

Um Modelo Para Redes Neuroniais Biologicamente Inspirado Baseado em Minimização de Divergência Local

Ewaldo Eder Carvalho Santana

Tese de Doutorado submetida à Coordenação do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande - Campus de Campina Grande como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Microeletrônica/Processamento da
Informação

Raimundo Carlos Silvério Freire, Dr.

Orientador

Allan Kardec Duailibe Barros Filho, Dr.

Orientador

Campina Grande, Paraíba, Brasil

©Ewaldo Eder Carvalho Santana, Novembro de 2009

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

S232m

2009 Santana, Ewaldo Eder Carvalho.

Um modelo para redes neuronais biologicamente inspirado baseado em minimização de divergência local / Ewaldo Eder Carvalho Santana.- Campina Grande, 2009.

87 f. : il.

Tese (Doutorado em Engenharia de Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Referências.

Orientadores: Prof. Dr. Raimundo Carlos Silvério Freire, Prof. Dr. Allan Kardec Duailibe Barros Filho.

1. Redes Neurais. 2. Córtex Visual Primário. 3. Plausibilidade Biológica. I. Título.

CDU – 004.032.26(043)

**UM MODELO PARA REDES NEURONIAIS BIOLÓGICAMENTE INSPIRADO
BASEADO EM MINIMIZAÇÃO DE DIVERGÊNCIA LOCAL**

EWALDO EDER CARVALHO SANTANA

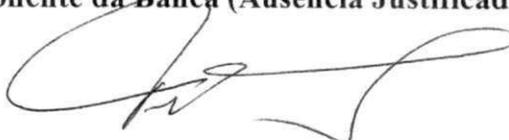
Tese Aprovada em 06.11.2009



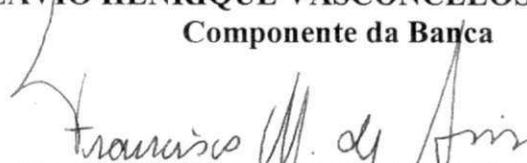
RAIMUNDO CARLOS SILVÉRIO FREIRE, Dr., UFCG
Orientador

ALLAN KARDEC DUAILIBE BARROS FILHO, Dr., UFMA
Orientador (Ausência Justificada)

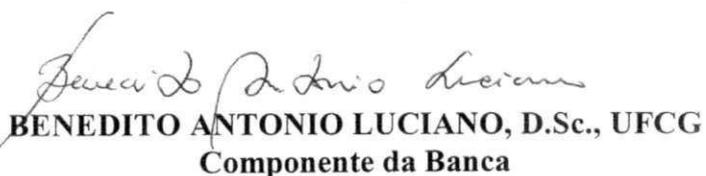
DRÁULIO BARROS DE ARAÚJO, Dr., UFRN
Componente da Banca (Ausência Justificada)



FLÁVIO HENRIQUE VASCONCELOS, Dr., UFMG
Componente da Banca



FRANCISCO MARCOS DE ASSIS, Dr., UFCG
Componente da Banca



BENEDITO ANTONIO LUCIANO, D.Sc., UFCG
Componente da Banca

CAMPINA GRANDE – PB
NOVEMBRO - 2009

Dedicatória

Este trabalho é dedicado aos pesquisadores do Laboratório para o Processamento da Informação Biológica - PIB, do Departamento de Engenharia da Eletricidade da Universidade Federal do Maranhão - UFMA.

Agradecimentos

Agradeço aos professores Raimundo Carlos Silvério Freire, Dr., Allan Kardec Duailibe Barros Filho, Dr., pela incansável orientação e colaboração, sem as quais não seria possível realizar este trabalho.

Agradeço ao Dr. José de Ribamar Fiquene, Presidente do CEAMA; à Professora Zenira Massoli Fiquene e à Professora Maria de Nazaré Ferraz Tomaz, Diretora Geral e Vice-Diretora, respectivamente, da Faculdade Atenas Maranhense - FAMA, pelo apoio a mim concedido durante toda a minha jornada de formação acadêmica.

Agradeço profundamente à Prof.^a Dra. Maria da Guia da Silva, do Departamento de Engenharia da Eletricidade da Universidade Federal do Maranhão - UFMA, pelos incentivos constantes.

Agradeço ao Alfranke, Deusdete, Ênio e Luís Cláudio pelo apoio precioso. Meu obrigado, também, aos colegas do LIMC - UFCG.

Meu muito obrigado ao professor Francisco F. Grangeiro, do Departamento de Matemática da Universidade Estadual do Maranhão, pelo auxílio na matemática. Agradecimentos aos colegas do DEMATI-UEMA: Anselmo, Fernanda, Magno, Mauro, Jackson, Neto e demais.

Agradecimentos especiais ao Éder Júnior, filho, parceiro e cientista.

Arthur e Tiago, papai é sempre grato à existência de vocês.

Também agradeço à FAPEMA pelo suporte financeiro que viabilizou a realização deste trabalho.

Resumo

Neste trabalho é proposto o desenvolvimento de uma rede neuronal com aprendizagem não supervisionada, para modelar a organização topográfica do córtex visual primário. Para isto, estuda-se o comportamento dos campos receptivos do córtex visual primário (V1), e, para o modelamento da rede utilizam-se os conceitos de divergência local e de interação entre neurônios vizinhos, bem como da característica de não linearidades dos neurônios. Para treinamento da rede desenvolveu-se um algoritmo de ponto fixo.

Abstract

In this work it is proposed an unsupervised neural network model, which seems biologically plausible in modeling the primary visual cortex (V1). It is, also, studied de behavior of the receptive fields of V1. In order to modeling the net it was used the concepts of local discrepancy and interactions between neighbor neurons, as well the non-linearity characteristics of neurons. It was designed a fixed-point algorithm to train the neural network.

Sumário

Introdução	1
1 Fundamentos de Neurociência	5
1.1 Introdução	5
1.2 Codificação Neuronal	6
1.3 Córtex visual	8
1.3.1 Córtex Visual Primário	11
1.4 Topografia no Sistema Visual	16
2 Redes Neuroniais Artificiais	19
2.1 Introdução	19
2.2 Redes Neuroniais Perceptron Multicamadas	20
3 Codificação Eficiente ou Esparsa	24
3.1 Introdução	24
4 Análise de Componentes Independentes	30
4.1 Introdução	30
4.2 Não gaussianidade e Independência Estatística	31
4.3 Medindo a Não Gaussianidade Através da Curtose	33
4.4 Um algoritmo de ponto fixo usando curtose	34
4.5 ICA aplicado a imagens naturais	36
4.6 O Modelo TICA	39

5	Desenvolvimento do Modelo	43
5.1	Introdução	43
5.2	Modelo Proposto	45
5.3	Validação	50
6	Conclusão, Publicações e Trabalhos Futuros	62
6.1	Conclusão	62
6.2	Lista de Publicações	63
6.2.1	Periódicos	63
6.2.2	Conferências	63
6.3	Trabalhos Futuros	64
6.3.1	Tamanho da vizinhança	64
6.3.2	Separação Cega de Fontes	64
6.3.3	Extração de Características Temporal	64
7	Anexo	65
7.1	Arquivos Matlab	65
7.2	Imagens Naturais	69
	Referências Bibliográficas	79

Lista de Figuras

1.1	Diagrama de três neurônios. A) Neurônio piramidal cortical. B) Uma célula de Purkinje do cerebelo. C) Uma célula estrelada do córtex cerebral.	7
1.2	Potencial de ação.	9
1.3	Sequência de potenciais de ação.	9
1.4	Caminho da sequência de potenciais de ação a partir da retina, através do núcleo geniculado lateral no tálamo, até o córtex visual primário no cérebro humano.	10
1.5	Córtex Visual. (a) As setas indicam as áreas V1, V2, V3, V4 e V5 do córtex visual. (b) Organização hierárquica e fluxo de informação bi-direcional do córtex visual primário.	11
1.6	Exemplos de campos receptivos de células simples do V1. Os marcadores X e Δ representam áreas que evocam respostas excitatórias e inibitórias, respectivamente.	13
1.7	Campos receptivos mapeados mediante o emprego da correlação reversa. Os contornos sólidos representam áreas excitatórias enquanto que as regiões com contornos tracejados são inibitórias. A representação tridimensional dos campos receptivos é mostrada ao lado.	14
1.8	Aproximação de campos receptivos através de <i>wavelets</i> de Gabor. Os parâmetros associados são $\sigma_x = 1$, $\sigma_y = 2$, $\frac{1}{k} = 0,56$ e (a) $\phi = 1 - \frac{\pi}{2}$ ou (b) $\phi = 1 - \pi$	15

1.9	Representação de estímulos no córtex visual. (a) Estimulação das células da retina e transmissão da informação ao V1. (b) Geração de uma representação neuronal em V1.	17
2.1	Modelo de McCulloch e Pitts.	20
2.2	Direções dos fluxos de um sinal em uma rede MLP: propagação direta dos sinais e retro-propagação dos sinais de erro.	23
2.3	função de ativação não linear para um neurônio.	23
3.1	Cem amostras aleatórias de uma variável gaussiana com média zero de desvio padrão 1.	25
3.2	Cem amostras aleatórias de uma variável esparsa ou supergaussiana.	26
3.3	Funções de densidade de probabilidades: Esparsa ou super-Gaussiana (linha vermelha); Gaussiana (linha azul).	27
3.4	Filtros gerados utilizando-se codificação eficiente.	28
4.1	Exemplos de imagens naturais em preto e branco.	36
4.2	Amostra de uma imagem natural disposta como combinação linear das funções (imagens) base. (a) amostras aleatórias da imagem, \mathbf{x} ; (b) componentes independentes s_i ; (c) funções base, \mathbf{a}_i	37
4.3	Funções bases aprendidas pelo modelo proposto por Olshausen e Field.	38
4.4	Modelo para TICA.	40
5.1	Modelo proposto: Primeira camada: Vetor de entrada. Segunda camada: Vetor dos pesos sinápticos. Terceira camada: funções não lineares. Quarta camada: interação entre a saída de um neurônio com as saídas dos seus vizinhos (divergência local).	46
5.2	Vizinhança de um neurônio k : O neurônio k é supostamente o neurônio central. Seus vizinhos são aqueles que estão ao seu lado, dentro do quadrado tracejado.	48
5.3	Função $f(u) = u \cdot atan(0, 5u) - Ln(1 + 0, 25u^2)$	51
5.4	Função $f(u) = Ln[cosh(0, 25u^2)]$	52

5.5	Situação onde se consegue maior transmissão da informação de um sinal u através de uma função $f(u)$	53
5.6	Derivada primeira da função $u \cdot \operatorname{atan}(0,5u) - \operatorname{Ln}(1 + 0,25u^2)$	54
5.7	Derivada primeira da função $\operatorname{Ln}[\operatorname{cosh}(0,25u^2)]$	55
5.8	Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k \cdot \operatorname{atan}(0,5u_k) - \operatorname{Ln}(1 + 0,25u_k^2)$ foi a não-linearidade utilizada.	56
5.9	Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = \operatorname{Ln}(\operatorname{cosh}(0,25u_k))$ foi a não-linearidade utilizada.	57
5.10	Histograma da estimação de uma imagem codificada. A componente obtida tem distribuição esparsa. Ou seja, a componente está esporadicamente ativa.	58
5.11	Os 24 vizinhos do neurônio central).	59
5.12	Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k \cdot \operatorname{atan}(0,5u_k) - \operatorname{Ln}(1 + 0,25u_k^2)$ foi a não-linearidade utilizada com vizinhança maior.	60
5.13	Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k^2$ foi a não-linearidade utilizada.	61
7.1	Imagem Natural 01	70
7.2	Imagem Natural 02	70
7.3	Imagem Natural 03	71
7.4	Imagem Natural 04	72
7.5	Imagem Natural 05	73
7.6	Imagem Natural 06	73
7.7	Imagem Natural 07	74
7.8	Imagem Natural 08	74
7.9	Imagem Natural 09	75
7.10	Imagem Natural 10	76
7.11	Imagem Natural 11	76

7.12 Imagem Natural 12	77
7.13 Imagem Natural 13	78

Introdução

A cada momento, os receptores sensoriais do nosso corpo enviam para o nosso cérebro uma vasta quantidade de informação sobre o meio ambiente que nos rodeia. Informações visuais, por exemplo, são processadas por cerca de dez milhões de células cones e cem milhões de bastonetes, em cada olho. Como o cérebro transforma estas informações sensoriais em uma forma tal que seja útil para o comportamento do sistema visual? Estudos nas áreas de neurofisiologia, neuroanatomia e imagens cerebrais, nas últimas décadas, têm ajudado a clarear esta questão, revelando partes da enorme complexidade que é o entendimento de como a informação sensorial é representada e processada pelos neurônios nos vários estágios dentro do cérebro [1,2].

É muito difícil abordar tal questão baseado apenas em experimentos com animais. Uma alternativa promissora é investigar modelos computacionais baseados em princípios de codificação eficiente [3,4]. Tais modelos levam em conta as propriedades estatísticas dos sinais do meio ambiente e tentam explicar os tipos de representações encontradas no cérebro em termos de modelos probabilísticos destes sinais [5–7]. Desta forma, em neurociência, a modelagem computacional é utilizada para descrever, reproduzir e fazer previsões sobre o comportamento dos diferentes componentes do sistema nervoso. Assim como em outras áreas das ciências, este procedimento tem-se mostrado eficiente no estudo e aprimoramento das teorias a respeito da função cerebral. Com o crescente aumento da capacidade computacional, modelos maiores e mais detalhados podem ser construídos com um grau de realismo biológico cada vez maior. Tal realismo é o que denomina-se de plausibilidade biológica.

Um exemplo deste desenvolvimento, são as redes neuronais artificiais, que apesar de serem originalmente inspiradas em biologia, são geralmente treinadas por meio de uma minimização de uma função de erro [8, 9], a qual pressupõe a existência de um sinal desejado. A plausibilidade biológica da retropropagação do erro é duvidosa [8–12]. Então, muitos pesquisadores têm sugerido que aprendizagem no cérebro deve ser pensada como uma aprendizagem não-supervisionada, complementada com reforço da aprendizagem [9]. Esta ideia é baseada no trabalho do neurofisiologista Donald Hebb [13], que estudando a comunicação entre neurônios enunciou um princípio que se tornou muito importante na área de computação neuronal. Hebb verificou que uma vez que um neurônio repetidamente excitava um outro neurônio, o limiar de ativação deste último diminuía, isto é, a comunicação entre eles era facilitada pela excitação repetida. Isto significa, equivalentemente, que o efeito da excitação do primeiro neurônio é amplificado [9], o que pode ser visto como uma maximização da correlação entre entrada e saída de um neurônio [14]. Além do mais, após Barlow ter relacionado percepção com teoria da informação [4], Linsker propôs o princípio da maximização da informação (INFOMAX), na qual pode-se observar que a maximização da informação mútua de uma variável aleatória gaussiana conduz à correlação [15]. Usando correlação como a base da aprendizagem, alguns pesquisadores desenvolveram vários algoritmos para imitar o processo de aprendizagem [16–18].

Um conceito que atualmente está sendo muito divulgado estabelece que o cérebro codifica informações ignorando ou reduzindo a informação mútua (medida da quantidade de informação que uma variável aleatória X contém acerca da variável aleatória Y) [7, 19, 20]. Este tipo de codificação inclui dependências de altas ordens, o que está diretamente relacionado com a ideia de independência estatística [21] e tem guiado vários pesquisadores a desenvolverem algoritmos que fornecem informação mútua com redundância mínima [22, 23].

Analisando a codificação sensorial, Barros e Ohnishi [24] propuseram a minimização de não linearidades aplicadas sobre o erro entre a entrada sensorial e um sinal de referência interna, a qual forneceu campos receptivos tipo *wavelets*

quando o algoritmo foi aplicado para codificar imagens naturais. Resultados similares foram obtidos em redes neuronais não supervisionadas quando aplicadas em redução de redundância.

A determinação do valor máximo da entropia de uma distribuição, um dos princípios da redução de redundância, quando realizada sob a restrição de uma média fixa proporciona funções de densidade de probabilidades super gaussianas centradas na média, ou seja, distribuições esparsas [19, 25, 26]. Desta forma, supondo que o objetivo do córtex visual Primário (V1) dos mamíferos é obter uma representação esparsa das cenas naturais, Olshausen e Field derivaram uma rede neuronal artificial que realiza uma filtragem orientada, localizada e passa banda para modelagem dos campos receptivos do V1 [20].

Explorando o córtex visual primário, Vinje e Gallant [27] provaram experimentalmente a representação esparsa da visão natural e assinalaram que isto é devido à modulação supressiva e excitatória do campo receptivo. Em modelagem de cenas naturais, Hyvärinen e seus colaboradores [28] definiram um modelo gerador para análise de componentes independentes (ICA) no qual as componentes não são completamente independentes, mas possuem uma dependência em energia que é definida em relação a uma certa topografia. Componentes próximas umas das outras, nesta topografia, têm uma maior co-dependência que aquelas que são distantes. Utilizando esta mesma ideia, Osindero e seus colaboradores [29] apresentaram um modelo de função de densidade de probabilidades baseado em energia, aplicável em conjuntos de dados que têm uma estrutura esparsa.

Além desses, existe uma grande quantidade de estudos em codificação esparsa e ICA que produziram algoritmos que proporcionam resultados semelhantes àqueles obtidos pelo córtex visual [6, 20, 30].

Uma das dificuldades de todos esses estudos é a computação sofisticada necessária para obter representação esparsa. Além do mais, muitos dos modelos propostos impõem uma série de restrições sobre os sinais analisados, de forma a adequar-se a uma análise matemática. Este comportamento inibe bastante a plausibilidade biológica do modelo.

Neste trabalho é apresentado um modelo de rede neuronal, com aprendizagem

não supervisionada, biologicamente plausível, para obtenção de organização topográfica, imitando os neurônios do córtex visual primário, V1. O modelo foi desenvolvido com base na iteração entre o sinal de um neurônio e os sinais oriundos dos seus vizinhos, definidos dentro de uma certa topografia.

Esta tese está organizada da seguinte maneira: no capítulo 1 foi feita uma revisão dos conhecimentos básicos de neurociência necessários para este estudo, tais como: codificação neuronal, córtex visual e córtex visual primário. Mostra-se, também, alguns modelos teóricos para emergência das características destas regiões corticais, e discute-se o conceito de organização topográfica no sistema visual. No capítulo 2, tem-se as Redes Neuroniais artificiais (RNA's), no qual é apresentado o modelo perceptron de múltiplas camadas e descreve-se algumas etapas do projeto de uma rede neuronal artificial. O conceito de Codificação esparsa e sua relação com o córtex visual primário é descrito no capítulo 3. No capítulo 4, faz-se uma introdução à Análise de Componentes Independentes (ICA), sua relação com não gaussianidade e sua medida através da utilização da curtose. Mostra-se, também, neste capítulo, um algoritmo para utilização em ICA de bastante uso na comunidade de processamento de sinais, faz-se a aplicação de ICA em imagens naturais e apresenta-se um modelo topográfico desenvolvido utilizando-se tais aplicações. No capítulo 5, desenvolve-se um modelo de rede neuronal artificial para emergência de características topográficas, apresenta-se o algoritmo para treinamento desta rede, aplica-se este modelo para codificação de imagens naturais, valida-se e discute-se o método utilizado. No capítulo 6 conclui-se esta tese, relacionam-se as publicações originadas durante o desenvolvimento da mesma e apresentam-se propostas de continuidade de trabalho. Finalmente, no capítulo 7, disponibiliza-se o algoritmo para treinamento da rede neuronal artificial apresentada, bem como as imagens que foram usadas para treinamento desta rede.

Capítulo 1

Fundamentos de Neurociência

1.1 Introdução

Neste capítulo é apresentada uma breve revisão de conhecimentos de neurociência necessários para a fundamentação deste trabalho.

A neurociência é o estudo da realização física do processo de informação no sistema nervoso humano e animal. O estudo da neurociência engloba três áreas principais: a neurofisiologia, a neuroanatomia e neuropsicologia.

A neurofisiologia é o estudo das funções do sistema nervoso. Ela utiliza eletrodos para estimular e gravar a reação das células nervosas ou de áreas maiores do cérebro.

A neuroanatomia é o estudo da estrutura do sistema nervoso, nos níveis microscópico e macroscópico. Os neuroanatomistas dissecam o cérebro, a coluna vertebral e os nervos periféricos fora dessa estrutura.

A neuropsicologia é o estudo da relação entre as funções neuronais e psicológicas. A principal pergunta da neuropsicologia é qual área específica do cérebro controla ou media as funções psicológicas. O principal método de estudo usado pelos neuropsicólogos é o estudo do comportamento ou mudanças cognitivas que acompanham lesões em partes específicas do cérebro. Estudos experimentais com indivíduos normais também são comuns.

1.2 Codificação Neuronal

Os neurônios são células animais diferenciadas em sua habilidade de transmitir sinais rapidamente através de grandes distâncias. Eles fazem isto gerando pulsos elétricos característicos, chamados de potenciais de ação, que podem ser transmitidos através de fibras nervosas. Os neurônios representam e transmitem informações pela geração de sequências de potenciais de ação com diferentes características temporais. O estudo da codificação neuronal envolve medir e caracterizar de que forma as propriedades de certos estímulos, tais como intensidade sonora ou luminosa, ou as propriedades de certas ações motoras, tais como direção de movimento de um braço, são representados pelos potenciais de ação.

A relação entre estímulos e respostas pode ser estudada a partir de dois pontos de vistas opostos: codificação neuronal e decodificação neuronal. A codificação neuronal refere-se ao mapeamento das respostas a partir dos estímulos. Por exemplo, pode-se catalogar de que forma neurônios respondem a uma certa variedade de estímulos e, em seguida, constrói-se modelos visando prever respostas para outros estímulos. Por outro lado, decodificação neuronal refere-se ao mapeamento inverso, ou seja, mapear os estímulos a partir das respostas. O desafio, neste campo de pesquisa, é reconstruir estímulos, ou certos aspectos destes estímulos, a partir de uma sequência de potenciais de ação por eles evocados.

Os neurônios são altamente especializados na geração de sinais elétricos em respostas a atividades químicas e de outras naturezas e os transmite para outras células. Algumas especializações morfológicas importantes, as quais podem ser vistas na Figura 1.1, são os dendritos e os axônios. Os dendritos recebem sinais provenientes de outros neurônios enquanto o axônio transmite os sinais de um neurônio para outros. A estrutura ramificada dos dendritos permitem ao neurônio receber sinais de vários outros neurônios através das conexões sinápticas. O neurônio piramidal cortical (Figura 1.1A) e o interneurônio cortical (Figura. 1.1C) recebem, cada um, milhares de sinais sinápticos, enquanto a célula de Purkinje (Figura 1.1B) recebe mais de cem milhões de sinais sinápticos.

Na Figura 1.1 não estão representadas as extensões verdadeiras dos axônios

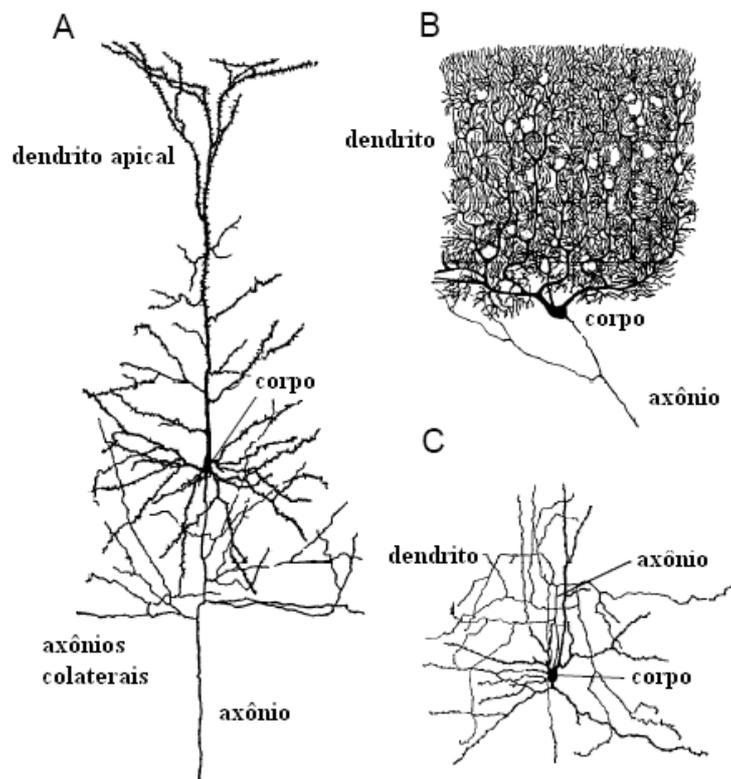


Figura 1.1: Diagrama de três neurônios. A) Neurônio piramidal cortical. B) Uma célula de Purkinje do cerebelo. C) Uma célula estrelada do córtex cerebral.

daqueles neurônios. Os axônios podem estender-se por várias partes do cérebro e, em alguns casos, até do corpo inteiro. O axônio é um grande canal de transmissão que finaliza em pequenas ramificações que transmitem informações do neurônio para outros. É estimado que um axônio faz em média 180 conexões sinápticas com outros neurônios por milímetro de comprimento em suas terminações.

Chamamos de sinapses as junções entre o axônio de uma célula e os dendritos de outras células. Os dendritos coletam pulsos que chegam de outras sinapses, converte-os em correntes elétricas e, após somar todos esses pulsos, enviam esta soma para o segmento inicial do axônio. Esta soma ocorre entre todos os dendritos durante o tempo. Cada membrana sináptica atua como um integrador com uma constante de tempo associada. A função do axônio é transmitir, sem atenuação, a amplitude da soma de correntes advinda dos dendritos para alvos distantes [31]. Quando o neurônio atinge certo limiar ele é ativado, isto é, despolarizado, o que significa que ele produz um pico de energia no seu axônio, ou seja, ele gera um potencial de ação, como pode ser visto na Figura 1.2. A ativação contém um período refratário de forma tal que um neurônio constantemente ativado produzirá uma sequência de potenciais de ação, que é chamada de trem de impulso, no seu axônio. Na Figura 1.3 temos uma representação esquemática de uma sequência de potenciais de ação.

Como as redes neuronais biológicas são treinadas não é sabido, mas muito do que se sabe sobre o treinamento é baseado no conceito de aprendizagem Hebbiana [13]. A lei da aprendizagem Hebbiana fortalece as sinapses (permitindo maior receptividade do neurônio pós-sináptico) quando dois neurônios são ativados ao mesmo tempo. Se existir uma correlação consistente entre as ativações dos dois neurônios, então o neurônio pré-sináptico deve ser, pelo menos parcialmente, responsável pela ativação do neurônio pós-sináptico.

1.3 Córtex visual

A conversão do estímulo luminoso em um sinal elétrico e, posteriormente, em uma sequência de potenciais de ação ocorre na retina. Esta conversão é feita por um

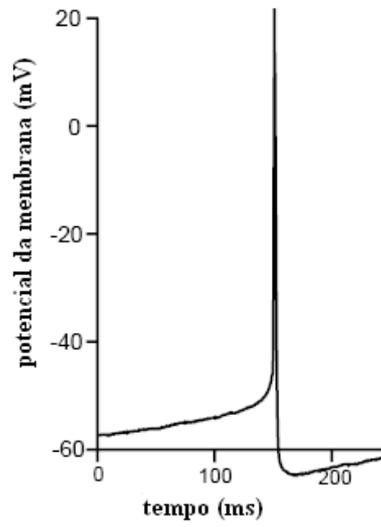


Figura 1.2: Potencial de ação.

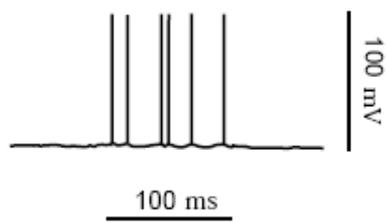


Figura 1.3: Sequência de potenciais de ação.

conjunto de células fotorreceptoras (cones e bastonetes). Os neurônios da camada de saída da retina são as células do gânglio retiniano. Os axônios destas células formam o nervo óptico. O nervo óptico conduz a sequência de potenciais de ação gerado nas células do gânglio retiniano para o núcleo geniculado lateral do tálamo, entre a retina e o córtex visual, como pode ser visto na Figura 1.4. Antes de atingir o núcleo geniculado lateral, algumas células do gânglio retiniano passam pelo quiasma óptico que faz com que os lados esquerdo e direito do campo visual de ambos os olhos sejam representados nos lados direito e esquerdo do cérebro, respectivamente.

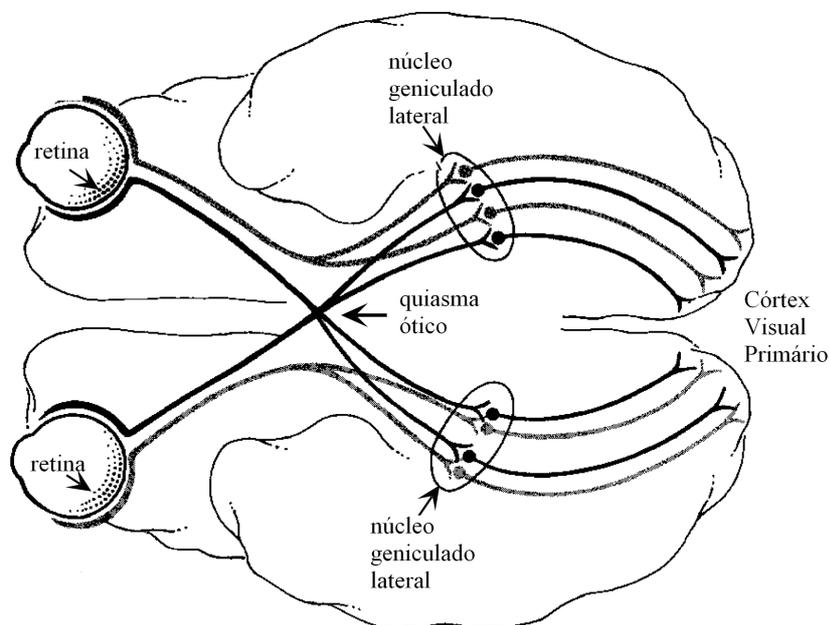


Figura 1.4: Caminho da sequência de potenciais de ação a partir da retina, através do núcleo geniculado lateral no tálamo, até o córtex visual primário no cérebro humano.

O córtex visual é localizado no lóbulo occipital, na parte posterior do cérebro, como pode ser visto na Figura 1.5.a. Este é responsável pelo processamento das informações visuais. A região que recebe informações diretamente do Núcleo Geniculado Lateral é o córtex visual primário, ou V1 ou córtex estriado. Em seguida, a informação visual passa por outras áreas de processamentos. Essas

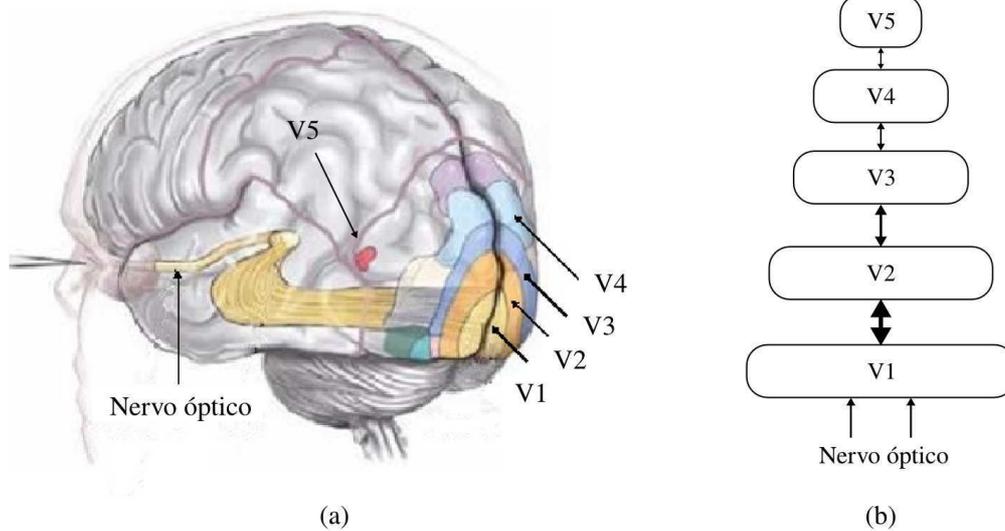


Figura 1.5: Córtex Visual. (a) As setas indicam as áreas V1, V2, V3, V4 e V5 do córtex visual. (b) Organização hierárquica e fluxo de informação bi-direcional do córtex visual primário.

áreas incluem o V2, V3, V4 e V5. O córtex visual primário, V1, e demais áreas, V2, V3, V4 e V5 são indicadas, na Figura 1.5b, pelo conjunto de setas.

O córtex visual possui uma organização hierárquica que realiza análise dos estímulos visuais em diferentes níveis de processamento ainda não completamente compreendidos [32]. Esta organização é representada na Figura 1.5.b em conjunto com o fluxo de informação bi-direcional do córtex visual.

1.3.1 Córtex Visual Primário

O córtex visual primário (V1) dos mamíferos tem sido objeto de vários estudos, principalmente nas últimas quatro décadas. No começo dos anos sessenta do século passado, os estudos de Hubel e Wiesel fizeram uma mudança de paradigma ao demonstrarem que o padrão de respostas dos neurônios, neste córtex, está relacionado a parâmetros de estímulos visuais [33, 34], tais como: orientação, contraste, frequência espacial, posição, fase, direção de movimento. Tal padrão pode ser comparado às funções bases da transformada de *wavelets* [35]. Uma

abordagem para entender as propriedades destas respostas dos neurônios visuais tem sido considerar o relacionamento do V1 com as estruturas estatísticas das imagens naturais em termos de codificação eficiente [3]. Então, uma série de algoritmos foi desenvolvida para treinar redes neuronais não supervisionadas sobre imagens naturais na tentativa de obterem-se modelos para os campos receptivos com propriedades semelhantes [3, 6, 7, 30, 36, 37].

A partir dos parâmetros dos estímulos visuais, foi possível mapear características ou estruturas desses estímulos que inibem ou excitam uma determinada célula do córtex visual. O mapa de inibição/excitação resultante é definido como o campo receptivo de uma célula do córtex.

Inicialmente, o processo de mapeamento dos campos receptivos foi realizado analisando a resposta das células de V1 à estimulação visual por meio de barras de luz orientadas [38]. Iluminando certas regiões do espaço foi possível observar o aumento (excitação) ou decaimento (inibição) da taxa de ativação das células estudadas. Mudando os parâmetros (orientação, posição, etc) associados com as barras de luz, foi possível definir a estrutura dos campos receptivos.

A partir dessa estrutura, células de V1 foram caracterizadas como simples ou complexas. Os campos receptivos das células simples foram definidos como possuindo regiões discretas de inibição e excitação e, também, por serem altamente seletivos aos parâmetros anteriormente citados. Além do mais, as contribuições do estímulo visual que incide em diferentes regiões do campo receptivo somam-se linearmente [34]. Na Figura 1.6 são mostrados exemplos de campos receptivos de células simples mapeados a partir de estimulação visual utilizando barras de luz orientadas.

Na Figura 1.6(a), o campo receptivo representado possui uma área de excitação central acompanhada por áreas inibitórias em cada lado. Isto significa que estímulo posicionado e orientado de acordo com esse campo receptivo irá aumentar ou inibir a atividade da célula correspondente. Um campo receptivo com estrutura complementar é mostrado na Figura 1.6(b). O campo receptivo da Figura 1.6(c) possui áreas com dimensões diferentes em relação aos dois anteriormente citados. De uma forma interessante, o campo receptivo mostrado na Figura 1.6(d) apresenta

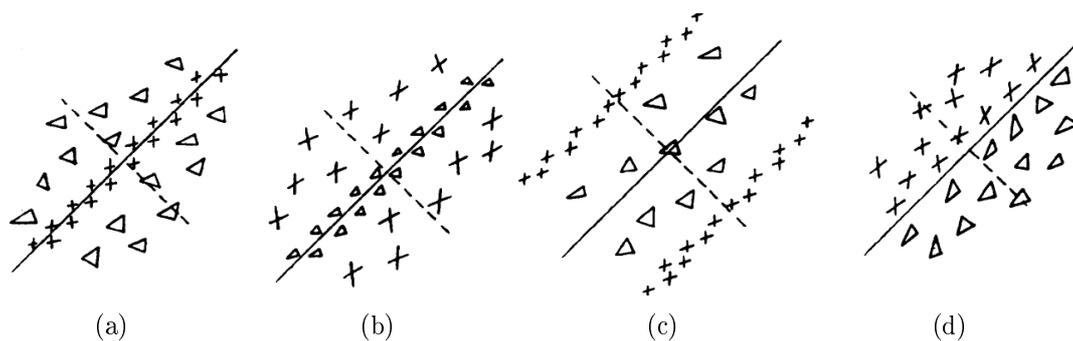


Figura 1.6: Exemplos de campos receptivos de células simples do V1. Os marcadores X e Δ representam áreas que evocam respostas excitatórias e inibitórias, respectivamente.

apenas duas regiões, uma excitatória e outra inibitória. Para este campo receptivo, o estímulo ótimo é uma borda ou duas regiões em diferentes níveis de brilho.

Quanto às células complexas, elas respondem a barras de luz orientadas e a estímulos que incidem em uma região limitada do campo receptivo. Mas, diferentemente das células simples, seu comportamento não pode ser explicado através de uma subdivisão ordenada do campo receptivo em regiões excitatória e inibitória [38]. Elas respondem mais fortemente à movimentos de barras de luz orientadas através do seu campo receptivo. Para obter-se uma sequência de potências de ação, este movimento deve ser orientado e bem posicionado no campo receptivo. Diferentemente das células simples, as contribuições do estímulo visual que incide em diferentes regiões do campo receptivo não podem ser linearmente combinados [39].

Uma forma alternativa para mapeamento dos campos receptivos é o método da correlação reversa.

O objetivo deste método é estimar a "resposta ao impulso" espaço-temporal $h(x, y, t)$ de uma célula do córtex. Primeiramente, o processo de mapeamento consiste em mensurar e gravar a resposta $y(t)$ de um neurônio a um estímulo momentâneo $s(x, y, t)$ apresentado repetidamente [40]. A correlação reversa:

$$C(x, y, t) = \int_0^T y(\tau) s(x, y, \tau - t) d\tau, \quad (1.1)$$

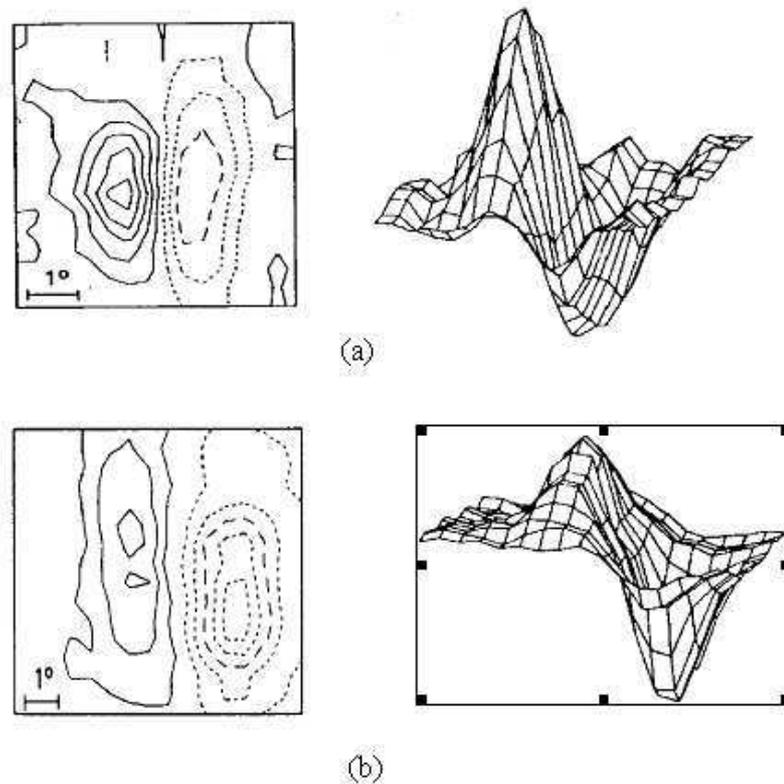


Figura 1.7: Campos receptivos mapeados mediante o emprego da correlação reversa. Os contornos sólidos representam áreas excitatórias enquanto que as regiões com contornos tracejados são inibitórias. A representação tridimensional dos campos receptivos é mostrada ao lado.

entre a resposta da célula e a sequência utilizada como estímulo visual é utilizada como uma relação "entrada-saída" do neurônio analisado. Considerando que o estímulo $s(x, y, t)$ é aproximado como ruído branco, a função de correlação cruzada $C(x, y, t)$ é igual à resposta ao impulso $h(x, y, t)$ da célula analisada. A resposta ao impulso é, então, utilizada como uma estimativa espaço-temporal do campo receptivo. Como exemplo, a estrutura espacial de dois campos receptivos estimada em instantes t específicos é mostrada na Figura 1.7.

Uma aproximação matemática da estrutura espacial mostrada na Figura 1.7 pode ser realizada mediante o emprego de transformadas *wavelets* de Gabor [41]. De acordo com a Equação 1.2, uma *wavelet* de Gabor $G(x, y)$ pode ser definida

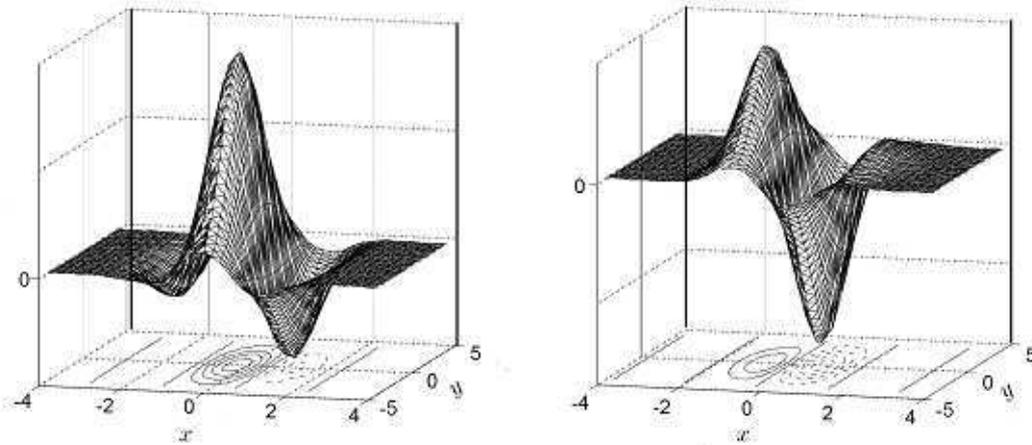


Figura 1.8: Aproximação de campos receptivos através de *wavelets* de Gabor. Os parâmetros associados são $\sigma_x = 1$, $\sigma_y = 2$, $\frac{1}{k} = 0,56$ e (a) $\phi = 1 - \frac{\pi}{2}$ ou (b) $\phi = 1 - \pi$.

como um produto entre uma função Gaussiana e uma função senoidal [32].

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(kx - \phi). \quad (1.2)$$

Os parâmetros de $G(x, y)$ determinam as propriedades da estrutura espacial de um campo receptivo. As extensões do campo receptivo nas direções x e y são determinadas por σ_x e σ_y , respectivamente. O parâmetro k determina a frequência espacial preferida, ou seja, a forma das oscilações entre áreas excitatórias e inibitórias. A fase das oscilações é determinada por ϕ . Exemplos de campos receptivos aproximados a partir de *wavelets* de Gabor são mostrados na Figura 1.8.

Os parâmetros dos campos receptivos das células simples de V1 variam ao longo da população de células. Dessa forma, cada célula específica responde apenas aos estímulos visuais que correspondam aos seus respectivos parâmetros.

Por exemplo, é possível observar que todos os campos receptivos mostrados

na Figura 1.6 apresentam orientação oblíqua. Assim, estímulos na mesma orientação deverão excitar ou inibir as células, enquanto que estímulos em diferentes orientações não deverão influenciar o padrão de ativação. Por outro lado, se estes estímulos possuírem uma outra característica que corresponda aos campos receptivos da célula, então eles deverão influenciar o padrão de ativação, indiferentemente à orientação.

Dessa forma, as células de V1 funcionam como um detector de características sendo ativadas por estruturas que correspondam à informação representada pelos seus respectivos campos receptivos. Este processo é representado na Figura 1.9, na qual pode-se ver que, inicialmente, um estímulo visual ativa ou inibi as células da retina. Tal informação é transmitida ao córtex visual primário através do nervo óptico. No córtex visual primário os campos receptivos que correspondem às características do estímulo são utilizados para geração de uma representação neuronal.

Um dos problemas mais antigos em neurociência visual é entender os princípios e o objetivo dessa representação neuronal criada pelos campos receptivos de V1. Uma hipótese é que células sensoriais têm como objetivo extrair a maior quantidade de informação do ambiente para sobrevivência do organismo. No entanto, para isso as células devem gerar uma representação não redundante ou eficiente do ambiente.

1.4 Topografia no Sistema Visual

Topografia é um princípio básico da organização cerebral. Por topografia entende-se, aqui, a existência de mapas ordenados, nos quais a localização espacial de uma célula, na superfície cortical, está relacionada, de uma maneira sistemática, com suas propriedades funcionais [34,39,42,43]. Ou seja, as vias visuais são organizadas de forma que pontos vizinhos no campo visual são projetados em localizações vizinhas no córtex, definindo assim uma organização topográfica. Especificamente, as fibras do nervo óptico conectam-se às células do corpo geniculado lateral e os axônios destas terminam no córtex visual primário. Estas conexões, dos olhos para o corpo geniculado lateral e deste para o córtex visual primário são organizadas

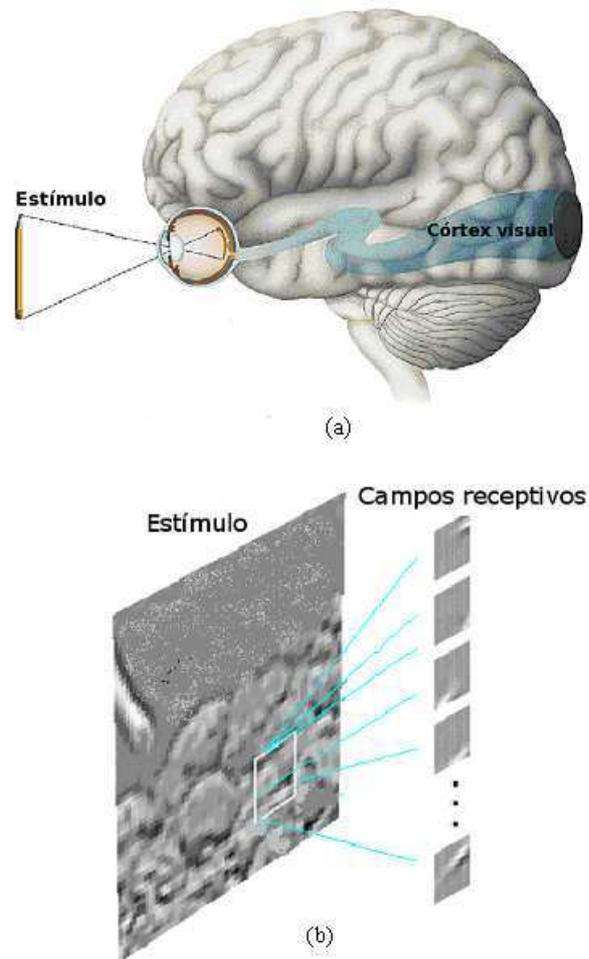


Figura 1.9: Representação de estímulos no córtex visual. (a) Estimulação das células da retina e transmissão da informação ao V1. (b) Geração de uma representação neuronal em V1.

topograficamente. O mapeamento de cada estrutura na próxima é sistemática: Movendo-se de um ponto para outro na retina, os pontos correspondentes no corpo geniculado lateral ou no córtex traça um caminho contínuo. Por exemplo, fibras do nervo ótico oriundas de uma pequena região da retina irão todas para uma região particular pequena no corpo geniculado lateral, e fibras de uma pequena região do corpo geniculado lateral irão todas para uma pequena região particular no córtex visual primário [39].

Modelos computacionais para representações topográficas têm uma longa história. Após os modelos iniciais de von der Malsburg [44], Kohonen desenvolveu seu famoso mapa auto-organizável [45] que tornou-se um arquétipo para os mapas topográficos.

Hoje em dia assume-se que a organização cortical é topográfica em virtude de certas restrições biológicas. O argumento básico é que esta organização visa diminuir a distância entre neurônios [46]. Ou seja, a organização topográfica é útil para reduzir o comprimento dos axônios dos neurônios que são conectados uns aos outros, supondo que apenas neurônios com propriedades funcionais similares precisam ser conectados. É claro que uma disposição deste tipo economiza espaço e energia.

Alguns pesquisadores afirmam que a organização topográfica é uma propriedade útil até mesmo na ausência de qualquer restrição biológica. No córtex visual primário, por exemplo, conjectura-se que o sistema sensorial deve utilizar a estrutura estatística dos sinais de estímulo para processar informações [2, 47, 48]. Estas pesquisas têm-se concentrado na propriedade da esparsificidade. Esta propriedade estatística será abordada no capítulo 3.

Capítulo 2

Redes Neuroniais Artificiais

2.1 Introdução

As Redes Neuroniais Artificiais (RNA's) foram desenvolvidas, originalmente pelo neurofisiologista Warren McCulloch e pelo matemático Walter-Pitts, que utilizando conceitos da cibernética, fizeram uma analogia entre células nervosas vivas e o processo eletrônico, em um trabalho publicado em 1943 [49]. O trabalho consistia em um modelo de resistores variáveis e amplificadores representando conexões sinápticas de um neurônio biológico.

Desde então, mais precisamente a partir do decênio de 1980, diversos modelos de redes neuronias artificiais têm surgido com o propósito de aperfeiçoar e aplicar esta tecnologia. Na Figura 2.1 pode-se ver a arquitetura de uma RNA.

As variantes de uma rede neuronal são muitas, e combinando-as, pode-se mudar a arquitetura conforme a necessidade da aplicação. Todas as RNA's possuem alguma regra de treinamento, em que os pesos de suas conexões são ajustados de acordo com os padrões apresentados, de tal forma que podemos dizer que a capacidade de aprendizagem é uma das suas propriedades.

A aprendizagem é feita por um processo iterativo de ajustes dos pesos . Quando a RNA atinge uma solução generalizada para uma classe de problemas diremos que houve aprendizagem.

Um conjunto de regras bem definidas para que a RNA aprenda a solução

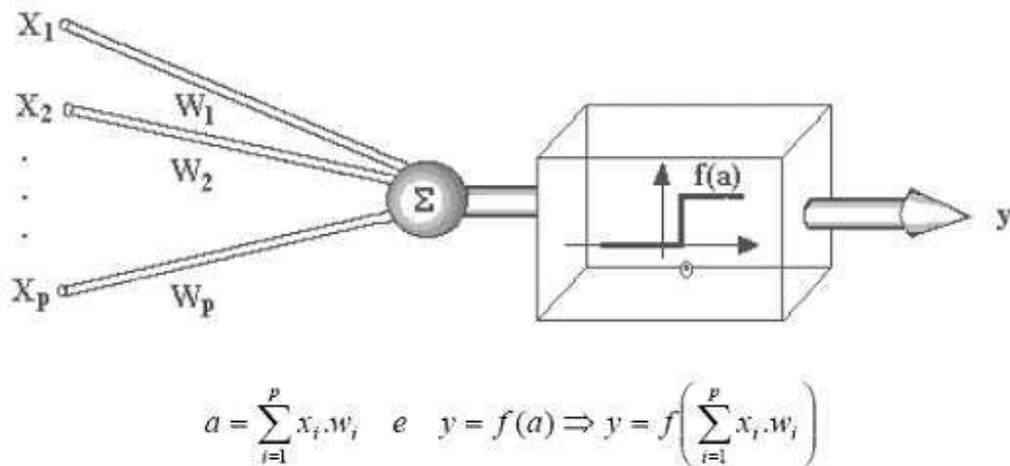


Figura 2.1: Modelo de McCulloch e Pitts.

de um problema é chamado de algoritmo de aprendizagem. Existem muitos algoritmos de aprendizagem específicos para determinados modelos de redes neuroniais. Estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. Outro ponto importante é a classificação das situações de aprendizagem das RNA's. Pode-se citar duas classes de aprendizagem [8,9]:

- Aprendizagem Supervisionada: quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- Aprendizagem Não Supervisionada: quando não existe um agente externo indicando a resposta desejada para os padrões de entrada.

2.2 Redes Neuroniais Perceptron Multicamadas

As Redes Neuroniais Perceptron Multicamadas (MLP, do inglês *Multilayer Perceptron*) são frequentemente usadas em reconhecimento de padrões. Uma RNA do tipo MLP é constituída por um conjunto de nós fontes, os quais formam a camada de entrada da rede, uma ou mais camadas escondidas e uma camada de saída.

O número de nós fontes na entrada da rede é determinado pela dimensionalidade do espaço de observações, que é responsável pela geração dos sinais de entrada. O número de neurônios na camada de saída é determinado pela dimensionalidade requerida na resposta desejada. Assim, o projeto de uma rede MLP requer a consideração de três aspectos:

1. A determinação do número de camadas escondidas;
2. A determinação do número de neurônios em cada uma das camadas escondidas;
3. A especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede.

Os aspectos 1 e 2 determinam a complexidade do modelo da RNA escolhida e, infelizmente, não há regras determinadas para tal especificação. A função das camadas escondidas em uma RNA é de influir na relação entrada-saída da rede de uma forma ampla. Uma RNA com uma ou mais camadas escondidas é apta a extrair as características de ordem superior de algum processo aleatório desconhecido, responsável pelo comportamento dos dados de entrada. Processo esse sobre o qual a rede está tentando adquirir conhecimento. A RNA adquire uma perspectiva global do processo aleatório, apesar de sua conectividade local, em virtude do conjunto adicional de pesos sinápticos e da dimensão adicional de interações neuronais proporcionada pelas camadas escondidas [8].

O terceiro aspecto envolve a utilização de algoritmos de treinamento supervisionados. As RNA's MLP têm sido aplicadas nas soluções de diversos e difíceis problemas através da utilização de tais algoritmos [8, 9]. Um algoritmo muito utilizado é o algoritmo da retropropagação, tecnicamente conhecido como *backpropagation*. Este algoritmo baseia-se na heurística do aprendizado por correção de erro, o qual pode ser visto como uma generalização do algoritmo LMS (*Least Mean Square*), desenvolvido por Bernard Widrow [50]. Basicamente, o algoritmo da retropropagação consiste de dois passos através das diferentes camadas do MLP: um passo direto e um passo reverso:

- No passo direto, um padrão de atividade do processo a ser aprendido (ou vetor de entrada) é aplicado aos nós de entrada da MLP e o seu efeito se propaga através da rede, camada por camada, produzindo na camada de saída a resposta da MLP à excitação aplicada (vetor de saída). Durante o passo direto os pesos sinápticos são todos fixos.

- No passo reverso, os pesos sinápticos são todos ajustados de acordo com a regra de aprendizagem por correção de erro. Especificamente, a resposta da MLP à excitação é subtraída de um padrão de resposta desejado para aquela excitação aplicada, de forma a produzir um sinal de erro, de forma semelhante ao algoritmo LMS, como pode ser visto na Figura 2.2. Este sinal de erro é, então, propagado de volta aos mesmos neurônios utilizados no passo direto, mas no caminho contrário ao do fluxo daquele sinal nas conexões sinápticas, daí o nome retropropagação. Os pesos sinápticos são, então, ajustados de forma que a resposta obtida da MLP aproxime-se mais do padrão de resposta desejado, de tal forma que

$$e_j(n) = d_j(n) - y_j(n), \quad (2.1)$$

na qual e_j é o sinal de erro, d_j é o sinal desejado, y_j é o sinal direto, j refere-se ao j -ésimo neurônio e n é a n -ésima amostra.

O modelo de cada neurônio MLP inclui uma função de ativação não linear. Esta não linearidade é suave, sendo a mais utilizada a função logística,

$$y_j = \frac{1}{1 + \exp(-v_j)}, \quad (2.2)$$

na qual v_j é chamado de potencial de ativação e é a soma ponderada de todas as entradas sinápticas do neurônio. Uma parte do gráfico da função dada pela equação 2.2 está representada na Figura 2.3,

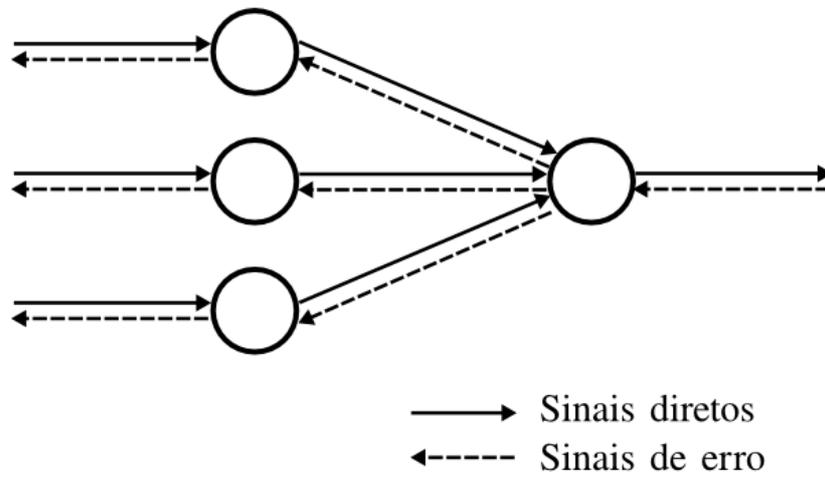


Figura 2.2: Direções dos fluxos de um sinal em uma rede MLP: propagação direta dos sinais e retro-propagação dos sinais de erro.

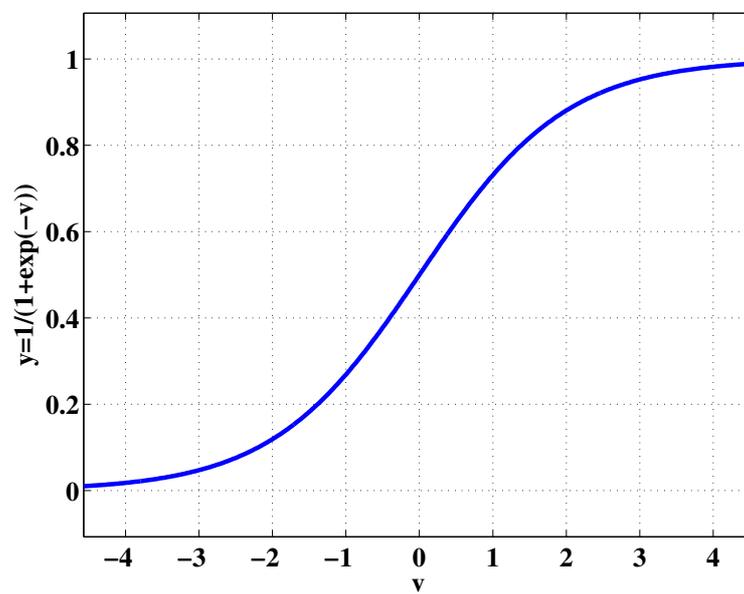


Figura 2.3: função de ativação não linear para um neurônio.

Capítulo 3

Codificação Eficiente ou Esparsa

3.1 Introdução

Supõe-se que uma das estratégias utilizadas pelo cérebro para representar informação seja o princípio da codificação eficiente. Este conceito foi proposto por Horace Barlow em 1961 [2] como um modelo teórico para a codificação das informações sensoriais pelo sistema nervoso. Para ele, um modelo eficiente seria aquele que minimizasse a quantidade de impulsos nervosos utilizados para transmitir a informação desejada.

No desenvolvimento de sua teoria, Barlow foi inspirado por conceitos da Teoria da Informação. Ele definiu que os caminhos neuroniais percorridos por informações sensoriais são similares a canais de comunicação. Através de conceitos como capacidade de canal e redundância, Barlow sugeriu que a codificação neuronal é realizada de forma a maximizar a capacidade de canal e reduzir a redundância na informação transmitida.

Biologicamente, o conceito de codificação eficiente pode ser descrito em duas perspectivas: para uma única célula e para múltiplas células. Na primeira, é considerado que o padrão de resposta de um neurônio deve ser esparsamente distribuído [20].

Uma variável aleatória é dita esparsa quando ela assume valores pequenos (absolutos) ou valores altos, com mais frequência do que uma variável aleatória

gaussiana assume. Para compensar, ela, a variável aleatória, raramente assume valores intermediários. Desta forma, ela é ativa, isto é, significativamente diferente de zero, apenas raramente. Uma variável aleatória esparsa também é dita supergaussiana ou leptocúrtica. Nas figuras 3.1 e 3.2 pode-se verificar estes fatos.

Uma distribuição de probabilidade esparsa pode ser caracterizada por possuir um pico em zero e longas extremidades. Uma curva que pode caracterizar a distribuição da resposta de um neurônio é mostrada na figura 3.3. As longas extremidades fazem com que as probabilidades de resposta da célula sejam pequenas. Dessa forma, para um dado estímulo, é provável que apenas um pequeno conjunto de células responda.

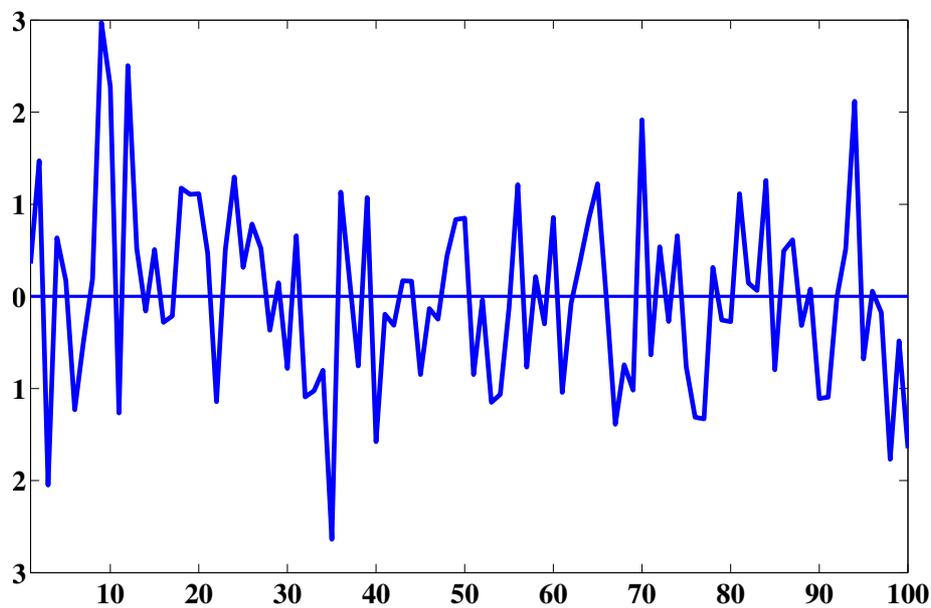


Figura 3.1: Cem amostras aleatórias de uma variável gaussiana com média zero e desvio padrão 1.

O fato de uma variável aleatória ser esparsa é independente da sua variância. Para medi-la, normaliza-se uma variável aleatória, digamos s_i , de forma a ficar com média zero e variância, $E[s_i^2]$, igual a uma constante dada. Desta forma, mede-se a esparsificidade como o valor esperado $E[f(s_i^2)]$, para uma função não linear, f , adequada. Tipicamente f é escolhida como sendo convexa, isto é, sua derivada segunda é positiva. Convexidade implica que o valor da expectância é

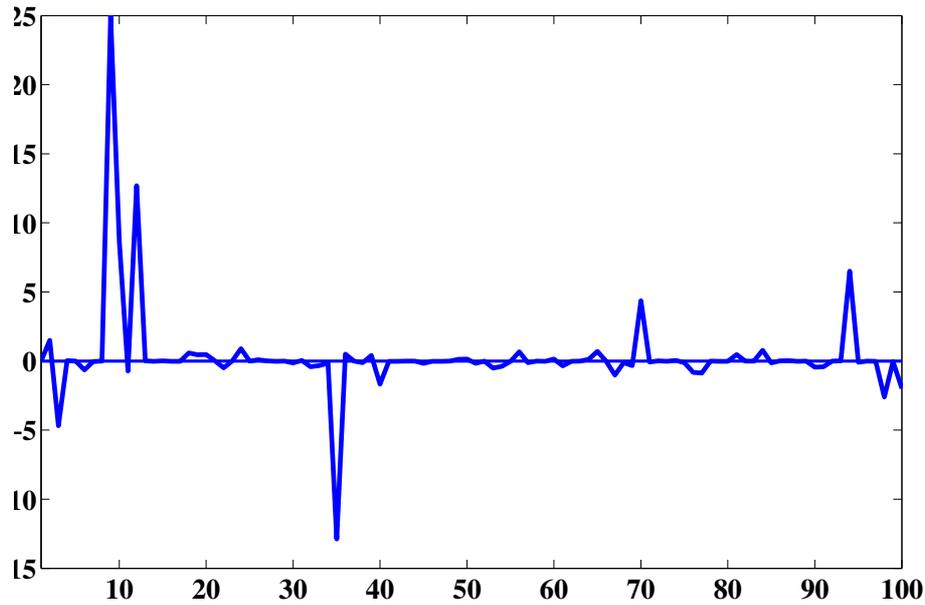


Figura 3.2: Cem amostras aleatórias de uma variável esparsa ou supergaussiana.

grande quando s_i^2 toma valores muito próximos de zero ou muito próximos de 1, ou seja, quando s_i é esparsa [21].

Sabe-se que células do córtex visual primário de primatas apresentam distribuições esparsas quando estimuladas com sequências de imagens [51]. Este comportamento também se estende ao córtex auditivo primário-A1 [52]. As células de A1 podem produzir um único disparo em resposta a um estímulo sonoro. Este comportamento foi caracterizado como uma ‘codificação binária’ porque tais células produzem ‘0 ou 1’ em resposta a um estímulo. No entanto, a probabilidade de disparo é muito pequena ao longo do tempo, o que é consistente com a ideia de esparsificidade. Pequenas probabilidades de disparo também são observadas em células do córtex somatosensorial de roedores [53]. Além disso, neurônios do sistema olfativo de insetos disparam somente até duas vezes para odores utilizados como estímulos [54]. Dessa forma, esparsificidade parece ser um dos princípios utilizados na codificação neuronal.

Na segunda perspectiva, a representação de informação sensorial é eficiente caso as respostas neuronais sejam estatisticamente independentes. Dessa forma, não existe "informação redundante" entre células da população.

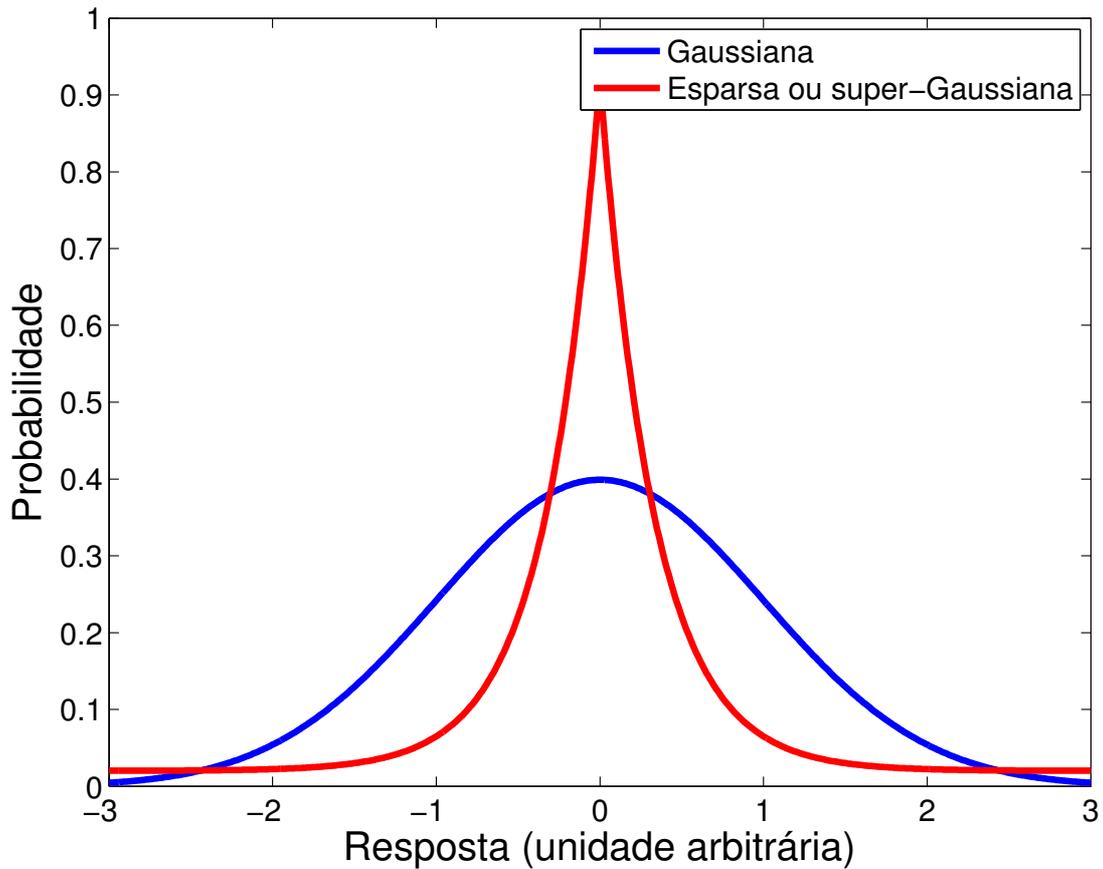


Figura 3.3: Funções de densidade de probabilidades: Esparsa ou super-Gaussiana (linha vermelha); Gaussiana (linha azul).

De fato, células do córtex visual primário emitem fortes respostas quando estimuladas com estruturas não redundantes, tais como, barras e bordas [38]. Isto sugere que estímulos visuais devam ser representados de forma eficiente no córtex visual primário.

O conceito de codificação eficiente também pode ser definido a partir de um modelo matemático para geração de imagens. Neste modelo, uma imagem \mathbf{X} é gerada a partir de um conjunto de filtros $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$, no qual os \mathbf{a}_i são vetores colunas, ativados independentemente por coeficientes $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_n]$. O modelo matemático é dado por

$$\mathbf{X} = \sum_{i=1}^n \mathbf{a}_i \mathbf{s}_i. \quad (3.1)$$

Olshausen e Field [30] mostraram que utilizando imagens naturais, este modelo é capaz de gerar filtros \mathbf{A} similares aos campos receptivos das células simples de V1. Na Figura 3.4 pode-se ver os filtros gerados a partir da utilização do algoritmo, produzido por estes pesquisadores.

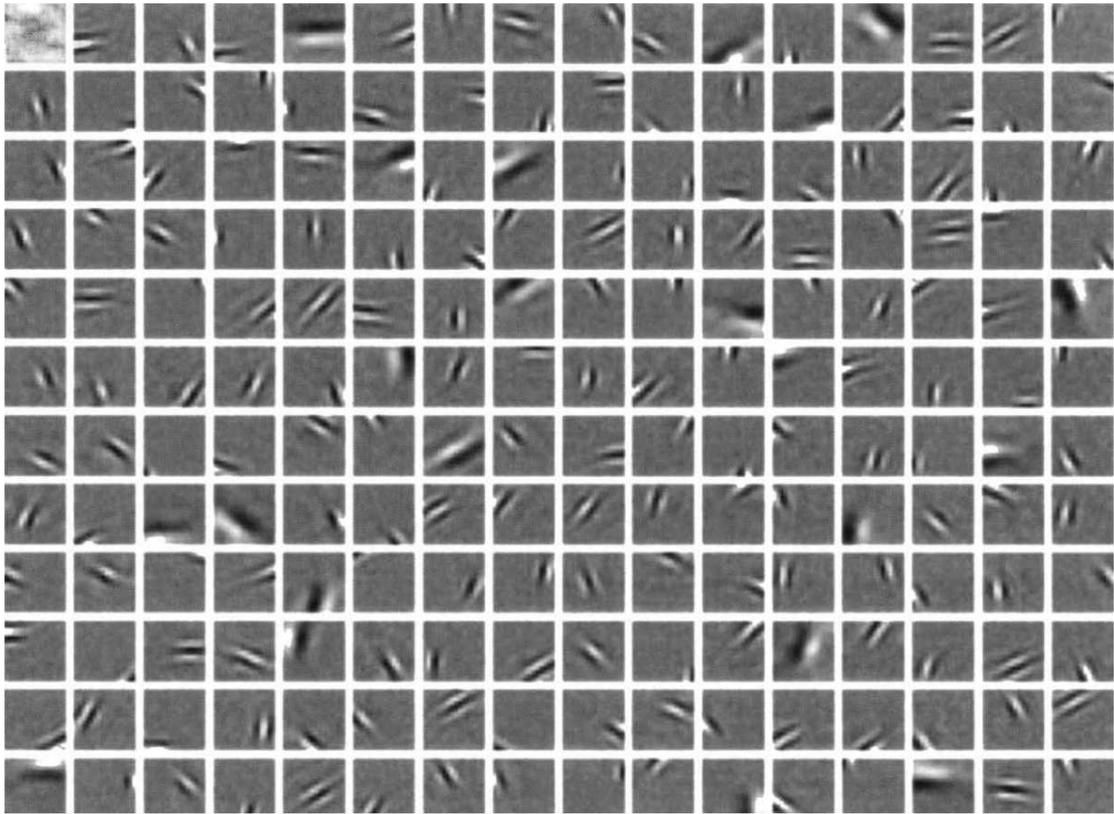


Figura 3.4: Filtros gerados utilizando-se codificação eficiente.

É possível observar que os filtros mostrados na Figura 3.4 apresentam características similares às *wavelets* de Gabor que correspondem aos campos receptivos de V1. A partir desses resultados, foi proposto por Olshausen e Field [20] que os campos receptivos das células simples de V1 são gerados a partir de um processo neuronal baseado em codificação eficiente.

O modelo proposto por aqueles autores supõe a determinação de funções base

que gerem as imagens através de combinações lineares que resultem em coeficientes estatisticamente independentes, tal como no modelo de Análise de Componentes Independentes.

Capítulo 4

Análise de Componentes Independentes

4.1 Introdução

Análise de Componentes Independentes (ICA) é uma técnica estatística e computacional para processamento de sinais cujo objetivo é representar um conjunto de variáveis aleatórias como uma combinação linear de componentes estatisticamente independentes [55]. Na forma mais simples de ICA, observam-se m variáveis aleatórias escalares x_1, x_2, \dots, x_m , as quais são supostas como combinações lineares de n componentes independentes desconhecidas, denotadas por s_1, s_2, \dots, s_n . Estas componentes independentes, s_i , são supostas como sendo mutuamente estatisticamente independentes. Colocando as variáveis observadas, x_i , em um vetor $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$, e as variáveis independentes, s_i , em outro, $\mathbf{s} = [s_1, s_2, \dots, s_n]^T$, então a combinação linear será dada por

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (4.1)$$

na qual \mathbf{A} é uma matriz desconhecida, de ordem $m \times n$, chamada de matriz de mistura.

O problema básico em ICA é estimar ambos: a matriz de mistura e as

realizações das componentes independentes, utilizando apenas as observações das combinações x_j .

A restrição fundamental no modelo é que só se podem estimar componentes independentes que tenham distribuição de probabilidades não gaussianas (a não ser que apenas uma delas seja gaussiana). Além do mais, não se pode estimar nem as energias, $E[s_i^2]$, nem os sinais (positivos ou negativos) das componentes independentes, porque, uma constante multiplicando qualquer uma das componentes independentes, na Equação (4.1) pode ser cancelada, dividindo-se a coluna respectiva da matriz pela mesma constante. Objetivando-se uma maior facilidade matemática, geralmente definem-se as componentes independentes como tendo variâncias unitárias. Este procedimento faz com que as componentes independentes (não gaussianas) sejam únicas (exceto pelo sinal) [56]. Observe que esta definição de ICA não implica em nenhuma ordem das componentes independentes.

4.2 Não gaussianidade e Independência Estatística

A partir do teorema do limite central, um resultado clássico da teoria da probabilidade, pode-se concluir que a distribuição de probabilidade da soma de duas ou mais variáveis aleatórias independentes tende para uma distribuição gaussiana, sob certas condições. Ou seja, a soma de duas ou mais variáveis aleatórias independentes têm uma distribuição mais próxima de uma distribuição gaussiana que qualquer uma das variáveis parcelas não gaussianas.

Para este estudo, supõe-se que o vetor das variáveis observadas (vetor de dados), \mathbf{x} , o qual segue o modelo básico de ICA, como na Equação (4.1), é uma combinação linear de componentes independentes, \mathbf{s} . Supõe-se, também, que tais componentes independentes tenham distribuições de probabilidades idênticas e médias zero. A estimação destas componentes pode ser feita encontrando-se a combinação linear correta das variáveis observadas tal que,

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}. \quad (4.2)$$

Desta forma, para estimar uma das componentes independentes pode-se considerar uma combinação linear dos x_i . denote-se tal combinação por $y = \mathbf{b}^T \mathbf{x} = \sum_i b_i x_i$, na qual \mathbf{b} é um vetor a ser determinado. Observe, também, que pode-se ter $y = \mathbf{b}^T \mathbf{A} \mathbf{s}$. Desta forma, tem-se que y é uma certa combinação linear dos s_i , com coeficientes dados por $\mathbf{b}^T \mathbf{A}$. Represente tais coeficientes por um vetor \mathbf{q} . Então,

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s} = \sum_i q_i s_i. \quad (4.3)$$

Se \mathbf{b} for uma das linhas da matriz inversa de \mathbf{A} , esta combinação linear, $\mathbf{b}^T \mathbf{x}$, será, realmente, igual a uma das componentes independentes.

Pretende-se, agora, utilizar o teorema do limite central para determinar \mathbf{b} de forma tal que seja igual a uma das linhas da matriz inversa de \mathbf{A} . Na prática, não se pode determinar tal \mathbf{b} exatamente, devido ao fato de não se ter nenhum conhecimento da matriz \mathbf{A} . Mas, pode-se encontrar um estimador que dê uma boa aproximação.

Faça-se os coeficientes do vetor \mathbf{q} variar e veja como a distribuição de $y = \mathbf{q}^T \mathbf{s}$ se comporta. A idéia fundamental, aqui, é que, desde que a distribuição de probabilidade da soma de variáveis aleatórias independentes é mais gaussiana que a distribuição de probabilidade de qualquer uma das variáveis originais não gaussianas, $y = \mathbf{q}^T \mathbf{s}$ deverá ser mais gaussiana que qualquer uma das componentes s_i e torna-se menos gaussiana quando, de fato, é igual a uma das s_i . Observe-se que isto só será estritamente verdadeiro se as componentes s_i tiverem distribuições de probabilidades idênticas, como está sendo suposto aqui. Neste caso, obviamente, apenas um dos elementos, q_i , do vetor \mathbf{q} , será diferente de zero.

Não se tem condições, na prática, de conhecer os valores de \mathbf{q} , mas isto não é necessário porque, como $\mathbf{q}^T \mathbf{s} = \mathbf{b}^T \mathbf{x}$ na definição de \mathbf{q} , pode-se deixar \mathbf{b} variar e atentar para a distribuição de $\mathbf{b}^T \mathbf{x}$.

Desta forma, pode-se considerar \mathbf{b} como um vetor que *maximiza a não gaussianidade de $\mathbf{q}^T \mathbf{x}$* . Tal vetor deverá, necessariamente, corresponder a $\mathbf{q} = \mathbf{A}^T \mathbf{b}$, que terá apenas um elemento diferente de zero. Isto significa que $y = \mathbf{b}^T \mathbf{x} =$

$\mathbf{q}^T \mathbf{s}$ será igual a uma das componentes independentes. Desta forma, a maximização da não gaussianidade de $\mathbf{b}^T \mathbf{x}$ dará uma das componentes independentes.

4.3 Medindo a Não Gaussianidade Através da Curtose

Para usar a não gaussianidade na estimação de ICA, deve-se ter uma medida quantitativa da não gaussianidade de uma variável aleatória y . Veja aqui, de uma forma breve, como ICA utiliza a curtose [57], para estimação de não gaussianidade. Curtose é o nome dado ao cumulante de quarta ordem de uma variável aleatória. A curtose de y , denotada por $Kurt(y)$, é definida por:

$$Kurt(y) = E[y^4] - 3(E[y^2])^2. \quad (4.4)$$

Lembre que todas as variáveis aleatórias, aqui, têm médias zero. No caso geral, a definição de curtose é um pouco mais complicada. Para facilitar as coisas, pode-se supor que y foi normalizado de forma tal que sua variância é unitária: $E[y^2] = 1$. Então, o lado direito da Equação (4.4) reduz-se a $E[y^4] - 3$. Vê-se, então, que a curtose é simplesmente uma versão normalizada do quarto momento, $E[y^4]$. Para uma variável aleatória, y , gaussiana, o quarto momento é igual a $3(E[y^2])^2$. Logo, a curtose é nula para uma variável aleatória gaussiana. Para grande parte das variáveis aleatórias não gaussianas a curtose é diferente de zero.

A curtose pode ser positiva ou negativa. Variáveis aleatórias que têm curtose negativa são chamadas de sub-gaussianas, e aquelas com curtose positiva são chamadas de super-gaussianas.

Tipicamente, a não gaussianidade é medida pelo valor absoluto da curtose. O seu quadrado também pode ser usado. Estas medidas são zero para uma variável gaussiana e maior que zero para uma não gaussiana.

A principal razão para usar curtose, ou seu valor absoluto, para medir não gaussianidade é a sua simplicidade tanto computacional quanto teórica.

Computacionalmente, a curtose pode ser estimada simplesmente usando o

quarto momento dos dados amostrais (Se a variância for constante). A análise teórica é simplificada por causa das seguintes propriedades lineares: se x_1 e x_2 são duas variáveis aleatórias independentes, então,

$$Kurt(x_1 + x_2) = kurt(x_1) + kurt(x_2). \quad (4.5)$$

e

$$Kurt(\alpha x_1) = \alpha^4 kurt(x_1), \quad (4.6)$$

na qual α é uma constante real.

Algumas palavras devem ser ditas, aqui, a respeito da curtose como medida de não gaussianidade. Existem algumas dúvidas sobre de que forma a curtose está relacionada à forma de uma distribuição de probabilidades. Geralmente, ela é de grande utilidade quando lida-se com distribuições aproximadamente simétricas. Muitos autores têm dito que curtose é uma medida do "achatamento" de uma distribuição o que não é estritamente verdadeiro [58]. É fácil confundir pequenos valores de curtose com baixa variância, mas distribuições com valores de curtoses idênticos podem diferir em variância e distribuições com iguais medidas de variância podem diferir em curtose. Atualmente, pesquisadores têm referido-se à curtose como, em parte, uma medida do alongamento da distribuição relativamente à variância total.

Mas, apesar disto, a curtose tem sido utilizada, em Análise de Componentes Independentes, com grande aproveitamento, como uma medida de não gaussianidade.

4.4 Um algoritmo de ponto fixo usando curtose

Considerando o modelo básico de ICA, dado pela equação (3.1), considere-se uma combinação linear dos sinais observados, \mathbf{x} , na forma $\mathbf{w}^T \mathbf{x}$, na qual o vetor \mathbf{w} é tal que $E[(\mathbf{w}^T \mathbf{x})^2] = 1$ [22]. Quando $\mathbf{w}^T \mathbf{x}$ for igual, a menos do sinal, a uma das componentes independentes, a curtose de $\mathbf{w}^T \mathbf{x}$ é localmente maximizada ou

minimizada [59, 60]. Esta propriedade é bastante utilizada no desenvolvimento de algoritmos em ICA. Para maximizar (ou minimizar) a curtose de $\mathbf{w}^T \mathbf{x}$, um algoritmo neuronal baseado no método da subida (ou da descida) mais íngreme pode ser utilizado. O gradiente da curtose de $\mathbf{w}^T \mathbf{x}$ em relação a \mathbf{w} é:

$$\nabla_{\mathbf{w}} kurt(\mathbf{w}^T \mathbf{x}) = 4\{E[\mathbf{x}(\mathbf{w}^T \mathbf{x})^3] - 3\mathbf{C}\mathbf{w}E[(\mathbf{w}^T \mathbf{x})^2]\}, \quad (4.7)$$

na qual $\mathbf{C} = E[\mathbf{x}\mathbf{x}^T]$ é a matriz de covariância dos sinais observados.

Após, algumas manipulações matemáticas obtêm-se o algoritmo seguinte [21].

$$\begin{aligned} \mathbf{W}_k &= E[\mathbf{x}(\mathbf{w}^T \mathbf{x})^3] - 3\mathbf{w}_{k-1} \\ \mathbf{w}_k &= \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}. \end{aligned} \quad (4.8)$$

Este algoritmo, denominado de FASTICA, tem sido largamente utilizado para separação cega de fontes [61–63], separação cega de fontes com estruturas temporais [64, 65] e modelamento do córtex visual primário através de imagens naturais [66, 67].

As características principais deste algoritmo são: (1) convergência cúbica (alta velocidade); (2) não necessita de um parâmetro para controlar a taxa de convergência (tamanho do passo, nos algoritmos que utilizam o método do gradiente). Isto faz com que o FASTICA seja de fácil implementação.

Geralmente, como uma etapa de pré-processamento, os algoritmos para ICA removem as correlações de segunda ordem nos sinais observados, proporcionando um espectro de potência que tenha mesma amplitude para todas as frequências. Esta etapa é chamada de branqueamento. Isto é feito removendo-se o vetor média, $E[\mathbf{x}]$, do vetor de observações, \mathbf{x} , e normalizando-se através de uma transformação linear \mathbf{U} de forma tal que a matriz de covariância dos dados branqueados, $\mathbf{z} = \mathbf{U}\mathbf{x}$, seja a matriz identidade. Este processo é similar à Análise em Componentes Principais (PCA, do inglês *Principal Component Analysis*). Esta etapa de pré-processamento reduz o espaço de busca da matriz formada pelos vetores \mathbf{w} para o subespaço das matrizes ortogonais [21, 56].

A seguir, veja como pode utilizar-se ICA para codificar imagens naturais.

4.5 ICA aplicado a imagens naturais

Imagens naturais podem ser definidas como imagens fotográficas tomadas do meio ambiente natural. Na Figura 4.1 tem-se alguns exemplos de imagens naturais.



Figura 4.1: Exemplos de imagens naturais em preto e branco.

O modelo estatístico para imagens naturais baseado em ICA é definido como superposições lineares de imagens bases, com coeficientes independentes e não gaussianos. Denotemos por \mathbf{x} um vetor aleatório cujas componentes são

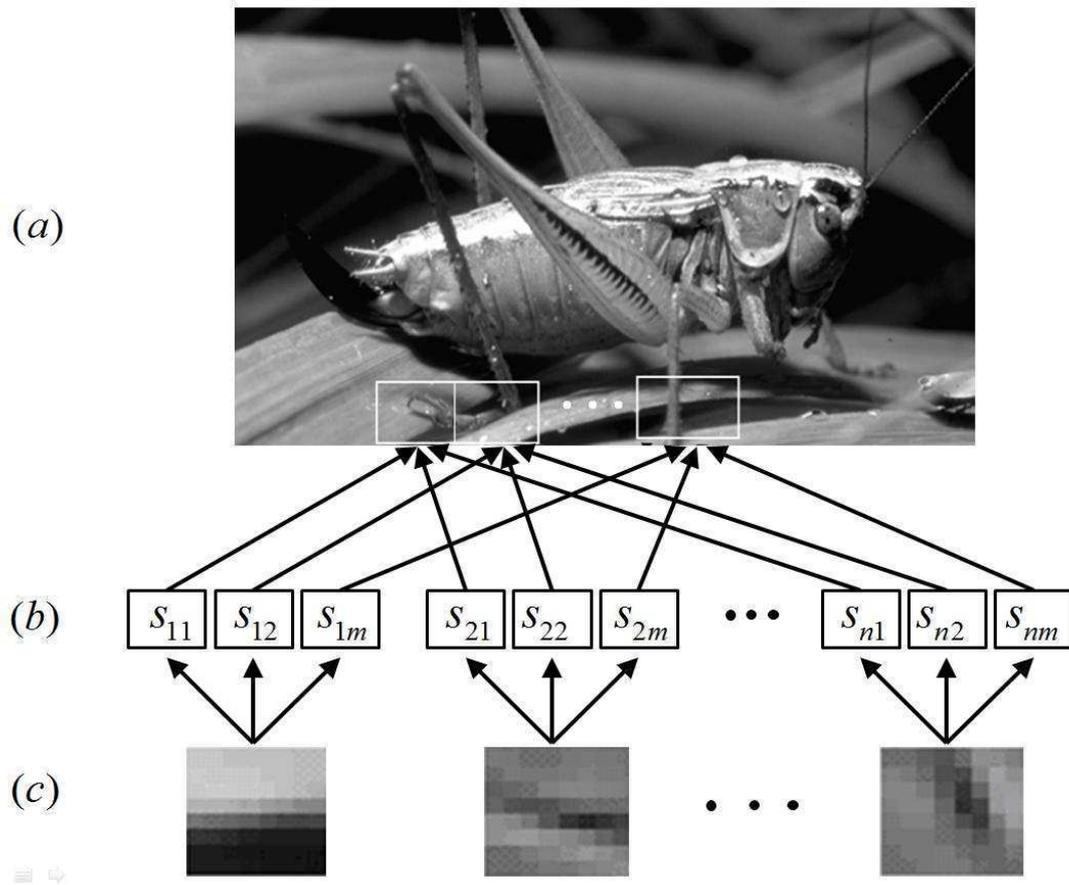


Figura 4.2: Amostra de uma imagem natural disposta como combinação linear das funções (imagens) base. (a) amostras aleatórias da imagem, \mathbf{x} ; (b) componentes independentes s_i ; (c) funções base, \mathbf{a}_i .

valores de pixels em uma escala de cinza (pontos de luminosidade) de partes de imagens. O modelo básico considerado aqui expressa uma parte de uma imagem monocromática estática, \mathbf{x} , como uma superposição linear de algumas características ou vetores bases \mathbf{a}_i :

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i, \tag{4.9}$$

na qual os s_i são coeficientes estocásticos, diferentes para cada imagem \mathbf{x} . Em uma interpretação cortical, os s_i modelam as respostas de células simples e os \mathbf{a}_i

estão relacionados com os campos receptivos destas células [20,68]. Neste trabalho, considera-se apenas o contraste da imagem, isto é, a média local (ou componente DC) é removida da imagem o que significa que as componentes s_i têm média zero.

Este modelo, primeiramente aplicado em imagens por Olshausen e Field [20], é ilustrado na Figura 4.2:

Quando as imagens base são otimizadas para modelar imagem natural, elas se tornam orientadas, localizadas e com características de filtro passa-faixa, podendo ser modeladas através das transformadas de *wavelets*. Na Figura 4.3, pode-se ver algumas imagens bases obtidas pela aplicação do algoritmo desenvolvido pelos pesquisadores Olshausen e Field.

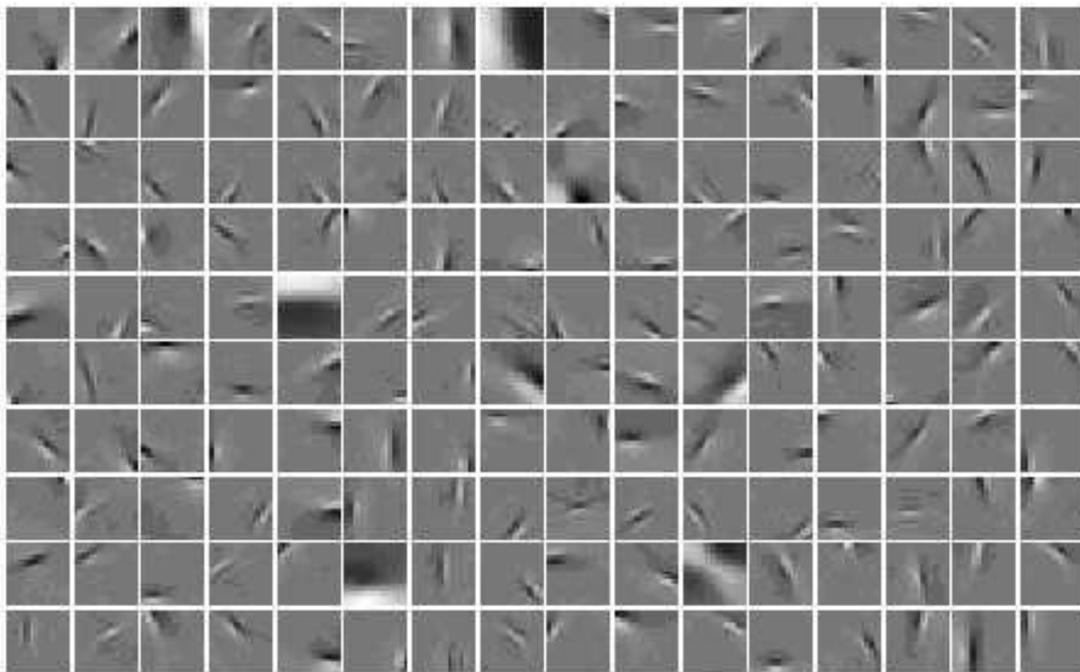


Figura 4.3: Funções bases aprendidas pelo modelo proposto por Olshausen e Field.

É importante observar que o número de funções base estimadas é maior que a dimensão do espaço das imagens. Esta característica permite a geração de um espaço onde pode-se ter uma representação conjunta das características de posição, de frequência e fase [69], permitindo, desta forma, uma maior flexibilidade na

representação, já que não existe nenhuma razão, a priori, para acreditar que o número de causas para imagens é menor que ou igual ao número de pixels [70].

Esta decomposição assemelha-se bastante às representações dadas pelas células simples no córtex visual primário dos mamíferos.

Em suma, pode-se dizer que representações de imagens dadas pelo modelamento em ICA tem alcançado grande sucesso em processamento de imagem e modelamento do córtex visual primário. Isto tem proporcionado métodos baseados em decomposição em *wavelets* devido ao fato de fornecer uma interpretação estatística ao modelo [7, 20, 26, 71].

4.6 O Modelo TICA

No modelo básico de Análise de Componentes Independentes as componentes são supostas completamente independentes e não apresentam nenhuma relação de ordem significativa [56]. Na prática, entretanto, as componentes estimadas não são de todo independentes. Aapo Hyvarinen e seus colaboradores propõem [28] que esta estrutura de dependência residual pode ser usada para definir uma ordem topográfica entre as componentes. Para isto foi desenvolvido o algoritmo TICA (do inglês, *Topographic Independent Component Analysis*), que é descrito a seguir de forma sucinta.

Os autores definem um modelo gerador para ICA no qual as componentes/fontes não são completamente independente mas apresentam uma dependência que é definida com relação a alguma topologia, tal como uma matriz toroidal - componentes próximas umas das outras, nesta topologia, têm maiores co-dependências do que aquelas mais afastadas.

O esquema para este modelo pode ser visto na figura 4.4. A primeira camada fornece uma combinação linear de variáveis definidas como "geradoras de variâncias", \mathbf{t} , na qual a matriz \mathbf{H} é a matriz de mistura. Em seguida, estas variáveis passam por uma não-linearidade, $\phi(\cdot)$, resultando em escalares positivos, $\sigma_i = \phi(\mathbf{H}_i^T \mathbf{t})$, que representam as variâncias de cada uma das fontes s_i . Estas fontes são, então, geradas de forma a terem distribuições de probabilidades

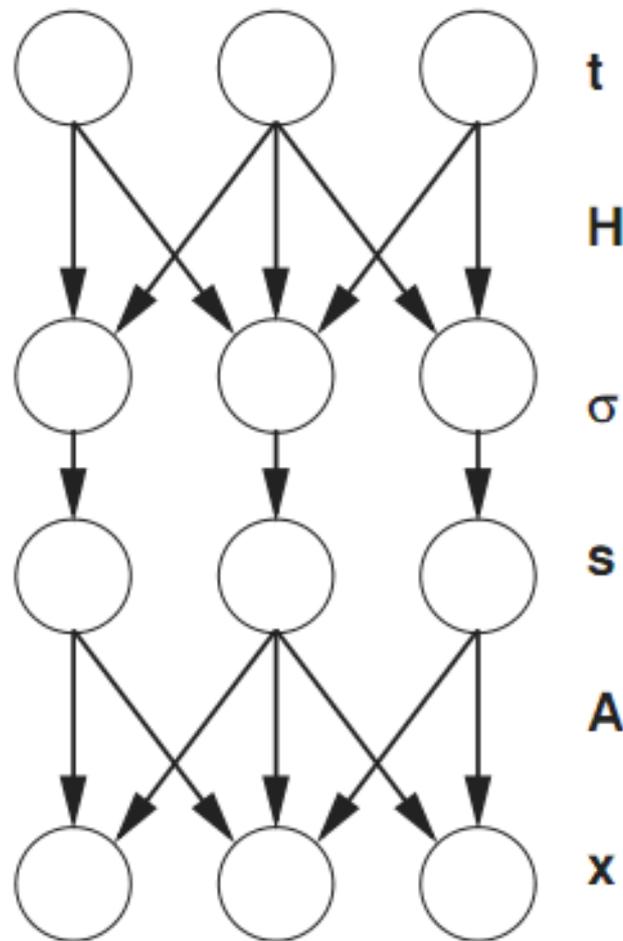


Figura 4.4: Modelo para TICA.

independentes, com média zero e com variância σ_i . Em seguida as fontes s_i são combinadas linearmente, pela matriz **A**, para fornecer os sinais observáveis x_i

Estes, por sua vez, são usados para gerar as variâncias das fontes s e, condicionados a estes escalares, as componentes da terceira camada são independentes. Estas são, finalmente, combinadas linearmente pela matriz **A**, resultando nos sinais observáveis, **x**.

A distribuição conjunta de s é dada por

$$P(s, t) = \prod p_{s_i} \left(\frac{s_i}{\phi(\mathbf{H}_i^T \mathbf{t})} \right) \frac{p_{t_i}(t_i)}{\phi(\mathbf{H}_i^T \mathbf{t})}, \quad (4.10)$$

e a função log-verossimilhança dos dados, sendo fornecidos os parâmetros, é dada por

$$L(\mathbf{B}) = \sum_{datax} \int \prod_i p_{s_i} \left(\frac{s_i}{\phi(\mathbf{H}_i^T \mathbf{t})} \right) \frac{p_{t_i}(t_i)}{\phi(\mathbf{H}_i^T \mathbf{t})} |\det(\mathbf{B})| dt, \quad (4.11)$$

na qual $\mathbf{B} = \mathbf{A}^{-1}$.

Como assinalado pelos autores a equação 4.11 é de difícil computação devido à intratabilidade desta integral. Este fato os levou a tomar uma abordagem que utiliza várias aproximações para determinar um limite inferior da função de verossimilhança.

Primeiro, eles restringem a função de distribuição de probabilidade de s a ser Gaussiana, \mathbf{t} e \mathbf{H} são supostos não negativos e $\phi(\cdot)$ é tomado como sendo $(\cdot)^{-1}$. Estas restrições conduzem à seguinte expressão para a distribuição marginal de s :

$$p(s) = \int \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} \sum t_k [\sum H_{ik} s_i^2]\right) \prod_i p_{t_i} \sqrt{\mathbf{H}_i^T \mathbf{t}} dt. \quad (4.12)$$

Esta expressão é, então, simplificada pela aproximação

$$\sqrt{\mathbf{H}_i^T \mathbf{t}} \approx \sqrt{H_{ii} t_i}. \quad (4.13)$$

Embora esta não seja uma boa aproximação, ela é um limitante inferior na função de verossimilhança. A função de verossimilhança aproximada final, obtida por estes autores é então:

$$\tilde{L}(\mathbf{B}) = \sum \left(\sum_{j=1}^d G \left(\sum_{i=1}^d H_{ij} (\mathbf{B}_i^T \mathbf{x})^2 \right) + \log |\det(\mathbf{B})| \right), \quad (4.14)$$

na qual a função escalar G é da forma

$$G(\tau) = \log \int \frac{1}{2\pi} \exp\left(\frac{1}{2} t\tau\right) p_t(t) \sqrt{H_{ii}} dt. \quad (4.15)$$

A partir daí, eles desenvolveram o algoritmo dado por

$$\Delta \mathbf{W} = \mathbf{E}[\mathbf{x}(\mathbf{W}_i^T \mathbf{x})r_i], \quad (4.16)$$

no qual r_i é uma determinada função.

Portanto, o que o algoritmo desenvolvido por aqueles autores resolve é a equação 4.14 e não o problema originalmente proposto.

Nesta tese apresenta-se um modelo de rede neuronal artificial que soluciona o problema originalmente proposto de forma bastante simplificada e com a utilização de um algoritmo de aprendizagem de ponto fixo com um menor custo computacional que aquele apresentados pelos autores do TICA. O algoritmo proposto naquele trabalho baseia-se no método da descida mais íngreme, dependente, portanto, da escolha perfeita do parâmetro que viabiliza a convergência [50].

Capítulo 5

Desenvolvimento do Modelo

5.1 Introdução

O desenvolvimento de modelos de Redes Neuroniais Artificiais que tenham plausibilidade biológica é um dos grandes anseios dos neurocientistas na tentativa de modelar o cérebro humano. Nesta tentativa, algumas modificações no modelo básico de ICA têm sido desenvolvidas para descrever estruturas estatísticas de imagens naturais e também explicar alguns aspectos do córtex visual primário de forma a apresentar uma maior plausibilidade biológica. Estas modificações usam uma decomposição linear, como no modelo básico de ICA, mas as componentes apresentam algum tipo de dependência [72].

Uma metodologia desenvolvida, utilizando estas modificações, é a análise em subespaço independente, no qual as componentes são divididas em grupos ou subespaços de forma tal que componentes em subespaços diferentes são independentes, mas componentes no mesmo subespaço não o são. Isto implica que componentes de um mesmo subespaço tendem a ser ativadas simultaneamente [19, 20]. Quando utilizados para modelar imagens naturais, estes modelos mostraram o aparecimento de propriedades de células complexas, em particular fase e invariância de translação junto com orientação e seletividade em frequência [72].

Neste trabalho, propõe-se uma nova abordagem na forma de modelar um

neurônio artificial, de forma a obter-se um número maior de estruturas de altas ordens. Para isto utilizar-se-á as iterações de um neurônio com seus vizinhos. Esta abordagem apresenta uma boa plausibilidade biológica, devido ao fato de que, no sistema nervoso, a resposta de um neurônio tende a ser correlacionada com a saída dos neurônios da sua vizinhança [18].

Como foi visto, um neurônio atua de forma não supervisionada, ou seja, não existe um sinal externo a ser comparado com a sua saída, gerando um sinal de erro [9]. O que propõe-se aqui, é que as interações locais entre um neurônio e os seus vizinhos podem ser descritas como uma divergência local, dada como a diferença entre a resposta de um neurônio e a resposta da sua vizinhança.

Para a geração de um modelo numérico iterativo computacional, obtém-se uma função, chamada função de auto-referência, a qual deverá ser maximizada (ou minimizada) [73]. Esta função é aplicada sobre a divergência local

No modelo clássico de ICA as componentes supostamente não apresentam nenhum ordenamento ou qualquer outro tipo de relação. Entretanto, na prática, é possível determinar certa relação entre as componentes "independentes" através de alguns critérios tais como não-gaussianidade ou contribuições à variância [74]. Estas relações de ordem são úteis em algumas aplicações, mas em geral, não são muito informativas. A falta de uma ordem inerente nas componentes independentes é devido ao fato da suposição de completa independência estatística. Baseando-se na divergência local entre um neurônio e a sua vizinhança, mostra-se, neste trabalho, que é possível definir uma ordem topográfica (organização espacial) para os filtros de ICA.

Sabe-se que neurônios ativos ao mesmo tempo são correlacionados em energia (correlações em energias é a forma mais simple de dependência porque pode ser interpretada como ativações simultâneas dos neurônios). Além do mais, uma importante motivação para este tipo de dependência pode ser encontrada em extração de características em imagens. Simoncelli e Schwartz [75] mostraram que a dependência predominante entre as saídas dos filtros tipo *wavelets* é a grande correlação em energia, e utilizaram essa propriedade para aplicações em eliminação de ruídos.

O que propõe-se aqui é que esta dependência pode ser estendida às estruturas de altas ordens dos sinais neuronais. Dependência de energia determina apenas dependência de baixas ordens, mas grandes características podem ser extraídas ao utilizarem-se as estatísticas de altas ordens. Desta forma, utiliza-se funções não-lineares que possibilitem capturar as características de altas ordens.

Com a utilização destas dependências, obtém-se uma ordem topográfica de forma a ordenar os vetores bases (\mathbf{w} , no modelo básico) de tal maneira que componentes que são próximas umas das outras, na representação topográfica, tendem a serem ativas simultaneamente. Isto pode ser considerado como uma generalização dos subespaços independentes, onde cada vizinhança corresponde a um subespaço.

5.2 Modelo Proposto

O modelo da rede neuronal, aqui proposto, pode ser visto na Figura 5.1.

A primeira camada do modelo recebe os estímulos previamente branqueados, o que facilitará a operação de busca do algoritmo [21]. A segunda camada, contém os filtros \mathbf{w} , que faz um processamento linear sobre os estímulos, fornecendo os sinais filtrados, u_k . Na terceira camada, aplica-se uma não-linearidade, apropriadamente escolhida, sobre os sinais filtrados objetivando extrair suas características de altas ordens. Cada saída desta camada, y_k interagirá com as saídas vizinhas com o intuito de gerar as divergências locais. Esta estrutura de modelamento neuronal apresenta uma ideia interessante ao mostrar que cada neurônio, na rede, buscará características diferentes das entradas, preservando a média, como se verá no desenvolvimento do algoritmo.

Antes do branqueamento têm-se os sinais $\mathbf{z} = [z_1, z_2, \dots, z_n]$, que são supostos como uma combinação linear dos estímulos $\mathbf{s} = [s_1, s_2, \dots, s_n]$, como no modelo básico de ICA:

$$\mathbf{z} = \mathbf{A}\mathbf{s}, \quad (5.1)$$

na qual \mathbf{A} é uma matriz desconhecida cujas colunas serão chamadas de funções

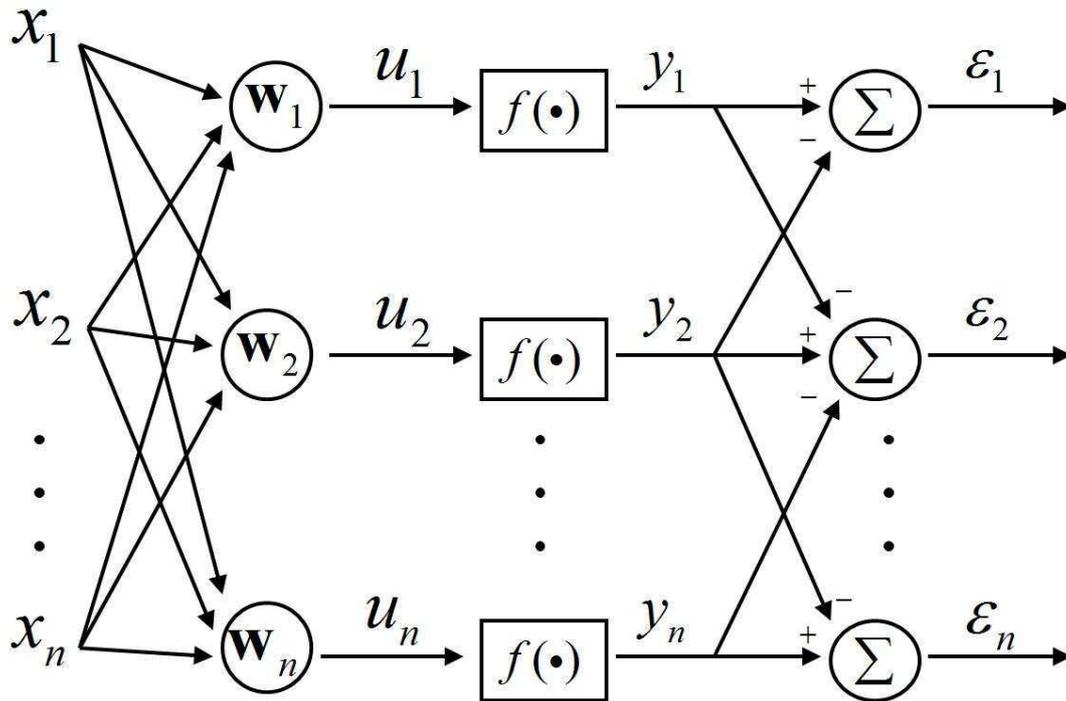


Figura 5.1: Modelo proposto: Primeira camada: Vetor de entrada. Segunda camada: Vetor dos pesos sinápticos. Terceira camada: funções não lineares. Quarta camada: interação entre a saída de um neurônio com as saídas dos seus vizinhos (divergência local).

base. Em seguida, faz-se o branqueamento destes sinais.

$$\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s}, \tag{5.2}$$

na qual a matriz de branqueamento \mathbf{V} pode ser computada como, por exemplo, $\mathbf{V} = (E[\mathbf{z}\mathbf{z}^T])^{-1/2}$. A raiz quadrada inversa é aqui definida pela decomposição em autovalores de $E[\mathbf{z}\mathbf{z}^T] = \mathbf{E}\mathbf{D}\mathbf{E}^T$ como $(E[\mathbf{z}\mathbf{z}^T])^{-1/2} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$. Alternativamente, pode-se fazer o branqueamento via PCA, $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T$, o que também permite fazer uma redução na dimensão dos dados. Estes sinais, \mathbf{x} , serão os sinais de entradas da rede proposta.

Cada sinal de entrada, x_j , conectado à entrada da sinapse j do neurônio k será multiplicado pelo peso sináptico w_{kj} e somados para produzir um sinal de saída, u_k , que será uma estimação de s_k ,

$$u_k = \sum_{j=1}^n w_{kj} x_j = \mathbf{w}_k^T \mathbf{x}. \quad (5.3)$$

Agora, aplica-se uma função não linear sobre o sinal u_k , obtendo-se o sinal y_k .

$$y_k = f(u_k), \quad (5.4)$$

na qual f é escolhida, como anteriormente citado, de forma a possibilitar a extração máxima de informação dos sinais de entrada [76].

Este sinal, y_k , interagirá com as saídas dos neurônios vizinhos produzindo, assim, a divergência local.

Para determinar-se a vizinhança de um neurônio define-se a topografia mediante o emprego de uma função, no plano bidimensional, chamada função de vizinhança, $h(i, j)$, a qual especifica a proximidade entre a i -ésima e a j -ésima componentes [28]. Baseando-se em Mapas de Kohonen [77], define-se, então, tal função como uma função monotonicamente decrescente de alguma medida de distância, o que significa, entre outras coisas, que ela é simétrica, $h(i, j) = h(j, i)$, e tem diagonal constante: $h(i, i) = \text{constante}$ para todo i . Neste trabalho, toma-se como vizinhos de um neurônio k , em um plano bidimensional, aqueles que estão imediatamente ao lado deste neurônio, como pode ser visto na Figura 5.2.

A ação da vizinhança do neurônio k será dada por

$$v_k = \sum_{i \in N} y_{ik} = \sum_{i \in N} f(u_{ik}), \quad (5.5)$$

na qual N é uma família de índices que descreve a vizinhança, ou seja, representa os índices dos neurônios vizinhos ao neurônio k . A divergência local será definida como

$$\varepsilon_k = y_k - v_k. \quad (5.6)$$

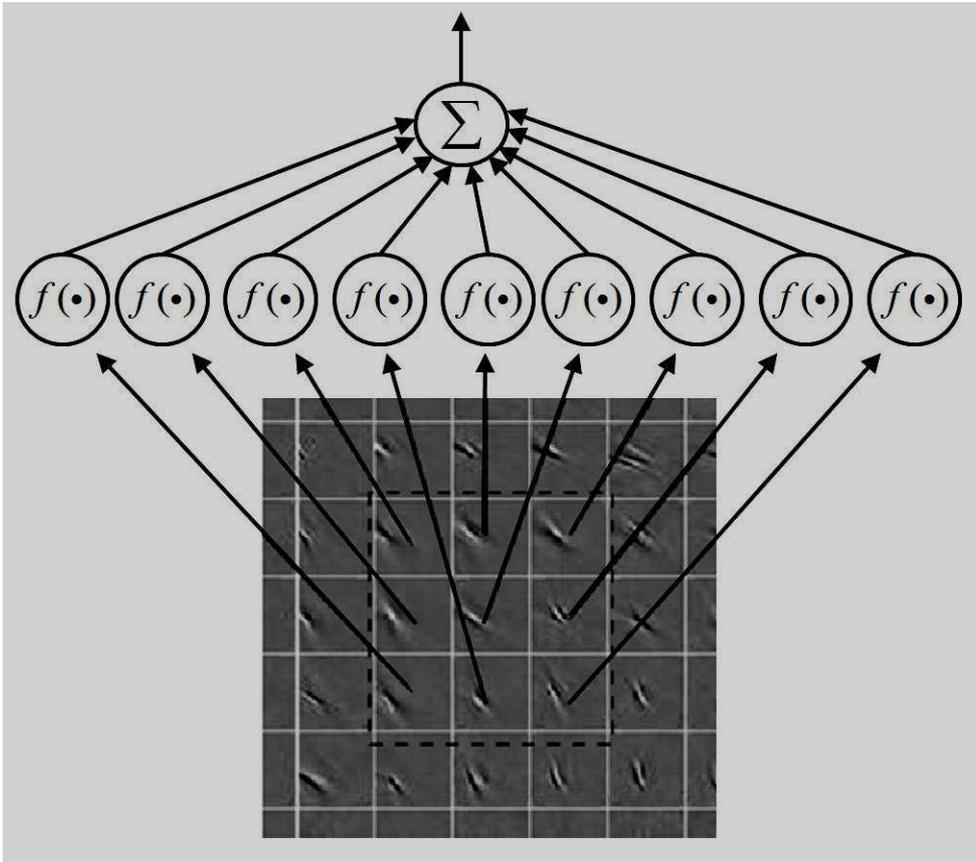


Figura 5.2: Vizinhança de um neurônio k : O neurônio k é supostamente o neurônio central. Seus vizinhos são aqueles que estão ao seu lado, dentro do quadrado tracejado.

Agora, baseado no algoritmo LMS [50, 78] define-se a seguinte função de auto-referência [79],

$$J_k = E[\varepsilon_k^2] = E[(y_k - v_k)^2], \quad (5.7)$$

a qual quer-se minimizar para ajustar os pesos. Para isto, faz-se a expansão da equação 5.7,

$$\begin{aligned}
 J_k &= E[y_k^2 - 2y_kv_k + v_k^2] \\
 &= E\left[\left(f(\mathbf{w}_k^T \mathbf{x})\right)^2 - 2f(\mathbf{w}_k^T \mathbf{x}) \sum_{i \in N} f(\mathbf{w}_i^T \mathbf{x}) + \left(\sum_{i \in N} f(\mathbf{w}_i^T \mathbf{x})\right)^2\right], \quad (5.8)
 \end{aligned}$$

determina-se a derivada parcial, $\frac{\partial J_k}{\partial \mathbf{w}}$, e iguala esta derivada a zero:

$$\begin{aligned}
 \frac{\partial J_k}{\partial \mathbf{w}} &= 2E[\mathbf{x}f(\mathbf{w}_k^T \mathbf{x})f'(\mathbf{w}_k^T \mathbf{x}) - \mathbf{x}f'(\mathbf{w}_k^T \mathbf{x}) \sum_{i \in N} f(\mathbf{w}_i^T \mathbf{x})] \\
 &= 2E[\mathbf{x}y_k f'(u_k) - \mathbf{x}v_k f'(u_k)] = 0 \quad (5.9)
 \end{aligned}$$

Utilizando multiplicadores de Lagrange, soluciona-se a equação 5.9 sob a restrição $\|\mathbf{w}_k\|^2 = 1$:

$$\begin{aligned}
 \beta \mathbf{w}_k &= \frac{\partial J_k}{\partial \mathbf{w}} \\
 \|\mathbf{w}_k\|^2 - 1 &= 0 \quad (5.10)
 \end{aligned}$$

Obtendo-se, desta forma, o seguinte algoritmo de dois passos:

$$\begin{aligned}
 \mathbf{w}_k &= E[\mathbf{x}y_k f'(u_k) - \mathbf{x}v_k f'(u_k)] \\
 \mathbf{w}_k &= \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}, \quad (5.11)
 \end{aligned}$$

que pode ser reescrita como

$$\begin{aligned}
 \mathbf{w}_k &= E[\mathbf{x}\varepsilon_k f'(u_k)] \\
 \mathbf{w}_k &= \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}, \quad (5.12)
 \end{aligned}$$

Este procedimento é realizado paralelamente para ajustar todos os pesos para

cada neurônio na rede, no qual adiciona-se uma etapa de ortogonalização [17]. Desta forma, obtém-se uma matriz \mathbf{W} que é uma estimativa da matriz inversa de \mathbf{A} .

Após a aprendizagem, a matriz \mathbf{A} pode ser computada invertendo-se o processo de branqueamento,

$$\mathbf{A} = (\mathbf{W}\mathbf{V})^{-1} = \mathbf{V}^{-1}\mathbf{W}^T. \quad (5.13)$$

5.3 Validação

Para validar nosso modelo, aplicou-se o algoritmo dado pela equação 5.11 para codificar imagens naturais, da mesma forma como feito em [7, 20, 28, 79].

Os experimentos foram definidos da seguinte maneira:

1) Foram escolhidas as seguintes não linearidades [80], para serem aplicadas sobre o sinal u_k ,

$$f_1(u) = u \cdot \text{atan}(0, 5u) - \text{Ln}(1 + 0, 25u^2) \quad (5.14)$$

$$f_2(u) = \text{Ln}[\text{cosh}(0, 25u)], \quad (5.15)$$

cujos gráficos podem ser vistos nas Figuras 5.3 e 5.4

A ideia estimuladora da escolha destas funções foi a de utilizar-se um resultado obtido por Laughlin [76], o qual afirma que ao aplicar uma função sobre um sinal, consegue-se a transmissão máxima da informação contida no sinal quando a parte deste, com maior probabilidade, está alinhada com a parte de maior inclinação da função. Veja-se Figura 5.5.

As funções escolhidas têm funções derivadas primeira crescentes monótonas ("sigmoidais") [81] dadas por

$$f_1'(u) = \text{atan}(0, 5u) \quad (5.16)$$

$$f_2'(u) = \frac{1}{4} \tanh(0, 25u), \quad (5.17)$$

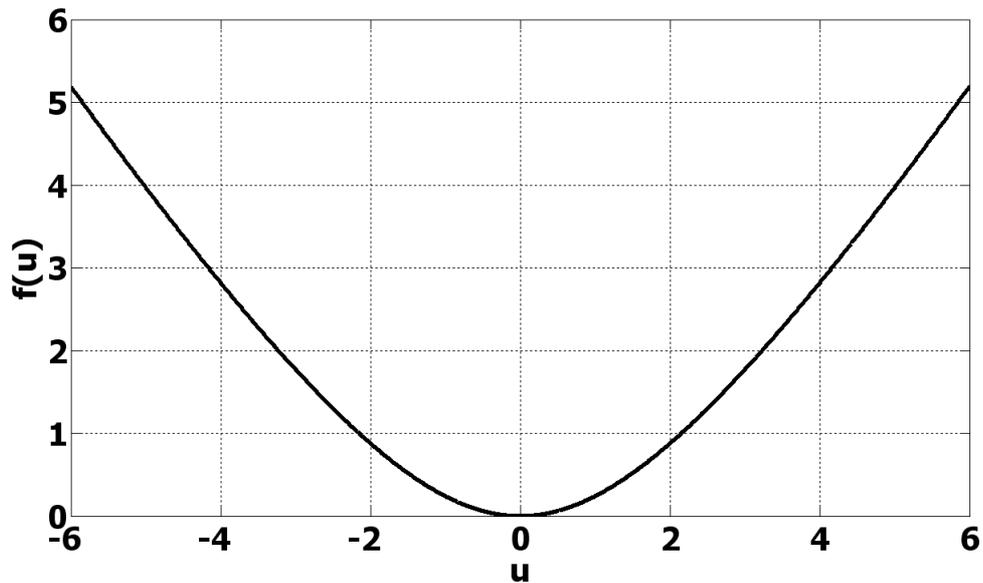


Figura 5.3: Função $f(u) = u \cdot \text{atan}(0,5u) - \text{Ln}(1 + 0,25u^2)$.

as quais são utilizadas no algoritmo dado pela equação (5.12). Os gráficos das funções derivadas podem ser visto nas Figuras 5.6 e 5.7, respectivamente.

Além do mais, expandindo-se em séries de Taylor as funções $f_1(u)$ e $f_2(u)$, obtém-se:

$$\begin{aligned} f_1(u) &\approx a_1u^{18} + a_2u^{16} + a_3u^{14} + a_4u^{12} + a_5u^{10} + a_6u^8 + a_7u^6 + a_8u^4 + a_9u^2 \\ f_2(u) &\approx b_1u^{18} + b_2u^{16} + b_3u^{14} + b_5u^{12} + b_6u^{10} + b_7u^8 + b_8u^6 + b_8u^4 + b_9u^2. \end{aligned} \tag{5.18}$$

Pode-se ver que as funções derivadas são combinações lineares de u^2 . Desta forma, estas funções têm um único mínimo global [82], garantindo a convergência do algoritmo.

2) Escolheu-se, aleatoriamente amostras de 12 x 12 pixels de um conjunto de 13 imagens naturais e cada amostra foi colocada como um vetor coluna de dimensão 144 x 1. Utilizou-se 100.000 versões branqueadas dessas amostras como sinal de entrada [83]. Ou seja $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_{100000}]$ é uma matriz de ordem 144×10^5

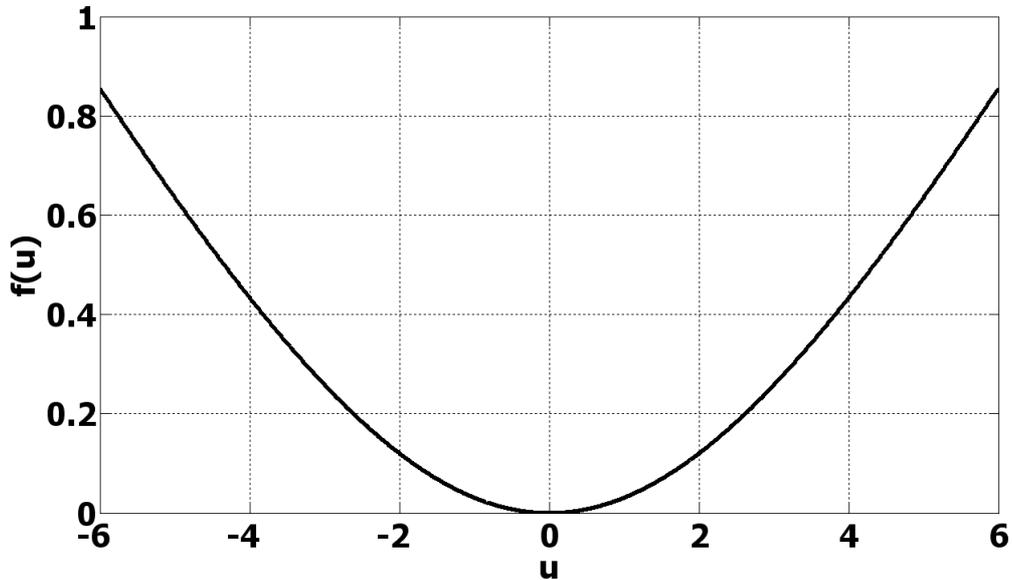


Figura 5.4: Função $f(u) = \text{Ln}[\cosh(0,25u^2)]$.

3) Para cada função escolhida, executou-se o algoritmo 2.000 vezes, obtendo-se estimações para a matriz \mathbf{A} .

Na Figura 5.8, tem-se funções bases estimadas quando utilizou-se a função $f(u_k) = u_k \cdot \text{atan}(0,5u_k) - \text{Ln}(1 + 0,25u_k^2)$, e, na Figura 5.9, tem-se funções bases estimadas quando utilizou-se a função $f(u_k) = \text{Ln}(\cosh(0,25u_k))$. A organização topográfica pode ser vista em cada uma destas figuras. Observe-se a pequena variação nas características de um pequeno quadrado (função base) com os quadrados vizinhos. Dois vetores bases vizinhos tendem a serem de mesmas orientação e frequência e suas respectivas localizações espaciais também estão próximas. Mas, em contraste, suas fases são muito diferentes. Isto significa que a vizinhança de cada um dos vetores bases (células simples) agem como uma célula complexa. Os sinais y_{ik} , que são somados na Equação 5.5, podem ser considerados como saídas de uma célula complexa, após uma transformação não linear [72].

A emergência de componentes esparsas, s , foi observada através do histograma das componentes estimadas, s . Na figura 5.10 está representado o histograma de uma das componentes.

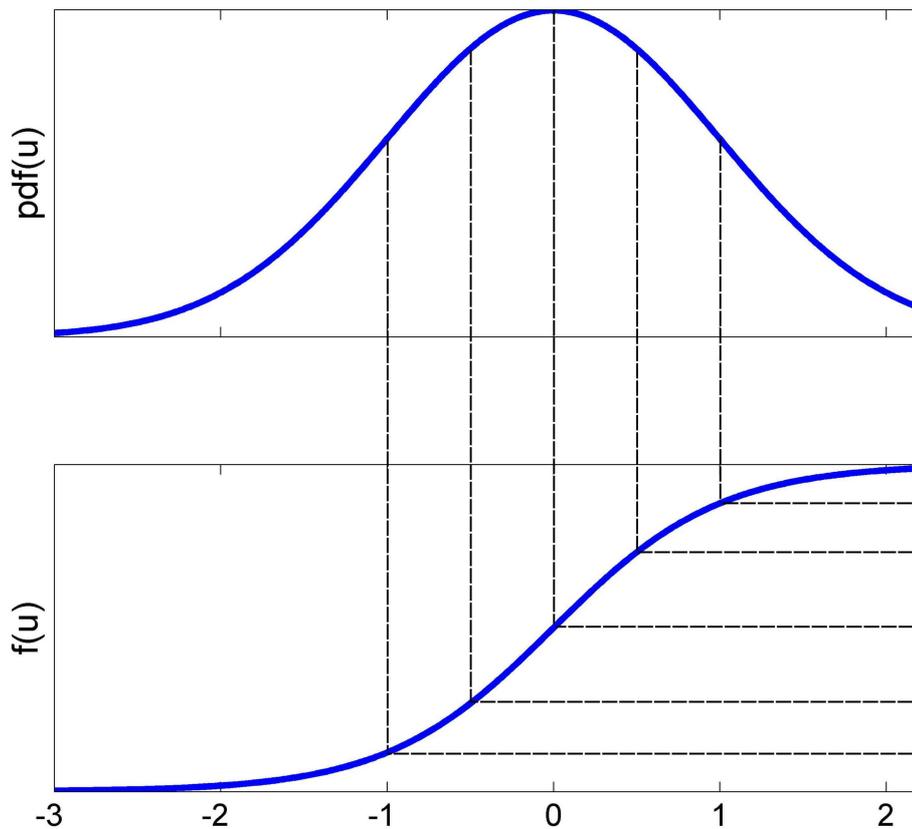


Figura 5.5: Situação onde se consegue maior transmissão da informação de um sinal u através de uma função $f(u)$.

Para estudar a influência do tamanho da vizinhança, experiências foram simuladas com outro tipo de vizinhança. Neste tipo consideramos como vizinhos os oito mais próximos, como podem ser vistos na Figura 5.2, mais aqueles que estão acima, abaixo e aos lados destes. Para este tipo de vizinhança, veja-se a Figura 5.11. As estimações das funções bases não foram muito boas, como pode ser visto na Figura 5.12, onde $f(u_k) = u_k \cdot \text{atan}(0.5u_k) - \text{Ln}(1 + 0.25u_k^2)$ foi a não-linearidade utilizada. Resultado similar foi observado quando utilizou-se a outra não-linearidade. Este resultado supostamente deve-se a que, neste tipo de vizinhança, têm-se uma quantidade maior de informações a serem processadas do que a capacidade do neurônio.

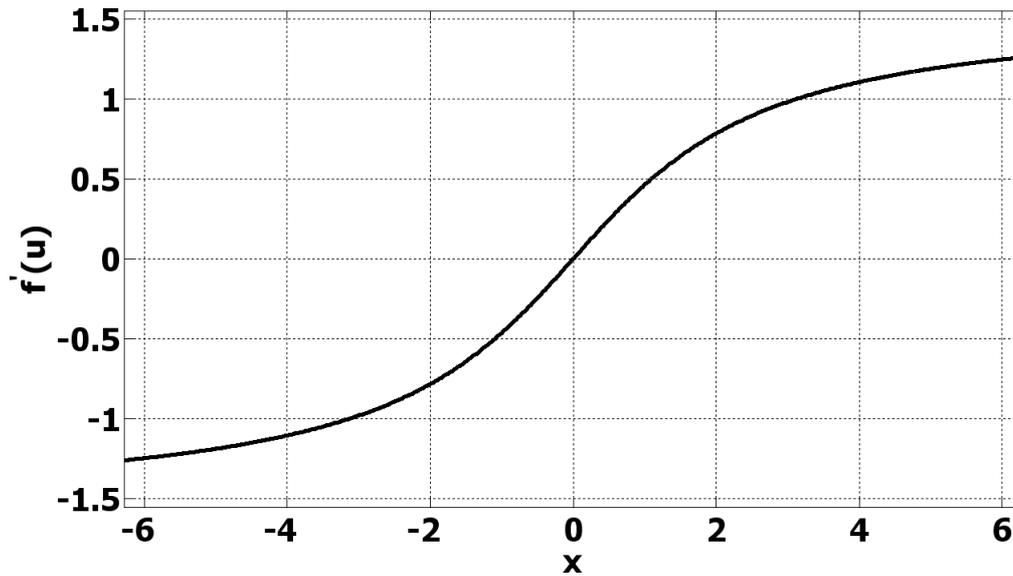


Figura 5.6: Derivada primeira da função $u \cdot \text{atan}(0,5u) - \text{Ln}(1 + 0,25u^2)$.

Para mostrar a influências das características de altas ordens dos estímulos, simulou-se experimentos utilizando-se a função $f(u_k) = u_k^2$ como a não linearidade a ser aplicada sobre o sinal u_k . As estimações das funções bases podem ser vista na Figura 5.13. Nelas, não tem-se uma organização topográfica bem definida, além do mais as funções bases não apresentam localização espacial, como pode ser visto pelo formato mais alongado das mesmas. Ou seja, a perda das informações de altas ordens, não apreendidas pela não linearidade, impossibilitou a emergência de filtros localizados no espaço.

É sabido que os campos receptivos de células do córtex visual primário (V1) dos mamíferos atuam como filtros (*wavelets*) de Gabor [20,84], isto é, a ativação de um neurônio pode ser simulada pelo método de correlação reversa entre as cenas visuais e o campo receptivo dos neurônios.. Desta forma, como afirmado por Marcelja [85], a informação sobre o mundo visual é representada como um conjunto de níveis de excitações de células diferentes. Neste trabalho, mostrou-se que a estimação da matriz \mathbf{A} é um banco de filtros de Gabor seletivos em frequência e em orientação, como pode ser visto nas Figuras 5.8 e 5.9. Cada pequeno quadrado, nas Figuras mencionadas, representa o campo receptivo de um

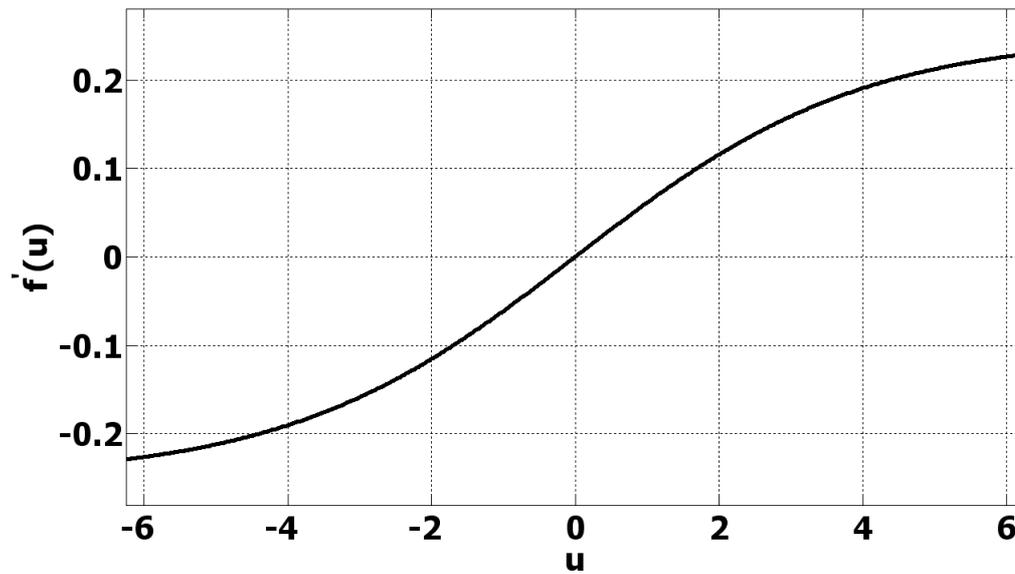


Figura 5.7: Derivada primeira da função $\text{Ln}[\cosh(0, 25u^2)]$.

único neurônio. Por inspeção visual, pode-se ver que a orientação e localização de cada um dos campos receptivos variam de forma suave em função da posição na malha topográfica.

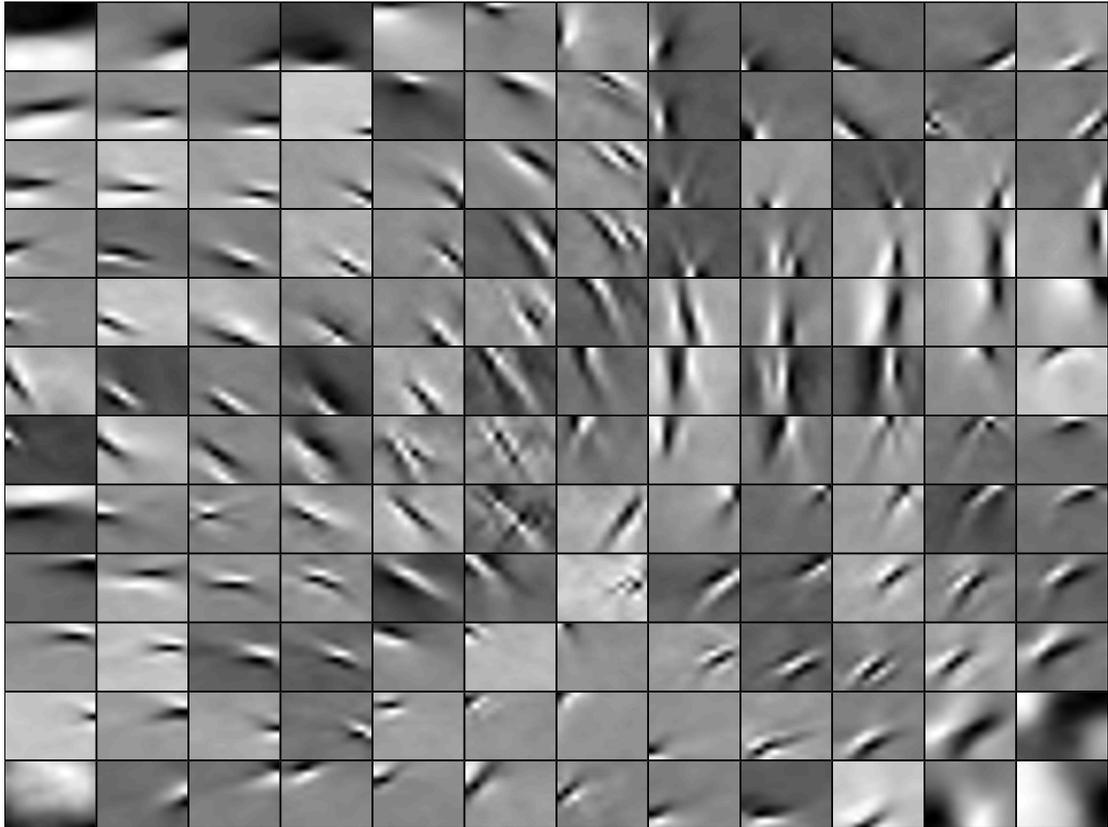


Figura 5.8: Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k \cdot \text{atan}(0.5u_k) - \text{Ln}(1 + 0.25u_k^2)$ foi a não-linearidade utilizada.

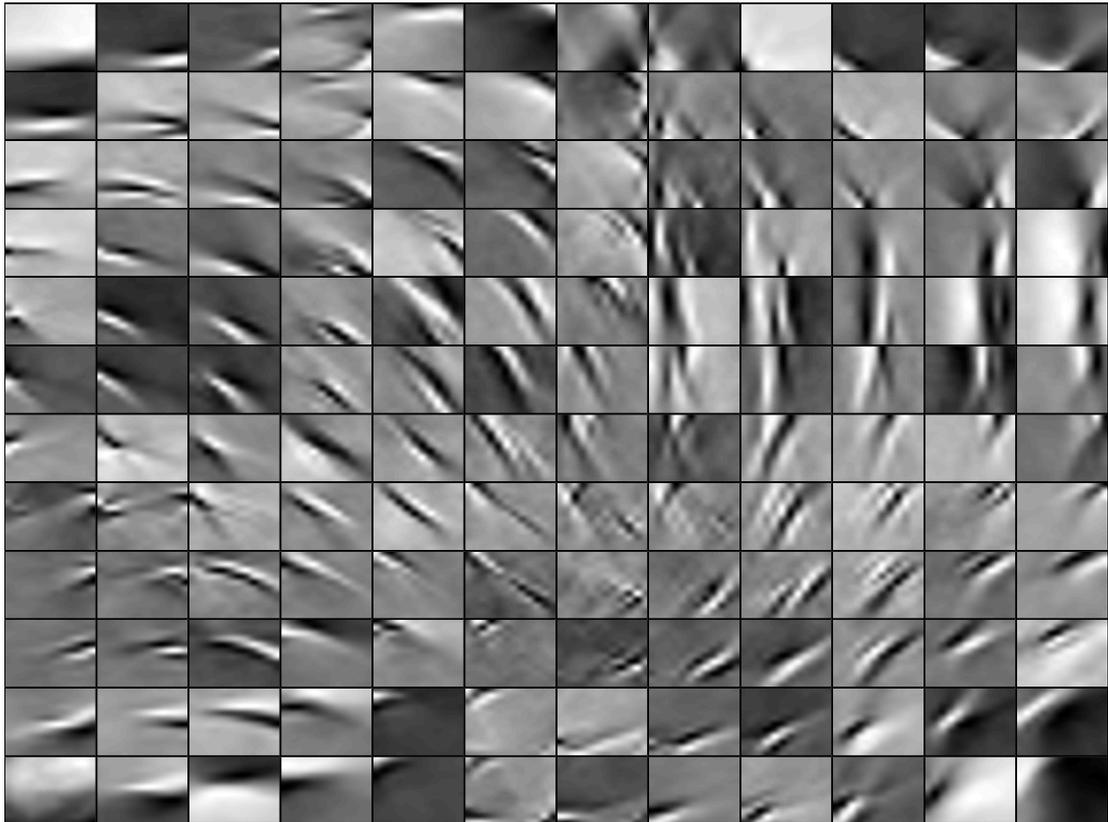


Figura 5.9: Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = Ln(\cosh(0.25u_k))$ foi a não-linearidade utilizada.

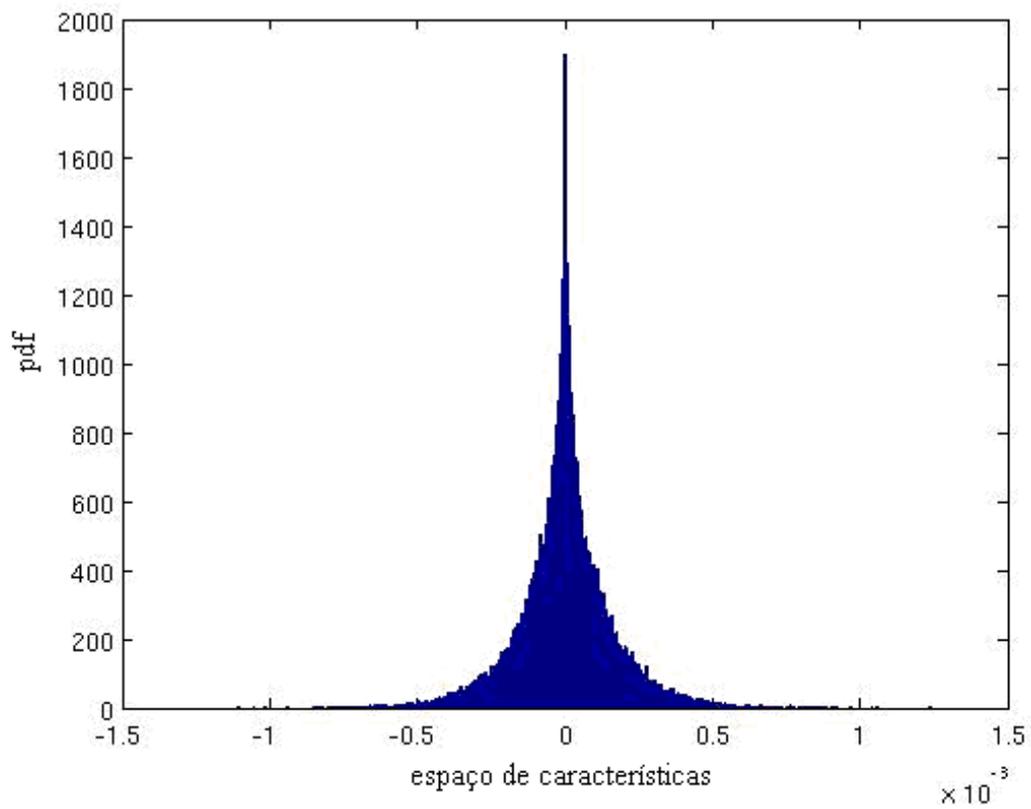


Figura 5.10: Histograma da estimação de uma imagem codificada. A componente obtida tem distribuição esparso. Ou seja, a componente está esporadicamente ativa.

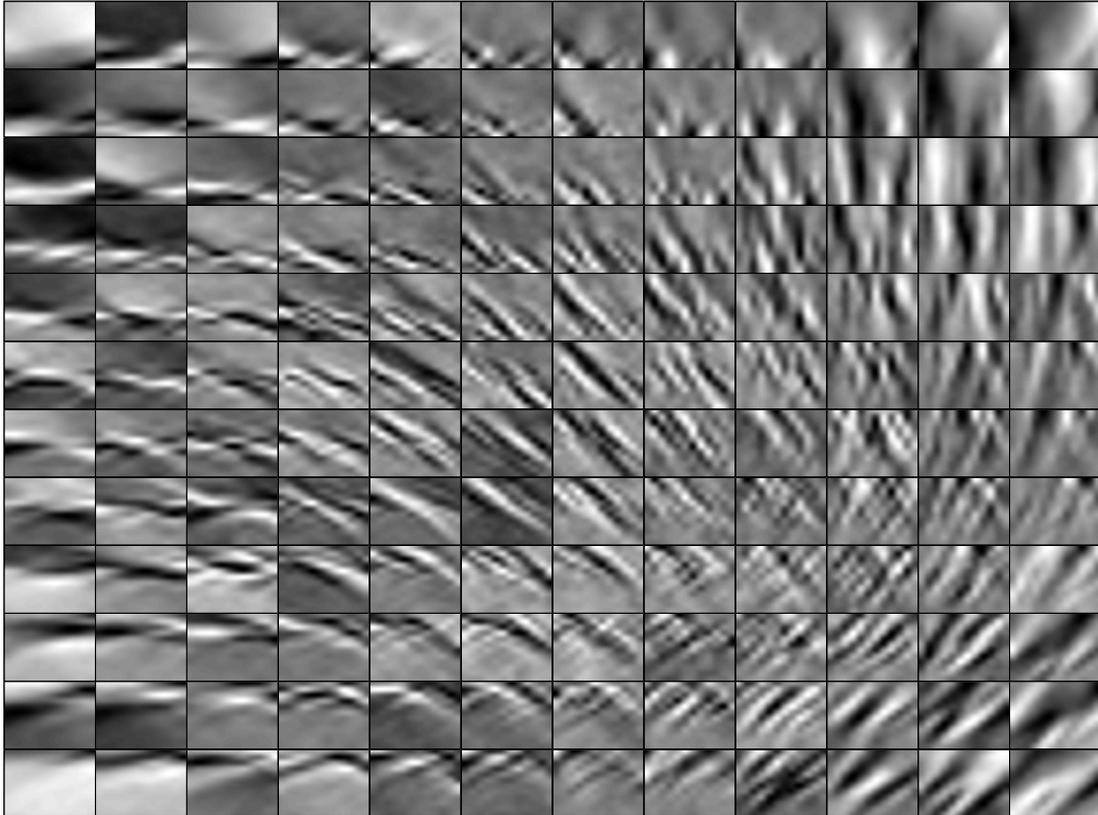


Figura 5.12: Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k \cdot \text{atan}(0.5u_k) - \text{Ln}(1 + 0.25u_k^2)$ foi a não-linearidade utilizada com vizinhança maior.

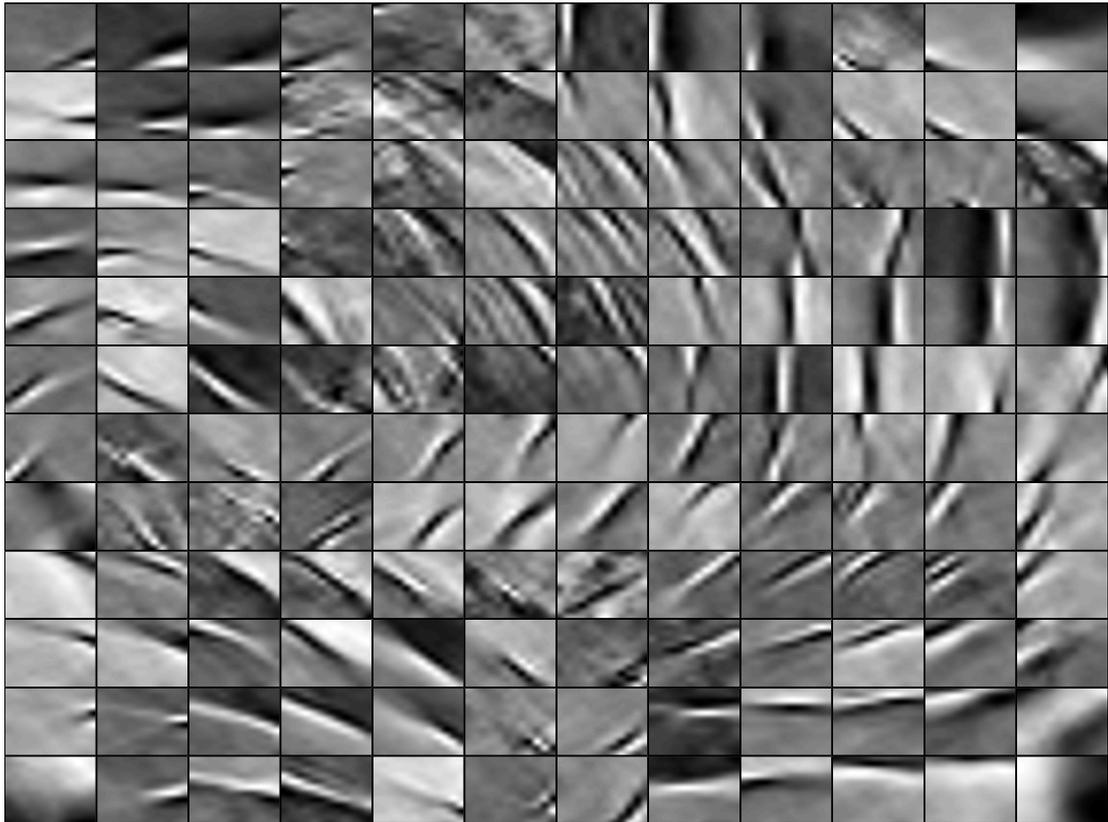


Figura 5.13: Funções bases estimadas (colunas da matriz \mathbf{A}) a partir da utilização de imagens naturais quando $f(u_k) = u_k^2$ foi a não-linearidade utilizada.

Capítulo 6

Conclusão, Publicações e Trabalhos Futuros

6.1 Conclusão

Neste trabalho, apresentou-se um modelo para uma rede neuronal com aprendizagem não supervisionada baseado no princípio de divergência local. Neste modelo (Figura 5.1), os sinais de entrada, após serem processados linearmente, são filtrados por uma função não linear capaz de extrair suas características de altas ordens. Em seguida, cada neurônio interagirá com as informações vindas dos neurônios vizinhos para gerar seu sinal de resposta. Com a utilização da Equação 5.12, desenvolveu-se um algoritmo de ponto fixo, de baixa custo computacional, para atualização dos filtros \mathbf{w} na rede neuronal. Este modelo foi aplicado para modelamento do córtex visual primário (V1), o qual mostrou-se satisfatório ao gerar filtros de Gabor (Figuras 5.8 e 5.9) como descrito na bibliografia consultada.

Foi definido um tipo específico de vizinhança para um neurônio (Figura 5.2), que mostrou-se eficaz no desenvolvimento do modelo.

6.2 Lista de Publicações

6.2.1 Periódicos

[1] Ewaldo Santana; Allan kardec Barros e Raimundo C. S. Freire. "On the Time Constant Under General Error Criteria". *IEEE Signal Processing Letters*, vol. 14, n. 8, pp. 533-536, 2006.

[2] Ewaldo Santana; André B. Cavalcante; Márcio de O. Santos; Allan kardec Barros e Raimundo C. S. Freire. "On-Line Algorithm for Extraction of Specific Signals With Temporal Structure". *Lectures Notes On Computer Sciences*. vol. 4985, pp. 21-29, 2007.

6.2.2 Conferências

[1] Ewaldo Santana; José Carlos Príncipe; Allan kardec Barros e Raimundo C. S. Freire. "An Adaptive Algorithm Based On The Sigmoidal Function. Proc. of the Ninth Brazilian Symposium On Neural Networks-SBRN'06. 2006.

[2] Ewaldo Santana; Allan kardec Barros; Y. Yasuda; F. Grangeiro e Raimundo C. S. Freire. "Analysis of the Time Constant For The Sigmoidal Algorithm Applied to Biomedical Signals". *International Workshop on Medical Measurement and Applications*, pp. 115-119. 2006.

[3] Ewaldo Santana; Éder Santana; Allan kardec Barros e Raimundo C. S. Freire. "A New Fixed-Point Algorithm For Non-Linear Hebbian Learning". *21º Congresso Brasileiro de Engenharia Biomédica*, pp. 1011-1014. 2008.

[4] Ewaldo Santana; José Carlos Príncipe; Allan kardec Barros; Éder Santana e Raimundo C. S. Freire. "Biologically Inspired Algorithm Based On Error Minimization". *8th Brain Inspired Cognitive Systems*. 2008.

[5] Ewaldo Santana; Éder Santana; Allan kardec Barros e Raimundo C. S. Freire. "Reproducing Kernel Hilbert Space Method for Blind Source Extraction". VIII Semetro. 2009.

6.3 Trabalhos Futuros

A seguir, temos algumas sugestões para trabalhos futuros no cenário da organização topográfica.

6.3.1 Tamanho da vizinhança

Foi mostrado, neste trabalho, que ao tentar-se modificar o tamanho da vizinhança, o modelo perdeu eficiência, sugerindo que estudos devam ser direcionados no sentido da determinação do tamanho da vizinhança que obtenha respostas coerentes, bem como na definição de funções de vizinhanças.

6.3.2 Separação Cega de Fontes

Como o modelo apresentou organização topográfica dos filtros, \mathbf{w} , pode-se aplicar tal modelo em separação cega de fontes baseado em ICA, onde deve-se obter as componentes independentes de forma organizada, ou, seja, componentes com características relacionadas deverão ser extraídas próximas umas das outras.

6.3.3 Extração de Características Temporal

Pode-se estender a ideia da organização topográfica para uma organização temporal, onde sinais com valores correlacionados, em tempos diferentes, possam ser separadamente extraídos [86].

Capítulo 7

Anexo

7.1 Arquivos Matlab

Disponibiliza-se, aqui, os programa código, desenvolvido em Matlab, para a estimação das funções bases.

Programa 01: Amostra as imagens naturais, determina o tipo de vizinhança.

```
% Amostragem das Imagens Naturais
```

```
clear, clc
rdim = 12;
p.xdim = rdim;
p.ydim = rdim;
p.maptype = 'standard'
; p.neighborhood = 2;
p.components =  $rdim^2$ ;
p.pca = 'n',
p.iter = 20000;
p.visual = 'y';
p.save = 'n';
cd D : \EWALDO\work\EWALDO\bipnet\data';
```

```
[X,whiteningMatrix,dewhiteningMatrix] = data(10e4,rdim,rdim.^2);
[A,W,Y] = newbipnet(X,p,whiteningMatrix,dewhiteningMatrix);
```

Programa 02: Estima as funções bases.

```
function [A,W,Y]=newbipnet( X,p,whiteningMatrix,dewhiteningMatrix )

disp('Parameters: ')
p

[nn,M]=size(X);

if p.pca == 'y'
if nn>M,
X=X';
[nn, M] = size(X);
end
X=double(X) - double( mean(X')'*ones([1,M]) );

fprintf('Calculatingcovariance...\n');
covarianceMatrix = X*X'/size(X,2);
[E, D] = eig(covarianceMatrix);
[dummy, order] = sort(diag(-D));
E = E(:, order(1 : p.components));
d = diag(D);
d = real(d.^(-0.5));
D = diag(d(order(1 : p.components))) ;
X = D * E' * X;

whiteningMatrix = D * E';
```

```

dewhiteningMatrix =  $E * D^{(-1)}$ ;
N = size(X, 2);
end
X1=X;

    B = rand( size(X,1), p.components);
B = B * real((B' * B)(-0.5));
[NB] = GenerateNB(p);

    iter = 0;

    fprintf('CalculatingComponents : \n\n');

    c = 0.5; while iter < p.iter,
iter = iter + 1;

    if rem(iter,100)==0, clc;end

    fprintf ('(
    U = B' * X;
Y = U * atan(0.5 * U) - log(0.25 * U.2 + 1);
V = NB * Y;

    B = X * (Y * atan(0.5 * U))' - X * (V * atan(0.5 * U))';

    B = B * real((B' * B)(-0.5));
A = dewhiteningMatrix * B;
if rem(iter, 10) == 0 & p.visual == 'y'
    fprintf ('\n');
    pause(0.1)
end

```

```

    ifrem(iter,100) == 0 & p.save == ' y'
cd matSTD
eval(['save' fname, int2str(iter/100)'AYiter']);
cd..
end

end
W = B' * whiteningMatrix;
Y = W * X; function [NB] = GenerateNB( p )

NB = zeros(p.xdim*p.ydim*[1 1]);

k = p.neighborhood;

ind = 0;
for y=1:p.ydim
for x=1:p.xdim

ind = ind+1;

[xn,yn] = meshgrid( (x-k):(x+k), (y-k):(y+k) );
xn = reshape(xn, [1(length(xn)^2)]);
yn = reshape(yn, [1(length(yn)^2)]);

ifstrcmp(p.matype,'torus')

i = find(yn < 1); yn(i) = yn(i) + p.ydim;
i = find(yn > p.ydim); yn(i) = yn(i) - p.ydim;
i = find(xn < 1); xn(i) = xn(i) + p.xdim;
i = find(xn > p.xdim); xn(i) = xn(i) - p.xdim;

```

```
elseif strcmp(p.mapttype,'standard')

    i = find(yn >= 1 & yn <= p.ydim & xn >= 1 & xn <= p.xdim);
    xn = xn(i);
    yn = yn(i);

    else
    error('Nosuchmaptype!');
end

    NB( ind, (yn-1)*p.xdim + xn )= 1;
for count = 1:p.xdim*p.ydim
NB(count,count) = 0;
end

    end
end
return;
```

7.2 Imagens Naturais

A seguir, tem-se as 13 imagens naturais utilizadas para treinamento da rede desenvolvida nesta tese. Estas imagens podem ser adquiridas no seguinte endereço eletrônico: www.cis.hut.fi/project/ica/data/images/



Figura 7.1: Imagem Natural 01



Figura 7.2: Imagem Natural 02



Figura 7.3: Imagem Natural 03

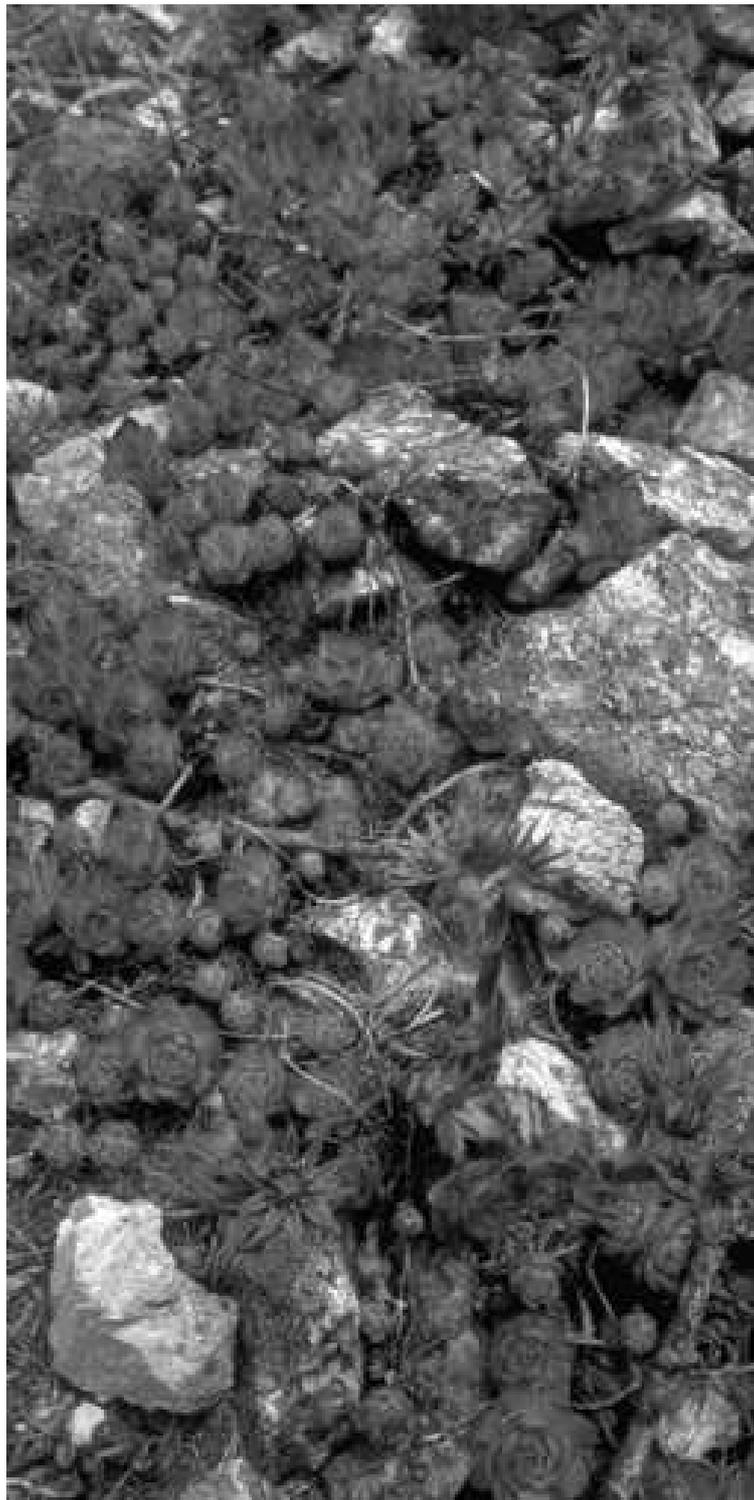


Figura 7.4: Imagem Natural 04



Figura 7.5: Imagem Natural 05

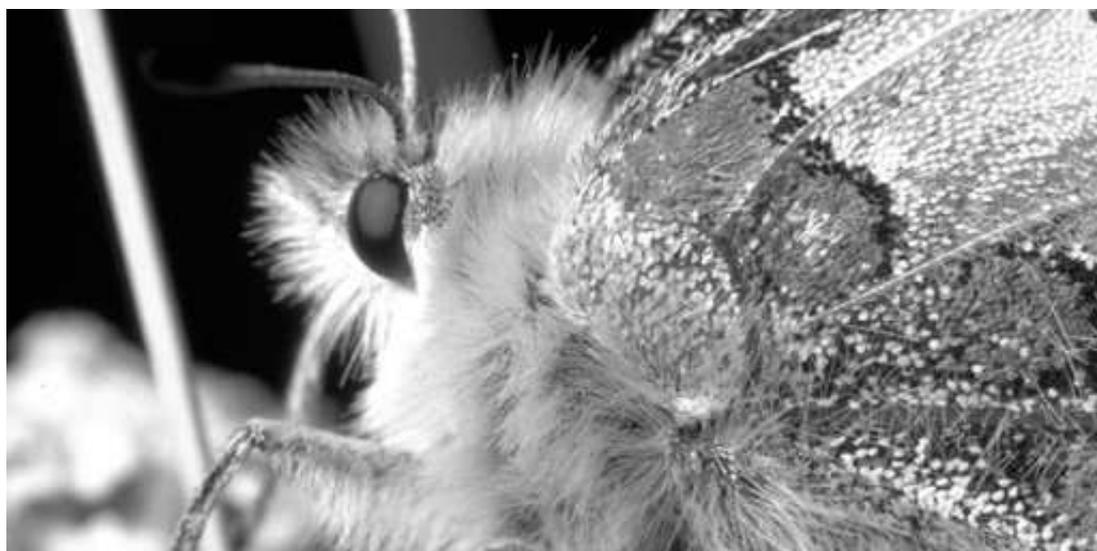


Figura 7.6: Imagem Natural 06



Figura 7.7: Imagem Natural 07



Figura 7.8: Imagem Natural 08



Figura 7.9: Imagem Natural 09



Figura 7.10: Imagem Natural 10



Figura 7.11: Imagem Natural 11



Figura 7.12: Imagem Natural 12



Figura 7.13: Imagem Natural 13

Referências Bibliográficas

- 1 F. Atteneave. Some informational aspects of visual perception. *Psychological review*, 61, 1954.
- 2 H. B. Barlow. Possible principles underlying the transformations of sensory messages. *Sensory Communication*, 1961.
- 3 J. J. Atick. Could information theory provide an ecological theory of sensory processing ? *Network*, 3, 1992.
- 4 H. B. Barlow. Unsupervised learning. *Neural Computation*, 1, 1989.
- 5 J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20, 1980.
- 6 D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Optical Society of America*, 4(12):2379–2394, 1987.
- 7 Anthony J. Bell e terrence J. Sejnowski. The independent components of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- 8 S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Englewood Cliffs, NJ., 1999.
- 9 José C. Principe; Neil R. Euliano and Curt W. Lefebvre. *Neural and Adaptive Systems - Fundamentals Through Simulations*. John Wiley & sons Inc, Englewood Cliffs, NJ., 1999.

- 10 David E. Rumelhart e McClelland. On learning the past tenses of english verbs. *Parallel Distributed Processing*, 1986.
- 11 David G. Stork. Is backpropagation biologically plausible? *International Joint Conference on Neural Networks - IJCNN*, 1989.
- 12 Rafal Bogacz; Malcolm W. Browne and Christophe G.-C. Ifrequency-based error back-propagation in a cortical network. *International Joint Conference on Neural Networks, IJCNN*, 1995.
- 13 D.O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.
- 14 José C. Principe; D. Erdogmuz and Kenneth E. Hild III. Do hebbian synapses estimate entropy? *Proc. Of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002.
- 15 R. Linsker. Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Computation*, 4, 1992.
- 16 E. Oja. Neural networks, principal components and linear neural networks. *Neural Networks*, 5, 1989.
- 17 P. Foldiak. Forming sparse representations by local anti-hebbian learning. *Biological Cybernetics*, 64, 1990.
- 18 K. I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks: Theory and Applications*. John Willey and sons, New York, 1996.
- 19 H. Barlow. Redundancy reduction revisited. *Network*, 12, 2001.
- 20 B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse coding for natural images. *Nature*, 381, 1996.
- 21 Aapo Hyvarinen; J. Karhunen and E. Oja. *Independent Component Analysis*. John Willey and sons, New York, 2001.

- 22 Aapo Hyvarinen; Patrik O. Hoyer e Mika Inki. Topographic ica as a model of v1 receptive fields. *Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 3, 2000.
- 23 Anthony J. Bell and terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Networks*, 3, 1995.
- 24 A. K. Barros and N. Ohnishi. Wavelet like receptive fields emerges by non-linear minimization of neuron error. *International Journal of Neural Systems*, 13(2):87–91, 2003.
- 25 E. R. Kandel; J. H. Schwartz and T. M. Jessel. *Principles Of Neural Science*, 4th edition. McGraw Hill, New York, 2000.
- 26 E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representations. *Annu. Rev. Neurosci*, 24, 2001.
- 27 W. E. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287, 2000.
- 28 Aapo Hyvarinen; Patrik O. Hoyer and Mika Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001.
- 29 S. Osindero; M. Welling and G. E. Hinton. Topographic products models applied to natural scene statistics. *Neural Computation*, 18(2):381–414, 2006.
- 30 B. O. Olshausen and D. J. Field. Natural images statistics and efficient coding. *Network: Computation in Neural Science*, 7, 1996.
- 31 W.J. Freeman. Tutorial on neurobiology: From single neurons to brain chaos. *International Journal of Bifurcation and Chaos*, 2(3):451–482, 1992.
- 32 P. Dayan. and L. F. Abbott. *Theoretical Neuroscience - Computational and Mathematical Modeling of Neural Systems*. MIT Press, New York., 2001.

- 33 D. H. Hubel e T. N. Viesel. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiol.*, 148, 1959.
- 34 D. H. Hubel e T. N. Viesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiol.*, 195, 1968.
- 35 S. Mallat. *Wavelet Tour of Signal Processing*. Academic Press, New York, 1999.
- 36 C. H. Anderson e D. C. van Essen B.A. Olshausen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience*, 13, 1997.
- 37 M. V. Srinivasan; S. B. Laughlin e A. Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B*, 216, 1982.
- 38 D. H. Hubel e T. N. Viesel. Receptive fields, binocular interactions and functional architecture in teh cat's visual cortex. *Journal of Physiol.*, 160, 1962.
- 39 David H. Hubel. *Eye, Brain, and Vision*. Scientific American Libray, New York., 1988.
- 40 G. C. DeAngelis; I. Ohzawa and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex i. general characteristics and postnatal development. *J. Neurophysiology*, 69(4):1091–1117, 1993.
- 41 S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 7, 1989.
- 42 G. G. Blasdel. Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12(8):3139–3161, 1992.
- 43 R. B. H. Tootell; M. S. Silverman; S. L. Hamilton; E. Switkes and R. L de Valois. Functional anatomy of macaque striate cortex versus spatial frequency. *Journal of Neuroscience*, 8, 1988.

- 44 von der Malsburg. Self-organization of orientation-sensitive cell in the striate cortex. *Biological Cybernetics*, 14, 1973.
- 45 T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):56–69, 1982.
- 46 R. Durbin and G. Mitchison. A dimension reduction framework for understanding cortical maps. *Nature*, 343, 1990.
- 47 H. B. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1, 1972.
- 48 D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6, 1994.
- 49 Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanents in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 1943.
- 50 B. Widrow and S. D. Stearns. *Adaptive signal Processing*. Englewood Cliffs, New Jersey, 1985.
- 51 R. Baddeley; L. F. Abbott; M. C. Booth; F. Sengpiel; T. Freeman; E. A. Wakeman and E. T. Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. London B*, 264, 1998.
- 52 M. Deweese; M. Wehr and A. Zador. Binary spiking in auditory cortex. *The Journal of Neuroscience*, 23(21):7940–7949, 2003.
- 53 M. Brecht and B. Sakmann. Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *The Journal of Physiology*, 543, 2002.
- 54 J. Perez Orive; O. Mazor; G. C. Turner; S. Cassenaer; R. I. Wilson and G. Laurent. Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297, 2002.

- 55 C. Jutten and J. Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24, 1991.
- 56 P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36, 1994.
- 57 Athanasios Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill series in electrical engineering, Communications and signal processing, New York, 1991.
- 58 Kevin P. Balanda e H. L. MacGillivray. Kurtosis: A critical review. *The American Statistician.*, 42(2):111–119, 1988.
- 59 N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45, 1995.
- 60 Aapo Hyvarinen and E. Oja. One-unit learning rules for independent component analysis. *NIPS'96*, 1996.
- 61 Aapo Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 64, 1999.
- 62 Aapo Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions On Neural Network*, 10(3):626–634, 1999.
- 63 S. Comani e Giovanna Alleva. Fetal cardiac time intervals estimated on fetal magnetocardiograms: single cycle analysis versus average beat inspection. *Physiological Measurement*, 28, 2007.
- 64 Aapo Hyvarinen. A unifying modelo for blind separation of independent sources. *Signal Processing*, 85(7):1419–1427, 2005.
- 65 Aapo Hyvarinen and J. Hurri. Blind separation of sources that have spatiotemporal dependencies. *Signal Processing*, 84(2):247–254, 2004.

- 66 Aapo Hyvarinen; P. O. Hoyer; J. Hurri and M. Gutmann. Statistical models of images and early vision. *Proceedings of the Int. Symposium on Adaptive Knowledge Representation and Reasoning*, 2005.
- 67 Aapo Hyvärinen; J. Hurri and J. Väyrynen. Bubbles: a unifying framework for low-level statistical properties of natural image sequences. *Journal of the Optical Society of America A*, 20(7):1237–1252, 2003.
- 68 B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37, 1997.
- 69 E. P. Simoncelli; W. T. Freeman; E. H. Adelson e D. J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, 1982.
- 70 M. S. Lewicki e T. J. Sejnowski. Learning overcomplete representation. *Neural Computation*, 12, 2000.
- 71 H. van Hateren e A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Society London, ser. B*, 265, 1998.
- 72 A. Hyvärinen and P.O. Hoyer. Emergence of phase and shift-invariant features by decomposition of natural images into independent features subspaces. *Neural Computation*, 12(7):1705–1720, 2000.
- 73 Gene H. Golub; and Charles F. Van Loan. *Matrix Computations, 3th edition*. The John Hopkins university Press, Baltimore, 1996.
- 74 Aapo Hyvarinen. Survey on independent component analysis. *Neural Computing Series*, 2, 1999.
- 75 Eero P. Simoncelli and Adelia Schwartz. Modeling surround suppression in v1 neurons with a statistically-derived normalization model. *NIPS*, 1999.
- 76 Simon Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*, 36, 2002.

- 77 T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlim, 2001.
- 78 Ewaldo Santana; André B. Cavalcante; Márcio de O. Santos; Allan Kardec Barros e Raimundo C. S. Freire. On-line algorithm for extraction of specific signals with temporal structure. *Lecture Notes On Computer Science*, 4985, 2008.
- 79 Ewaldo Santana; José C. Principe; Allan Kardec Barros; Eder Santana e Raimundo C. S. Freire. Biologically inspired algorithm based on error minimization. *Proc. of the 8th. Brain Spired Cognitive Systems - BICS*, 2008,
- 80 Ewaldo Santana; Allan Kardec Barros; Y. Yasuda; F. Grangeiro e Raimundo C. S. Freire. Analysis of time constant for the sigmoidal algorithm applied to biomedical signals. *IEEE International Workshop on Medical Measurements and Applications*, 2006.
- 81 Ewaldo Santana; José C. Principe; Allan Kardec Barros e Raimundo C. S. Freire. An adaptive algorithm based on the sigmoidal function. *Proc. of the 9th. Brazilian Symposium on Neural Networks - SBRN*, 2006.
- 82 A. Gersho. Some aspects of linear estimation with non-mean square error criteria. *Proc. Asilomar Ckts. and Systems Conf.*
- 83 Ewaldo Santana; Eder Santana; Allan Kardec Barros e Raimundo C. S. Freire. A new fixed-point algorithm for non-linear hebbian learning. *21º Congresso Brasileiro de Engenharia Biomédica*, 2008.
- 84 J. A. Movshon; I. D. Thompson e D. J. Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. of Physiology.*, 283, 1978.
- 85 S. Marcelja. Mathematical description of the response of simple cortical cells. *J. Opt. Soc. Am.*, 70(11):1297–1300, 1980.

- 86 A. Cichocki e S.-I. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley and Sons, New York, 2002.