of Σ_X, and Z to be the $(m-n)$x$(m-n)$ diagonal matrix containing the remaining $(r-n)$ non-zero eigenvalues as its first $(r-n)$ diagonal entries, and zeros elsewhere. Then from (E.5)

$$\Sigma_Y = \begin{bmatrix} A & O \end{bmatrix} \begin{bmatrix} B & O \\ O^T & Z \end{bmatrix} \begin{bmatrix} A^T \\ O^T \end{bmatrix} = ABA^T \tag{E.6}$$

Now since A and B are both of full rank, then so is the right hand side of (E.6), giving that Σ_Y is also of full rank. This gives

$$QT^T = I \tag{E.7}$$

as the only possible solution to (E.4). Now substituting (E.2a) into (E.7) we have

$$QQ^T[(I-W)^{-1}]^T = I \tag{E.8}$$

Since W is initialised symmetrically, and according to (E.1a) receives only symmetrical updates, $I-W$ is symmetric, and so is its inverse. Note that if the inverse does not exist, then the weight update algorithm fails in trying to evaluate T, so we only consider the case where it does exist. Thus (E.8) becomes

$$QQ^T = I - W$$

which is the desired result.

BIBLIOGRAPHY

Adelson, E. H. & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America: A*, *2*(2), 284–99.

Albrecht, D. G., de Valois, R. L., & Thorell, L. G. (1980). Visual cortical neurons: Are bars or gratings the optimal stimuli ? *Science*, *207*(4426), 88–90.

Albrecht, D. G. & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, *48*(1), 217–37.

Anderson, J. A., Pellionisz, A., & Rosenfeld, E. (Eds.). (1990). *Neurocomputing 2: Directions for Research.* Cambridge, Massachusetts; London, England: MIT Press.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–51.

Andrews, B. W. & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology*, *287*, 163–76.

Atick, J. J. & Redlich, A. N. (1990a). Mathematical model of the simple cells in the visual cortex. *Biological Cybernetics*, *63*(2), 99–109.

Atick, J. J. & Redlich, A. N. (1990b). Towards a theory of early visual processing. *Neural Computation*, *2*(3), 308–20.

Atick, J. J. & Redlich, A. N. (1991). Predicting ganglion and simple cell receptive field organizations. *International Journal of Neural Systems*, *1*(4), 305–16.

Atick, J. J. & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*(2), 196–210.

Attneave, F. (1954). Informational aspects of visual perception. *Psychological Review*, *61*, 183–93.

Bargmann, V., Butera, P., Girardello, L., & Klauder, J. R. (1971). On the completeness of the coherent states. *Reports on Mathematical Physics*, *2*(4), 221–8.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *Mechanisation of Thought Processes.* National Physical Laboratory Symposium, H.M. Stationery Office, London.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. (Ed.), *Sensory Communication*. Cambridge, Massachusetts; London, England: MIT Press.

Barlow, H. B. (1969a). Pattern recognition and the responses of sensory neurons. *Annals of the New York Academy of Sciences, 156*(2), 872–81.

Barlow, H. B. (1969b). Trigger features, adaptation, and economy of impulses. In Leibovic, K. N. (Ed.), *Information Processing in the Nervous System*, chapter 11, pages 209–26. Berlin, New York, London, Tokyo: Springer-Verlag.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*, 371–94.

Barlow, H. B. (1981). Critical limiting factors in the design of the eye and visual cortex. *Proceedings of the Royal Society of London: B, 212*, 1–34. The 1980 Ferrier Lecture.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanisms of binocular depth discrimination. *Journal of Physiology, 193*, 327–42.

Barlow, H. B., Hawken, M., Kaushal, T. P., & Parker, A. J. (1987). Human contrast discrimination and the contrast discrimination of cortical neurons. *Journal of the Optical Society of America: A, 4*, 2366–71.

Bastiaans, M. (1980). Gabor's expansion of a signal into Gaussian elementary functions. *Proceedings of the IEEE, 68*, 538–9.

Bazaraa, M. S. & Shetty, C. M. (1979). *Nonlinear Programming: Theory and Algorithms*. New York, London, Sydney: John Wiley & Sons.

Beck, H., Bergondy, D., Brown, J., & Sari-Sarraf, H. (1990). Multiresolution segmentation of forward looking IR and SAR imagery using neural networks. In *Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, volume 1381 of *Proceedings of the SPIE*, pages 600–9.

Ben-Israel, A. & Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications*. Wiley-Interscience: Pure & Applied Mathematics. New York, London, Sydney: John Wiley & Sons.

Berman, N. J., Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1991). Mechanisms of inhibition in cat visual cortex. *Journal of Physiology, 440*, 697–722.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science, 252*(5014), 1854–7.

Bishop, P. O. & Henry, G. H. (1972). Striate neurons: Receptive field concepts. *Investigative Opthalmology, 11*(5), 346–54.

Blakemore, C. & Campbell, J. G. (1969). On the existence in the human visual system of neurons selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237–60.

Bonds, A. B. (1992). Spatial and temporal nonlinearities in receptive fields in the cat striate cortex. In Pinter & Nabet (1992), chapter 12, pages 329–52.

Bossomaier, T. & Snyder, A. W. (1986). Why spatial frequency processing in the visual cortex? *Vision Research, 26*(8), 1307–9.

Bouzerdoum, A. (1991). *Nonlinear Lateral Inhibitory Neural Networks: Analysis and Application to Motion Detection.* Seattle, usa, Department of Electrical Engineering, University of Washington, Department of Electrical Engineering.

Bouzerdoum, A. & Pattison, T. R. (1993a). Constrained quadratic optimisation using neural networks. In Leong & Jabri (1993), pages 10–3.

Bouzerdoum, A. & Pattison, T. R. (1993b). Neural network for quadratic optimization with bound constraints. *IEEE Transactions on Neural Networks, 4*(2), 293–304.

Bovik, A. C. (1991). Analysis of multichannel narrow-band filters for image texture segmentation. *IEEE Transactions on Signal Processing, 39*(9), 2025–43.

Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 12*(1), 55–73.

Bracewell, R. N. (1986). *The Fourier Transform and its Applications* (Second Revised ed.). McGraw-Hill Series in Electrical Engineering. New York, London, Sydney: McGraw-Hill International Book Company.

Brown, T. H., Kairiss, E. W., & Keenan, C. L. (1990). Hebbian synapses: Biophysical mechanisms. *Annual Review of Neuroscience, 13*, 475–511.

Buhmann, J., Lange, J., & von der Malsburg, C. (1989). Distortion invariant object recognition by matching heirarchically labeled graphs. In *Proceedings of the IEEE International Conference on Neural Networks*, volume 1, pages 155–9.

Camarda, R. M., Peterhans, E., & Bishop, P. O. (1985a). Simple cells in cat striate cortex: Responses to stationary flashing and to moving light bars. *Experimental Brain Research, 60*, 151–8.

Camarda, R. M., Peterhans, E., & Bishop, P. O. (1985b). Spatial organization of subregions in receptive fields of simple cells in cat striate cortex as revealed by stationary flashing bars and moving edges. *Experimental Brain Research*, *60*(1), 136–50.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology*, *203*(1), 223–35.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551–66.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *PAMI-8*(6), 679–98.

Carlson, A. B. (1986). *Communications Systems: An Introduction to Signals and Noise in Electrical Communication* (Third ed.). McGraw-Hill Series in Electrical Engineering. New York, London, Sydney: McGraw-Hill International Book Company. International Student Edition.

Chen, J., Shanblatt, M. A., & Maa, C.-Y. (1992). Improved neural networks for linear and non-linear programming. *International Journal of Neural Systems*, *2*(4), 331–9.

Chhabra, A. K. & Grogan, T. A. (1989). Depth from stereo: Variational theory and a hybrid analog-digital network. In *Image Understanding & the Man-Machine Interface II*, volume 1076 of *Proceedings of the SPIE*, pages 131–8.

Chhabra, A. K. & Grogan, T. A. (1990). Vision, analog networks, and the minimum norm constraint. In *International Joint Conference on Neural Networks*, volume II, pages 973–979.

Cichocki, A. & Unbehauen, R. (1992). Neural networks for solving systems of linear equations and related problems. *IEEE Transactions on Circuits & Systems*, *CAS-39*(2), 124–38.

Clark, M. & Bovik, A. C. (1989). Experiments in segmenting texton patterns using localized spatial filters. *Pattern Recognition*, *22*(6), 707–17.

Cohen, D. & Shawe-Taylor, J. (1990). Daugman's Gabor transform as a simple generative back-propagation network. *Electronics Letters*, *26*(16), 1241–3.

Crawford, M. L. J., Andersen, R. A., Blake, R., Jacobs, G. H., & Neumeyer, C. (1990). Interspecies comparisons in the understanding of human visual perception. In Spillmann & Werner (1990), chapter 4, pages 23–52.

Crick, F. & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In Rumelhart & McClelland (1986b), chapter 20, pages 333–71.

Culhane, A. D., Peckerar, M. C., & Marrian, C. R. K. (1989). A neural network approach to discrete Hartley and Fourier transforms. *IEEE Transactions on Circuits & Systems, CAS-36*(5), 695–703.

Curtis, S. R., Shitz, S., & Oppenheim, A. (1989). Reconstruction of two-dimensional signals from zero-crossings. *IEEE Transactions on Acoustics, Speech, & Signal Processing.*

Dale, H. H. (1935). Pharmacology and nerve endings. *Proceedings of the Royal Society of Medicine, 28*, 319–32.

Daubechies, I., Grossman, A., & Meyer, Y. (1986). Painless nonorthogonal expansions. *Journal of Mathematical Physics, 27*(5), 1271–83.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*, 847–56.

Daugman, J. G. (1983). Six formal properties of two-dimensional anisotropic visual filters: Structural principles and frequency/orientation selectivity. *IEEE Transactions on Systems, Man & Cybernetics, SMC-13*(5), 882–7.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America: A, 2*(7), 1160–9.

Daugman, J. G. (1988a). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, & Signal Processing, 36*(7), 1169–79.

Daugman, J. G. (1988b). Pattern and motion vision without Laplacian zero crossings. *Journal of the Optical Society of America: A, 5*(7), 1142–8.

Daugman, J. G. (1989a). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering, BME-36*(1), 107–14.

Daugman, J. G. (1989b). Non-orthogonal wavelet representations in relaxation networks: Image encoding and analysis with biological visual primitives. In *New Developments in Neural Computing*, pages 233–50. IOP Publishing.

Daugman, J. G. (1989c). Relaxation neural network for complete discrete 2-D Gabor transforms. In Pietikäinen, M. & Röning, J. (Eds.), *Proceedings of the Sixth Scandinavian Conference on Image Analysis.*

Daugman, J. G. (1990). An information-theoretic view of analog representation in striate cortex. In Schwartz (1990), pages 403–23.

Daugman, J. G. (1993). Quadrature-phase simple-cell pairs are appropriately described in complex analytic form. *Journal of the Optical Society of America: A*, *10*(2), 375–377.

Daugman, J. G. & Kammen, D. M. (1986). Image statistics, gases, and visual neural primitives. In *Proceedings of the IEEE International Conference on Neural Networks*, volume IV, pages 163–174.

Davis, B. & Pattison, T. R. (1992). Error in proof of exponential convergence. *Neural Networks*, *5*(6), 869.

de Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in Macaque visual cortex. *Vision Research*, *22*(5), 545–59.

de Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in Macaque visual cortex. *Vision Research*, *22*(5), 531–44.

Dehay, C., Douglas, R. J., Martin, K. A. C., & Nelson, C. (1991). Excitation by geniculocortical synapses is not 'vetoed' at the level of dendritic spines in cat visual cortex. *Journal of Physiology*, *440*, 723–34.

den Broeder, C. G. J. & Charnes, A. (1957). Contributions to the theory of generalized inverses for matrices. Technical report, Purdue University, Lafayette, Indiana. Reprinted 1962: *Office of Naval Research* Res. Memo 39, Northwestern University, Evanston Illinois.

Douglas, R. J. & Martin, K. A. C. (1991). A functional microcircuit for cat visual cortex. *Journal of Physiology*, *440*, 735–69.

Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1988). Selective responses of visual cortical cells do not depend on shunting inhibition. *Nature*, *332*, 642–4.

du Buf, J. M. H. (1990). Gabor phase in texture discrimination. *Signal Processing*, *21*, 221–240.

Eckmiller, R. (Ed.). (1990). *Advanced Neural Computers.* North Holland: Elsevier Science Publishers B.V.

Emerson, R. C. & Citron, M. C. (1988). How linear and nonlinear mechanisms contribute to directional selectivity in simple cells of cat striate cortex. *Investigative Opthalmology and Visual Science*, *29*, 23. Annual Meeting Abstract Issue.

Emerson, R. C. & Citron, M. C. (1989). Linear and nonlinear mechanisms of motion selectivity in single neurons of the cat's visual cortex. In Kleinman, D. L. (Ed.), *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 448–53, Cambridge, MA.

Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *Journal of Neuroscience*, *8*(4), 1172–80.

Ferster, D. & Koch, C. (1987). Neuronal connections underlying orientation selectivity in cat visual cortex. *Trends in Neuroscience*, *10*(12), 487–92.

Ferster, D. & Lindström, S. (1983). An intracellular analysis of geniculocortical connectivity in area 17 of the cat. *Journal of Physiology*, *342*, 181–215.

Ferster, D. & Lindström, S. (1985). Synaptic excitation of neurones in area 17 of the cat by intracortical axon collaterals of cortico-geniculate cells. *Journal of Physiology*, *367*, 233–52.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America: A*, *4*(12), 2379–94.

Field, D. J. & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London: B*, *228*, 379–400.

Finlay, B. L., Schiller, P. H., & Volman, S. F. (1976). Meridional differences in orientation sensitivity in monkey striate cortex. *Brain Research*, *105*(2), 350–2.

Flaton, K. A. & Toborg, S. T. (1989). An approach to image recognition using sparse filter graphs. In *Proceedings of the IEEE International Conference on Neural Networks*, volume I, pages 313–20.

Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminators. *Biological Cybernetics*, *61*, 103–13.

Földiák, P. (1989). Adaptive network for optimal linear feature extraction. In *Proceedings of the IEEE International Conference on Neural Networks*, volume I, pages 401–5.

Földiák, P. (1992). Models of sensory coding. Technical report, Physiological Laboratory, University of Cambridge. *email*:peter@psy.oxford.ac.uk.

Foster, K. H., Gaska, J. P., Marčelja, S., & Pollen, D. A. (1983). Phase relationships between adjacent simple cells in the feline visual cortex. *Journal of Physiology*, 22P.

Freeman, R. D. & Ohzawa, I. (1990a). Binocular mechanisms in the normal and abnormal visual cortex of the cat. In Blakemore, C. (Ed.), *Vision: Coding and Efficiency*, chapter 27, pages 291–301. Cambridge, New York, Melbourne: Cambridge University Press.

Freeman, R. D. & Ohzawa, I. (1990b). On the neurophysiological organization of binocular vision. *Vision Research*, *30*(11), 1661–76.

Gabor, D. (1946). Theory of communication. *Journal of the Institute of Electrical Engineers*, *93*(21), 429–57. Part III.

Galatsanos, N. P. & Katsaggelos, A. K. (1992). Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, *1*(3), 322–36.

Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics.* New York, London, Sydney: John Wiley & Sons.

Geisler, W. S. & Hamilton, D. B. (1986). Sampling theory analysis of spatial vision. *Journal of the Optical Society of America: A*, *3*(1), 62–70.

Gellert, W., T.S., G., Hellwich, M., Kastner, H., & Kustner, H. (Eds.). (1989). *The VNR Concise Encyclopedia of Mathematics* (Second ed.). New York: Van Nostrand Reinhold Company.

Gilbert, C. D. (1983). Microcircuitry of the visual cortex. *Annual Review of Neuroscience*, *6*, 217–47.

Glezer, V. D., Gauzelman, V. E., & Yakovlev, V. V. (1989). Spatial organization of subfields in receptive fields of cells in cat striate cortex. *Vision Research*, *29*(7), 777–88.

Glezer, V. D., Tsherbach, T. A., Gauselman, V. E., & Bondarko, V. M. (1980). Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex: A model of the field. *Biological Cybernetics*, *37*(3), 195–208.

Glezer, V. D., Tsherbach, T. A., Gauselman, V. E., & Bondarko, V. M. (1982). Spatio-temporal organization of receptive fields of the cat striate cortex: The receptive fields as the grating filters. *Biological Cybernetics*, *43*(1), 35–49.

Golden, R. M. (1992). Stability and optimization analyses of the generalized brain-state-in-a-box neural network model. *Journal of Mathematical Psychology.* In press.

Golub, G. H. & Van Loan, C. F. (1989). *Matrix Computations* (Second ed.). Johns Hopkins Series in the Mathematical Sciences. Baltimore, London: The Johns Hopkins University Press.

Gopal, N., Bovik, A. C., & Ghosh, J. (1990). Multiple channel surface orientation from texture. In *Human Vision and Electronic Imaging: Models, Methods, and Applications*, volume 1249 of *Proceedings of the SPIE*, pages 366–75.

Greenbaum, A. & Rodrigue, G. H. (1989). Optimal preconditioners of a given sparsity pattern. *BIT*, *29*, 610–34.

Greville, T. N. E. (1960). Some applications of the pseudoinverse of a matrix. *SIAM Review*, *2*, 15–22.

Grimson, W. E. L. (1981). *From Images to Surfaces: A Computational Study of the Human Early Visual System.* Cambridge, Massachusetts; London, England: MIT Press.

Grossberg, S. (1969). On learning and energy-entropy in recurrent and non-recurrent signed networks. *Journal of Statistical Physics*, *1*, 319–50.

Grossberg, S., Mingolla, E., & Todorovič, D. (1989). A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering*, *BME-36*(1), 65–84.

Grzywacz, N. M. & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London: B*, *239*, 129–61.

Gutschow, T. & Hecht-Nielsen, R. (1991). Processing complexity of two approaches to object detection and recognition. In *Artificial Neural Network Approaches in Guidance and Control*, NATO Advisory Group for Aerospace Research and Development: Lecture Series 179, chapter 4, pages 1–11. Monterey, U.S.; Kjeller, Norway; Neubiberg, Germany: NATO.

Hager, W. W. (1988). *Applied Numerical Linear Algebra.* New Jersey: Prentice Hall.

Hamilton, D. B., Albrecht, D. G., & Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: Spatial and temporal phase transfer function. *Vision Research*, *29*(10), 1285–1308.

Hammond, P. (1991). Binocular phase specificity of striate cortical neurons. *Experimental Brain Research*, *87*(3), 615–23.

Hammond, P. & Pomfrett, C. J. D. (1990). Visual cortical neurones in the anaesthetised cat: Binocular mismatch of spatial frequency tuning and directionality. *Journal of Physiology*, *426*, 109P.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, *121*, 400–15.

Hata, Y., Tsumoto, T., Sato, H., Hagihara, K., & Tamura, H. (1988). Inhibition contributes to orientation selectivity in visual cortex of cat. *Nature*, *335*(6193), 815–7.

Hawken, M. J. & Parker, A. J. (1987). Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal Society of London: B*, *231*, 251–88.

Hebb, D. O. (1949). *The Organisation of Behaviour*. New York, London, Sydney: John Wiley & Sons.

Heeger, D. (1987). A model for the extraction of image flow. *Journal of the Optical Society of America: A*, *4*(8), 1455–71.

Heggelund, P. (1986a). Quantitative studies of enhancement and suppression zones in the receptive field of simple cells in cat striate cortex. *Journal of Physiology*, *373*, 293–310.

Heggelund, P. (1986b). Quantitative studies of the discharge fields of single cells in cat striate cortex. *Journal of Physiology*, *373*, 277–92.

Heggelund, P. & Albus, K. (1978). Orientation selectivity of single cells in the striate cortex of cat. *Vision Research*, *18*(8), 1067–71.

Heggelund, P., Krekling, S., & Skottun, B. C. (1983). Spatial summation in the receptive field of simple cells in the cat striate cortex. *Experimental Brain Research*, *52*(1), 87–98.

Heil, C. E. & Walnut, D. F. (1989). Continuous and discrete wavelet transforms. *SIAM Review*, *31*(4), 628–66.

Henry, G. H. (1977). Receptive field classes of cells in the striate cortex of the cat. *Brain Research*, *133*, 1–28.

Henry, G. H. (1985). Physiology of cat striate cortex. In Peters, A. & Jones, E. G. (Eds.), *The Cerebral Cortex*, volume 3: Visual Cortex, pages 119–55. New York, London: Plenum Press.

Henry, G. H. (1991). Afferent inputs, receptive field properties and morphological cell types in different laminae of the striate cortex. In Leventhal (1991), pages 223–245.

Henry, G. H. (1993). Personal communication. Visual Neurosciences Group, John Curtin School of Medical Research, ANU, Canberra, Australia.

Henry, G. H. & Bishop, P. O. (1972). Striate neurons: Receptive field organization. *Investigative Opthalmology*, *11*(5), 357–36?

Henry, G. H., Dreher, B., & Bishop, P. O. (1974). Orientation selectivity of cells in cat striate cortex. *Journal of Neurophysiology*, *37*(6), 1394–1409.

Henry, G. H., Goodwin, A. W., & Bishop, P. O. (1978). Spatial summation of responses in receptive fields of single cells in cat striate cortex. *Experimental Brain Research*, *32*(2), 245–66.

Henry, G. H., Mustari, M. J., & Bullier, J. (1983). Different geniculate inputs to B and C cells of cat striate cortex. *Experimental Brain Research*, *52*, 179–189.

Higgins, J. R. (1977). *Completeness and Basis Properties of Sets of Special Functions.* Cambridge, New York, Melbourne: Cambridge University Press.

Hlawatsch, F. & Boudreaux-Bartels, G. F. (1992). Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 21–67.

Hodgkin, A. L. & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.

Honavar, V. (1990). Perceptual development and learning: From behavioral, neurophysiological, and morphological evidence to computational models. Anonymous ftp from archive.cis.ohio-state.edu/pub/neuroprose/ honavar.develop-learn.ps.Z, Computer Sciences Department, University of Wisconsin-Madison.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science*, *81*, 3088–92.

Hopfield, J. J. & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, *52*, 141–52.

Hopfield, J. J. & Tank, D. W. (1986). Computing with neural circuits: A model. *Science*, *233*, 625–33.

Horn, R. A. & Johnson, C. R. (1988). *Matrix Analysis.* Cambridge, New York, Melbourne: Cambridge University Press.

Horn, R. A. & Johnson, C. R. (1991). *Topics in Matrix Analysis.* Cambridge, New York, Melbourne: Cambridge University Press.

Hubel, D. H. & Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *Journal of Neurophysiology, 7*, 3378–415.

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, 148*, 574–91.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *Journal of Physiology, 160*, 106–54.

Hubel, D. H. & Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate areas (18 and 19) of the cat. *Journal of Neurophysiology, 28*, 229.

Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*(1), 215.

Huggins, W. H. & Licklider, J. C. R. (1951). Place mechanisms of auditory frequency analysis. *Journal of the Acoustical Society of America, 23*, 290–9.

Hungenahally, S. K. (1991). *Discriminant Functions: Emulation and Generalization of Neuronal-Morphology of Biological Visual Receptive Fields.* PhD thesis, Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, Canada.

Hungenahally, S. K., Harrison, H. B., & Agnew, G. (1992). Emulation of the neuro-morphological functions of biological visual receptive fields for medical image processing. In Leong & Jabri (1992), pages 131–4.

Hungenahally, S. K., Postula, A., & Jain, L. C. (1993). Neuro-morphology of biological vision: Fractional discriminant functions in the emulation of visual receptive fields for remote sensed images. In Leong & Jabri (1993), pages 146–9.

Ikeda, H. & Wright, M. J. (1975a). Retinotopic distribution, visual latency and orientation tuning of 'sustained' and 'transient' cortical neurones in area 17 of the cat. *Experimental Brain Research, 22*(4), 385–98.

Ikeda, H. & Wright, M. J. (1975b). Spatial and temporal properties of 'sustained' and 'transient' neurones in area 17 of the cat's visual cortex. *Experimental Brain Research, 22*(4), 363–83.

Jain, A. K. & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Psychological Review, 24*(12), 1167–86.

Jones, J. P. (1991). Personal communication.

Jones, J. P. & Palmer, L. A. (1987a). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1233–58.

Jones, J. P. & Palmer, L. A. (1987b). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1187–211.

Jones, J. P., Stepnoski, A., & Palmer, L. A. (1987). The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1212–32.

Juang, B., Kung, S., & Kamm, C. A. (Eds.). (1991). *Neural Nets and Signal Processing: Proceedings of the 1991 IEEE Workshop*. IEEE Press.

Kennedy, M. P. & Chua, L. O. (1988). Neural networks for linear and non-linear programming. *IEEE Transactions on Circuits & Systems*, *CAS-35*, 554–62.

Klein, S. A. & Beutter, B. (1992). Minimizing and maximizing the joint space – spatial frequency uncertainty of gabor-like functions: Comment. *Journal of the Optical Society of America: A*, *9*(2), 337–340.

Koenderink, J. J. & van Doorn, A. J. (1990a). Receptive field families. *Biological Cybernetics*, *63*(4), 291–7.

Koenderink, J. J. & van Doorn, A. J. (1990b). Receptive field taxonomy. In Eckmiller (1990), pages 295–301.

Kreyszig, E. (1983). *Advanced Engineering Mathematics* (5th ed.). New York, London, Sydney: John Wiley & Sons.

Ku, T.-K. & Kuo, C.-C. J. (1992). Design and analysis of Toeplitz preconditioners. *IEEE Transactions on Signal Processing*, *40*(1), 129–41.

Kulikowski, J. J. & Bishop, P. O. (1981a). Fourier analysis and spatial representation in the visual cortex. *Experientia*, *37*(2), 160–3.

Kulikowski, J. J. & Bishop, P. O. (1981b). Linear analysis of the responses of simple cells in the cat visual cortex. *Experimental Brain Research*, *44*, 371–85.

Kulikowski, J. J., Bishop, P. O., & Kato, H. (1981). Spatial arrangements of responses by cells in the cat visual cortex to light and dark bars and edges. *Experimental Brain Research*, *44*, 371–85.

Kulikowski, J. J., Marčelja, S., & Bishop, P. (1982). Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics*, *43*, 187–98.

Kulikowski, J. J., Movshon, J. A., & Robson, J. G. (1980). Personal communication to Marčelja (1980).

Lancaster, P. & Tismenetsky, M. (1985). *The Theory of Matrices: Second Edition with Applications.* Computer Science and Applied Mathematics. San Diego, London, Sydney, Tokyo: Academic Press Inc.

LaSalle, J. P. (1968). Stability theory for ordinary differential equations. *Journal of Differential Equations, 4*, 57–65.

LaSalle, J. P. (1976). *The Stability of Dynamical Systems.* Philadelphia: SIAM.

Lawton, T. B. (1989). Outputs of paired Gabor filters summed across the background frame of reference predict the direction of movement. *IEEE Transactions on Biomedical Engineering, BME-36*(1), 130–9.

Lee, D., Papageorgiou, A., & Wasilkowski, G. W. (1988). Computational aspects of determining optical flow. In *Proceedings of the Second International Conference on Computer Vision*, pages 612–7.

Leipnik, R. (1959). Entropy and the Uncertainty Principle. *Information and Control, 2*, 64–79.

Leong, P. & Jabri, M. (Eds.). (1992). *Proceedings of the Third Australian Conference on Neural Networks*, Australian National University, Canberra.

Leong, P. & Jabri, M. (Eds.). (1993). *Proceedings of the Fourth Australian Conference on Neural Networks*, University of Melbourne, Australia.

Leung, M.-T., Engeler, W. E., & Frank, P. (1990). Fingerprint processing using back-propagation neural networks. In *International Joint Conference on Neural Networks*, volume I, pages 15–20, San Diego, California. IEEE/INNS, IEEE Neural Networks Council.

LeVay, S. & Voigt, T. (1988). Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience, 1*, 395–414.

Leventhal, A. G. (Ed.). (1991). *The Neural Basis of Visual Function*, volume 4 of *Cronly-Dillon, J. R. (Ed.), Vision and Visual Dysfunction.* London: The Macmillan Press Ltd.

Levick, W. R. (1993). Personal communication. Visual Neurosciences Group, John Curtin School of Medical Research, ANU, Canberra, Australia.

Linsker, R. (1986a). From basic network principles to neural architecture: Emergence of orientation columns. *Proceedings of the National Academy of Science, 83*, 8779–83.

Linsker, R. (1986b). From basic network principles to neural architecture: Emergence of orientation-selective cells. *Proceedings of the National Academy of Science, 83*, 8390–4.

Linsker, R. (1986c). From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proceedings of the National Academy of Science, 83*, 7508–12.

Linsker, R. (1988). Self-organisation in a perceptual network. *Computer*, 105–17.

Linsker, R. (1989). Designing a sensory processing system: What can be learned from Principal Component Analysis? Technical report, IBM Research Division, T.J.Watson Research Center, Yorktown Heights, NY 10598.

Linsker, R. (1990). Perceptual neural organization: Some approaches based on network models and information theory. *Annual Review of Neuroscience, 13*, 257–81.

Logan, B. F. (1977). Information in the zero-crossings of band-pass signals. *The Bell Systems Technical Journal, 56*, 487–510.

Lund, J. S. (1988). Anatomical organization of *macaque* monkey striate visual cortex. *Annual Review of Neuroscience, 11*, 253–88.

Maa, C.-Y. & Shanblatt, M. A. (1992a). Linear and quadratic programming neural network analysis. *IEEE Transactions on Neural Networks, 3*(4), 580–594.

Maa, C.-Y. & Shanblatt, M. A. (1992b). A two-phase optimization neural network. *IEEE Transactions on Neural Networks, 3*(6), 1003–1009.

MacLennan, B. (1991). Gabor representation of spatiotemporal visual images. Technical Report CS-91-144, Computer Science Department, University of Tennessee, Knoxville, TN 37996. *email*:maclennan@cs.utk.edu.

MacLennan, B. (1992a). Field computation in the brain. Technical Report CS-92-174, Computer Science Department, University of Tennessee, Knoxville, TN 37996. *email*:maclennan@cs.utk.edu.

MacLennan, B. (1992b). Information processing in the dendritic net. Technical Report CS-92-180, Computer Science Department, University of Tennessee, Knoxville, TN 37996. *email*:maclennan@cs.utk.edu.

MacLennan, B. (1993a). Field computation in the brain. In Pribram (1993). Adapted from (MacLennan, 1992a).

MacLennan, B. (1993b). Information processing in the dendritic net. In Pribram (1993). Adapted from (MacLennan, 1992b).

Maffei, L. & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, *13*(7), 1255–67.

Mallat, S. G. (1989a). Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, *37*(12), 2091–110.

Mallat, S. G. (1989b). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *11*(7), 674–93.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, *70*(11), 1297–1300.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London*, *275*(942), 483–524.

Marr, D. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Research Program Bulletin*, *15*(3), 470–88. Neuronal Mechanisms in Visual Perception.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman & Company.

Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London: B*, *207*, 187–217.

Marr, D., Poggio, T., & Ullman, S. (1979). Bandpass channels, zero-crossings, and early visual information processing. *Journal of the Optical Society of America*, *69*(6), 914–6.

Martens, J.-B. (1990). The Hermite transform – theory. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, *38*(9), 1595–606.

Martin, K. A. C. & Whitteridge, D. (1984). Form, function and intracortical projections of spiny neurones in the striate visual cortex of the cat. *Journal of Physiology*, *353*, 463–504.

Maske, R., Yamane, S., & Bishop, P. O. (1984). Binocular simple cells for local stereopsis: Comparison of receptive field organizations for the two eyes. *Vision Research*, *24*(12), 1921–30.

Maske, R., Yamane, S., & Bishop, P. O. (1985). Simple and B-cells in cat striate cortex: Complementarity of responses to moving light and dark bars. *Journal of Neurophysiology*, *53*(3), 670–85.

McGuire, B. A., Hornung, J.-P., Gilbert, C. D., & Wiesel, T. N. (1984). Patterns of synaptic input to layer 4 of cat striate cortex. *Journal of Neuroscience*, *4*(12), 3021–33.

McLean, J. & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research*, *29*(6), 675–9.

Mohanty, N. (1986). *Random Signals Estimations and Identifications: Analysis and Applications.* Van Nostrand Reinhold Electrical/Computer Science and Engineering Series. New York: Van Nostrand Reinhold Company.

Moré, J. J. & Toraldo, G. (1991). On the solution of large quadratic programming problems with bound constraints. *SIAM Journal of Optimization*, *1*(1), 93–113.

Movshon, J. A. (1975). The velocity tuning of single units in cat striate cortex. *Journal of Physiology*, *249*, 445–68.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's striate cortex. *Journal of Physiology*, *283*, 101–20.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology*, *283*, 53–77.

Mullikin, W. H., Jones, J. P., & Palmer, L. A. (1984). Periodic simple cells in cat area 17. *Journal of Neurophysiology*, *53*(2), 372–87.

Murphy, P. C. & Sillito, A. M. (1987). Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, *329*(6141), 727–9.

Nabet, B. & Pinter, R. B. (1991). Sensory neural networks: Lateral inhibition. In Nabet, B. & Pinter, R. B. (Eds.), *Sensory Neural Networks: Lateral Inhibition*, chapter 3, pages 19–26. Boca Raton: CRC Press.

Nomura, M., Matsumoto, G., & Fujiwara, S. (1990). A binocular model for the simple cell. *Biological Cybernetics*, *63*(3), 237–42.

Ogata, K. (1987). *Discrete-Time Control Systems.* New Jersey: Prentice Hall.

Ohzawa, I. & Freeman, R. D. (1986). The binocular organization of simple cells in cat's visual cortex. *Journal of Neurophysiology*, *56*(1), 221–42.

Orban, G. A. (1991). Quantitative electrophysiology of visual cortical neurones. In Leventhal (1991), pages 173–222.

Paler, K. & Bowler, I. W. (1986). Gabor filters applied to electronic speckle pattern interferometer images. Technical report, Rutherford Appleton Laboratories, UK.

Palmer, L. A. & Davis, T. L. (1981a). Comparison of responses to moving and stationary stimuli in cat striate cortex. *Journal of Neurophysiology*, *46*(2), 277–95.

Palmer, L. A. & Davis, T. L. (1981b). Receptive-field structure in cat striate cortex. *Journal of Neurophysiology*, *46*(2), 260–76.

Palmer, L. A., Jones, J. P., & Stepnoski, R. A. (1991). Striate receptive fields as linear filters: Characterization in two dimensions of space. In Leventhal (1991), chapter 10, pages 246–265.

Papoulis, A. (1968). *Systems and Transforms with Aplications in Optics.* New York, London, Sydney: McGraw-Hill International Book Company.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes* (Second ed.). New York, London, Sydney: McGraw-Hill International Book Company.

Pattison, T. R. (1992). Relaxation network for Gabor image decomposition. *Biological Cybernetics*, *67*(1), 97–102.

Pece, A. E. C. (1992). Redundancy reduction of a Gabor representation: A possible computational role for feedback from primary visual cortex to lateral geniculate nucleus. In Aleksander, I. & Taylor, J. (Eds.), *Artificial Neural Networks II: Proceedings of the International Conference on Artificial Neural Networks.*

Peterhans, E., Bishop, P. O., & Camarda, R. M. (1985). Direction selectivity of simple cells in cat striate cortex to moving light bars: I. Relation to stationary flashing bar and moving edge responses. *Experimental Brain Research*, *57*(3), 512–22.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. a. (1967). Binocular interaction on single units in cat striate cortex: Simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research*, *6*(4), 391–410.

Pinter, R. B. & Nabet, B. (Eds.). (1992). *Nonlinear Vision: Determination of Neural Receptive Fields, Function, and Networks.* CRC Press.

Poggio, G. F. (1980). Central neural mechanisms in vision. In Mountcastle, V. B. (Ed.), *Medical Physiology* (Fourteenth ed.). The C.V Mosby Company.

Poggio, G. F. & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving Rhesus monkey. *Journal of Neurophysiology*, *40*(6), 1392–1405.

Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience*, *8*(12), 4531–50.

Poggio, T. & Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analogue networks. *Proceedings of the Royal Society of London: B*, *226*, 303–23.

Pollen, D. & Ronner, S. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, *212*(4501), 1409–11.

Pollen, D. A., Foster, K. H., & Gaska, J. P. (1985). Phase-dependent response characteristics of visual cortical neurons. In Rose, D. & Dobson, V. (Eds.), *Models of the Visual Cortex*, pages 281–91. New York, London, Sydney: John Wiley & Sons.

Pollen, D. A. & Gaska, J. P. (1987). Vision, visual cortex, and frequency analysis. In Adelman, G. (Ed.), *Encyclopedia of Neuroscience*, volume II, pages 1262–3. Boston, Basel, Stuttgart: Birkhäuser.

Pollen, D. A., Lee, J. R., & Taylor, J. H. (1971). How does the striate cortex begin the reconstruction of the visual world? *Science*, *173*, 74–7.

Pollen, D. A. & Ronner, S. F. (1982). Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Research*, *22*, 101–18.

Pollen, D. A. & Ronner, S. F. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man & Cybernetics*, *SMC-13*(5), 907–16.

Pollen, D. A. & Taylor, J. H. (1974). The striate cortex and the spatial analysis of visual space. In Schmitt, F. O. & Worden, F. G. (Eds.), *The Neurosciences: Third Study Program*, pages 239–47. Cambridge, Massachusetts; London, England: MIT Press.

Porat, M. & Zeevi, Y. Y. (1985). The generalised Gabor scheme of image representation. *Acta Polytechnica Scandinavica: Applied Physics Series*, *150*, 166–69.

Porat, M. & Zeevi, Y. Y. (1988). The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *10*(4), 452–68.

Porat, M. & Zeevi, Y. Y. (1989). Localized texture processing in vision: Analysis and synthesis in the Gaborian space. *IEEE Transactions on Biomedical Engineering*, *BME-36*(1), 115–29.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, New York, Melbourne: Cambridge University Press.

Pribram, K. (Ed.). (1993). *Rethinking Neural Networks: Quantum Fields and Biological Data.* Hillsdale, New Jersey: Lawrence E. Erlbaum Associates. In press.

Ramoa, A. S., Shadlen, M., Skottun, B. C., & Freeman, R. D. (1986). A comparison of inhibition in orientation and spatial frequency selectivity of cat visual cortex. *Nature, 321*, 237–9.

Regan, D. (1991). A brief review of some of the stimuli and analysis methods used in spatiotemporal vision research. In Regan, D. (Ed.), *The Neural Basis of Visual Function*, volume 10 of *Cronly-Dillon, J. R. (Ed.), Vision and Visual Dysfunction*, pages 1–42. London: The Macmillan Press Ltd.

Regan, D., Frisby, J. P., Poggio, G. F., Schor, C. M., & Tyler, C. W. (1990). The perception of stereodepth and stereo-motion: Cortical mechanisms. In Spillmann & Werner (1990), pages 317–347.

Reid, R. C., Soodak, R. E., & Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proceedings of the National Academy of Science, 84*(23), 8740–4.

Reinhard, H. (1986). *Differential Equations: Foundations and Applications.* London: North Oxford Academic Publishers.

Robson, J. G. (1983). Frequency domain visual processing. In Braddick, O. J. & Sleigh, A. C. (Eds.), *Physical and Biological Processing of Images*, Springer Series in Information Science, pages 73–87. The Rank Prize Funds, Springer-Verlag:Berlin, New York, London, Tokyo. Proceedings of an International Symposium.

Rodríguez-Vázquez, A., Domínguez-Castro, R., Rueda, A., Huertas, J. L., & Sánchez-Sinencio, E. (1990). Nonlinear switched-capacitor "neural" networks for optimization problems. *IEEE Transactions on Circuits & Systems, 37*, 384–398.

Rose, D. & Blakemore, C. (1974). An analysis of orientation selectivity in the cat's visual cortex. *Experimental Brain Research, 20*(1), 1–17.

Rotem, D. & Zeevi, Y. (1986). Image reconstruction from zero-crossings. *IEEE Transactions on Acoustics, Speech, & Signal Processing, ASSP-34*, 1269–77.

Rubenstein, B. S. & Sagi, D. (1990). Spatial variability as a limiting factor in texture discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America: A*, 7(9), 1632–43.

Rubner, J. & Schulten, K. (1990). Development of feature detectors by self-organisation: a network model. *Biological Cybernetics*, *62*, 193–9.

Rumelhart, D. E. & McClelland, J. L. (Eds.). (1986a). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1: Foundations. Cambridge, Massachusetts; London, England: MIT Press.

Rumelhart, D. E. & McClelland, J. L. (Eds.). (1986b). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 2: Psychological and Biological Models. Cambridge, Massachusetts; London, England: MIT Press.

Sakitt, B. & Barlow, H. B. (1982). A model for the economical encoding of the visual image in cerebral cortex. *Biological Cybernetics*, *43*, 97–108.

Sanger, T. D. (1990). Analysis of the 2-dimensional receptive-fields learned by the Generalized Hebbian Algorithm in response to random input. *Biological Cybernetics*, *63*(3), 221–8.

Schach, C. (1992). An analogue VLSI implementation of a recurrent neural network. Honours project report, University of Adelaide, Dept. Electrical & Electronic Eng. Adelaide 5005, South Australia.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley-Interscience. John Wiley & Sons.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976a). Quantitative studies of single-cell properties in monkey striate cortex: I. spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, *39*(6), 1288–1319.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976b). Quantitative studies of single-cell properties in monkey striate cortex: II. orientation specificity and ocular dominance. *Journal of Neurophysiology*, *39*(6), 1320–1333.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976c). Quantitative studies of single-cell properties in monkey striate cortex: III. spatial frequency. *Journal of Neurophysiology*, *39*(6), 1334–51.

Schwartz, E. (Ed.). (1990). *Computational Neuroscience.* Bradford/MIT Press.

Sejnowski, T. J. (1986). Open questions about computation in cerebral cortex. In Rumelhart & McClelland (1986b), chapter 21, pages 372–89.

Sereno, M. I. (1988). The visual system. In von Seelen, W., Shaw, G., & Leinhos, U. (Eds.), *Organisation of Neural Networks: Structures and Models*, pages 167–84. Weinheim, Cambridge, New York: VCH Verlagsgesellschaft mbH.

Shapley, R., Caelli, T., Grossberg, S., Morgan, M., & Rentschler, I. (1990). Computational theories of visual perception. In Spillmann & Werner (1990), chapter 15, pages 417–448.

Shapley, R. & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, *8*, 547–83.

Shepherd, G. M. (1988). *Neurobiology* (Second ed.). New York, Oxford: Oxford University Press.

Shepherd, G. M. (1990). The significance of real neuron architectures for neural network simulations. In Schwartz (1990), chapter 8.

Sillito, A. M. (1975). The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *Journal of Physiology*, *250*, 305–29.

Sillito, A. M., Kemp, J. A., Milson, J. A., & Berardi, N. (1980). A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain Research*, *194*(2), 517–20.

Skottun, B. C. & Freeman, R. D. (1984). Stimulus specificity of binocular cells in the cat's visual cortex: Ocular dominance and the matching of left and right eyes. *Experimental Brain Research*, *56*, 206–16.

Spillmann, L. & Werner, J. S. (Eds.). (1990). *Visual Perception: The Neurophysiological Foundations.* San Diego, London, Sydney, Tokyo: Academic Press Inc.

Spitzer, H. & Hochstein, S. (1985). Simple- and complex-cell response dependences on stimulation parameters. *Journal of Neurophysiology*, *53*(5), 1244–65.

Stone, J., Dreher, B., & Leventhal, A. (1979). Heirarchical and parallel mechanisms in the organisation of the visual cortex. *Brain Research Review*, *1*, 345–94.

Stork, D. G. & Wilson, H. R. (1990). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *Journal of the Optical Society of America: A*, *7*(8), 1362–73.

Stratford, K., Mason, A., Larkman, A., Major, G., & Jack, J. (1989). The modelling of pyramidal neurones in the visual cortex. In Durbin, R., Miall, C., & Mitchison, G. (Eds.), *The Computing Neuron*, pages 296–321. Wokingham, England: Addison-Wesley.

Sudharsanan, S. I. & Sundareshan, M. K. (1991). Exponential stability and a systematic synthesis of a neural network for quadratic minimization. *Neural Networks*, *4*(5), 599–613.

Tan, T. N. & Constantinides, A. G. (1990). Texture analysis based on a human visual model. In *Proceedings of the International Conference on Acoustics, Speech, & Signal Processing*, volume 4. IEEE Press.

Tank, D. W. & Hopfield, J. J. (1986). Simple "neural" optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits & Systems*, *CAS-33*(5), 533–41.

Taylor, J. G. (1990). Neural modelling of vision and olfaction. In Eckmiller (1990), pages 313–22.

Teuner, A. & Hosticka, B. J. (1993). Adaptive gabor transformation for image processing. *IEEE Transactions on Image Processing*, *2*(1), 112–7.

Theimer, W. M. & Mallot, H. A. (1992). Binocular vergence control and depth reconstruction using a phase method. Technical report, Institut für Neuroinformatik, Ruhr-Universität, W-4630 Bochum.

Thorpe, S. J. & Imbert, M. (1989). Biological constraints in connectionist modelling. In Pfeifer, R., Schreter, Z., Fogelman-Soulié, F., & Steels, L. (Eds.), *Connectionism in Perspective*, pages 63–91. North Holland: Elsevier Science Publishers B.V.

Tolhurst, D. J. & Dean, A. F. (1987). Spatial summation by simple cells in the striate cortex of the cat. *Experimental Brain Research*, *66*(3), 607–20.

Toyama, K., Kimura, M., & Tanaka, K. (1981). Cross-correlation analysis of interneuronal connectivity in cat visual cortex. *Journal of Neurophysiology*, *46*(2), 191–201.

Toyama, K., Matsunami, K., Ohno, T., & Tokashiki, S. (1974). An intracellular study of neuronal organization in the visual cortex. *Experimental Brain Research*, *21*, 45–66.

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience*, *6*(4), 1160–70.

Tsumoto, T., Eckart, W., & Creutzfeldt, O. D. (1979). Modification of orientation sensitivity of cat visual cortex neurons by removal of GABA-mediated inhibition. *Experimental Brain Research*, *34*(2), 351–63.

Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, *55*, 71–82.

Ullman, S. (1986). Artificial intelligence and the brain: Computational studies of the visual system. *Annual Review of Neuroscience*, *9*, 1–26.

Victor, J. D. (1992). Nonlinear systems analysis in vision: Overview of kernel methods. In Pinter & Nabet (1992), chapter 1, pages 2–34.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striata cortex. *Kybernetik*, *14*, 85–100.

von der Malsburg, C. (1990). Considerations for a visual architecture. In Eckmiller (1990), pages 303–12.

Wang, H. & Yan, H. (1992). Efficient computation of the Gabor transform using a neural network. In Leong & Jabri (1992), pages 256–9.

Watkins, D. W. & Berkley, M. A. (1974). The orientation selectivity of single neurons in cat striate cortex. *Experimental Brain Research*, *19*(4), 433–46.

Watson, A. B. (1987). Efficiency of a model human image code. *Journal of the Optical Society of America: A*, *4*(12), 2401–17.

Watson, A. B. & Ahumada, A. J. J. (1983). A look at motion in the frequency domain. Technical Memorandum 84352, NASA.

Watson, A. B. & Ahumada, A. J. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America: A*, *2*(2), 322–41.

Webster, M. A. & de Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America: A*, *2*(7), 1124–32.

Wechsler, H. (1990). *Computational Vision.* San Diego, London, Sydney, Tokyo: Academic Press Inc.

Westheimer, G. (1984). Spatial vision. *Annual Review of Psychology*, *35*, 201–26.

Wexler, J. & Raz, S. (1990). Discrete Gabor expansions. *Signal Processing*, *21*, 207–220.

Weyl, H. (1932). *Theory of Groups and Quantum Mechanics.* New York: Dutton.

White, E. L. (1989). *Cortical Circuits: Synaptic Organization of the Cerebral Cortex — Structure, Function, and Theory.* Boston, Basel, Stuttgart: Birkhäuser.

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J., & de Valois, R. (1990). The perception of form: Retina to striate cortex. In Spillmann & Werner (1990), chapter 10, pages 231–72.

Wörgötter, F. & Koch, C. (1991). A detailed model of the primary visual pathway in the cat: Comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *Journal of Neuroscience*, *11*(7), 1959–79.

Wright, M. J. & Ikeda, H. (1973). Processing of spatial and temporal information in the visual system. In Schmidt, F. O. & Worden, F. G. (Eds.), *The Neurosciences: Third Study Program*, chapter 11, pages 115–22. Cambridge, Massachusetts; London, England: MIT Press.

Yamane, S., Bishop, P. O., & Maske, R. (1985). Direction selectivity of simple cells in cat striate cortex to moving light bars: II. Relation to moving dark bar response. *Experimental Brain Research*, *57*(3), 523–36.

Yan, H. (1991a). Image compression using Hopfield neural networks. In Jabri, M. (Ed.), *Proceedings of the Second Australian Conference on Neural Networks*, pages 267–70, University of Sydney, Australia.

Yan, H. (1991b). Stability and relaxation time of Tank and Hopfield's neural network for solving LSE problems. *IEEE Transactions on Circuits & Systems*, *CAS-38*, 1108–10.

Yan, H. & Gore, J. C. (1990). Weight adjustment rule of neural networks computing discrete 2D Gabor transforms. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, *38*(9), 1654–6.

Yang, J. (1992). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields?: Comment. *Journal of the Optical Society of America: A*, *9*(2), 334–336.

Young, R. A. (1985). The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles. Technical Report GMR-4920, General Motors Research.

Yuille, A. L., Kammen, D. M., & Cohen, D. S. (1989). Quadrature and the development of orientation selective cortical cells by Hebb rules. *Biological Cybernetics*, *61*, 183–94.

Zeevi, Y. Y. & Porat, M. (1984). Combined frequency-position scheme of image representation in vision. *Journal of the Optical Society of America: A*, *1*(12), 1248. Abstract from the 1984 Annual Meeting.

Zetzsche, C. & Caelli, T. (1989). Invariant pattern recognition using multiple filter image representations. *Computer Vision, Graphics, & Image Processing*, *45*, 251–62.

Zhou, Y.-T. & Chellappa, R. (1992). *Artificial Neural Networks for Computer Vision*, volume 5 of *Research Notes in Neural Computing*, chapter 5, pages 63–82. Berlin, New York, London, Tokyo: Springer-Verlag.