

1993/10
S 795



DANIEL ADRIÁN STARIOLO

Redes de Neurônios e outros Sistemas Dinâmicos Não-lineares

Tese de Doutorado

Centro Brasileiro de Pesquisas Físicas
Rio de Janeiro - 1993

003.85

A Bea

Agradecimentos

Aos meus pais, Rosa e Homero, pelo seu esforço e apoio ao longo de muitos anos.

A Constantino Tsallis, pela sua orientação profissional e amizade.

A Francisco Tamarit, em parte co-autor desta tese, com o qual compartilhamos muitas horas de trabalho e amizade.

A Evaldo Curado e Aglaé Magalhães pelo seu apoio constante.

Aos meus companheiros no CBPF, Celia, Hélio, Marcelo, Ladário, Welles, Edson, Ana, André, Zé Arnaldo, Isaías e Ricardo com os quais compartilhamos muitos bons momentos.

A Vanda, Almério e Myriam pela sua ajuda e paciência.

A Carla e Eliane, pela sua importante ajuda na parte computacional dos trabalhos desta tese.

Aos meus amigos de tantos anos, Mirian Bracco, Gabriel Fabricius, Jorge Shitu, Guillermo Cecci, Sandra Sautú, Ramón Méndez, Sergio Cannas, Vicky, Marcelo Argüello, Sergio Uriñuela, ao meu irmão Marcelo, e especialmente ao meu amigo Daniel Barci com o qual compartilhamos o gosto pela física desde o princípio. A Guillermo Bozzolo, quem me mostrou o prazer de estudar e fazer física.

Ao LAFEX, por permitir-me usar sua infraestrutura computacional.

À Universidad Nacional de La Plata e ao Centro Brasileiro de Pesquisas Físicas, responsáveis pela minha formação profissional. Ao CNPq pelo apoio financeiro.

Resumo

Nesta tese estudam-se modelos de redes de neurônios e outros sistemas dinâmicos não-lineares, analiticamente como também mediante simulações por computador.

No capítulo 1 fazemos um resumo da biologia do cérebro e de como, sendo um sistema físico, ele pode ser considerado e analisado como um sistema estatístico com uma dinâmica própria.

No capítulo 2 consideramos redes com diluição das sinapses. Estudamos o efeito de cortar sinapses de forma simétrica e assimétrica em modelos de redes de neurônios analógicos. Encontramos que as propriedades de reconhecimento dos modelos são pouco afetadas para graus de diluição razoavelmente elevados, de modo similar ao que acontece em modelos discretos. No caso de diluição assimétrica, simulações por computador revelam uma fase caótica na dinâmica da rede.

No capítulo 3 estudamos a estabilidade das memórias numa família de regras de aprendizado hebbianas. Consideramos memórias com superposição não nula e regras assimétricas. Uma análise sinal-ruído das memórias revela que, no limite de baixa atividade, a regra ótima é simétrica, enquanto que regras assimétricas são mais robustas para valores gerais da atividade.

No capítulo 4 apresentamos um estudo analítico das possibilidades de generalização numa rede neuronal analógica. Encontramos que o sistema é capaz de reconhecer categorias, desde que seja apresentado um número mínimo de exemplos das mesmas.

No capítulo 5 estudamos, por meio de simulações, a interação entre a parábola logística e um autômato celular determinista. Encontramos uma dinâmica que apresenta uma forma de ruído não-trivial: bandas múltiplas.

Abstract

In the present thesis we study neural network models and other dynamical systems, both analytically and by means of computer simulations.

In chapter 1 we make a review of the biology of the brain, and of how it can be considered and analyzed as a statistical system with its own dynamics, taking into account that it is a physical system.

In chapter 2 we consider networks with dilution of synapses. We study the effect of symmetric as well as asymmetric dilution in models of analog neural networks. We find that the retrieval properties of the models are mildly affected for relatively high levels of dilution, similarly to what occurs in discrete models. In the case of asymmetric dilution, computer simulations reveal a chaotic phase in the dynamics of the network.

In chapter 3 we study the stability of the memories in a family of hebbian learning rules. We consider memories with finite overlap and asymmetric rules. A signal-to-noise analysis reveals that, in the limit of low activity, the optimal rule is a symmetric one. However asymmetric rules are more robust for general values of the activity.

In chapter 4 we present an analytical study of the possibilities of generalization in an analog neural network. We find that the system is capable of recognizing categories, if a critical minimum number of examples of them are presented to the network.

In chapter 5 we study, by means of computer simulations, the interaction between the logistic map and a deterministic cellular automaton. We find a dynamics that presents an unusual type of noise, namely multiple bands.

Sumário

Resumo	iv
Abstract	v
Sumario	vi
Lista de Figuras	ix
Lista de Tabelas	xii
Introdução	1
1 Redes Neurais: elementos físicos e biológicos para a modelagem de funções do cérebro	5
1.1 A Física e o Estudo do Cérebro	5
1.2 Os átomos do cérebro: os Neurônios	7
1.2.1 Transmissão da informação no sistema nervoso	8
1.3 Da Biologia à Modelagem do Processamento da Informação	13

1.3.1	Fluxo da Informação em Redes Neurais Analógicas	13
1.3.2	Os Atratores da Dinâmica e o Processo de Reconhecimento	17
1.4	Aprendizado em Redes Neurais	20
2	Redes Neurais com Conectividade Restrita: o Papel da Diluição	26
2.1	Diluição em Redes Neurais	26
2.2	Diluição Simétrica numa Rede Neuronal Analógica	28
2.2.1	Ruído Gaussiano nas Sinapses	29
2.2.2	Analogia com uma Diluição simétrica nas sinapses: performance dos modelos	30
2.3	Diluição Assimétrica	33
2.3.1	Simulando uma Rede Neuronal Analógica	35
2.3.1.1	Diagramas de Fases	36
2.3.1.2	Fases de Reconhecimento	39
2.3.1.3	Caos em Redes Neurais	42
2.4	Conclusões	43
2.5	Apêndice: análise estatística da rede analógica com ruído gaussiano nas sinapses	44
3	Padrões com Atividade e Regras Hebbianas com Assimetria	48
3.1	Motivações	48
3.2	Padrões com Atividade	50
3.3	Estabilidade dos Padrões : Análise Sinal-Ruído	50
3.3.1	A Regra de Hopfield com padrões independentes	50
3.3.2	A Regra de Hopfield com padrões com atividade	51
3.3.3	Armazenamento eficiente de padrões com atividade e o limite de baixa atividade	52
3.4	Regras Hebbianas Generalizadas e o Papel da Assimetria	55
3.5	Conclusões	66
4	Aprendizado e Generalização em Redes de Neurônios	68

4.1	Aprendizado e Generalização em Redes de Neurônios tipo Perceptron	69
4.2	Redes de Neurônios com Estrutura Hierárquica das Memórias	74
4.2.1	Alguns exemplos	75
4.3	Generalização numa Rede de Neurônios Analógica	78
4.3.1	O Modelo	79
4.3.2	Análise Estatística e Equações dos Pontos Fixos	81
4.3.3	O limite $g \rightarrow \infty$	83
4.3.4	Resultados para g finito	86
4.3.5	Conclusões	88
4.4	Discussão e Perspectivas	88
5	Sistemas Dinâmicos Complexos: Mapas Unidimensionais e Automata Celulares	91
5.1	A parábola logística	93
5.2	Automata Celulares	95
5.3	Interação entre a Parábola Logística e um Autômato Celular	98
5.3.1	A Dinâmica da Parábola Logística com Perturbações	98
5.3.2	A Parábola Logística Controlada por um Autômato Celular Determinista: uma classe de ruído não trivial	100
5.4	Conclusões	106
6	Conclusões	108

Lista de Figuras

1.1	Versão simplificada de um neurônio	8
1.2	Esquema de uma junção sináptica química, o terminal do axônio pré-sináptico e o dendrito do neurônio pós-sináptico	9
1.3	Evolução temporal de um potencial de ação , V_R indica o valor de repouso	11
1.4	Função de transferencia contínua para diferentes valores do ganho . . .	15
1.5	Representação esquemática da metáfora de vales e colinas para um processo mental	18
2.1	Superposição crítica vs. grau de diluição , numa rede neuronal analógica, para diferentes valores do ganho g	32
2.2	Capacidade de armazenamento crítica vs. grau de diluição , numa rede neuronal analógica, para diferentes valores do ganho g	33
2.3	Diagrama de fases <i>diluição vs.ganho</i> para uma rede de neurônios analógica com $\alpha=0.01$. Os quadrados são o resultados de simulações numa rede com 480 neurônios e a linha cheia é uma aproximação teórica para o limite $N \rightarrow \infty$ e $\alpha \rightarrow 0$ (ver texto).	37

2.4	Diagrama de fases <i>diluição vs. ganho</i> para uma rede de neurônios analógica com $\alpha = 0.06$. Os quadrados são resultados de simulações numa rede de 480 neurônios. As regiões com linhas curtas assinalam as possíveis fronteiras entre as fases não-estacionária, vidro de spin e reconhecimento. As linhas cheias são uma guia para o olho.	38
2.5	<i>Superposição vs. ganho</i> para uma rede de neurônios analógica com $\alpha = 0.06$. Os círculos são resultado de simulações numa rede com 480 neurônios, e um nível de diluição de 10%. Os triângulos correspondem a uma diluição de 50%. As barras correspondem à dispersão estatística dos resultados.	41
2.6	<i>Superposição vs. diluição</i> numa rede de neurônios analógica para $\alpha = 0.06$ e três tamanhos da rede: $N=50$ (quadrados), $N=100$ (círculos) e $N=480$ (triângulos), e um valor fixo do ganho $g=10$	42
3.1	Capacidade de armazenamento crítica α_c vs. parâmetro de assimetria γ de uma análise sinal-ruído ($\gamma = 1$ corresponde a regras simétricas): (a) $a=1$; (b) $a=0.6$; (c) $a=0.2$; (d) $a=0.05$	62
3.2	Capacidade de armazenamento crítica α_c vs. parâmetro de assimetria γ obtido simulando uma rede com 400 neurônios analógicos para $a=0.2$. Os quadrados vazios (triângulos cheios) correspondem a $g=100$ ($g=2.5$).	64
4.1	Esquema de uma rede tipo perceptron.	70
4.2	Esquema de uma árvore hierárquica.	76
4.3	Organização hierárquica das memórias na rede analógica	81
4.4	Erro de generalização ϵ_g vs. número de exemplos s no limite $g \rightarrow \infty$, para $b=0.2$ e $\alpha = 0, 0.1, 0.5 e 1$	84
4.5	As curvas críticas $b(s)$ no limite $g \rightarrow \infty$, para $\alpha_c = 0.05, 0.1, 0.5 e 1$. O sistema começa a generalizar acima da correspondente curva.	84
4.6	Erro de generalização ϵ_g vs. b no limite $g \rightarrow \infty$, para $s=50$ e $\alpha = 0, 0.1, 0.5 e 1$	85
4.7	Erro de generalização ϵ_g vs. número de exemplos s , para $g=1, b=0.2$ e $\alpha = 0, 0.1, 0.5 e 1$	86

4.8	As curvas críticas $b(s)$, acima das quais o sistema generaliza, para $g=1$ e $g \rightarrow \infty$, com $\alpha = 0.1$	87
4.9	Erro de generalização ϵ_g vs. número de exemplos s , para $g = 1$ e $g \rightarrow \infty$, sendo $b = 0.1$ e $\alpha = 0.1$	87
5.1	Diagramas espaço-temporais para algumas regras com $k=5$ e $r=2$. No sentido vertical, a evolução temporal.	97
5.2	Representação esquemática dos atratores das eqs.(5.12) e (5.13), para três níveis de ruído diferentes	101
5.3	Diagramas espaço-temporais dos atratores das eqs.(5.12) e (5.13) para as primeiras 100 iterações depois do transiente (5000 iterações); círculos cheios correspondem a $\sigma_i(t) = 1$, e os vazios a $\sigma_i(t) = 0$. No detalhe, as correspondentes distribuições de frequências, para 10000 iterações após o transiente (o intervalo $[0,1]$ foi particionado em 2000 subintervalos iguais). (a)(a,b)=(2.95,0.05).	103
5.4	(b)(a,b)=(2.38,0.1).	104
5.5	(c)(a,b)=(3.14,0.1).	105
5.6	Distribuição de frequências para três casos típicos que satisfazem $a + b = 4$. Superposto ao caso $b=0$, a distribuição exata $1/\pi\sqrt{x(1-x)}$	106

Lista de Tabelas

3.1	Parâmetros experimentais de sinapses excitatórias.	56
3.2	Parâmetros teóricos para as sinapses	57
5.1	As regras C=90 e C=22 decompostas em forma binária.	96

Introdução

Desde o último decênio, as mais diversas áreas das ciências vêm juntando esforços para estudar problemas de caráter notadamente multidisciplinar. Problemas tais como a evolução das espécies, o funcionamento do sistema imunológico, crescimento e evolução de populações, a física de sistemas magnéticos e o estudo de funções do cérebro, entre muitos outros, estão pondo em evidência a necessidade de que as diferentes ciências saiam do isolamento para aportar cada uma suas técnicas e experiências no estudo destes temas.

Um dos empreendimentos mais ambiciosos é a compreensão das funções do cérebro onde, embora muito avanço tenha sido feito neste século, muito ainda é desconhecido. A medicina, a biologia, a psicologia e a química têm feito grandes aportes ao estudo do cérebro. Na segunda metade deste século, com o aparecimento dos computadores, surgiu uma linha de pesquisa conhecida hoje como “inteligência artificial”. Os pesquisadores nesta área, fundamentalmente engenheiros e matemáticos, começaram a tentar simular o comportamento lógico do cérebro com máquinas de estado sólido, os computadores. Em 1943 Mc Culloch e Pitts [1] introduziram o primeiro modelo de neurônio artificial: simularam o neurônio como uma simples unidade lógica. Ele podia estar em um dentre

dois estados: ativo ou em repouso. Por sua vez, cada neurônio estaria ligado a outros, dos quais receberia e aos quais enviaria sinais elétricas, formando desta forma um circuito, ou *rede de neurônios*. O físico-matemático von Neumann, na década dos 50, foi dos primeiros a estudar as possibilidades lógicas das “máquinas inteligentes”. Na década dos 60 muito trabalho foi feito em modelos conhecidos como “perceptrons”. Estes são redes formadas por uma ou varias camadas de neurônios, onde a informação flui numa direção preferencial. O avanço nesta linha de pesquisa foi interrompido pelo aparecimento do livro “Perceptrons”, de autoria de Minsky e Papert [38], no qual se fazia uma dura crítica sobre as possibilidades destes modelos no avanço da inteligencia artificial. Em 1974 aparece um trabalho de Little [2], muito importante conceitualmente, no qual, pela primeira vez, a dinâmica de uma rede de neurônios é considerada como a de um sistema termodinâmico à temperatura fixa.

O grande avanço feito no final da década dos 70 na teoria de sistemas desordenados, e particularmente na teoria de campo médio de certos sistemas magnéticos conhecidos como vidros de spin [70,54,3], fez aos físicos pensar que certas funções do cérebro podem ser modeladas considerando o proprio cérebro um sistema desordenado. Cada neurônio seria o análogo ao spin e o valor das sinapses que conectam os nerônios seria equivalente ao valor da interação magnética. Esta interessante analogia, percebida pela primeira vez por J. Hopfield [4], abriu caminho para uma nova linha de pesquisa em física, a das redes de neurônios. A analogia baseia-se na possibilidade de identificar cada um dos múltiplos estados de baixa energia de um sistema desordenado com uma memória, ou conceito na rede neuronal. Este descobrimento, além de dar novo impulso à antiga linha ligada à inteligencia artificial, abriu a possibilidade de modelar certas funções do cérebro biológico. Na década dos 80, poderosas ferramentas da física-matemática foram utilizadas no estudo e desenvolvimento de modelos de redes de neurônios, principalmente no estudo da memória: armazenamento e reconhecimento de padrões . Os modelos introduzidos, ainda longe de ser realistas, vão aperfeiçoando-se continuamente desde a proposta original de Hopfield. Dentre as contribuições mais importantes, Amit, Gutfreund e Sompolinsky [12,13,14] resolveram a mecânica estatística do modelo de Hopfield, e E.Gardner [39] introduziu uma análise no espaço das interações (sinapses)

que permite se tornar independente de um modelo particular.

Nos últimos anos, novos problemas, além do da memória, vêm sendo estudados, tais como possíveis mecanismos de aprendizado e a possibilidade de generalizar a partir de exemplos. Dentre as extensões e generalizações dos modelos originais, aparecem como particularmente interessantes modelos nos quais as variáveis dinâmicas, ou seja o estado do neurônio, são contínuas [16,18,19], conhecidos como *redes de neurônios analógicos*. Estes modelos são de mais fácil implementação em forma de circuitos e permitem modelar outros aspectos da dinâmica das redes, já que se baseiam na codificação de frequências de disparo dos neurônios e não na codificação de pulsos discretos como no caso dos modelos tipo Hopfield.

Na presente tese vamos introduzir e analisar alguns aspectos de modelos de redes de neurônios analógicos. O capítulo 1 está dedicado a uma introdução , necessariamente muito limitada, à biologia das redes de neurônios, à codificação e transmissão da informação , à relação entre a dinâmica do fluxo de informação e o processo de memorização e reconhecimento e finalmente, a uma discussão sobre o problema da dinâmica do aprendizado.

Nos capítulos 2, 3, 4 e 5 são apresentadas as contribuições originais da tese. No capítulo 2 estudamos como as propriedades de reconhecimento numa rede de neurônios analógica são afetadas quando é cortada uma certa fração das sinapses da rede [55]. Os modelos assim incorporam um ingrediente realista, isto é , conectividade restrita por neurônio. Numa primeira parte estudamos o efeito de diluir de forma simétrica as sinapses e, para a análise, nos aproveitamos de uma interessante equivalência entre diluir e adicionar um termo de ruído na matriz de sinapses. Numa segunda parte consideramos as conseqüências de cortar sinapses de forma assimétrica. Nos dois casos calculamos os diagramas de fases com as capacidades críticas de armazenamento em função do grau de diluição .

No capítulo 3 introduzimos uma família de regras de aprendizado Hebbianas [68], que permitem analisar uma grande variedade de condições nas atividades dos neurônios pré e pós-sinápticos. Na forma mais geral as regras são assimétricas. Consideramos padrões com atividade e , usando uma análise sinal-ruído, estudamos a estabilidade das

memórias para diferentes regras, simétricas e assimétricas, bem como para diferentes valores da atividade dos padrões . Consideramos o limite, interessante do ponto de vista biológico, de padrões com baixa atividade, e procuramos a regra que otimize as condições de reconhecimento em função do grau de assimetria da regra e do nível de atividade dos padrões . Utilizando simulações por computador verificamos os resultados da análise sinal-ruído.

No capítulo 4 deixamos o problema de reconhecimento de padrões para estudar as possibilidades de generalização em redes de neurônios. Depois de fazer uma revisão do trabalho feito em redes tipo perceptron, usamos uma rede de neurônios analógica com uma estrutura hierárquica simples das memórias, e investigamos a capacidade do modelo para criar um conceito representativo de uma série de exemplos do mesmo apresentados à rede no processo de aprendizado [81]. Fechamos o capítulo com uma discussão sobre a possibilidade de criar máquinas “inteligentes”, com uma definição operacional de inteligência.

O capítulo 5 está dedicado a um estudo numérico da dinâmica de interação entre um mapa unidimensional (parábola logística) e um autômato celular determinista [115]. Estes sistemas dinâmicos apresentam propriedades similares às das redes de neurônios no que diz respeito à presença de pontos fixos, ciclos limite e atratores caóticos nas suas dinâmicas. É interessante a possibilidade de definir grandezas comuns, tais como expoentes de Liapunov, para caracterizar os atratores em sistemas tão diversos como a parábola logística (definida pela iteração de uma variável contínua), e um autômato celular (sistema com muitas variáveis discretas em interação).

Capítulo 1

Redes Neurais: elementos físicos e biológicos para a modelagem de funções do cérebro

1.1 A Física e o Estudo do Cérebro

Tradicionalmente objeto de estudo da medicina, biologia, psicologia, o funcionamento do cérebro tem interesse tanto prático como teórico. As redes neurais sugerem novas estruturas para sistemas computacionais e novos métodos de aprendizado. No entanto, a maior ambição no estudo das redes neurais é a compreensão do sistema nervoso. No despontar deste século, no auge da física como mãe das ciências da natureza, Sir E.Rutherford dizia: *“Ciência é física, o resto é colecionar selos”*. Certamente a realidade da ciência na atualidade é outra e outras ciências, que não a física, tomaram

seu lugar na vanguarda da aventura científica. Talvez a biologia tenha sido responsável pelos descobrimentos científicos mais importantes da segunda metade deste século XX, desde a descoberta da estrutura do ADN, o código genético, até o alto status que a Teoria da Evolução tem atingido nos últimos anos. Tentando aportar alguma coisa nova, graças aos seus poderosos métodos analíticos, nos últimos anos a física tem-se juntado a outras ciências num empreendimento muito ambicioso: explorar o cérebro e tentar compreender como compreendemos. Certamente este empreendimento deverá ser multidisciplinar: a biologia, como ciência pioneira no estudo do cérebro, continuará a aportar muitos dos avanços na área, principalmente na descrição da morfologia e da fisiologia do sistema nervoso; a ciência da computação terá muito a dizer no que diz respeito às propriedades computacionais do cérebro; a física aplicará sua experiência considerando o cérebro como um sistema físico, com massa, temperatura, campos; e assim por diante com muitos outros ramos científicos que serão imprescindíveis para obter uma ideia clara de como o cérebro funciona.

A física ingressou definitivamente no estudo das funções cerebrais no começo da década de 80 com o trabalho pioneiro do neurobiólogo John Hopfield[4]. Partindo de modelos **conexionistas**, ele explorou uma analogia entre as redes neuronais e certos materiais magnéticos conhecidos dos físicos como *vidros de spin*. Modelos conexionistas já eram conhecidos nas neurociências desde varias décadas antes dos trabalhos de Hopfield. Partem do pressuposto de que o comportamento do modelo é consequência da interação de um número grande de unidades “conectadas” entre si, e não somente das propriedades de cada unidade. Esta filosofia faz-nos lembrar à física estatística a qual é capaz de analisar sistemas formados por um grande número de unidades simples que, no entanto, apresentam comportamentos de caráter cooperativo quando interagem entre si. É assim que a física deverá considerar o estudo do cérebro em dois níveis diferentes: primeiro enquanto sistema físico comum, determinando a natureza das reações químicas, registrando a resposta dos neurônios aos estímulos elétricos, analisando os caminhos dos sinais elétricos na rede celular, etc., e segundo, como sistema físico altamente complexo, deverá estudar propriedades que não estão presentes em sistemas com níveis menores de organização . Algumas das propriedades essencialmente coletivas, próprias do com-

portamento cognitivo, são : armazenamento de memórias, reconhecimento de padrões, associação, generalização, aprendizado, resolução de problemas, ..., e finalmente os problemas das percepções , sensações e consciencia. A partir do trabalho original de Hopfield podem-se distinguir duas grandes linhas de pesquisa que se diferenciam por suas motivações : por uma parte está o projeto de desenvolver máquinas inteligentes, capazes de substituir o trabalho do homem, inclusive o intelectual, empreendimento similar, ainda que mais ambicioso, do que o da revolução industrial no século passado; por outra parte está o interesse científico de compreender o funcionamento do cérebro real, biológico. Obviamente, ambas linhas de pesquisa estão fortemente entrelaçadas e os avanços numa delas é imediatamente aproveitado na outra área.

Para poder discutir alguns dos desenvolvimentos na área de redes neuronais nos últimos anos, vamos começar com uma resenha muito simplificada dos elementos que formam as redes neuronais, os **neurônios**, sua estrutura e a forma como eles carregam a informação que será responsável pelos comportamentos extraordinariamente complexos próprios do cérebro.

1.2 Os átomos do cérebro: os Neurônios

Vamos começar esta seção chamando a atenção para o fato de que em qualquer resenha simplificada do sistema nervoso, intentos de fazer afirmações gerais é difícil e perigoso. Podem-se achar tratados autorizados da biologia do sistema nervoso nas Refs.[5]- [8].

Na Fig.1.1 vemos uma representação esquemática de uma célula do sistema nervoso: o **neurônio**, dos quais existem $\approx 10^{10}$ no cérebro. Um neurônio está formado basicamente por três partes: o **corpo celular** ou **soma**, que contem o núcleo da célula e é o responsável pela atividade metabólica normal. Saíndo do corpo celular, apresentam-se umas foliações , são fibras chamadas **dendritos**, que podemse ramificar repetidamente porém mantendo-se perto do corpo celular. A terceira parte importante na morfologia de um neurônio é uma fibra longa chamada **axônio**, o qual pode-se estender por uma distância considerável desde o corpo celular até se ramificar. Levando em conta que a escala de distâncias característica do mundo neuronal é o micron $\mu = 10^{-6} m$, os

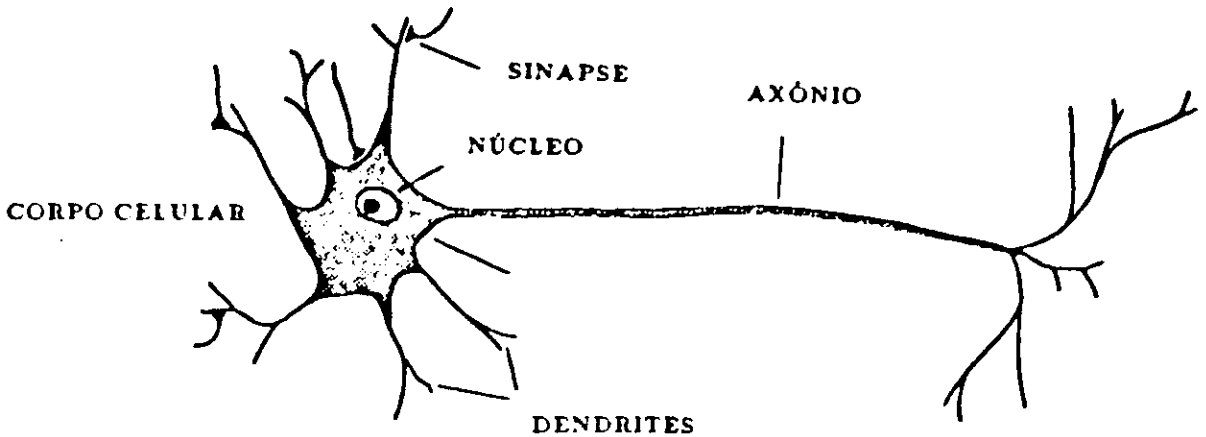


Figura 1.1: Versão simplificada de um neurônio

diâmetros do corpo celular variam entre $5\text{-}100\ \mu$, a largura do axônio entre $1\ \mu$ chegando a $20\ \mu$ em casos extremos e seu comprimento varia entre $\approx 0.1\text{mm}$ até $\approx 1\text{m}$ no homin. A variedade nos padrões dos dendritos são ainda mais impressionantes.

1.2.1 Transmissão da informação no sistema nervoso

O grau de conectividade no cérebro (um neurônio estará conectado em média a $\approx 10^4$ outros neurônios numa intrincada rede neuronal) pode nos dar uma idéia da incrível complexidade dos processos de transmissão da informação no sistema nervoso. Cada uma das partes canônicas de uma célula nervosa—árvore dendrítica, corpo celular e axônio— é especializada para uma função particular no processamento da informação . A informação é recebida pelos dendritos ou o corpo celular, e vem em forma de estímulos elétricos, que podem provir de fontes externas ou de outros neurônios. Estes neurônios comunicam-se com o neurônio em questão via junções sinápticas ou simplesmente **sinapses** (ver Fig. 1.2). As sinapses são formadas pela proximidade do terminal de um axônio com uma dendrite ou o corpo celular de outro neurônio. A comunicação é

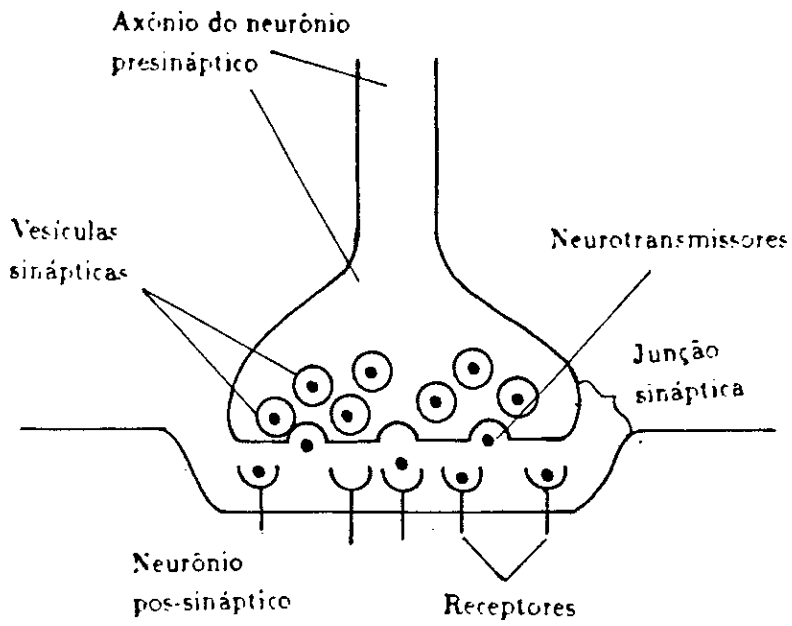


Figura 1.2: Esquema de uma junção sináptica química, o terminal do axônio pré-sináptico e o dendrito do neurônio pós-sináptico

feita em geral mediante um processo eletro-químico. Enquanto a membrana celular não recebe estímulo algum, é mantida através dela uma diferença de potencial, denominada *potencial de repouso*. Este é produto do equilíbrio nas concentrações de diferentes íons dentro e fora da célula. Em condições de equilíbrio, não temos fluxo líquido de corrente iônica através da membrana e resulta um potencial de repouso de $\approx -70mV$, o sinal negativo indica que o interior da membrana em repouso está negativamente carregado com respeito ao fluido exterior. Quando está ativo, o neurônio pré-sináptico libera a través da membrana do seu axônio, um composto químico conhecido como **neurotransmissor**. As moléculas neurotransmissoras são difundidas através da sinapse em direção à membrana pós-sináptica onde se fixam a receptores. Estes são grandes proteínas que formam canais que atravessam a parede da membrana pós-sináptica, e alteram a sua permeabilidade com respeito a certos íons, principalmente sódio Na^+ e potássio K^+ . São geradas pequenas correntes iônicas, que produzem flutuações no potencial da membrana com respeito ao seu valor de repouso. Estas flutuações chamadas **potenciais pós-sinápticos (PPS)** podem ser positivos (correspondendo a um estímulo excitatório), ou negativos (correspondendo a um estímulo inibitório), dependendo da natureza da substância neurotransmissora e da natureza dos receptores na membrana pós-sináptica. Um PPS excitatório implica num fluxo de Na^+ para o interior da célula, e um PPS inibitório uma corrente local de K^+ do interior para o exterior da célula. O primeiro caso “depolariza” a membrana (seu potencial torna-se menos negativo) e o segundo caso “hiperpolariza” a membrana.

O corpo celular, além de manter os processos metabólicos da célula, também é responsável pela integração dos estímulos que chegam ao neurônio. Os PPS, originados localmente nos diversos pontos dos dendritos e do corpo celular, decaem com a distância percorrida e com o tempo. O efeito conjunto é finalmente integrado na região onde o axônio encontra-se com o corpo celular. O resultado da integração pode ser um aumento do potencial de membrana, uma depolarização excitatória, ou uma diminuição do potencial ou hiperpolarização inibitória.

Se o potencial de membrana cresce acima de um certo limiar, acontece um evento catastrófico, como causa do desbalanceamento das concentrações internas e externas

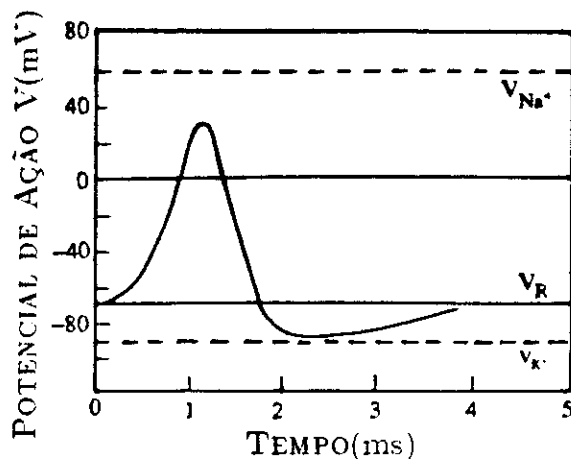


Figura 1.3: Evolução temporal de um potencial de ação, V_R indica o valor de repouso

de sódio e potássio. Quando o potencial cresce até valores positivos e o interior da membrana carrega-se positivamente, fecham-se os canais de sódio e são abertos os canais de potássio, permitindo que íons potássio carregados positivamente saiam da célula, e assim se recomponha o valor de repouso do potencial. Esta grande e rápida flutuação do potencial de membrana é chamada **potencial de ação** e está esquematicamente representado na Fig.1.3.

A excitação (ou pulso), é transmitida ao longo do axônio rapidamente e sem atenuação, a velocidade de transmissão variando entre 1-120 m/s em mamíferos.

A transmissão de informação entre os neurônios tem varias escalas de tempo, que podem ser importantes na modelagem dos processos neuronais. Se um neurônio dispara um potencial de ação, deve aguardar um certo tempo até poder efetuar um novo disparo, independentemente da magnitude do estímulo. Este tempo morto é conhecido como *período refratário absoluto*. Sua duração é de uns poucos milissegundos. Também existe um período mais longo, durante o qual o neurônio é menos excitável do que normalmente, de forma tal que um potencial de ação é produzido somente quando o

estímulo ultrapassa um limiar maior do que o normal. Este é o chamado *período refratário relativo* e pode durar até $\approx 50ms$. Uma consequência da natureza refratária dos neurônios é a observação da **unidirecionalidade** da transmissão dos pulsos na rede neuronal: se afastando do corpo celular e descendo pelo axônio até seus terminais. Também a natureza assimétrica das sinapses é responsável por um outro aspecto do fluxo unidirecional da informação em sistemas neuronais. A influência da assimetria em modelos de redes neuronais será analisada em detalhe nos capítulos 2 e 3. O período refratário fornece uma escala de tempos para os eventos neuronais, outra escala é fornecida pelo *retardamento sináptico* entre a chegada de um potencial de ação ao terminal de um axônio pré-sináptico e a produção de um PPS na membrana pós-sináptica. Este retardamento pode levar entre uns décimos de milissegundo até 3–4 ms.

Resumindo, os principais pontos na resposta unitária de um neurônio são os seguintes:

- Depois de receber estímulos de uma ou mais células (ou de alguma fonte externa), o neurônio pode apresentar uma resposta local, que decai no tempo, de magnitude dependente do estímulo recebido. Exemplos são potenciais pós-sinápticos excitatórios ou inibitórios.
- Se um neurônio recebe um estímulo despolarizador que é suficientemente grande num tempo relativamente curto, ele pode emitir um pulso elétrico, ou potencial de ação, uma unidade de informação sempre a mesma para a mesma célula, e com uma amplitude independente da magnitude do estímulo.

Baseado neste comportamento, se diz que o potencial de ação é um processo “tudo ou nada”, e que o neurônio funciona como um elemento binário com um certo limiar de disparo. Esta forma de raciocínio leva-nos a pensar que a codificação da informação numa rede neuronal é feita pela presença ou ausência de um pulso, ou seja, uma série de zeros e uns ou uma série de intervalos entre pulsos (potenciais de ação). Fala-se então de **codificação de pulso**.

Como todos os potenciais de ação de um neurônio específico são iguais, os pulsos individuais nada podemos dizer acerca do sinal que está sendo recebido, exceto que sua intensidade é maior que um certo limiar. No entanto, como é observado usualmente

em sistemas sensoriais, um neurônio podemos dizer que está “sentindo” um estímulo maior, disparando pulsos com uma frequência maior. De acordo com este comportamento, a frequência de pulsos num neurônio pode ser considerada como uma medida da intensidade do estímulo. Fala-se então de **codificação de frequência** da informação neuronal. O aumento da frequência dos pulsos pode ser consequência de outros fatores além do aumento da intensidade do estímulo, mas não vamos nos aprofundar mais na análise desta questão .

Codificação de frequência é evidentemente mais apropriada na descrição da atividade neuronal quando o neurônio recebe estímulos prolongados no tempo, e os potenciais de ação são intensos e rápidos. Porém, codificação de pulso faz maior sentido quando os estímulos são transitórios ou esporádicos, e o tráfego de potenciais de ação é esparso. Neste sentido as duas descrições da resposta neuronal são complementares.

1.3 Da Biologia à Modelagem do Processamento da Informação

1.3.1 Fluxo da Informação em Redes Neurais Analógicas

Na presente tese vamos estudar propriedades de modelos simples de redes neurais, baseados na *codificação de frequência*, ou seja, as variáveis básicas dos modelos tomarão valores reais num intervalo contínuo[9]. Estes modelos são conhecidos como **redes neurais analógicas**. No entanto, modelos baseados na *codificação de pulso* tem sido os mais populares na última década. Nós não vamos considerá-los em muito detalhe: resenhas muito interessantes e atualizadas podem ser encontradas no livro de Daniel Amit[11] “*Modelling Brain Function*” e na Tese de Doutorado de Francisco A. Tamarit[54].

Consideraremos modelos nos quais o processamento da informação dependerá somente de dois grupos de variáveis:

1. A distribuição de frequências entre os neurônios.
2. O catálogo de eficácias sinápticas.

Especificamente: vamos supor que não há estímulo externo presente na rede. Num

intervalo de tempo da ordem de 1-2ms podemos registrar a freqüência de disparo de cada neurônio. Isso dará a distribuição instantânea de freqüências. O catálogo de sinapses, na ausencia de aprendizado, fixa a “arquitetura” ou “diagrama de conectividade” da rede, representando as interações. A freqüência instantânea determinará a liberação de neurotransmissores pelos axônios pré-sinápticos. Os valores das eficácias sinápticas determinarão quanto cada neurônio irá receber. Esta é a informação essencial que um neurônio necessitará para computar seu estado próximo (variação da sua freqüência de disparo), e isto é o que mantém o processamento de informação em andamento.

De acordo com o expressado anteriormente, numa rede neuronal consistente de N neurônios e as correspondentes sinapses, a saída (output) do neurônio i -ésimo, x_i , será uma variável contínua e sua dinâmica pode ser expressa pelo seguinte conjunto de equações diferenciais ordinarias acopladas[9,10]:

$$\tau_i \frac{dx_i}{dt} = -x_i + F \left(\sum_{j=1}^N J_{ij} x_j \right), \quad i = 1 \dots N, \quad (1.1)$$

onde N é o número de neurônios na rede e τ_i é uma constante de tempo, que basicamente fixa o tempo de decaimento do potencial pós-sináptico. Pode-se observar que em ausencia do último termo a freqüência de disparo irá decair exponencialmente com uma constante de tempo dada por τ_i . A entrada (input) ao neurônio i -ésimo vem dada pela expressão entre parêntesis, que nada mais é do que o potencial de ação :

$$h_i = \sum_{j=1}^N J_{ij} x_j \quad (1.2)$$

É uma soma linear dos estímulos recebidos dos restantes neurônios que estão ligados ao i -ésimo pesados com as correspondentes eficácias sinápticas J_{ij} . Finalmente o soma faz uma integração não-linear da entrada via a *função de transferencia* F . A função F apresenta geralmente uma não-linearidade em torno da origem e satura para valores muito afastados da origem. Uma escolha muito comum é a tangente hiperbólica $F(z)=\tanh(gz)$, onde g mede o grau de não-linearidade (ver Fig.1.4). A interpretação da função de transferencia é a seguinte: quando o PPS toma valores muito positivos, a freqüência de disparo satura, o neurônio está emitindo pulsos na máxima freqüência possível. Se o PPS toma valores muito negativos a freqüência é muito baixa e o neurônio

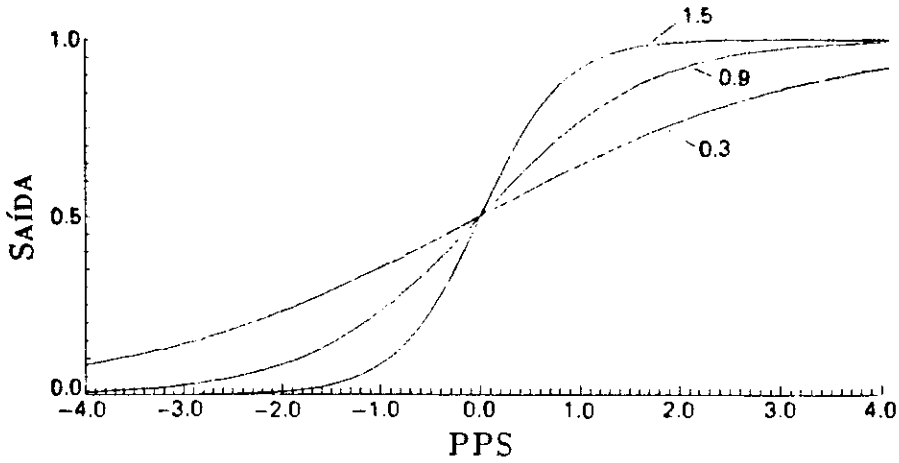


Figura 1.4: Função de transferência contínua para diferentes valores do ganho

está em repouso. Quando o grau de não-linearidade ou *ganho* g é muito grande, a frequência flutua entre máxima e nula, e pode-se fazer uma equivalência completa com os modelos discretos, baseados na codificação de pulso. Se a matriz de sinapses J_{ij} é simétrica, a dinâmica das eqs.(1.1) sempre atinge um estado de equilíbrio no qual $x_i(t)$ deixa de flutuar, então $dx_i/dt = 0$ para todo i , e todos os neurônios obedecem :

$$x_i = F \left(\sum_{j=1}^N J_{ij} x_j \right), \quad i = 1 \dots N. \quad (1.3)$$

As suas soluções serão os atratores da dinâmica. Se a matriz de conexões é simétrica ($J_{ij} = J_{ji}$) pode-se definir uma função energia ou função de Liapunov para o sistema de equações (1.1) :

$$L(t) = \frac{1}{2} \sum_{i,j} J_{ij} x_i(t)x_j(t) + \sum_{i=1}^N G_i(x_i(t)), \quad (1.4)$$

$$G_i(x_i) \equiv \int_0^{x_i} F_i^{-1}(z) dz \quad (1.5)$$

e pode-se demonstrar que os atratores são somente pontos fixos[9].

Estas equações são inteiramente equivalentes as equações que descrevem um circuito eletrônico. Mediante adequadas mudanças de variáveis podemos reescrever as eqs.(1.1)

da forma:

$$C_i \frac{dh_i}{dt} = \sum_{j=1}^N J_{ij} g(h_j) - \frac{h_i}{R_i} \quad (1.6)$$

Agora as variáveis dinâmicas são os potenciais pós-sinápticos h_i que vão mudar de valor devido a dois fatores:

1. Correntes induzidas pela atividade de neurônios pré-sinápticos.
2. Correntes de difusão devido à resistividade finita da membrana.

Uma terceira fonte de entrada podem ser correntes externas; estas não serão tidas em consideração. O lado direito da equação contém as correntes que irão contribuir para carregar a capacitancia C_i da membrana i -ésima pelo potencial dh_i . O primeiro termo representa as contribuições devidas às atividades de todos os outros neurônios com as sinapses J_{ij} fazendo o papel de resistencias, e o segundo representa as correntes difusivas devidas à resistencia R_i , entre o interior e o exterior da membrana. Na ausencia do primeiro termo o potencial h_i decai a zero com uma constante de tempo igual a $R_i C_i$. A representação em termos de circuitos eletrônicos é muito conveniente para as implementações em hardware de redes neuronais, e oferecem um esquema muito próximo do que é um circuito biológico real.

As equações (1.1) evoluem em tempo contínuo. No entanto na presente tese consideraremos o conjunto de mapas acoplados (1.3) como o sistema que define a dinâmica básica da rede neuronal, com **tempo discreto e atualização em paralelo**. A atualização em paralelo é mais natural e conveniente para fazer grandes simulações computacionais. Não será considerada a atualização seqüencial. Notamos que nenhuma das duas prescrições de atualização tradicionais (seqüencial e paralela) são boas aproximações ao fluxo de informação numa rede biológica. No entanto é interessante notar que a escala de tempos típica de um ciclo neuronal é de uns poucos milisegundos, o que é aproximadamente um milhão de vezes mais lento que num chip de silicone, embora o cérebro possa processar informação de forma muito mais rápida em funções como visão, controle motor e decisões baseadas em dados incompletos ou com ruído, problemas estes que estão muito além da capacidade de um supercomputador Cray. Obviamente

isto é possível graças ao fato de que bilhões de neurónios operam simultaneamente, em paralelo, de uma forma otimizada por milhões de anos de evolução . O problema importante de retardamento na transmissão de informação entre neurónios, já foi considerado por alguns autores[18,19,20]. Uma diferença importante entre as duas dinâmicas (tempo contínuo ou discreto), é que enquanto no caso das eqs.(1.1) os atratores são sempre pontos fixos (sempre que as sinapses sejam simétricas), a dinâmica paralela no sistema de mapas acoplados pode gerar ciclos de período 2. Porém todos os ciclos podem ser eliminados, diminuindo o ganho g para satisfazer a seguinte condição [16]:

$$\frac{1}{g} > -\lambda_{min} \quad (1.7)$$

na qual λ_{min} é o mínimo autovalor da matriz de sinapses J_{ij} . No entanto, se a matriz de eficácias sinápticas é **assimétrica**, as características da dinâmica podem mudar completamente. A frequência $x_i(t)$ pode oscilar, e no caso de conexões completamente aleatórias, pode-se apresentar caos na dinâmica como veremos no capítulo 2.

Uma característica importante com referencia às eqs. (1.3), é que se a função de transferencia é a tanh, as equações são as mesmas que as soluções para a magnetização média num sistema estocástico de spins de Ising com interações de longo alcance. A frequência x_i corresponderá ao valor medio do spin $\langle s_i \rangle$ no modelo estocástico, o potencial pós-sináptico corresponderá ao campo local médio h_i , e o ganho g à inversa da temperatura β . Conseqüentemente podemos utilizar uma rede neuronal com variáveis contínuas para resolver o problema de campo medio de um sistema estocástico a temperatura finita, com interações de longo alcance. Esta analogia muito interessante estará presente de forma geral nas soluções de modelos de redes neuronais com memórias associativas [12,16].

1.3.2 Os Atratores da Dinâmica e o Processo de Reconhecimento

Um ponto fixo estável da dinâmica será identificado com uma atividade neuronal importante, se ele manifesta-se como um período no qual alguns neurónios apresentam frequências de disparo perto do máximo, enquanto que outros apresentam frequências muito baixas. A identificação de uma atividade neuronal importante com pontos fixos

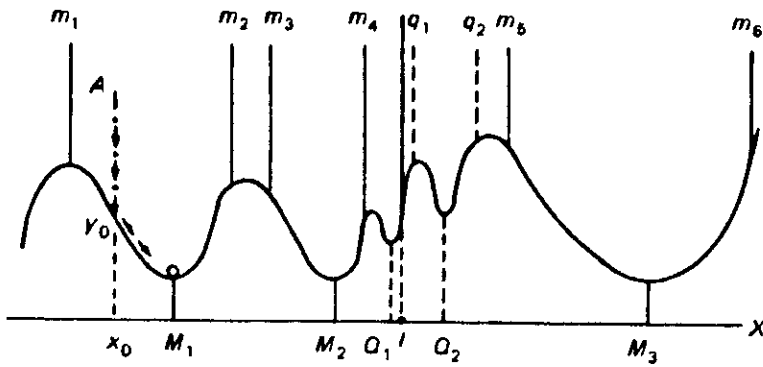


Figura 1.5: Representação esquemática da metáfora de vales e colinas para um processo mental atratores, sugere a imagem de superfícies com vales e colinas, sobre as quais o processo dinâmico executa uma caminhada, que termina no fundo de um dos vales. Essa imagem é muito atrativa na física, onde são comuns os processos de relaxação até mínimos de energia. Como nós estamos interessados em funções neuronais tais como *memorização*, *reconhecimento*, *correção de erros*, *classificação*, etc., vamos considerar metaforicamente como um desses processos pode ser representado. Vamos supor que desejamos armazenar memórias, elas estarão representadas pelas coordenadas do fundo dos vales M_1 , M_2 , M_3 na Fig.1.5. Um estímulo “cai” como uma gota verticalmente sobre a superfície, e faz contato com ela num ponto de coordenada x_0 . De aqui em diante, a gota escorrega sobre a superfície, descendo sempre que puder até que depois de um tempo ela ficará no fundo do vale localizado estritamente embaixo do ponto onde ela fez contato com o vale. A gota atingiu um “ponto fixo atrator”, um mínimo no seu caminho pela superfície. Os mínimos representados pelas coordenadas M_1 , M_2 , M_3 , Q_1 e Q_2 serão as **memórias** reconhecidas pelos estímulos. Todos os estímulos entre m_1 e m_2 reconhecerão a mesma memória M_1 . Este fato de todos eles reconhecerem a mesma memória é conhecido com o nome de **reconhecimento associativo**. O intervalo entre m_1 e m_2 é a **bacia atratora** da memória M_1 . Desta forma um estímulo imperfeito, ou com um certo número de erros, fluirá dinamicamente até a memória mais próxima. O

sistema associa muitos estímulos diferentes (configurações numa certa bacia atratora), com uma única memória, a que reflete mais de perto as características dos estímulos; fala-se de **memória associativa**. É simples de ver que uma rede neuronal deste tipo é capaz também de *corrigir erros*. Partindo de uma informação deficiente ou parcial de um certo conceito, o próprio processo dinâmico “purifica” o estímulo de partida, finalizando no conceito desejado. Uma outra questão interessante é que no processo de geração da superfície (por exemplo no processo de aprendizado que será discutido na próxima seção), em geral são gerados vales e colinas de diferentes profundidades e alturas, geralmente os vales mais fundos serão identificados com as memórias do sistema, no nosso exemplo os vales M_1 , M_2 e M_3 , e os restantes serão estados “indesejáveis” em geral, ou **estados espúrios**. Existem formas de eliminar os estados espúrios, por exemplo calibrando a temperatura num sistema estocástico ou o valor do ganho numa rede analógica[21]. No entanto no capítulo 4 será estudada a possível utilidade dos estados espúrios na capacidade de generalização de uma rede neuronal analógica. Esta metáfora da superfície não tem validade geral na descrição de um sistema dinâmico, inclusive no caso dele apresentar atratores. Ela é válida somente naqueles sistemas onde pode ser definida uma função das variáveis dinâmicas do mesmo, que decresce a cada passo da dinâmica. Esta função é a chamada *função de Liapunov* do sistema; em sistemas físicos, uma boa função de Liapunov é geralmente a energia potencial.

Os atratores podem ser visualizados efetivamente, seguindo a evolução temporal das **superposições** entre as configurações instantâneas da rede neuronal, e todos os padrões memorizados. As superposições são uma medida da distância entre configurações da rede neuronal. Um sistema que armazene p padrões poderá ser caracterizado mediante p parâmetros de superposição (ver próxima seção):

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu x_i(t) \quad (\mu = 1, \dots, p) \quad (1.8)$$

Estes parâmetros tomarão valores reais entre -1 e $+1$. Se o estado da rede está descorrelacionado completamente com um certo padrão, seu correspondente m^μ será nulo. Se a correlação for completa (estado da rede coincidente com um certo padrão), essa superposição particular será um. Em qualquer circunstância, quando o sistema atinge

um ponto fixo atrator, os valores das superposições ficam fixas.

1.4 Aprendizado em Redes Neurais

Vamos considerar o problema do aprendizado como um processo dinâmico da rede neuronal, mediante o qual é alterada sua estrutura de conexões por causa da interação da rede com estímulos externos. Uma discussão geral do problema do aprendizado [79,23] sai fora do marco da presente tese. Até agora temos considerado a matriz de sinapses, J_{ij} , como um simples parâmetro resistivo no funcionamento da rede neuronal. Na verdade ela joga um papel fundamental no processo de aprendizado da rede, especialmente no que diz respeito ao seu funcionamento como memória associativa, já que nas sinapses serão armazenadas as memórias que o sistema vai incorporando, por exemplo mediante aprendizado. Esta idéia foi sugerida num tratado do neurofisiologista Donald Hebb[24]:

“Quando um axônio de uma célula A está suficientemente perto para excitar uma célula B, e repetidamente ou de forma persistente participa no seu disparo, tem lugar algum processo de crescimento ou mudança metabólica em alguma ou ambas células, de forma tal que a eficácia de A, como uma das células que causam o disparo de B, é aumentada”.

Mediante este processo as memórias são armazenadas de uma forma distribuída na rede, pequenos pedaços de informação de cada memória são codificados em pequenas mudanças em muitas sinapses. É claro que este mecanismo é muito difícil de ser verificado experimentalmente, no entanto um número crescente de experiências em sistemas neuronais estão dando crédito à proposta de Hebb [95,26,27]. No capítulo 3 será feita uma discussão detalhada das possibilidades das chamadas *regras de aprendizado Hebbianas*. Agora vamos descrever brevemente alguns dos mais conhecidos algoritmos de aprendizado em redes neuronais, tais como a regra de Hopfield, a do perceptron, a pseudo-inversa e uma versão local dela:

1. Talvez a mais popular, por ser a mais simples, é a proposta feita originalmente por Hopfield [4], que, mesmo estando longe de ser uma regra realista, apresenta

uma serie de propriedades muito interessantes. Vamos discuti-la como protótipo em algum detalhe. O problema pode ser colocado da seguinte forma: queremos armazenar na rede neuronal um conjunto de p padrões , ou configurações estáveis da dinâmica neuronal. Os padrões serão memorizados no sentido de que cada uma das configurações da rede:

$$x_i = \xi_i^\mu , \quad i = 1 \dots N,$$

para cada um dos p padrões numerados pela letra μ , será um ponto fixo da dinâmica. Então ξ_i^μ representa o valor da atividade do neurônio i -ésimo na memória μ -ésima. No modelo proposto por Hopfield, os padrões são configurações aleatórias, $\xi_i^\mu = \pm 1$ com igual probabilidade. Uma matriz de sinapses que satisfaz as condições anteriores é :

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu , \quad J_{ii} = 0. \quad (1.9)$$

A diagonal igual a zero implica que nenhum neurônio forma sinapses com ele mesmo. O produto das atividades dos neurônios pré-sináptico j e pós-sináptico i nada mais é do que uma forma simplificada de realizar matematicamente a proposta de Hebb, apresentada acima. O comportamento do modelo como memória associativa depende da *capacidade de armazenamento* α , que se define como a razão do número de padrões armazenados pela conectividade efetiva por neurônio. No caso do modelo de Hopfield, $\alpha = p/N \approx 0.14$. Nos modelos se procura que o número de padrões a serem armazenados seja o maior possível, ou seja, maximizar o número de configurações estáveis (pontos fixos) da dinâmica, ou, tecnicamente, fazer α o maior possível. Algumas das características importantes da regra de Hopfield são as seguintes:

- A matriz de sinapses é *simétrica* e completamente conectada. Cada par de neurônios apresenta sinapses nas duas direções , então o diagrama de conectividade da rede apresenta $N(N-1)/2$ sinapses diferentes. A simetria assegura a existencia de uma função de Liapunov para a dinâmica. A relaxação destas

condições (simetria e completa conectividade), certamente não realistas, e suas conseqüências serão analisadas em detalhe nos capítulos 2 e 3.

- A regra é *local* num sentido topológico; ela depende somente das atividades dos neurônios pré-sináptico e pós-sináptico, embora os neurônios possam estar metricamente muito afastados, devido às interações de longo alcance. Regras não-locais tem sido estudadas, por exemplo a *pseudo-inversa*[29], que consegue armazenar padrões com uma correlação arbitrária, ao preço de perder a localidade que é uma característica desejável.
- A regra é incremental, cada nova memória a ser incorporada adiciona-se linearmente às anteriores, o que simula um processo de aprendizado.
- Todas as memórias tem a mesma hierarquia, não interessando a ordem na qual elas foram apresentadas à rede. No entanto, comportamentos muito interessantes relacionados com memórias de curto alcance surgem quando as memórias são pesadas diferentemente [32].
- Como $\xi_i^{\mu} = \pm 1$ com igual probabilidade, então aproximadamente a metade dos neurônios num padrão estão ativos. É conhecido experimentalmente, que a atividade de uma rede neuronal é geralmente muito inferior a 50%, quando reconhece um padrão. O problema relacionado à memórias com valores arbitrários da atividade será discutido no capítulo 3.
- Cada neurônio possui, em média, igual número de sinapses excitatórias (positivas) e inibitórias (negativas), emergindo do seu axônio. Existe um consenso geral no sentido de que neurônios típicos possuem só um tipo de sinapses emergentes dos seus axônios, é a chamada Lei de Dale [33,34].
- A regra é susceptível de modificações simples que dão lugar ao estudo de fenômenos novos e interessantes, como generalizações que levem em conta o caráter excitatório ou inibitório das sinapses [35], os efeitos de cortar sinapses no processo de aprendizado (ver capítulo 2), efeitos de considerar sinapses com valores discretos [41], saturação nos valores das sinapses [51,41,36] e muitos outros.

2. Nos algoritmos do tipo **perceptron** se inicializa a rede com sinapses aleatórias, e no processo de aprendizado só são modificadas aquelas sinapses que produzem erros no reconhecimento de um padrão . A estabilidade local dos padrões é garantida impondo-se a seguinte condição :

$$h_i \xi_i^\mu \geq T > 0; \quad i = 1, \dots, N; \quad \mu = 1, \dots, p. \quad (1.10)$$

Cada vez que a condição não é satisfeita para o sitio i -ésimo do padrão μ , as sinapses são atualizadas para todos os $j \neq i$ segundo a regra:

$$J_{ij}^\mu = J_{ij}^{\mu-1} + \frac{1}{N} \xi_i^\mu \xi_j^\mu \quad (1.11)$$

O parâmetro de estabilidade T está diretamente relacionado com as bacias atratoras das memórias, quanto maior for T maior será a bacia atratora do correspondente padrão [37]. O Teorema do Perceptron [38] assegura que o procedimento converge a um conjunto de acoplamentos que serão os melhores possíveis com as condições impostas (1.10). Para padrões sem correlação isto implica que é possível atingir uma capacidade crítica $\alpha_c = 2$. O algoritmo é também aplicável a padrões com correlação e no limite de correlação completa o sistema pode apresentar uma capacidade de armazenamento ilimitada ($\alpha_c = \infty$) [39]. Se bem o Teorema do Perceptron assegura convergencia do processo iterativo num número finito de passos, o número aumenta linearmente com N , o número de neurônios, o que implica uma convergencia lenta.

3. Também foram propostas regras de aprendizado não-locais, que são capazes de armazenar padrões com uma correlação arbitraria, e reconhecê-los sem erros [28, 29,30]. É a chamada regra de aprendizado pseudo-inversa ou do **projektor**. Nesta regra a matriz de sinapses J_{ij} é escolhida como sendo a solução do conjunto de equações :

$$\sum_{j=1}^N J_{ij} \xi_j^\mu = \lambda \xi_i^\mu; \quad i = 1, \dots, N; \quad \mu = 1, \dots, p, \quad (1.12)$$

sendo λ uma constante positiva. Uma possível solução para $\lambda = 1$ é :

$$J_{ij} = \frac{1}{N} \sum_{\mu, \nu=1}^p \xi_i^\mu \xi_j^\nu A_{\mu\nu}^{-1} \quad (1.13)$$

onde $A_{\mu\nu}^{-1}$ é a inversa da matriz de correlações entre os padrões :

$$A_{\mu\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu, \quad \mu, \nu = 1, \dots, p. \quad (1.14)$$

Esta solução requer padrões linearmente independentes o que limita o número de memórias armazenadas à N , o número de neurônios ($\alpha_c = 1$). Este número é muito maior do que o permitido pela regra de Hopfield, no entanto as bacias de atração dos padrões somem quando $p=N/2$. Claramente a matriz de correlações é a responsável pela não localidade da regra, fato este que a torna pouco plausível biologicamente, embora não seja impossível que uma sinapse dependa, direta ou indiretamente, da história de um conjunto de neurônios diferentes dos vizinhos imediatos.

4. Para evitar o problema da não-localidade da pseudo-inversa, foi proposta uma regra local que converge exponencialmente para a matriz do projetor [31], e não requer nenhuma condição sobre os padrões a serem aprendidos. Cada padrão é memorizado em um evento, que consiste em uma ou mais apresentações do padrão ao sistema. Um novo padrão ξ^μ é apresentado n_l vezes à rede reajustando a matriz de sinapses de acordo com a regra:

$$J_{ij}^\mu = J_{ij}^{\mu-1} + \sum_{m=1}^{n_l} (\Delta_m)_{ij} \quad (1.15)$$

com

$$(\Delta_m)_{ij} = \frac{k^{m-1}}{N} (\xi_i^\mu - h_i) (\xi_j^\mu - h_j) \quad (1.16)$$

Δ_m representa a mudança na matriz de sinapses na m -ésima apresentação, e o campo h_i é calculado depois que $m - 1$ mudanças foram feitas, ou seja:

$$h_i = \sum_{j=1}^N \left[J_{ij}^{\mu-1} + \sum_{q=1}^{m-1} (\Delta_q)_{ij} \right] \xi_j^\mu \quad (1.17)$$

O parâmetro k é chamado “coeficiente de memória” já que quanto maior for k menor será o número de passos necessários a fim de memorizar os padrões. Esta regra soma às vantagens da pseudo-inversa o fato de ser local, não impor restrições sobre a natureza dos padrões e ter convergência rápida se comparada com regras tipo perceptron.

Na presente tese consideraremos modificações da regra original proposta por Hopfield, adicionando alguns dos elementos discutidos previamente procurando tornar os modelos mais realistas do ponto de vista biológico.

Podemos concluir esta seção resumindo a filosofia do aprendizado no contexto da presente tese:

APRENDIZADO É UM PROCESSO NO QUAL A REDE NEURONAL AJUSTA DINAMICAMENTE SUAS EFICÁCIAS SINÁPTICAS COM O OBJETIVO DE ACOMODAR UM CERTO PADRÃO ξ^{μ} COMO UM ATRATOR.

Capítulo 2

Redes Neurais com Conectividade Restrita: o Papel da Diluição

2.1 Diluição em Redes Neurais

A simplificação de supor completa conectividade entre os N neurônios numa rede neuronal é certamente artificial e, em vários aspectos, até indesejável. Do ponto de vista da biologia, o número de sinapses por neurônio no córtex cerebral ($\approx 10^4$) é bem inferior ao número de neurônios ($\approx 10^{10}$). Vários processos podem ser responsáveis pela diminuição da conectividade: desde o nascimento até atingir um estágio maduro, uma certa fração das sinapses originais são cortadas, “desativadas”. O processo de morte de sinapses também pode acontecer devido a doenças ou acidentes cerebrais. Em uma série de interessantes trabalhos M.A.Virasoro[40] investigou as conseqüências de cortar

aleatoriamente uma certa fração das sinapses na capacidade dos modelos em reconhecer categorias, problema ligado às doenças conhecidas como agnosia e prosopagnosia.

No desenho de redes neuronais artificiais, a total conectividade das unidades (neurônios) é um problema de engenharia. Consegue-se uma grande simplificação se um sistema com conectividade restrita conserva as suas propriedades essenciais de reconhecimento. Sob a abordagem da física, o problema da diluição em redes neuronais apresenta diferentes aspectos:

A Estrutura de Conectividade da Rede : partindo de uma rede completamente conectada, a maioria dos modelos estudados consideram diluição *aleatória* das sinapses [41]-[44]. Neste caso, o conjunto de sinapses a serem eliminadas é escolhido aleatoriamente do total. Uma outra possibilidade é considerar estruturas *locais* na conectividade [45,46]. Por exemplo modelos definidos num espaço de dimensão finita, onde cada neurônio interage com um conjunto de vizinhos próximos. Esta possibilidade é muito interessante para o desenvolvimento de redes de neurônios artificiais, já que estruturas locais são de fabricação mais simples.

A Simetria das Sinapses : podem-se estudar os efeitos de uma diluição simétrica [41,45] ou assimétrica [43,47]. A escolha mais simples é diluir simultaneamente as sinapses J_{ij} e J_{ji} . Desta forma é preservada a simetria da matriz de eficácias sinápticas, e conseqüentemente, o modelo pode ser analisado com os métodos da mecânica estatística de equilíbrio. Já um modelo com motivação biológica deverá apresentar diluição assimétrica das sinapses, porém esta escolha implica na inexistência de uma função de Liapunov para o modelo, e a possibilidade de usar as ferramentas tradicionais da física no seu estudo veem-se severamente limitadas, como será discutido em detalhe na seção 2.3.

O Grau de Diluição : numa rede com N neurônios apresentam-se dois casos

1. Após o processo de diluição, o número de conexões por neurônio é igual a cN . c é uma constante que representa o grau de diluição e permanece finita quando $N \rightarrow \infty$. Neste caso o sistema permanece densamente conectado no limite termodinâmico [41,48].

2. No limite chamado de *diluição extrema* e representa o número médio de conexões por neurônio e quando $N \rightarrow \infty$, $c/N \rightarrow 0$, ou seja, o sistema é fortemente diluído [43,49].

A Estratégia de Diluição : talvez seja o ponto mais importante no que diz respeito à otimização da performance do modelo diluído como memória associativa. Num trabalho muito interessante, K.Kürten [50] compara a performance de modelos diluídos seguindo três estratégias diferentes: diluição aleatória, diluição das sinapses de menor valor absoluto, e diluição seletiva de sinapses. Simulações por computador mostram que a estratégia ótima consiste em escolher seletivamente aquelas sinapses que melhor estabilizam um conjunto de padrões já que, além de obter um sistema com estabilidade máxima, consegue-se uma grande economia na conectividade. Além disso, o método permite armazenar padrões com atividade não necessariamente 1/2 (ver capítulo 3).

2.2 Diluição Simétrica numa Rede Neuronal Analógica

Consideremos uma rede neuronal com N neurônios analógicos, com atividades $x_i(t)$, que evoluem no tempo de acordo com a dinâmica:

$$x_i(t+1) = \tanh(gh_i(t)), \quad i = 1, \dots, N. \quad (2.1)$$

h_i é o potencial pós-sináptico no neurônio i -ésimo, originado das atividades de outros neurônios j que interagem com o i -ésimo via as eficácias sinápticas J_{ij} . Se seus efeitos adicionam-se linearmente, então :

$$h_i(t) = \sum_{j=1}^N J_{ij} x_j(t), \quad j \neq i. \quad (2.2)$$

Finalmente, o parâmetro g é o ganho da função de transferência na origem e mede o grau de não-linearidade da mesma.

As eficácias sinápticas J_{ij} serão determinadas pelo aprendizado. Neste caso consideramos a regra Hebbiana:

$$J_{ij} = \frac{d_{ij}}{(1-d)N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu}, \quad J_{ii} = 0, \quad (2.3)$$

onde os ξ^μ 's formam um conjunto de p padrões estatisticamente independentes (i.e., sem correlação alguma) e $\xi_i^\mu = \pm 1$ com probabilidade $1/2$. O parâmetro d_{ij} será responsável pela diluição das eficácias sinápticas. Pode valer 0 ou 1 com a seguinte probabilidade:

$$P(d_{ij}) = d\delta(d_{ij}) + (1-d)\delta(d_{ij} - 1), \quad (2.4)$$

e $0 \leq d \leq 1$. Desta forma a conectividade média por neurônio é $(1-d)N$.

Vamos considerar o caso de *diluição simétrica*, ou seja, $d_{ij} = d_{ji}$. De acordo com esta prescrição, cada par de neurônios terá ambas conexões ou não terá nenhuma, modelo claramente não realista, mas que apresenta a vantagem de ser analiticamente solúvel. Na solução do modelo vamos nos aproveitar de uma analogia descoberta por H.Sompolinsky[41,51] entre um ruído gaussiano adicionado à matriz de sinapses, e o efeito da diluição simétrica.

2.2.1 Ruído Gaussiano nas Sinapses

É natural supor que, no processo de aprendizado, estejam presentes perturbações de diversas origens, que vão afetar a efetividade do mesmo. A presença de termos, na matriz de eficácias sinápticas, que não são efeitos puramente do aprendizado foi sugerida por G.Toulouse et al.[5] para dar conta do fato de que, no começo do processo de aprendizado, o cérebro já apresenta um diagrama de conexões interneuronais de origem provavelmente genética. A suposição de uma condição inicial "tabula rasa", como no modelo de Hopfield, é claramente não realista. Inspirados no desenvolvimento da teoria de vidros de spin e no modelo de Hopfield, esses autores sugerem que inicialmente a rede neuronal pode apresentar uma superfície de energia tipo vidro de spin, com uma estrutura complexa de vales de diferentes tamanhos e profundezas. No posterior processo de aprendizado, mediante estímulos externos, seriam selecionados certos vales que viriam a ser reforçados, aprofundados, e se convertiriam nos estados fundamentais, de mais baixa energia, do sistema.

Uma forma possível de introduzir estes efeitos é mediante um ruído gaussiano adicionado, por exemplo, à regra de Hebb:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu + \delta_{ij} \quad (2.5)$$

δ_{ij} é descorrelacionado dos padrões e tem uma distribuição gaussiana com valor médio e variância respectivamente:

$$\overline{\delta_{ij}} = 0 \quad \overline{\delta_{ij}^2} = \frac{\Delta^2}{N}. \quad (2.6)$$

Se Δ é da ordem de um, então a componente aleatória nos J_{ij} é da ordem de $N^{-1/2}$.

Mediante uma análise estatística dos potenciais pós-sinápticos h_i , similar à considerada por C.M.Marcus et al.[16] na análise do modelo de Hopfield com neurônios analógicos, podemos obter um conjunto de equações auto-consistentes para a superposição com um padrão escolhido ν : m^ν . No apêndice mostramos os detalhes da derivação das equações. Estas são:

$$\begin{aligned} m^\nu &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh(gH^\nu) \\ q &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh^2(gH^\nu) \\ R &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh'(gH^\nu) = g(1 - q) \end{aligned} \quad (2.7)$$

onde $\tanh'(x)$ é a derivada da função, e da análise (ver apêndice) surge que H^ν é uma variável aleatória com distribuição gaussiana:

$$P(H_i^\nu) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \left[\frac{(H_i^\nu - m^\nu)^2}{2\sigma^2} \right] \quad (2.8)$$

e

$$\sigma^2 = \frac{\alpha q}{(1 - R)^2} + \Delta^2 q \quad (2.9)$$

Os parâmetros q e R , conhecidos da teoria de campo médio do modelo de vidro de spin de Sherrington-Kirkpatrick [70], descrevem o congelamento de spins não correlacionados espacialmente e o ruído gaussiano efetivo, respectivamente [11]. Na próxima seção vamos estudar as soluções destas equações, explorando a analogia entre o ruído gaussiano Δ e o nível de conectividade $(1-d)N$ numa rede diluída simetricamente.

2.2.2 Analogia com uma Diluição simétrica nas sinapses: performance dos modelos

A regra de aprendizado 2.3 pode ser reescrita na forma:

$$J_{ij} = T_{ij} + \delta T_{ij} \quad (2.10)$$

$$\text{onde} \quad T_{ij} \equiv \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad (2.11)$$

$$\delta T_{ij} \equiv \left[\frac{d_{ij}}{(1-d)} - 1 \right] T_{ij} \quad (2.12)$$

H.Sompolinsky[41,51] demonstrou que, no limite $N \rightarrow \infty$, δT_{ij} é assintoticamente equivalente a um ruído gaussiano com média e variância:

$$\overline{\delta T_{ij}} = 0 \quad (2.13)$$

$$\overline{\delta T_{ij}^2} \approx \frac{\Delta^2}{N} \equiv \frac{\alpha d}{(1-d)N} \quad (2.14)$$

Na ref.[41], esta importante idéia foi aplicada ao estudo da termodinâmica de uma rede neuronal com neurônios discretos, que podem estar em um dentre dois estados possíveis, ativos ou inativos. Na presente tese, a analogia será aplicada ao estudo das propriedades de reconhecimento de uma rede *analógica diluída*, definida pelo sistema de equações (2.1). Fazendo uso do resultado obtido na seção anterior, reescrevemos as equações (2.7) para o caso $\Delta^2 = \alpha d/(1-d)$:

$$\begin{aligned} m^\nu &= \int_{-\infty}^{+\infty} Dz \tanh \left[g \left(m^\nu + z \sqrt{\frac{\alpha q}{(1-R)^2} + \frac{d\alpha q}{(1-d)}} \right) \right] \\ q &= \int_{-\infty}^{+\infty} Dz \tanh^2 \left[g \left(m^\nu + z \sqrt{\frac{\alpha q}{(1-R)^2} + \frac{d\alpha q}{(1-d)}} \right) \right] \\ R &= g(1-q) \\ Dz &= \frac{dz}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) \end{aligned} \quad (2.15)$$

Pode-se notar que os potenciais pós-sinápticos apresentam três contribuições :

1. Um termo proporcional a m^ν , ou *senal*, que favorece o reconhecimento do padrão ξ^ν .
2. O primeiro termo na raiz quadrada, que provem das superposições aleatórias com os restantes $p-1$ padrões .
3. O segundo termo na raiz quadrada é efeito da diluição aleatória das sinapses.

Os dois últimos são ruídos gaussianos que desestabilizam o processo de reconhecimento. Conseqüentemente, veremos que a capacidade crítica de armazenamento $\alpha_c = p/N$,

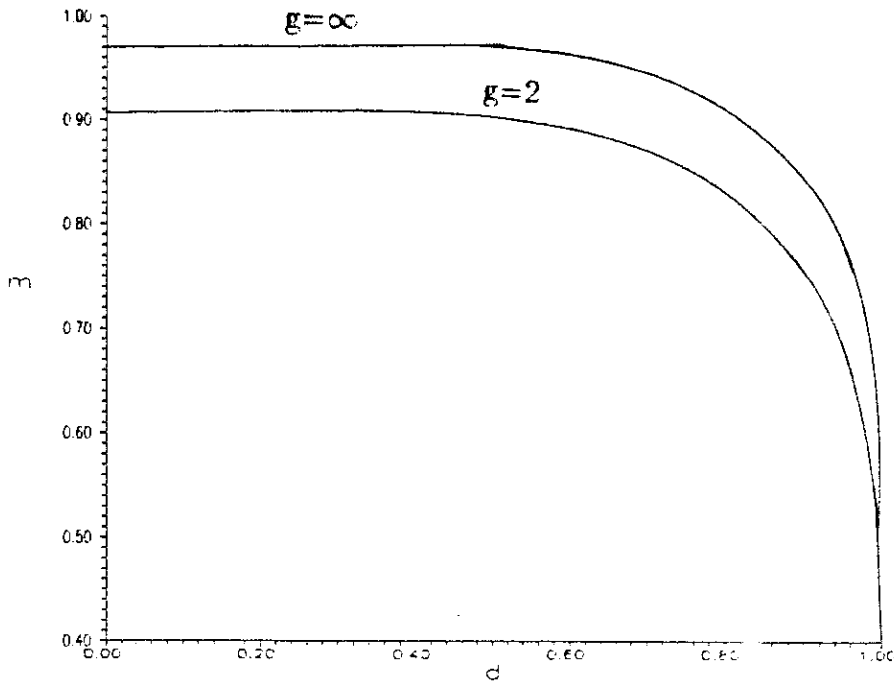


Figura 2.1: Superposição crítica vs. grau de diluição, numa rede neuronal analógica, para diferentes valores do ganho g .

diminui com valores crescentes da diluição d . No entanto, é interessante notar que, para um número finito de padrões ($\alpha = 0$), o ruído interno é zero, independente do nível de diluição, e o sistema reconhece sem erros.

As eqs.(2.16) para a superposição m^v do sistema definido pelos mapas acoplados (2.1), são as mesmas da solução de campo médio encontradas por H.Sompolinsky[41] para a termodinâmica de uma rede de neurónios de 2 estados, onde o parâmetro g em (2.16) corresponde à inversa da temperatura $\beta = 1/T$ na ref. [41]. Nas figs. 2.1 e 2.2 mostramos as soluções das eqs.(2.16). Pode-se observar que a performance do modelo é otimizada para valores de $g \rightarrow \infty$, ou seja, o limite no qual o estado do neurônio pode tomar só dois valores. Para $g \rightarrow \infty$, todos os padrões são reconhecidos com uma superposição maior que 90%, para níveis de diluição de até 80%. Neste caso cada neurônio permanece conectado só com 20% dos restantes neurónios. Com uma diluição de 40% o número de erros no reconhecimento é de 2.5%, e a capacidade de reconhecimento crítica α_c , cai de 0.14 para 0.1.

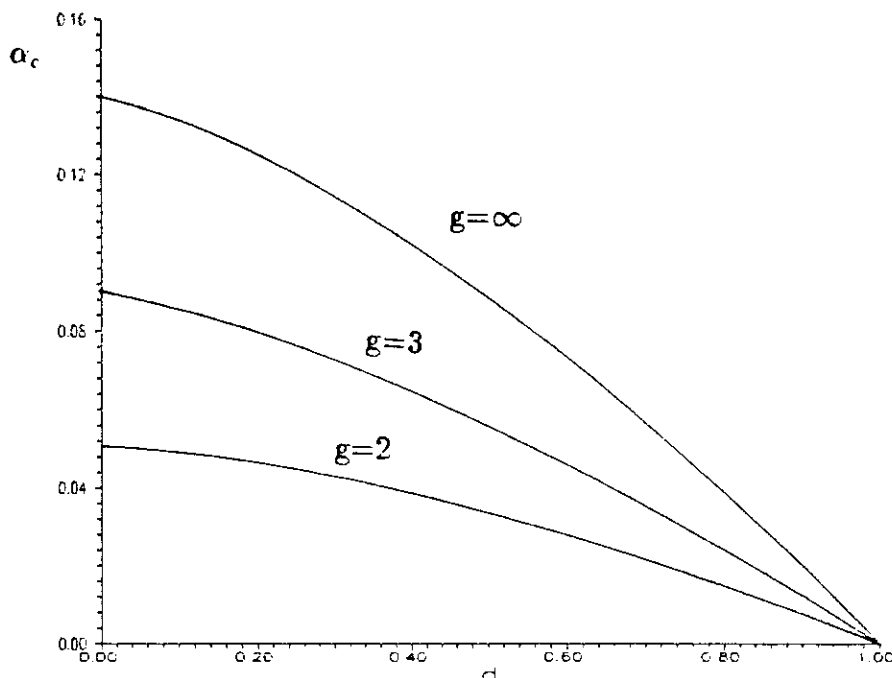
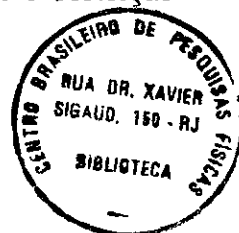


Figura 2.2: Capacidade de armazenamento crítica vs. grau de diluição, numa rede neuronal analógica, para diferentes valores do ganho g .

Para $g=2$ a rede diluída a 80% reconhece com uma superposição maior que 0.8. O α_c cai para 0.05 para níveis de diluição entre 0–40%. O nível de erros no reconhecimento é de 6.5% para diluição de 40%. A performance continua sendo muito boa.

2.3 Diluição Assimétrica

Depois de ter analisado em detalhe as implicações de limitar a conectividade numa rede neuronal diluindo simetricamente uma fração das sinapses, o passo seguinte procurando tornar o modelo mais realista é considerar *diluição assimétrica*. Isto é equivalente a considerar, no modelo introduzido na seção 2.2, que d_{ij} não necessariamente será igual a d_{ji} . No caso mais geral de diluição assimétrica aleatória, d_{ij} e d_{ji} serão variáveis aleatórias independentes, e então um par de neurónios i - j , poderá ter ambos canais J_{ij} e J_{ji} diferentes de zero [42], só um deles (J_{ij} ou J_{ji}) diferente de zero, ou ambos nulos, em cujo caso o par de neurónios estará completamente desconectado. Liberar a restrição



de simetria nas sinapses tem conseqüências muito interessantes, já que, além de tornar os modelos mais perto da realidade biológica (ver capítulo 3), a assimetria pode ser responsável por comportamentos dinâmicos muito complexos. O fato de ter $J_{ij} \neq J_{ji}$ implica na inexistência de uma **função energia** ou **função de Liapunov** para o modelo. No caso dos modelos estatísticos, a impossibilidade de definir uma energia implica que não é possível utilizar as técnicas da mecânica estatística de equilíbrio para resolver os modelos, como foram usadas na solução do modelo de Hopfield [12,13,14]. Então o caminho a seguir para estudar estes modelos é resolver explicitamente a dinâmica dos mesmos, empreendimento em geral muito difícil por causa das correlações existentes na dinâmica de um sistema de muitos corpos interagentes. Outra possibilidade é extrair informações úteis mediante simulações por computadores. Importantes resultados analíticos tem sido obtidos no limite de diluição extrema definido na seção 2.1. Mais especificamente, se $c \ll \ln N$, o conjunto de neurônios que, desde o tempo $t=0$, influenciam o estado de um neurônio qualquer $x_i(t)$ no tempo t , é diferente do conjunto que influencia o estado de qualquer outro neurônio $x_j(t)$. Em outras palavras: com uma alta probabilidade, a dinâmica dos neurônios da rede é descorrelacionada, e portanto acessível a uma solução analítica completa. Num trabalho clássico, B.Derrida, E.Gardner e A.Zippelius [43] estudaram o modelo de Hopfield com diluição assimétrica aleatória no limite de diluição extrema, para os casos de dinâmica seqüencial e paralela. A equação que define os atratores da dinâmica resulta ser a mesma para ambas dinâmicas. Nos limites $N \rightarrow \infty$, $p \rightarrow \infty$, $c \rightarrow \infty$, ($p/c=\alpha$) nessa ordem, com $c \ll \ln N$, a equação para a superposição com uma memória é :

$$m = \int_{-\infty}^{+\infty} Dz \tanh \left[\beta \left(z\sqrt{\alpha} + m \right) \right], \quad Dz = \frac{dz}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) \quad (2.16)$$

onde $\beta = 1/T_0$, $T_0 = T/c$.

Os principais resultados são os seguintes [43]:

- As transições entre as fases de reconhecimento ($m \neq 0$) e não reconhecimento ($m=0$) com α finito são contínuas para qualquer valor de β , diferentemente do que acontece com o modelo de Hopfield completamente conectado. Como conseqüência, a deterioração das propriedades de memória aumenta de forma contínua.

- Para $\beta = \infty$ a capacidade crítica de reconhecimento é :

$$\alpha_c = \frac{2}{\pi} \approx 0.637, \quad (2.17)$$

sensivelmente superior ao valor $\alpha_c = 0.138$ do modelo de Hopfield. Ainda neste limite extremo, a rede não perde suas propriedades essenciais de boa memória associativa.

Em outro trabalho nesta linha, foi estudada uma rede neuronal fortemente diluída com interações multi-neurônio, ou seja, com eficácias sinápticas efetivas entre um conjunto de R neurônios [53,54]. Novamente os resultados são melhores do que aqueles obtidos no caso de redes com conectividade completa.

Para níveis gerais de diluição, resultados analíticos tem sido obtidos [48] usando idéias introduzidas por E.Gardner. Na seção seguinte serão apresentados resultados obtidos mediante simulações numéricas, numa rede neuronal analógica com diluição aleatória [55].

2.3.1 Simulando uma Rede Neuronal Analógica

Tendo como objetivo investigar os efeitos de diluir assimetricamente uma rede neuronal, mantendo a conectividade por neurônio finita, simulamos no computador uma rede neuronal definida pelas eqs. (2.1) e (2.2) [55]. A regra de aprendizado é similar à (2.3), exceto na normalização, por razões históricas. Nas simulações escolhemos:

$$J_{ij} = \frac{d_{ij}}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} \quad (2.18)$$

Diferentemente do caso considerado na Seção 2.2, agora o parâmetro d_{ij} é independente de d_{ji} . Novamente o parâmetro de diluição é escolhido de acordo com a distribuição de probabilidades (2.4). Claramente, a fonte de assimetria nas sinapses é a independência nos valores de d_{ij} e d_{ji} , ou seja, diferentemente do que acontece nos modelos que consideraremos no capítulo 3, onde a assimetria é consequência da própria regra de aprendizado, no presente modelo ela é causada pela diluição das sinapses.

Os principais resultados das simulações foram obtidos com redes de 480 neurônios, embora também foram realizadas simulações em redes de 50 e 100 neurônios, para estudar os efeitos do tamanho finito da rede. Os diagramas de fases que serão apresentados

nas seções seguintes foram obtidos com condições iniciais para os estados dos neurônios nos vértices de um hipercubo $x_i(0) = \pm 1, i = 1 \dots N$. Para eliminar efeitos de transientes foram descartadas as primeiras 500 iterações. Os valores finais das grandezas calculadas correspondem a uma média sobre 100 diferentes realizações da matriz de sinapses J_{ij} . A identificação das fases foi feita da seguinte forma: quando o sistema atingia um ponto fixo na sua dinâmica calculávamos a superposição entre o estado assintótico da rede e os p padrões ξ_i^μ :

$$m^\mu = \frac{1}{N} \sum_{i=1}^N x_i \xi_i^\mu \quad (2.19)$$

Foram considerados três casos:

1. Se $m^\mu < 10^{-5}$ o estado foi considerado *origem*, ou seja, todos os neurônios no estado $x_i = 0$.
2. Se $m^\mu \approx O(1/\sqrt{N})$ o estado foi considerado *vidro de spin*.
3. Se $m^\mu \approx O(1)$ o estado foi considerado de *reconhecimento*.

No caso de ciclos limite ou caos m^μ não está definida. Nos estados origem, vidro de spin, e reconhecimento, e longe de fronteiras críticas, o sistema converge para um atrator no máximo em 20 iterações. Antes de descrever os resultados das simulações é bom lembrar que a dinâmica da rede é paralela, ou seja, todos os neurônios são atualizados simultaneamente.

2.3.1.1 Diagramas de Fases

O espaço de parâmetros do modelo é tridimensional; os três parâmetros livres são: o nível de diluição d , a razão de armazenamento $\alpha = p/N$, e o ganho da função de transferência na origem g . Nós obtivemos cortes bidimensionais para dois valores particulares do parâmetro α . Para valores pequenos do parâmetro $\alpha = 0.01$, o diagrama diluição vs. ganho (fig.2.3) apresenta somente duas fases: se g é suficientemente pequeno a rede atinge um atrator no qual $x_i = 0, i = 1 \dots N$, ou seja, a origem é um ponto fixo estável. Ao aumentarmos g , chegamos a um valor crítico acima do qual o estado estável

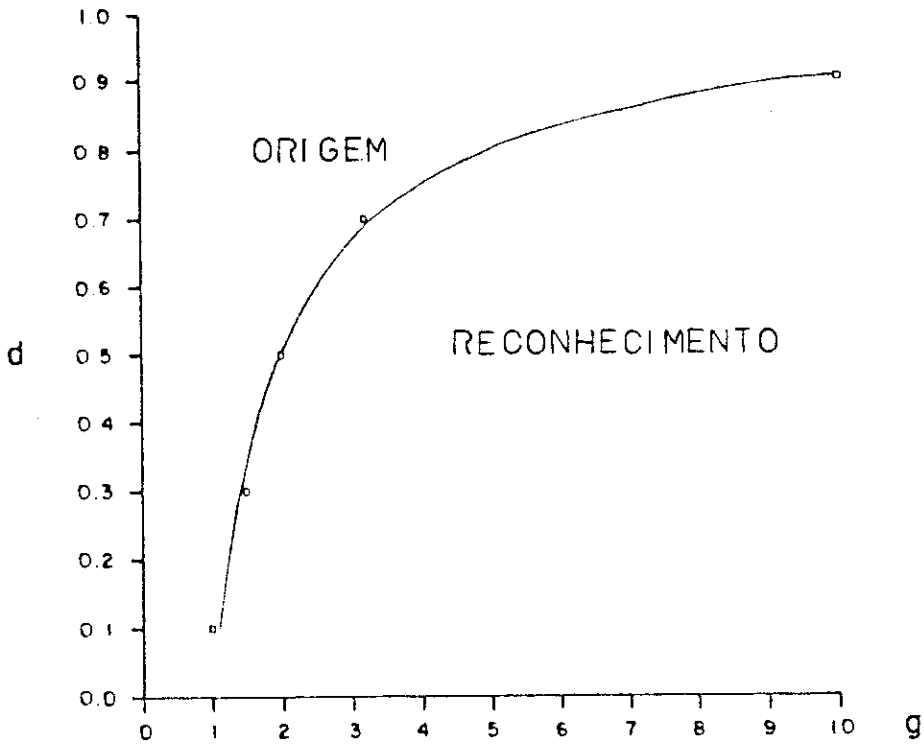


Figura 2.3: Diagrama de fases *diluição vs. ganho* para uma rede de neurônios analógica com $\alpha=0.01$. Os quadrados são o resultados de simulações numa rede com 480 neurônios e a linha cheia é uma aproximação teórica para o limite $N \rightarrow \infty$ e $\alpha \rightarrow 0$ (ver texto).

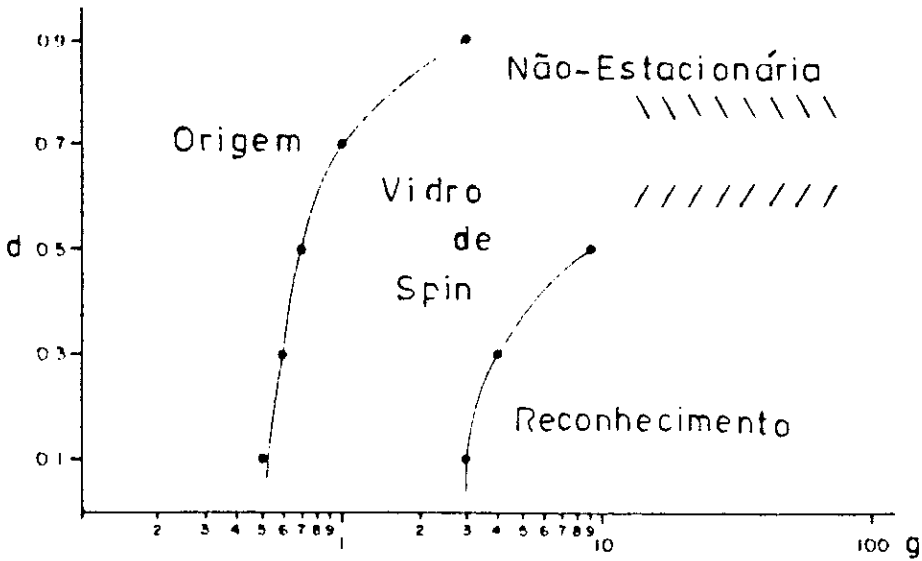


Figura 2.4: Diagrama de fases *diluição* vs. *ganho* para uma rede de neurónios analógica com $\alpha = 0.06$. Os quadrados são resultados de simulações numa rede de 480 neurónios. As regiões com linhas curtas assinalam as possíveis fronteiras entre as fases não-estacionária, vidro de spin e reconhecimento. As linhas cheias são uma guia para o olho.

é um ponto fixo de reconhecimento, ou seja, a dinâmica da rede converge a um estado que apresenta uma superposição macroscópica com um dos padrões armazenados. Na mesma figura observa-se que conforme aumenta o nível de diluição, a origem apresenta um domínio de estabilidade maior, isto é, o reconhecimento é desestabilizado. Neste sentido pode-se dizer que a rede reconhece melhor quanto maior for o valor do ganho, independentemente da diluição, efeito já presente no caso de diluição simétrica da seção 2.2.2.

Este comportamento pode ser aproximado teoricamente fazendo uma suposição simples: assumindo que quando $\alpha \rightarrow 0$, o efeito da diluição é simplesmente uma renormalização da matriz de sinapses,

$$J_{ij} = \frac{(1-d)}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu}. \quad (2.20)$$

Fazendo uma análise de estabilidade linear no sistema de mapas acoplados (2.1), encontramos que a fronteira onde a fase origem perde estabilidade é determinada pela

condição :

$$g(1 - d)\lambda_{max} = 1, \quad (2.21)$$

onde

$$(1 - d)\lambda_{max} = (1 - d)(1 + 2\sqrt{\alpha}) \quad (2.22)$$

é o autovalor de maior valor absoluto da matriz (2.20). Quando $\alpha \rightarrow 0$ podemos definir uma curva para o ganho crítico:

$$g_c \approx \frac{1}{1 - d}(1 - 2\sqrt{\alpha}) \quad (2.23)$$

Este resultado é uma ótima aproximação para os resultados das simulações , como pode-se conferir na fig. 2.3. Para $\alpha = 0.06$ o diagrama *diluição vs.ganho* apresenta quatro fases, como podemos observar na fig. 2.4: para níveis de diluição entre 0.1 e 0.5 o sistema apresenta uma fase vidro de spin, entre as fases origem e reconhecimento, as únicas presentes para $\alpha = 0.01$. Num estudo do número de estados metaestáveis de uma rede de neurônios com diluição assimétrica, A.Treves e D.Amit [47] acharam uma banda grande de estados com superposição quase nula com os padrões . Este resultado similar ao nosso está , no entanto, em contraposição com outros trabalhos, que sugerem que uma quantidade arbitrariamente pequena de assimetria seria suficiente para suprimir os estados vidro de spin [56]. As regiões com traços pequenos assinalam que para $d = 0.7$ a rede não reconhece pois só estão presentes as fases origem e vidro de spin. Para $d = 0.9$ não existem soluções de reconhecimento nem vidro de spin. Para um valor do ganho $g \approx 3$, a origem perde estabilidade, e apresenta-se uma nova fase que chamamos *não-estacionária* pelo fato de o sistema não atingir um atrator periódico em 500 iterações . As simulações foram feitas para g no intervalo $[0.1,100]$ e passos $\Delta g = 0.5$. Com esta precisão não foi encontrada evidencia de fases vidro de spin ou não-estacionária para $\alpha = 0.01$.

2.3.1.2 Fases de Reconhecimento

Para estudar as propriedades de reconhecimento do modelo, foram feitas simulações com uma condição inicial (estímulo) coincidente com um dos padrões . Na fig. 2.5 pode-se observar um diagrama *superposição vs.ganho*, ou seja, a variação da superposição

entre o estado assintótico da rede e um dos padrões armazenados, em função da não-linearidade da função de transferência na origem, para $\alpha = 0.06$. Até diluições de 50% a rede reconhece com um valor de $m \geq 0.9$ se $g > 50$. A performance do modelo para valores grandes do ganho é similar a do modelo com diluição simétrica da Seção 2.2. Para valores pequenos do ganho a comparação é difícil, já que o α_c diminui e as simulações correspondem a um valor fixo $\alpha = 0.06$. Pode-se notar que ao aumentar a diluição, a fase de reconhecimento aparece para valores cada vez maiores de g , e já para $d=0.7$ o sistema não reconhece. Neste sentido o modelo simétrico é melhor que o presente, porém é importante assinalar que a perda das propriedades de reconhecimento no modelo assimétrico vem acompanhada do aparecimento de uma fase não-estacionária, que pode ter interessantes consequências numa análise mais abrangente das funções cerebrais que podem ser consideradas com os modelos físicos.

Nossos resultados para valores grandes de g são consistentes com simulações apresentadas por D.Amit [57], para um modelo de Hopfield com diluição assimétrica. No seu modelo, um nível de diluição de 100% corresponde a conexões unidirecionais, e a capacidade de armazenamento crítica diminui com a diluição. Também proliferam trajetórias (definidas pelo conjunto de mapas acoplados (2.1)) não periódicas conforme aumenta d . A comparação entre ambos modelos tem que ser feita com cuidado, levando em conta a diferente forma na qual a diluição é incorporada, assim como o fato da dinâmica do modelo de Hopfield ser sequencial, e a do nosso paralela. Na fig. 2.6 são analisados os possíveis efeitos de tamanho finito no modelo para redes com 50, 100 e 480 neurônios, observando a variação da superposição em função da diluição, para um valor fixo do ganho $g=10$. Exceto por uma importante flutuação no valor da superposição para $d=0.5$ e $N=480$, não observamos uma forte dependência dos resultados com o tamanho da rede. É interessante notar que para uma diluição de 90%, foram obtidos pequenos valores da superposição ($m \approx 0.15$) para redes com $N=50$ e $N=100$. Já para $N=480$ o sistema não converge para um atrator periódico como foi discutido na seção anterior.

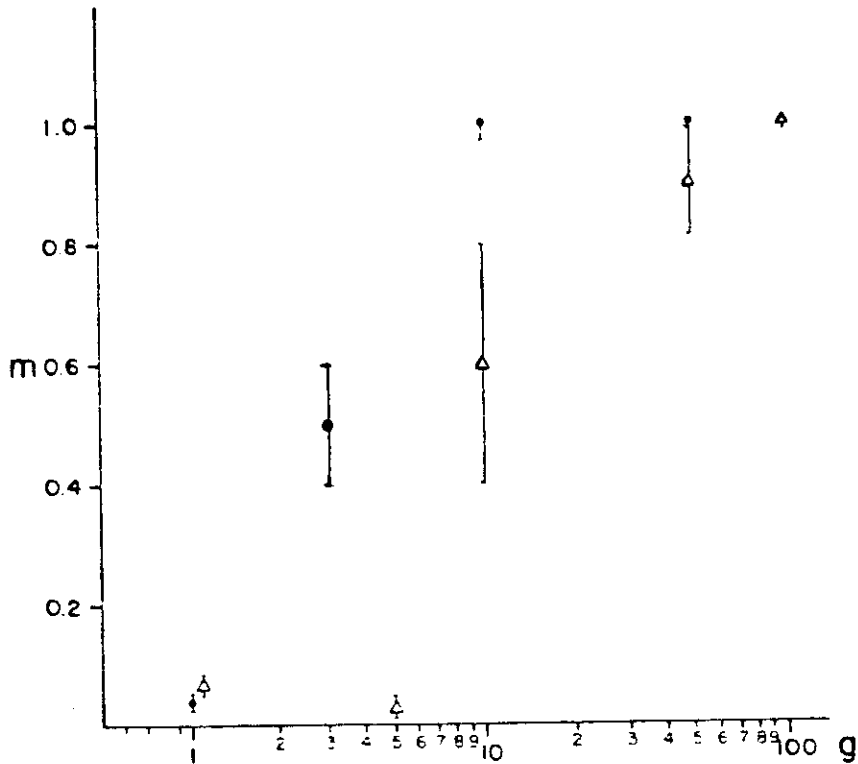


Figura 2.5: *Superposição vs.ganho* para uma rede de neurônios analógica com $\alpha = 0.06$. Os círculos são resultado de simulações numa rede com 480 neurônios, e um nível de diluição de 10%. Os triângulos correspondem a uma diluição de 50%. As barras correspondem à dispersão estatística dos resultados.

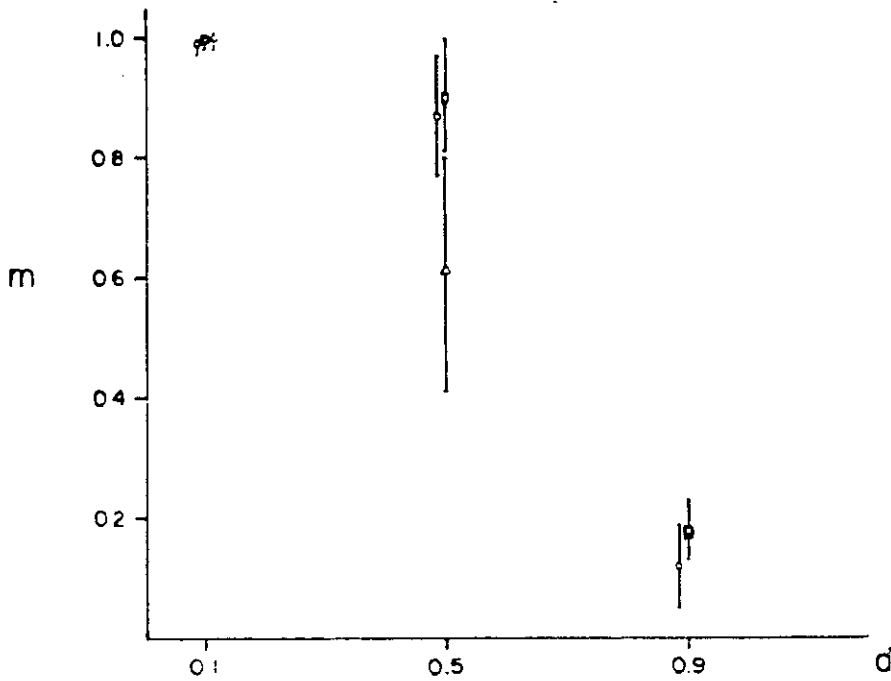


Figura 2.6: Superposição vs. diluição numa rede de neurônios analógica para $\alpha = 0.06$ e três tamanhos da rede: $N=50$ (quadrados), $N=100$ (círculos) e $N=480$ (triângulos), e um valor fixo do ganho $g=10$.

2.3.1.3 Caos em Redes Neurais

O modelo estudado nas seções anteriores apresenta, além das fases usuais de reconhecimento e vidro de spin, uma nova fase que chamamos de não-estacionária, e que apresenta uma dinâmica caótica para níveis de diluição grandes. Como foi indicado na seção 2.3, como consequência da assimetria da matriz de sinapses, não é possível definir uma função de Liapunov para o modelo. Nas simulações achamos que para sistemas com 50 e 100 neurônios interagentes existe um regime intermediário, separando as fases estacionária (pontos fixos) da caótica. Ou seja, mudando o valor do parâmetro g o sistema apresenta diferentes atratores: pontos fixos, ciclos limites e finalmente caos. Os ciclos limites tornam-se progressivamente mais complexos com o aumento da não-linearidade g . No entanto, a rede com 480 neurônios apresenta uma transição direta de uma fase com pontos fixos a outra caótica, sem passar por ciclos limites. Este comportamento já foi observado em outros trabalhos[59], e sugere que, com o aumento da dimensionalidade do sistema, a janela de periodicidade é cada vez mais estreita, até colapsar acima de

uma certa dimensionalidade crítica (que pode ser infinita).

Caos em redes neuronais analógicas foi encontrado num modelo com dinâmica em tempo contínuo[58], nos limites $N \rightarrow \infty$, e diluição extrema discutido na seção 2.1. Nesse trabalho, utilizando uma técnica de campo médio dinâmico[59], foi calculada a função de correlação temporal dos potenciais pós-sinápticos, e existe um regime no qual a correlação vai para zero, indicando a presença de uma fase caótica. De fato, no limite $N \rightarrow \infty$, o sistema apresenta uma transição de um estado estacionário (para valores pequenos de g) para um estado caótico (para valores grandes de g), quando varia o grau de não-linearidade g no sistema. Surpreendentemente, no limite $g \rightarrow \infty$, a capacidade crítica de armazenamento não concorda com o valor encontrado para um sistema de neurônios discretos por Derrida, Gardner e Zippelius (eq. 2.17). O valor para o limite no sistema analógico é $\alpha_c = 0.87$ [58], um pouco maior do $\alpha_c = 0.637$ da ref.[43]. Sugere-se [58] que a diferença é devida às diferentes dinâmicas, mesmo no limite $g \rightarrow \infty$.

Na filosofia de considerar as redes neuronais como sistemas computacionais, é indesejável a presença de dinâmicas caóticas nas mesmas. Porém, no que diz respeito aos neurônios biológicos, é um fato experimental a presença de respostas caóticas a estímulos periódicos. Por exemplo, foi achado experimentalmente em axônios de lulas gigantes, e teoricamente nas equações de Hodgkin-Huxley [61], que a resposta de membranas de neurônios em estado de repouso a excitações periódicas nem sempre são periódicas, e as respostas aparentemente não-periódicas podem ser entendidas como caos determinístico (ver ref. [60] e referências nesse trabalho). Estudos levando em consideração propriedades biológicas dos neurônios, como respostas analógicas, períodos refratários relativos e soma espaço-temporal dos estímulos [60], sugerem que o caos determinístico pode ter um papel funcional no comportamento de redes neuronais biológicas.

2.4 Conclusões

No presente capítulo foi considerado em detalhe o efeito causado, nas propriedades de reconhecimento de uma rede neuronal analógica, pela diluição de uma certa fração das conexões entre os neurônios da mesma. Numa primeira parte considerou-se o efeito da

diluição simétrica de sinapses. Esta aproximação biologicamente artificial nos permitiu fazer um estudo analítico do modelo. Comprovamos que até níveis de diluição muito elevados, a rede não sofre uma grande degradação em sua performance como memória associativa. Comparada com a rede completamente conectada, a rede diluída apresenta uma capacidade de armazenamento de padrões um pouco menor, porém isto é compensado pela grande economia no número de conexões. Este fato é muito importante na hora de fabricar redes neuronais artificiais, que apresentam o problema da limitação do espaço físico no qual a rede deve ser fabricada. Do ponto de vista funcional, também é importante que a rede continue a reconhecer satisfatoriamente mesmo sofrendo cortes aleatórios em algumas das sinapses.

Com o objetivo de tornar o modelo mais realista do ponto de vista biológico, estudou-se numa segunda parte o efeito de uma *diluição assimétrica* das sinapses. Também neste caso comprovamos que a rede apresenta uma performance muito boa, para níveis de diluição razoavelmente grandes. A grande novidade da assimetria em redes analógicas é a presença de caos na dinâmica dos neurônios. Ainda não é claro se a dinâmica caótica pode ser de importância nas funções computacionais das redes neuronais, embora exista evidência experimental de dinâmicas caóticas em neurônios biológicos.

2.5 Apêndice: análise estatística da rede analógica com ruído gaussiano nas sinapses

De acordo com (2.1), e como $\xi_i^\nu = \pm 1$, a superposição

$$m^\nu = \frac{1}{N} \sum_{i=1}^N \xi_i^\nu x_i \quad (2.24)$$

pode ser escrita:

$$m^\nu = \frac{1}{N} \sum_{i=1}^N \tanh(g \xi_i^\nu h_i) \quad (2.25)$$

ou

$$m^\nu = \overline{\tanh(g H_i^\nu)}, \quad (2.26)$$

onde $H_i^\nu = \xi_i^\nu h_i$. A barra superior indica uma média sobre o índice i , $\bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$. No limite $N \rightarrow \infty$ esta média pode ser escrita como uma integral na distribuição de

probabilidades do H^ν :

$$m^\nu = \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh(gH^\nu) \quad (2.27)$$

De acordo com (2.2) e (2.5), o potencial pós-sináptico no neurônio i -ésimo é

$$h_i = \frac{1}{N} \sum_{j=1}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu x_j + \sum_{j=1}^N \delta_{ij} x_j \quad (2.28)$$

Vamos considerar soluções de reconhecimento da forma $m^\nu \approx O(1)$, $m^\mu \approx O(N^{-1/2})$ se $\mu \neq \nu$.

$$H_i^\nu = m^\nu + \xi_i^\nu \sum_{\mu \neq \nu}^p \xi_i^\mu m^\mu + \xi_i^\nu \sum_{j=1}^N \delta_{ij} x_j \quad (2.29)$$

Devido a independência dos ξ_i^μ 's entre eles, e dos ξ_i^μ 's com os δ_{ij} 's, no limite $N \rightarrow \infty$, $p \rightarrow \infty$, $p/N = \alpha$, a variável H_i^ν terá uma distribuição gaussiana, com valor médio igual à soma dos valores médios dos três termos. Temos que tomar a média na distribuição dos ξ_i^μ 's, e na distribuição dos δ_{ij} 's. Só o primeiro termo tem valor médio não nulo, então

$$\overline{H_i^\nu} = m^\nu \quad (2.30)$$

A variância será a soma das variâncias. A variância do segundo termo é

$$\begin{aligned} \sigma_2^2 &= \ll \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \sum_{\mu \neq \nu}^p \sum_{\rho \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_i^\rho \xi_k^\rho x_j x_k \gg \\ &= \sum_{\mu \neq \nu}^p \sum_{\rho \neq \nu}^p m^\mu m^\rho \ll \xi_i^\mu \xi_i^\rho \gg, \end{aligned} \quad (2.31)$$

onde $\ll \dots \gg$ representa uma média com respeito à distribuição dos ξ_i^μ 's. Só os termos com $\mu = \rho$ terão valor médio não nulo, então

$$\sigma_2^2 = \sum_{\mu \neq \nu}^p (m^\mu)^2 \quad (2.32)$$

Para calcular m^μ levamos em conta que:

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \tanh(gh_i) \quad (2.33)$$

Se A_i e B_i são variáveis independentes, então o quadrado da média do produto é igual à média do produto dos quadrados:

$$\left(\frac{1}{N} \sum_{i=1}^N A_i B_i \right)^2 = \frac{1}{N} \sum_{i=1}^N A_i^2 B_i^2 \quad (2.34)$$

No nosso caso, ξ_i^μ está correlacionada com $\tanh(gh_i)$ pelo termo que contém ξ_i^μ no potencial pós-sináptico. Então temos que analisar este termo separadamente:

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \tanh \left[g \left(\xi_i^\mu m^\mu + \sum_{\rho \neq \mu}^p \xi_i^\rho m^\rho \right) \right] \quad (2.35)$$

Como $m^\mu \approx O(N^{-1/2})$ é possível expandir a \tanh em potências de $\xi_i^\mu m^\mu$. Até primeira ordem em m^μ :

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \left\{ \tanh \left(g \sum_{\rho \neq \mu}^p \xi_i^\rho m^\rho \right) + \xi_i^\mu m^\mu \tanh' \left(g \sum_{\rho \neq \mu}^p \xi_i^\rho m^\rho \right) \right\} \quad (2.36)$$

Definimos:

$$R \equiv \frac{1}{N} \sum_{i=1}^N \tanh'(gh_i) = \overline{\tanh'(gh_i)} \quad (2.37)$$

A soma nos argumentos da \tanh e da \tanh' , na eq.(2.36), difere de h_i num termo de $O(N^{-1/2})$, então vamos considerá-la como o próprio h_i .

$$m^\mu (1 - R) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \tanh \left(g \sum_{\rho \neq \mu}^p \xi_i^\rho m^\rho \right) \quad (2.38)$$

Agora no lado direito da expressão anterior, temos uma soma de um produto de variáveis independentes,

$$(m^\mu)^2 (1 - R)^2 = \frac{1}{N^2} \sum_{i=1}^N \tanh^2(gh_i) \quad (2.39)$$

Definimos:

$$q = \frac{1}{N} \sum_{i=1}^N \tanh^2(gh_i) = \overline{\tanh^2(gh_i)} \quad (2.40)$$

Finalmente, se $p \gg 1$:

$$\sigma_2^2 = \frac{\alpha q}{(1 - R)^2} \quad (2.41)$$

O terceiro termo de H_i^p tem uma variância:

$$\sigma_3^2 = \left[\sum_{j=1}^N \sum_{k=1}^N \delta_{ij} \delta_{ik} x_j x_k \right] \quad (2.42)$$

Como δ_{ij} é independente de δ_{ik} , então $\overline{\delta_{ij} \delta_{ik}} = 0$, se $j \neq k$. $[\dots]$ representa uma média na distribuição das variáveis δ_{ij} .

$$\sigma_3^2 = \left[\sum_{j=1}^N (\delta_{ij} x_j)^2 \right] \quad (2.43)$$

Fazemos uma análise similar a do termo anterior:

$$\begin{aligned}
\delta_{ij} x_j &= \delta_{ij} \tanh \left[g \left(\frac{1}{N} \sum_{k=1}^N \sum_{\mu=1}^p \xi_j^\mu \xi_k^\mu x_k + \sum_{k=1}^N \delta_{jk} x_k + \delta_{ji} x_i \right) \right] \\
&= \delta_{ij} \left\{ \tanh \left[g \left(\frac{1}{N} \sum_{k=1}^N \sum_{\mu=1}^p \xi_j^\mu \xi_k^\mu x_k + \sum_{k=1}^N \delta_{jk} x_k \right) \right] \right. \\
&\quad \left. + \delta_{ji} x_i \tanh' \left[g \left(\frac{1}{N} \sum_{k=1}^N \sum_{\mu=1}^p \xi_j^\mu \xi_k^\mu x_k + \sum_{k=1}^N \delta_{jk} x_k \right) \right] \right. \\
&\quad \left. + O(\delta_{ji}^2 x_i^2) \right\}
\end{aligned} \tag{2.44}$$

Como $\delta_{ij} = \delta_{ji}$, em segunda ordem em δ_{ij} temos:

$$(\delta_{ij} x_j)^2 = \delta_{ij}^2 \tanh^2(gh_j) \tag{2.45}$$

$$\sigma_3^2 = \sum_{j=1}^N [\delta_{ij}^2] \tanh^2(gh_j) = \Delta^2 q \tag{2.46}$$

A variância do H_i^ν será :

$$\sigma^2 = \frac{\alpha q}{(1-R)^2} + \Delta^2 q \tag{2.47}$$

Então H_i^ν terá uma distribuição de probabilidades:

$$P(H_i^\nu) = \frac{1}{\sqrt{2\pi\sigma}} \exp - \left[\frac{(H_i^\nu - m^\nu)^2}{2\sigma^2} \right] \tag{2.48}$$

As equações auto-consistentes para a superposição com a memória ν serão :

$$\begin{aligned}
m^\nu &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh(gH^\nu) \\
q &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh^2(gH^\nu) \\
R &= \int_{-\infty}^{+\infty} dH^\nu P(H^\nu) \tanh'(gH^\nu) = g(1-q)
\end{aligned} \tag{2.49}$$

Capítulo 3

Padrões com Atividade e Regras Hebbianas com Assimetria

3.1 Motivações

Nos capítulos prévios têm sido considerados modelos de memórias associativas nos quais os padrões apresentavam, em média, o mesmo número de neurônios ativos e inativos. Isto quer dizer que quando a rede atingia um ponto fixo atrator, 50% dos neurônios ficavam no valor +1 e 50% ficavam no -1, como consequência de escolher $\xi_i^\mu = \pm 1$ com probabilidade 1/2. No entanto, é conhecido experimentalmente que a atividade média espacial no córtex é relativamente baixa. Num ponto fixo atrator, a percentagem de neurônios que ficam ativos ($\xi_i^\mu = +1$) é aproximadamente 4-5% [62]. Este fato faz necessário modificar os modelos previos, baseados em igual percentagem de +1's e -1's,

se quisermos nos aproximar da realidade biológica. Mesmo pensando em problemas como reconstrução de imagens, por exemplo, é comum ter um fundo uniforme (-1 's na nossa codificação), e sobre ele uma imagem que carrega a informação útil que nos interessa. O interesse por considerar padrões com atividade também provem da necessidade de armazenar memórias correlacionadas, ou inclusive estruturadas em árvores, que permitam ao sistema formar *categorias* como será estudado no capítulo 4. O baixo nível de atividade encontrado no cérebro reforçou a procura de modelos que tivessem uma grande capacidade de armazenamento nesses níveis baixos. Em particular, num importante trabalho, E.Gardner [39] calculou um limite teórico para a capacidade de armazenamento de uma rede neuronal com atividade arbitrária. Supreendentemente, ela obteve que a capacidade pode aumentar indefinidamente no limite de baixa atividade:

$$\alpha_c \equiv \frac{1}{(1 - |a|)(\ln(1 - |a|))} \quad (3.1)$$

(a mede a atividade da rede e será definida com precisão na próxima seção. O limite de baixa atividade corresponde a $|a| \rightarrow 1$). O resultado é independente de qualquer regra de aprendizado particular; então a partir daí o problema foi encontrar regras de aprendizado que aceitassem padrões com atividade e tivessem, no caso ideal, a performance ótima encontrada por Gardner. Uma candidata natural é, certamente, a pseudo-inversa, que pode armazenar até N padrões com correlação arbitrária, desde que sejam linearmente independentes, sendo N o número de neurônios, como se viu no capítulo 1. Embora a pseudo-inversa resolva o problema da correlação, ela tem uma capacidade crítica $\alpha_c = 1$, bem maior do que o modelo de Hopfield ($\alpha_c \approx 0.138$), mas ainda muito longe do possível ∞ do cálculo de Gardner. Outra possibilidade explorada com maior sucesso foi voltar às regras locais, e fazer modificações na regra de Hopfield. Na próxima seção vamos ver o que acontece com a regra de Hopfield quando é quebrada a condição de independência estatística entre os padrões, e alguns caminhos possíveis para resolver as dificuldades que se apresentam.

3.2 Padrões com Atividade

A forma mais simples de considerar padrões com atividade é mantendo sua independência estatística, porém pesando as configurações com um maior número de +1's do que -1's ou vice-versa. Na linguagem dos sistemas magnéticos, isto é equivalente a considerar configurações com uma *magnetização* fixa. Especificamente, podemos escolher os padrões com a seguinte distribuição de probabilidades:

$$P(\xi_i^\mu) = \frac{1+a}{2} \delta(\xi_i^\mu - 1) + \frac{1-a}{2} \delta(\xi_i^\mu + 1) \quad (3.2)$$

onde $-1 \leq a \leq 1$. A *atividade média* num padrão é $(1+a)/2$. Por exemplo, o caso $a=0$ corresponde ao modelo de Hopfield (50% de atividade media). No caso $a=1$ todos os padrões são idênticos com todos os neurônios no valor +1. A média para o estado de um neurônio num padrão qualquer será :

$$\langle\langle \xi_i^\mu \rangle\rangle = a \quad (3.3)$$

e a superposição entre dois padrões quaisquer será :

$$\langle\langle \xi_i^\mu \xi_i^\nu \rangle\rangle = a^2 \quad (3.4)$$

O símbolo $\langle\langle \dots \rangle\rangle$ representa uma média tomada com respeito à distribuição (3.2). Outras distribuições deverão ser usadas se quisermos armazenar padrões com maior estrutura interna, ou mesmo padrões estruturados hierarquicamente; nós não vamos considerar essas possibilidades na presente tese. É importante comentar que freqüentemente, na literatura de redes de neurônios, fala-se de padrões correlacionados quando se quer indicar padrões que, mesmo sendo estatisticamente independentes, apresentam uma superposição não nula, como o caso que acabamos de considerar. Claramente, esta identificação é imprópria.

3.3 Estabilidade dos Padrões : Análise Sinal-Ruído

3.3.1 A Regra de Hopfield com padrões independentes

Nesta seção vamos fazer um estudo da estabilidade dos padrões armazenados com a regra de Hopfield. Vamos exigir como condição de estabilidade dinâmica para um padrão

qualquer ν , que satisfaça

$$\xi_i^\nu h_i > 0 \quad i = 1, \dots, N \quad (3.5)$$

o que nada mais é do que o alinhamento do estado do neurônio na memória ν com seu potencial pós-sináptico (PPS), ou campo local. Se o sistema atingiu um atrator, por exemplo a memória ξ_i^ν , o campo local no sitio i para a regra de Hopfield (1.9) será :

$$h_i = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \quad (3.6)$$

Temos uma soma de $p(N-1)$ termos; separando os termos que correspondem a $\mu = \nu$ obtemos:

$$h_i = \frac{N-1}{N} \xi_i^\nu + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \quad (3.7)$$

O primeiro termo da soma, que contem contribuições do padrão ν -ésimo (o qual estamos analisando), é conhecido como **sinal**. O segundo termo representa um **ruído** gerado pelos $p - 1$ padrões restantes. No limite $N \rightarrow \infty$ o sinal é igual a ξ_i^ν e, como os padrões são independentes, o ruído fica representado por uma caminhada aleatória com $(N-1)(p-1)$ passos de valor $+1$ ou -1 . O valor médio do ruído será conseqüentemente zero, e apresentará uma dispersão igual a $\sqrt{p/N}$. Podemos ver então que, se mantemos p fixo e $N \rightarrow \infty$, o ruído é essencialmente nulo, e assim o sinal do estado do neurônio será igual ao sinal do seu campo local (PPS), portanto a próxima configuração da rede será igual à atual, o que quer dizer que o sistema está num ponto fixo da sua dinâmica (i.e., a estabilidade do padrão está assegurada). Mesmo se p crescer proporcionalmente a N , de forma que $p = \alpha N$, o valor absoluto do ruído só irá ultrapassar o sinal se $\sqrt{\alpha} > 1$, o que implica que, segundo esta análise estatística, a capacidade crítica de armazenamento da rede é $\alpha_c \approx 1$.

3.3.2 A Regra de Hopfield com padrões com atividade

Agora vamos ver o que acontece com a capacidade de armazenamento da regra de Hopfield, se armazenamos padrões com atividade $(1+a)/2$, dada pela distribuição (3.2). Podemos decompor o campo local da mesma forma como fizemos na seção anterior. O

sinal (S) novamente será igual a ξ_i^ν , porém aguarda-nos uma surpresa ao calcular o termo de ruído (R). Em virtude das correlações (3.3) e (3.4):

$$R = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu = a^2 \sum_{\mu \neq \nu}^p \xi_i^\mu \quad (3.8)$$

O maior valor que pode tomar a soma no último fator é $(p - 1)$ e então os padrões são desestabilizados quando:

$$(p - 1) a^2 = 1 \quad (3.9)$$

Este resultado é catastrófico para o modelo já que, por exemplo, se uns 5% dos neurônios estão ativos (no valor +1), $a \approx -0.9$, e p não pode exceder de 2 independente do N . Certamente, uma rede que só pode armazenar 2 memórias não é muito interessante como rede associativa.

3.3.3 Armazenamento eficiente de padrões com atividade e o limite de baixa atividade

Numa tentativa de reduzir o ruído R (eq.3.8) e, conseqüentemente, aumentar a estabilidade dos padrões, Amit, Gutfreund e Sompolinsky [15] introduziram uma modificação simples na matriz sináptica do modelo de Hopfield, que permite armazenar um número de padrões proporcional a N , com atividade arbitrária:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p (\xi_i^\mu - a)(\xi_j^\mu - a) \quad (3.10)$$

Se os padrões são gerados de acordo com a distribuição (3.2), a análise sinal-ruído permite demonstrar que, no limite $N \rightarrow \infty$, o valor médio do ruído é zero e

$$\alpha_c(a) = (1 - |a|)^2 \alpha_c(0) \quad (3.11)$$

Este resultado implica que α_c é sempre menor do que o correspondente valor para o modelo de Hopfield ($a=0$), e se anula no limite $|a| = 1$. Este resultado ainda não corresponde as expectativas previstas no cálculo de Gardner. O modelo introduzido na ref. [15] ainda apresenta outras dificuldades como, por exemplo, proliferação de estados espúrios a medida que a cresce. Tanto este problema como a diminuição do α_c com a

são resolvidos modificando a dinâmica do modelo, limitando o número de configurações dinâmicas acessíveis a aquelas que satisfaçam $\frac{1}{N} \sum_{i=1}^N x_i = a$. Porém, a capacidade crítica do modelo, ainda que com a dinâmica modificada, cai a zero no limite $a \rightarrow 1$, contrariamente ao desejado.

Em dois trabalhos independentes, Buhmann, Divko e Schulten [63] e Tsodyks e Feigel'man [64], reformularam o modelo de Hopfield voltando as variáveis $\{0,1\}$, mais naturais do ponto de vista biológico do que as $\{-1,1\}$, na representação de estados neuronais passivos e ativos. As novas variáveis dinâmicas S' , e padrões η , são definidos em termos dos antigos S , e ξ , pelas transformações

$$x'_i = \frac{(x_i + 1)}{2} \quad (3.12)$$

$$\eta_i^\mu = \frac{(\xi_i^\mu + 1)}{2} \quad (3.13)$$

Esta mudança, a priori formal, acarreta uma mudança nos limiares de disparo dos neurônios, o que, por sua vez, traz como consequência uma melhora importante na capacidade de armazenamento da rede. Lembremos que, no modelo de Hopfield (a temperatura nula), cada neurônio apresenta um PPS

$$h_i = \sum_{j \neq i}^N J_{ij} \frac{(x_j + 1)}{2} \quad (3.14)$$

e sua dinâmica é governada pela equação

$$x_i(t+1) = \text{signal}(h_i(t) - U_i), \quad (3.15)$$

com o limiar de disparo $U_i = \frac{1}{2} \sum_{j \neq i}^N J_{ij}$. Se, diferentemente, escolhermos U_i constante, a performance do modelo muda. Especificamente, escolhem-se p padrões com a distribuição

$$P(\eta_i^\mu) = a \delta(\eta_i^\mu - 1) + (1 - a) \delta(\eta_i^\mu) \quad (3.16)$$

O modelo de Hopfield corresponde a $a=1/2$. Baixa atividade corresponde a valores pequenos de a . A menos de fatores constantes, em ambos trabalhos a matriz de sinapses é

$$T_{ij} = \frac{1}{N} \sum_{\mu=1}^p (\eta_i^\mu - a)(\eta_j^\mu - a) \quad (3.17)$$

Uma análise sinal-ruído do campo local evidencia a necessidade de introduzir um limiar uniforme $U = 1/2 - a$, de forma tal que o campo local sobre o neurônio i , no padrão ν , será

$$h_i = \sum_{j \neq i}^N T_{ij} \eta_j^\nu - U \quad (3.18)$$

O resultado é interessante:

- Padrões sem superposição ($a=1/2$) são desestabilizados para $\alpha_c = 1/2$. A capacidade do modelo, calculada mediante uma análise sinal-ruído, é a metade da do modelo de Hopfield ($\alpha_c = 1$).
- Com a diminuição da atividade a , o ruído diminui e a capacidade cresce como

$$\alpha_c = \frac{1}{4a} \quad (3.19)$$

No limite de baixa atividade $a \rightarrow 0$, a capacidade do modelo diverge.

Uma análise de campo médio do modelo permite obter um valor mais preciso para a capacidade crítica:

$$\alpha_c \approx \frac{1}{a \ln a} \quad (3.20)$$

Este valor está muito próximo do limite teórico obtido por Gardner (eq. 3.1). Inspirados nestes trabalhos, C. Perez-Vicente e D.Amit [65] generalizaram a transformação (3.12), introduzindo um novo parâmetro b :

$$x'_i = \frac{(x_i + b)}{2}, \quad (3.21)$$

além do limiar uniforme U , e voltando as antigas variáveis $\{-1, 1\}$, tanto para as variáveis dinâmicas x_i como para os padrões ξ , que são escolhidos, novamente, de acordo com a distribuição (3.2). O principal resultado deles é que é possível escolher um valor adequado de b que otimize as propriedades de reconhecimento da rede, obtendo também uma divergência no α_c no limite de baixa atividade ($a \rightarrow 1$). Mediante uma análise sinal-ruído, os autores obtêm uma aproximação para o α_c

$$\alpha_c = \frac{1}{(1 - a^2)} \quad (3.22)$$

Uma análise de campo médio do modelo melhora a previsão (qualitativamente correta) da análise sinal-ruído, e permite obter uma capacidade crítica [65]

$$\alpha_c \approx \frac{k(a)}{(1-a)\ln(1-a)} \quad (3.23)$$

O resultado, similar aos previos, também está perto do limite de Gardner.

Todos os trabalhos anteriores têm o mérito de atingir grandes capacidades de armazenamento de padrões no limite de baixa atividade. No entanto, as regras de aprendizado são , todas, pouco plausíveis biologicamente. Elas são simples modificações da regra de Hopfield, e apresentam, entre outros pontos fracos do ponto de vista biológico, o fato de serem simétricas. Na próxima seção vamos estudar uma generalização das regras de aprendizado Hebbianas que irá nos permitir considerar regras assimétricas e, em particular, recuperar os modelos descritos previamente para valores específicos dos parâmetros.

3.4 Regras Hebbianas Generalizadas e o Papel da Assimetria

D.O.Hebb publicou, num livro agora clássico [24], uma idéia que impulsionou a procura de uma teoria da memória: *“a eficácia de uma sinapse excitatória aumenta quando os dois neurônios que se unem estão ativos”*. Experimentos feitos por Levy e colaboradores [66] no hipocampus de ratos mostram evidencias de que, quando os neurônios pré-sináptico e pós-sináptico estão ativos simultaneamente, a sinapse que os une é reforçada fortemente. Esse efeito pode ser modelado pelo incremento ΔJ_{ij} da eficácia sináptica (excitatória)

$$\Delta J_{ij} \propto x'_i \cdot x'_j \quad (3.24)$$

$x'_i = 1$ se o neurônio está ativo, e $x'_i = 0$ se está em repouso. A equação (3.24) representa a regra de Hebb, e mais: ela diz que a eficácia sináptica J_{ij} não é modificada sempre que qualquer um dos neurônios, o pré-sináptico ou o pós-sináptico, estiverem em repouso. Em particular, os experimentos de Levy mostram que, quando o neurônio pré-sináptico j está em repouso e o pós-sináptico i está ativo, a eficácia sináptica excitatória sofre um decréscimo importante. Então , as sinapses excitatórias não só podem armazenar

neurônio pós-sináptico i	neurônio pré-sináptico j	$x_j = -1$	$x_j = +1$
	$x_i = -1$	0	0
$x_i = +1$		a_{cs}^- (antiHebb) (< 0)	a_{cs}^+ (Hebb) (> 0)

Tabela 3.1: Parâmetros experimentais de sinapses excitatórias.

atividades em conjunção como também anti-conjunção de atividades. Em lugar da eq.(3.24) a regra para a variação da eficácia sináptica pode ser

$$\Delta J_{ij} \propto x_i \cdot x_j \quad (3.25)$$

$x_i = 2x'_i - 1$ é a transformação (3.12). Claramente, a regra (3.25) leva em consideração mais informação sobre as atividades simultâneas de x_i e x_j e, conseqüentemente, o sistema de armazenamento de memórias na rede deve ser mais eficiente. A regra (3.25) é, basicamente, a regra de Hopfield. No que diz respeito à observação experimental, a regra (3.25) ainda não é satisfatória. Alguns dos seus pontos fracos já foram discutidos antes em diversos pontos da tese. No contexto da presente discussão, a regra tem como conseqüência indesejável que a eficácia sináptica é reforçada quando os dois neurônios que se unem estão em repouso ($x_i = x_j = -1$). Experiências feitas por Rauschecker e colaboradores [67] no córtex visual de gatos apresentam evidências de que as eficácias das sinapses excitatórias não são modificadas quando o neurônio pós-sináptico i está em repouso, qualquer que seja a atividade do neurônio pré-sináptico j . Na tabela 3.1 podemos observar os parâmetros experimentais de sinapses excitatórias. Podemos levar em conta este último ponto modificando a regra (3.25) da forma

$$\Delta J_{ij} \propto (x_i + 1) \cdot x_j \quad (3.26)$$

As considerações anteriores correspondem a sinapses excitatórias. Muito pouco se sabe da plasticidade de sinapses inibitórias, e muitos investigadores sugerem que as sinapses inibitórias são muito menos plásticas do que as sinapses excitatórias.

$i \backslash j$	$X_j = -1$	$X_j = +1$
$X_i = -1$	a^{--}	a^{-+}
$X_i = +1$	a^{+-}	a^{++}

Tabela 3.2: Parâmetros teóricos para as sinapses

Sem dúvida, devido à complexidade do cérebro, com seus diferentes módulos especializados em diversas funções, seus vários tipos de neurônios (piramidais, -stellate-, etc.), as diferentes arquiteturas de conectividade dos neurônios nas diversas partes do cérebro, uma grande variedade de regras devem ser necessárias na modelagem dos processos de aprendizado. Muitas destas regras podem ser introduzidas por meio de quatro parâmetros (ver tabela 3.2) que podem se relacionar diretamente com os parâmetros experimentais discutidos acima. A questão é encontrar uma região no espaço destes parâmetros no qual o sistema apresente boas propriedades como memória associativa.

Vamos introduzir agora uma classe de regras de aprendizado Hebbianas generalizadas que incorporam uma série de ingredientes biologicamente motivados como, por exemplo, **assimetria estrutural e padrões com atividade** [68]. Por outro lado, estas regras recuperam, em casos especiais, diversas regras de aprendizado conhecidas na literatura. Estas regras vão ser utilizadas em sistemas formados por N neurônios $\{x_i\}_{i=1,\dots,N}$, que armazenam p padrões com atividade $(1+a)/2$ escolhidos de acordo com a distribuição (3.2). Assumindo que cada novo padrão memorizado modifica cada sinapse *somente* por meio do neurônio pré-sináptico j e do neurônio pós-sináptico i , temos que a forma analítica mais geral que não mistura os padrões para a variação da eficácia sináptica pode ser escrita

$$\Delta J_{ij}(\xi_i^\mu, \xi_j^\mu) = A + B\xi_i^\mu + C\xi_j^\mu + D\xi_i^\mu \xi_j^\mu \quad (3.27)$$

A , B , C e D são quatro constantes, a priori arbitrárias, que podem ser relacionadas de

forma muito simples, com os parâmetros da tabela (3.2)

$$\begin{aligned}
 \Delta J_{ij}(-1, -1) &= a^{--} = (A - B - C + D) \\
 \Delta J_{ij}(-1, +1) &= a^{-+} = (A - B + C - D) \\
 \Delta J_{ij}(+1, -1) &= a^{+-} = (A + B - C - D) \\
 \Delta J_{ij}(+1, +1) &= a^{++} = (A + B + C + D)
 \end{aligned} \tag{3.28}$$

e conseqüentemente

$$\begin{aligned}
 A &= \frac{1}{4}(a^{--} - a^{-+} - a^{+-} + a^{++}) \\
 B &= \frac{1}{4}(a^{--} - a^{-+} + a^{+-} - a^{++}) \\
 C &= \frac{1}{4}(a^{--} + a^{-+} - a^{+-} - a^{++}) \\
 D &= \frac{1}{4}(a^{--} + a^{-+} + a^{+-} + a^{++})
 \end{aligned} \tag{3.29}$$

Desta forma, dado qualquer conjunto de parâmetros com motivação experimental $\{a^{++}, a^{-+}, a^{+-}, a^{--}\}$, obtemos um conjunto $\{A, B, C, D\}$, e a correspondente regra de aprendizado.

Sem levar em consideração fenômenos de saturação no valor das sinapses, parece razoável supor que os padrões aprendidos adicionam-se linearmente (como na regra de Hopfield), e então obtemos a seguinte classe de regras Hebbianas generalizadas [26]:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P (A + B\xi_i^{\mu} + C\xi_j^{\mu} + D\xi_i^{\mu} \xi_j^{\mu}) \tag{3.30}$$

O fator $1/N$ foi introduzido para efeitos de normalização da relação sinal-ruído das regras. Claramente, sempre que $B \neq C$ estas regras são assimétricas. Esta assimetria é uma propriedade intrínseca da regra e então vamos denominá-la **assimetria estrutural**, para distingui-la da assimetria gerada por um processo de diluição, caso que foi tratado no capítulo (2). Tomando valores particulares dos parâmetros A, B, C e D podemos recuperar diversos modelos conhecidos. Por exemplo, se $A=B=C=0$ e $D=1$, recuperamos o modelo de Hopfield. O caso, $A=B=C=D=1/4$ corresponde à proposta original de Hebb: só aqueles neurônios que estiverem *simultaneamente* ativos contribuirão para a modificação (reforço) da correspondente sinapse. Já se considerarmos padrões com atividade $(1+a)/2$, a eleição $A = a^2, B = C = -a$ e $D = 1$ corresponde à regra introduzida por Amit, Gutfreund e Sompolinsky [15] para armazenar padrões com superposição não nula entre eles.

As regras anteriores são todas elas simétricas. Nós esperamos que, a priori, regras mais gerais, com $B \neq C$, poderiam apresentar uma performance melhor, no que diz respeito à capacidade de armazenamento, do que as simétricas. Com a intenção de estudar esta possibilidade, fizemos uma análise sinal-ruído das regras (3.30), procurando uma regra no espaço dos parâmetros $\{A,B,C,D\}$ que permitisse otimizar a capacidade de armazenamento da rede.

Considerou-se que o neurônio pós-sináptico i recebe um estímulo (PPS) ou campo local

$$h_i = \sum_{j=1}^N J_{ij} (x_j - b) - U \quad (3.31)$$

Os parâmetros b e U modificam os limiares de disparo neuronais e, como iremos ver, irão nos permitir otimizar as propriedades de armazenamento da rede, além da otimização permitida pelos parâmetros da regra $\{A,B,C,D\}$. O parâmetro U (um campo uniforme sobre cada neurônio), irá intervir no cancelamento de termos perturbativos na parte do sinal no campo local. Como vamos requerer uma média nula para a parte de ruído no campo local, o parâmetro b também será útil.

Vamos supor agora que a rede tem evoluído até o padrão $\{\xi^1 \equiv \xi_i^1\}$, então

$$h_i = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu=1}^p (A + B\xi_i^\mu + C\xi_j^\mu + D\xi_i^\mu \xi_j^\mu)(\xi_j^1 - b) - U \quad (3.32)$$

Da mesma forma que fizemos com a regra de Hopfield, ao analisar a estabilidade de um padrão, podemos decompor o campo local em duas partes:

$$h_i = S + R \quad (3.33)$$

O termo de sinal S será produzido pelo padrão selecionado,

$$S = \frac{1}{N} \sum_{j \neq i}^N (A + B\xi_i^1 + C\xi_j^1 + D\xi_i^1 \xi_j^1)(\xi_j^1 - b) - U \quad (3.34)$$

Os restantes $(p - 1)$ padrões vão gerar um ruído R , que irá desestabilizar o padrão selecionado,

$$R = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu=2}^p (A + B\xi_i^\mu + C\xi_j^\mu + D\xi_i^\mu \xi_j^\mu)(\xi_j^1 - b) \quad (3.35)$$

Sem perda de generalidade (sempre que $D \neq 0$) podemos fixar $D=1$. O nosso objetivo será então otimizar a razão sinal-ruído $\rho = S/R$ no espaço dos parâmetros $\{A, B, C, U, b, a\}$.

De acordo com a distribuição (3.2) o valor médio para o sinal é

$$\langle\langle S \rangle\rangle = A(a - b) + C(1 - ab) + [B(a - b) + D(1 - ab)] \xi_i^1 - U \quad (3.36)$$

Na expressão anterior tem se mantido fixo o valor de ξ_i^1 , já que representa o nível de atividade do neurônio i -ésimo no padrão cuja estabilidade queremos analisar. Claramente, o valor de U que otimiza o sinal é

$$U = A(a - b) + C(1 - ab) \quad (3.37)$$

Da mesma forma calculamos a média do ruído e obtemos

$$\langle\langle R \rangle\rangle = (a - b) \left[A(p - 1) + aC(p - 1) + (B + a) \sum_{\mu=2}^p \xi_i^\mu \right] \quad (3.38)$$

Parece natural impor que $\langle\langle R \rangle\rangle = 0$ a fim de maximizar a estabilidade do padrão ξ^1 , então devemos escolher $b=a$. Como foi apontado antes, o parâmetro b permite-nos anular a média do ruído causado pelos $p - 1$ padrões restantes. Vamos considerar agora o efeito da variância do ruído. Temos

$$\langle\langle R^2 \rangle\rangle = \alpha(1 - a^2) \left\{ p \left[(A + a^2) + a(B + C) \right]^2 + B^2 + C^2 + 2a(aBC + B + C) + 1 \right\} \quad (3.39)$$

onde definimos $\alpha \equiv p/N$ (com $p \rightarrow \infty$ e $N \rightarrow \infty$) como o parâmetro que mede a capacidade de armazenamento da rede. Vamos querer que, para um valor de p arbitrário, $\langle\langle R^2 \rangle\rangle$ seja pequeno. Então o termo entre colchetes proporcional a p deve se anular

$$A + a^2 + a(B + C) = 0 \quad (3.40)$$

A regra simétrica da referência [65] cancela automaticamente este termo. No entanto, como nós estamos interessados em regras mais gerais (assimétricas), vamos tentar otimizar $\langle\langle R^2 \rangle\rangle$ impondo uma assimetria simples entre B e C :

$$B = \gamma C, \quad (3.41)$$

para valores gerais de γ ($\gamma = 1$ recupera o caso simétrico). Da eq.(3.40) obtemos:

$$C = -\frac{a^2 + A}{a(\gamma + 1)} \quad (3.42)$$

Substituindo na eq.(3.39),

$$\langle\langle R^2 \rangle\rangle = \alpha(1 - a^2) \left[A^2(2 + q/a^2) + 2A(3a^2 - 1 + q) + a^2q + 2a^4 + (a^2 - 1)^2 \right], \quad (3.43)$$

onde

$$q \equiv \frac{(\gamma^2 + 1)(1 - 3a^2) - 4\gamma a^2}{(\gamma + 1)^2}. \quad (3.44)$$

Para valores gerais de a , temos que maximizar a razão sinal- ruído ρ , como função de duas variáveis: A e a assimetria γ . Levando em conta que o sinal médio otimizado não depende nem de A nem de γ , só temos que minimizar a variância do ruído. Ela apresenta um mínimo absoluto nos valores $\gamma = 1$ e $A = a^2$. Chegamos ao resultado de que, dentre o grande número de regras definidas pelos parâmetros A , B , C e D , a ótima (no sentido de maximizar a razão sinal-ruído) é a regra simétrica $A = a^2$, $B = C = -a$, $D=1$, estudada na ref. [65].

Quando a intensidade do ruído torna-se comparável com a do sinal ($\rho \approx 1$), os padrões começam a ser desestabilizados e, conseqüentemente, podemos estimar a capacidade de armazenamento crítica α_c , fazendo uso das eqs.(3.36) e (3.43) e impondo $\rho = 1$. No caso particular $\gamma = 1$ e $a=0$ (regra de Hopfield), obtemos $\rho = a^{-1/2}$ e então $\alpha_c \approx 1$.

Na figura (3.1) podemos ver, para diversos valores de a , α_c como função do parâmetro de assimetria γ , resultado da análise sinal-ruído anterior, fixando $\rho = 1$ com a condição (ótima) $A = a^2$. No limite de baixa atividade ($a \rightarrow 1$), α_c diverge no valor ótimo $\gamma = 1$. Claramente, esta solução não é robusta frente a assimetria. De fato, uma pequena assimetria provoca o colapso das capacidades de armazenamento da rede, como pode se observar na fig.3.1(a). Nas figuras 3.1(b)–(d) podemos ver que, quando a diminui, as regras assimétricas apresentam uma performance cada vez melhor (no que diz respeito a sua robustez), exceto na vizinhança de $\gamma = -1$ onde aparece um mínimo $\alpha_c = 0$. No limite $a \rightarrow 0$, vemos que $\alpha_c \rightarrow 1$, independentemente do valor de γ . Seguindo as

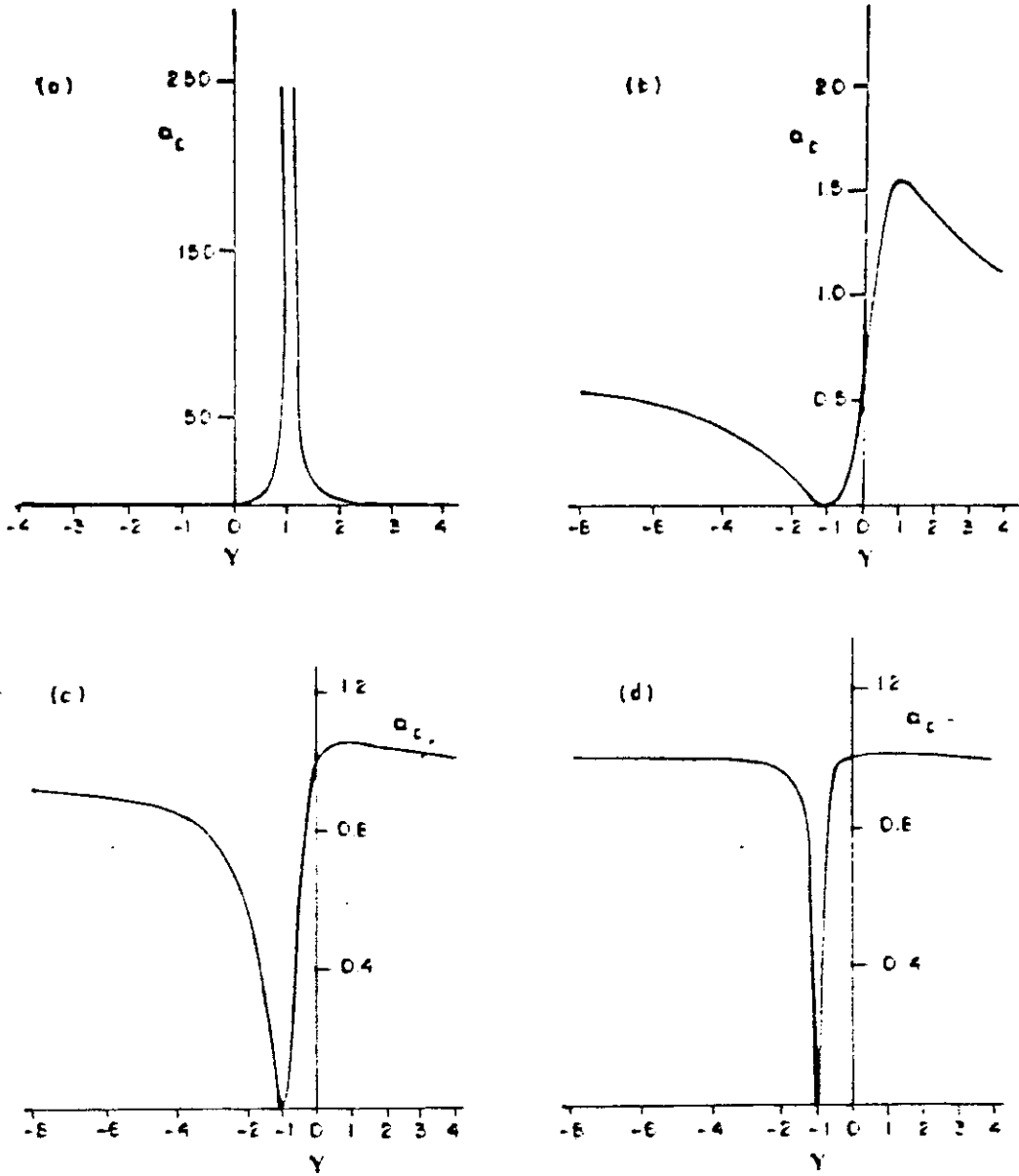


Figura 3.1: Capacidade de armazenamento crítica α_c vs. parâmetro de assimetria γ de uma análise sinal-ruído ($\gamma = 1$ corresponde a regras simétricas): (a) $a=1$; (b) $a=0.6$; (c) $a=0.2$; (d) $a=0.05$.

figuras podemos observar que o mínimo em $\gamma = -1$ sempre está presente e produz uma convergência não uniforme ao valor $\alpha_c = 1$ no limite $a \rightarrow 0$.

Levando em conta que uma análise sinal-ruído só representa uma primeira aproximação ao comportamento de um modelo, temos testado os resultados obtidos mediante simulações de computador. Com esse fim, simulamos uma rede neuronal analógica composta por $N=400$ neurônios. Os estados dos neurônios evoluem de acordo com uma dinâmica determinista, definida pelo seguinte sistema de mapas acoplados (similar ao introduzido no capítulo 1 e estudado no capítulo 2 em relação a sistemas diluídos)

$$x_i(t+1) = \tanh \left[g \left(\sum_{j=1}^N J_{ij} (x_j(t) - b) - U \right) \right], \quad i = 1, \dots, N. \quad (3.45)$$

Como foi discutido no capítulo 1, $x_i(t)$ representa a frequência de disparo do neurônio i -ésimo no tempo t , e pode tomar valores contínuos no intervalo $[-1, 1]$. Ambos extremos representam o neurônio em repouso e disparando com frequência máxima, respectivamente. O parâmetro g mede o grau de não-linearidade da função de transferência; no limite $g \rightarrow \infty$ recupera-se um sistema com variáveis discretas binárias; no limite $g \rightarrow 0$ podemos expandir a \tanh até primeira ordem em g , obtendo como único estado estável aquele com $x_i=0$ ($i=1, \dots, N$), ou seja, o sistema perde suas propriedades de reconhecimento de memórias. J_{ij} , a matriz de sinapses, é definida pelas eqs.(3.2) e (3.30). Os parâmetros U e b foram fixados nos seus valores otimizados. Foi escolhida uma atualização em *paralelo* dos estados dos neurônios, ou seja, todas as unidades são atualizadas simultaneamente. Esta forma de atualização da rede apresenta o problema da presença de ciclos limite em certas regiões do espaço de parâmetros. No entanto, ela é mais adequada na hora de fazer simulações muito grandes em processadores em paralelo (ver discussão no capítulo 1). De qualquer forma, nem a dinâmica paralela, nem a seqüencial são boas aproximações para a dinâmica de uma rede neuronal biológica.

A evolução do processo dinâmico foi o seguinte:

1. Fixamos um valor de α e geramos p padrões com a distribuição (3.2) para $a = 0.2$.
2. Como estudamos a capacidade de armazenamento da rede, a condição inicial para as simulações sempre foi um dos padrões armazenados.

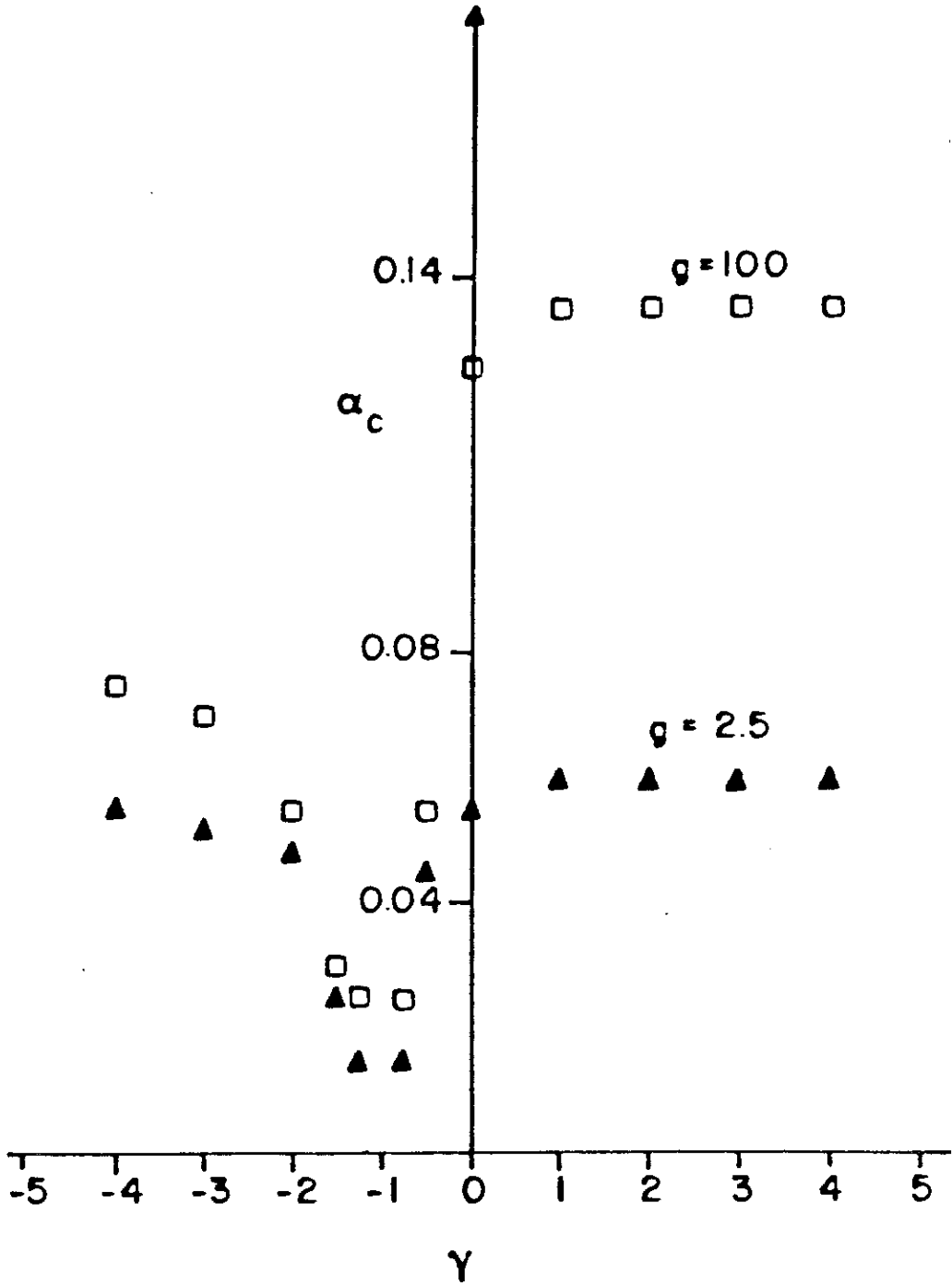


Figura 3.2: Capacidade de armazenamento crítica α_c vs. parâmetro de assimetria γ obtido simulando uma rede com 400 neurônios analógicos para $a=0.2$. Os quadrados vazios (triângulos cheios) correspondem a $g=100$ ($g=2.5$).

3. Deixamos evoluir a configuração (padrão) escolhida de acordo com a dinâmica (3.45).
4. Cada vez que o sistema atingiu um ponto fixo, calculamos e armazenamos a superposição final com o padrão usado como estímulo inicial.
5. As etapas 2–4 são repetidas para a metade dos padrões armazenados em cada uma de 40 realizações diferentes deles (o que dá um total de $20 \cdot p$ rodadas).
6. Finalmente calculamos a superposição final “média”, formada por todas as rodadas que acabaram em um ponto fixo. As rodadas que acabaram em ciclos foram deixadas fora da média, embora tenham sido contabilizadas separadamente.
7. Consideramos o sistema como estando numa fase de reconhecimento (para um α particular) sempre que a superposição final média for maior que 95% (valor convencional) e o número de ciclos for menor do que 5% do número total de rodadas. α_c foi definido com o valor de α para o qual uma das condições anteriores não for satisfeita.

Para $g = 100$ perturbamos em torno da regra ótima encontrada antes, a procura de uma direção no plano B–C na qual alguma regra tivesse um α_c maior do que o da regra simétrica. Esta busca foi feita para diversos valores de a sem sucesso, confirmando a predição da análise sinal-ruído. Também estudamos numericamente o efeito do parâmetro de assimetria γ (mantendo os outros parâmetros fixos nos seus valores otimizados). Podemos ver os resultados na figura 3.2 para $a = 0.2$, e dois valores do ganho: $g=100$ e $g=2.5$.

Podemos notar que, no caso $g = 100$, quando o sistema é praticamente discreto, o valor de saturação de α_c , para $\gamma \geq 1$, está perto do valor para o modelo de Hopfield ($\alpha_c \approx 0.14$). Isto parece estar em contradição com a predição da análise sinal-ruído ($\alpha_c \approx 1$). Porém esta diferença não é inesperada, já que, como foi apontado, a análise sinal-ruído só provê uma descrição qualitativa. Na mesma figura também podemos ver que o efeito de reduzir o ganho é basicamente uma redução da capacidade

de armazenamento. É importante notar que, para qualquer valor de a e g , α_c se anula para $\gamma = -1$ (caso extremamente assimétrico).

Voltando à análise geral, é interessante destacar que os resultados encontrados por Peretto [26], para $a=0$, podem ser melhorados muito mediante a introdução de um campo otimizador $U=C$. Porém, neste caso, o máximo α_c para $a = 0$ continua sendo o do modelo de Hopfield, como se segue da análise geral.

3.5 Conclusões

É importante notar que, apesar de que a análise feita por Gardner no espaço das interações admite que sistemas de redes de neurônios com sinapses assimétricas possuam excelentes qualidades de armazenamento, até hoje só foram propostas, em nosso conhecimento, regras simétricas. Nós encontramos que as regras assimétricas estudadas previamente, apesar de não serem ótimas no que diz respeito as suas capacidades de armazenamento, são mais robustas do que as regras simétricas, para valores da atividade não muito baixos. Além disso, a introdução dos quatro parâmetros A , B , C e D permite uma relação direta com valores experimentais em regras de aprendizado, como foi discutido no começo da seção anterior.

Nós limitamos nossa análise principalmente ao estudo do efeito da assimetria na capacidade de armazenamento da rede. Seria interessante estudar, para esta classe de regras, a influencia de γ (bem como do resto dos parâmetros) em outras quantidades de interesse como, por exemplo, a quantidade de informação das regras quando padrões com atividade são armazenados, os tempos de recuperação de padrões, o tamanho das bacias atratoras e o diagrama de fases.

As regras Hebbianas generalizadas estudadas neste capítulo podem ser ainda generalizadas, considerando termos que misturem as memórias na matriz de sinapses. Por exemplo, termos do tipo $\xi_i^\mu \xi_j^\nu$ para μ e ν arbitrarios. O caso particular cíclico $\xi_i^\mu \xi_j^{\mu+1}$ já tem sido considerado no estudo da recuperação de seqüências temporais [11].

Finalmente, no caso $A=B=C=U=b=0$, $D=1$, o modelo determinístico com um ganho g pode ser mapeado num modelo estocástico com um parâmetro β que faz o papel

do inverso de uma temperatura [16]. Seria interessante verificar se a equivalência é preservada no modelo mais geral.

Capítulo 4

Aprendizado e Generalização em Redes de Neurônios

Até agora temos estudado propriedades de redes de neurônios, fundamentalmente no que diz respeito ao seu comportamento como memórias associativas. No entanto, se nosso objetivo final é modelar o cérebro real, biológico, estes modelos deverão ter a capacidade de realizar outras funções , além de possuir boas memórias. Mesmo pensando em aplicações tecnológicas, é um desafio muito interessante e importante poder reproduzir o melhor possível uma variedade de funções próprias do cérebro humano. No presente capítulo vamos estudar o problema do *aprendizado e generalização* em modelos de redes de neurônios. A grande maioria dos trabalhos relacionados com aprendizado e generalização concentram-se num tipo particular de arquitetura da rede neuronal, uma arquitetura na qual a informação flui numa única direção , diferentemente do que

acontece com os modelos até agora estudados. O protótipo dos modelos com fluxo unidirecional de informação é o *perceptron*, cuja regra de aprendizado foi introduzida na seção (1.4). A arquitetura do perceptron mais simples, que foi introduzido por Rosenblatt [77] em 1962, consiste em uma camada de neurônios que recebem o estímulo inicial e um único neurônio de saída. Versões posteriores mais elaboradas incluem um maior número de camadas internas entre a camada inicial (input) e a camada de saída (output), que não necessariamente consta de um único neurônio. São as chamadas *redes multicamadas*, nas quais a informação flui da camada de entrada até a camada de saída, sem ter retro-alimentação nos neurônios. Em poucos trabalhos [78,81,79,80] estudou-se o problema de generalização em *redes com atratores*, como no modelo de Hopfield. Na seção (4.1) vamos fazer um resumo dos trabalhos em redes tipo perceptron, as quais historicamente foram as primeiras a serem estudadas. Na seção (4.2) vamos discutir brevemente as *redes hierárquicas*, que de alguma forma introduziram o problema de *categorização* em redes com atratores, problema estreitamente relacionado ao de generalização. Na seção (4.3) vamos introduzir um modelo de rede de neurônios analógica, e analisar em detalhe a sua capacidade de generalização. Finalmente na seção (4.4) vamos discutir a possibilidade de unificar o estudo do problema de aprendizado e generalização em redes tipo perceptron e redes com atratores tipo Hopfield.

4.1 Aprendizado e Generalização em Redes de Neurônios tipo Perceptron

Nos últimos anos, muito esforço tem sido aplicado na tentativa de desenvolver uma mecânica estatística para atacar o antigo problema do aprendizado e generalização em redes tipo perceptron [82]–[92]. Nós vamos somente fazer um resumo do formalismo desenvolvido, bastante geral e poderoso, e marcar os principais logros nesta linha de pesquisa. Consideremos uma rede neuronal com M neurônios na camada de entrada, S_i ($i = 1, \dots, M$), N sinapses W_i ($i = 1, \dots, N$), e um único neurônio de saída $\sigma = \sigma(\mathbf{W}; \mathbf{S})$. Para cada conjunto das N sinapses \mathbf{W} , o sistema define um mapeamento de \mathbf{S} para σ . Neste contexto, a arquitetura, determinada pelos valores de N e M , mais

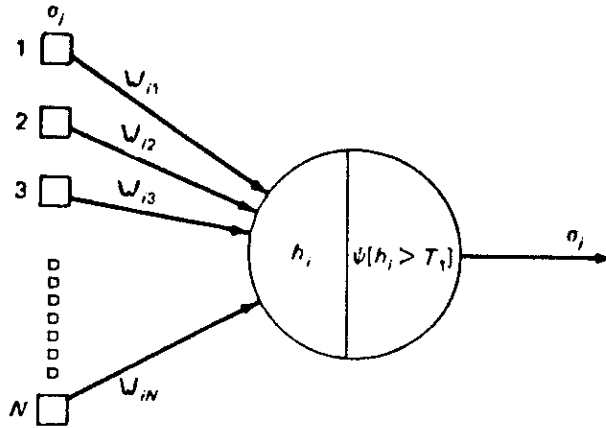


Figura 4.1: Esquema de uma rede tipo perceptron.

os valores das sinapses \mathbf{W} , definem uma rede (ver fig. 4.1). O aprendizado pode ser visto como uma busca, no espaço das sinapses, a procura de uma rede que possua as propriedades desejadas, por exemplo, que satisfaça uma certa regra, definida por um conjunto de padrões de entrada e de saída. No caso do *aprendizado supervisionado*, as sinapses são ajustadas de forma a satisfazer o melhor possível a regra desejada $\sigma_0(\mathbf{S})$. Uma forma de atingir este objetivo é prover à rede com uma série de *exemplos*. Estes exemplos irão consistir de um conjunto de P pares de entrada-saída, $(\mathbf{S}^l, \sigma_0(\mathbf{S}^l))$, com $l = 1, \dots, P$. O caso de aprendizado não supervisionado não será considerado. Vamos supor que o conjunto de exemplos de entrada \mathbf{S}^l é escolhido, dentre o conjunto total de possíveis entradas, de acordo com alguma medida $d\mu(\mathbf{S})$. Os exemplos podem ser utilizados para construir uma *energia de treinamento*

$$E(\mathbf{W}) = \sum_{l=1}^P \epsilon(\mathbf{W}; \mathbf{S}^l) \quad (4.1)$$

A *função erro* $\epsilon(\mathbf{W}; \mathbf{S}^l)$ representa uma medida do desvio da saída real da rede $\sigma(\mathbf{W}; \mathbf{S})$ com respeito a saída desejada $\sigma_0(\mathbf{S})$. A função erro deve ser nula sempre que ambas saídas sejam coincidentes, e positiva em qualquer outro caso. Uma escolha comum é a função quadrática

$$\epsilon(\mathbf{W}; \mathbf{S}) = \frac{1}{2} [\sigma(\mathbf{W}; \mathbf{S}) - \sigma_0(\mathbf{S})]^2 \quad (4.2)$$

A etapa de treinamento consiste, geralmente, em minimizar a energia, por exemplo, via um decréscimo pelo gradiente (gradient descent):

$$\frac{\partial \mathbf{W}}{\partial t} = -\nabla_{\mathbf{W}} E(\mathbf{W}) \quad (4.3)$$

A energia de treinamento mede a performance da rede num conjunto limitado de exemplos, porém o objetivo último é encontrar uma rede que realize a regra desejada para qualquer entrada, não só para os exemplos do treinamento. A performance de uma rede \mathbf{W} no espaço de entrada completo é quantificada pela *função de generalização*. É definida pela média do erro da rede, no total do espaço de entrada,

$$\epsilon(\mathbf{W}) = \int d\mu(\mathbf{S}) \epsilon(\mathbf{W}; \mathbf{S}) \quad (4.4)$$

Pode-se distinguir entre *regras realizáveis* e *regras não-realizáveis*. As primeiras são aquelas nas quais a saída desejada $\sigma_0(\mathbf{S})$ pode ser completamente realizada pelo menos por uma das redes definidas no espaço das sinapses. Consequentemente, para uma regra realizável, existe um vector de sinapses \mathbf{W}^* que satisfaz

$$\epsilon(\mathbf{W}^*, \mathbf{S}) = 0, \quad \forall \mathbf{S} \quad (4.5)$$

ou, equivalentemente, $\epsilon(\mathbf{W}^*) = 0$. Uma regra não-realizável é aquela cuja saída verifica

$$\epsilon_{min} = \min_{\mathbf{w}} \epsilon(\mathbf{W}) > 0 \quad (4.6)$$

Regras não-realizáveis ocorrem, basicamente, em dois casos. No primeiro, quando os dados necessários para o treinamento apresentam ruído. A segunda situação possível é quando a arquitetura da rede impossibilita a realização completa da regra.

É possível generalizar a dinâmica de aprendizado (4.3) para considerar-se efeitos estocásticos. Uma possível regra de evolução das sinapses é

$$\frac{\partial \mathbf{W}}{\partial t} = -\nabla_{\mathbf{W}} E(\mathbf{W}) - \eta(t) \quad (4.7)$$

onde η representa um ruído branco, com variância

$$\langle \eta_i(t) \eta_j(t') \rangle = 2T \delta_{ij} \delta(t - t') \quad (4.8)$$

Esta dinâmica produz um decréscimo gradual da energia porém, as vezes, a energia pode aumentar devido ao efeito do ruído . A $T=0$, o termo de ruído desaparece, e a dinâmica se reduz à da eq.(4.3). As equações anteriores são apropriadas para sinapses que podem tomar um conjunto contínuo de valores. No caso de sinapses discretas, a dinâmica análoga à (4.7) é uma dinâmica de Monte Carlo com tempo discreto. Em geral, o efeito da temperatura é evitar que o sistema fique preso num mínimo local da energia. No entanto, no presente contexto, não é muito importante atingir o mínimo global da energia de treinamento e, sim, obter uma boa qualidade de generalização . Acontece que a temperatura pode ser importante para evitar *sobre-treinamento*, ou seja, evitar que a apresentação de um excessivo número de exemplos da regra produza uma diminuição da capacidade de generalização . Já que, a tempos longos, a eq.(4.7) produz uma distribuição de Gibbs, no espaço das sinapses \mathbf{W} , podem-se estudar as propriedades médias de equilíbrio do ensemble de soluções para a regra desejada.

$$P(\mathbf{W}) = Z^{-1} e^{-\beta E(\mathbf{W})} \quad (4.9)$$

onde agora a variância do termo de ruído no processo de treinamento será a temperatura $T = 1/\beta$ do sistema. O fator de normalização é a função de partição

$$Z = \int d\mu(\mathbf{W}) \exp[-\beta E(\mathbf{W})] \quad (4.10)$$

Agora com o formalismo da mecânica estatística, podem-se calcular médias térmicas com respeito a distribuição $P(\mathbf{W})$, que serão representadas por $\langle \dots \rangle_T$. Ainda depois de ter realizado a média térmica, existe uma dependência com os P exemplos \mathbf{S}^l . Como os exemplos são escolhidos estocasticamente e logo fixados, eles representam desordem *congelada* (quenched). Conseqüentemente, para estudar o comportamento típico dos sistemas, devemos realizar uma média na distribuição dos exemplos, representada por $\langle\langle \dots \rangle\rangle \equiv \int \prod_l d\mu(\mathbf{S}^l)$. Os erros médios de *treinamento* e *generalização* vêm dados pelas equações

$$\epsilon_t(T, P) \equiv P^{-1} \langle\langle E(\mathbf{W}) \rangle\rangle_T \quad (4.11)$$

$$\epsilon_g(T, P) \equiv \langle\langle \epsilon(\mathbf{W}) \rangle\rangle_T \quad (4.12)$$

respectivamente. A energia livre e a entropia da rede vêm dadas pelas equações

$$F(T, P) = -T \langle \langle \ln Z \rangle \rangle \quad (4.13)$$

$$S(T, P) = - \langle \langle \int d\mu(\mathbf{W}) P(\mathbf{W}) \ln P(\mathbf{W}) \rangle \rangle \quad (4.14)$$

e estão relacionadas pela identidade

$$F = P\epsilon_t - TS \quad (4.15)$$

Conhecendo F , pode-se calcular o erro de treinamento médio

$$\epsilon_t = \frac{1}{P} \frac{\partial(\beta F)}{\partial \beta} \quad (4.16)$$

e a entropia por meio de

$$S = - \frac{\partial F}{\partial T} \quad (4.17)$$

As quantidades relevantes para a performance da rede serão $\epsilon_g(T, P)$ e $\epsilon_t(T, P)$; são as chamadas *funções de aprendizado*. Este formalismo é estritamente válido no limite termodinâmico, ou seja, quando $N \rightarrow \infty$. Pode-se demonstrar que, neste limite, o número de exemplos P deve escalar com N linearmente, $P = \alpha N$, sendo α uma constante.

O formalismo descrito anteriormente pode ser aplicado a uma diversidade de regras, a mais estudada das quais é a do perceptron. A regra do perceptron consiste no mapeamento

$$\sigma = g \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N W_j S_j \right) = g \left(N^{-1/2} \mathbf{W} \cdot \mathbf{S} \right) \quad (4.18)$$

onde $g(x)$ é uma função sigmoidal de x . Os exemplos são gerados por outro perceptron, o “instrutor”, e o objetivo do aprendizado pelo perceptron “aluno” é inferir os valores das sinapses do instrutor. Os resultados mais importantes, referentes ao aprendizado e generalização supervisionados da regra do perceptron, são os seguintes:

- No aprendizado, a temperatura finita tem um efeito positivo. Para uma T fixa, a medida que o número de exemplos cresce as sinapses atingem seus valores ótimos, ou seja, os valores que minimizam o erro de generalização. Ainda quando o $\alpha_c(T)$ e o ϵ_g crescem com T , treinar a rede a T finita é satisfatório no que diz respeito

aos tempos de convergencia; os tempos de convergencia a $T=0$ são geralmente excessivamente longos [92,88].

- No caso de redes com sinapses contínuas, e no limite $\alpha \rightarrow \infty$, tanto ϵ_t quanto ϵ_g aproximam-se a um limite comum, ϵ_{min} , segundo uma lei de potencias ($\propto 1/\alpha$). Este resultado é válido para regras realizáveis e não-realizáveis [92].
- No caso de sinapses discretas e regras realizáveis, e para qualquer temperatura, para um certo número de exemplos $N\alpha_c(T)$, o perceptron treinado congela no estado do perceptron de referência, passa de uma fase de generalização pobre a uma de generalização perfeita. ou seja, o objetivo do aprendizado é completamente satisfeito para um número finito de exemplos por sinapse [89].
- A natureza, contínua ou discreta, das sinapses, é determinante na natureza das curvas de aprendizado. O caso de sinapses binárias é em certo sentido marginal, e apresenta-se uma transição brusca entre uma fase de generalização pobre a uma de generalização perfeita. Ao aumentarmos o número de possíveis valores que podem tomar as sinapses, o tamanho da descontinuidade na transição diminui substancialmente já para os casos de sinapses com 4 ou 6 valores possíveis [91].
- Finalmente, a forma das curvas de aprendizado não só dependem da natureza das sinapses, bem como da arquitetura da rede (que define regras realizáveis e não-realizáveis), e da regra particular a ser aprendida (reconhecimento e geração de voz, reconhecimento de escritura à mão, predição de estrutura de proteínas, etc.).

4.2 Redes de Neurônios com Estrutura Hierárquica das Memórias

O problema da generalização em redes com atratores, como o modelo de Hopfield, originou-se de forma diferente ao caso das redes tipo perceptron. Logo depois de compreender de forma bastante completa o funcionamento do modelo de Hopfield, ficou claro que a necessidade de armazenar padrões com superposição nula era uma limitação muito forte. Se os modelos deviam progredir simulando o mais possível o comportamento do

cérebro, então suas memórias deveriam possuir algum tipo de organização hierárquica interna. O cérebro não funciona com porções de informação desconexas, ele se aproveita das associações entre essas informações. Em relação a este ponto, é que surge a necessidade de organizar as memórias, estrutura-las, de acordo com suas semelhanças, em diferentes classes ou categorias. Outro problema ligado à necessidade de organizar as memórias em categorias é a possibilidade de controlar erros no reconhecimento, sejam erros devidos à presença de estados espúrios ou à saturação no número de padrões armazenados. Por exemplo, um animal não pode cometer certos erros mais de uma vez na sua vida; se suas memórias estiverem armazenadas em classes ou categorias, um erro não fatal seria confundir indivíduos da mesma classe, já se ele não conseguisse distinguir entre as classes, o erro poderia ser fatal. É o que acontece entre predador e presa. Uma possibilidade para contornar o problema das memórias ortogonais é organizá-las em uma árvore hierárquica. Essa é a forma de organização que apresentam os estados estáveis do modelo de vidro de spin de Sherrington e Kirkpatrick (SK) [69,70], conhecida como *ultrametricidade*. Se escolhermos ao acaso tres estados estáveis do modelo SK, e calculamos as superposições entre eles, os dois mais próximos ficam exatamente à mesma distância do terceiro, ou seja, formam um triângulo isósceles. A ultrametricidade nos permite, então, posicionar os estados nas “folhas” de uma árvore, logo a distância entre dois estados vem representada por quanto temos que subir desde as “folhas” até os correspondentes ramos convergirem (ver fig. 4.2). Por exemplo, na fig. 4.2, a distância entre os estados 1 e 4 é 2. Dada a semelhança entre os modelos de redes de neurônios com atratores e o modelo SK de vidros de spin, viu-se a possibilidade de modelar uma rede neuronal ultramétrica e, assim, organizar os padrões em categorias.

4.2.1 Alguns exemplos

O exemplo mais simples de árvore hierárquica consiste num padrão “ancestral” e um conjunto de “descendentes” ou “filhos”. O ancestral $\xi = \{\xi_i\}$ ($i = 1, \dots, N$) é um padrão que pode tomar os valores $\xi_i = \pm 1$ com igual probabilidade. Depois se escolhem os filhos que serão um conjunto de p padrões ξ^u . Cada bit de cada filho pode ser mais ou menos

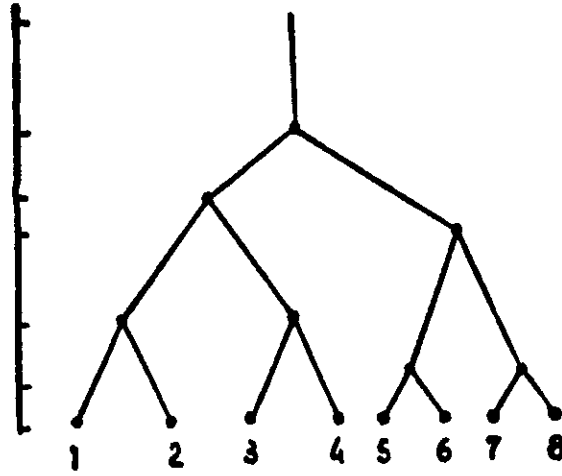


Figura 4.2: Esquema de uma árvore hierárquica.

o valor do correspondente bit do pai, de acordo com

$$\xi_i^\mu = \xi_i \eta_i^\mu, \quad (4.19)$$

com

$$P(\eta = \pm 1) = \frac{1}{2}(1 \pm a) \quad (4.20)$$

Conseqüentemente, a superposição entre dois filhos quaisquer será

$$\frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu = a^2 \quad (4.21)$$

ou seja, todos eles são equidistantes. Para poder armazenar eficientemente padrões com esta estrutura foi necessário modificar a regra de Hopfield para:

$$J_{ij} = \frac{1}{N} \xi_i \xi_j \left[1 + \frac{1}{\Delta} \sum_{\mu=1}^p (\eta_i^\mu - a)(\eta_j^\mu - a) \right] \quad (4.22)$$

Esta regra introduzida por Parga e Virasoro [71] foi estudada em detalhe por Feigel'man e Ioffe [72]. Os principais resultados são os seguintes:

- Escolhendo $\Delta = 1 - a^2$, o ancestral e os descendentes tem a mesma energia. Conseqüentemente, o modelo é equivalente a um modelo com $p + 1$ padrões, e a capacidade crítica é a usual do modelo de Hopfield, $\alpha_c \approx 0.14$.

- Mudando o valor de Δ quebra-se a degenerescência entre ancestral e descendentes.

O parâmetro pode variar no intervalo

$$1 - a^2 \leq \Delta < \frac{(1 - a^2)(1 + a)}{a} \quad (4.23)$$

e continuar mantendo os descendentes estáveis. No entanto, o ancestral é o padrão de mínima energia e os p descendentes são estados metaestáveis.

- Quando é quebrada a degenerescência, existem duas linhas críticas $\alpha(\Delta)$: $\alpha_c^1(\Delta)$ e $\alpha_c^2(\Delta)$. Acima da primeira não existem estados de reconhecimento; entre as duas somente o pai é um atrator, e embaixo da segunda podem ser reconhecidos ambos, ancestral e descendentes. Para $\Delta = 1 - a^2$ as duas linhas coincidem.

Este último ponto é, claramente, um indício de classificação de padrões. O sistema apresenta três regimes: num deles reconhece aos indivíduos dentro da classe, no outro só reconhece à classe, e no terceiro o sistema é incapaz de reconhecer ainda a classe. O modelo pode ser facilmente generalizável para estruturas de árvores mais complexas, com várias gerações. O comportamento, porém, é similar ao descrito, a diferença consistindo em possibilidades de classificação muito mais ricas quanto maior for o número de gerações na árvore.

Foram feitas várias outras propostas para armazenagem de padrões hierarquicamente organizados, por exemplo, Gutfreund [73] introduziu um modelo hierárquico no qual as diferentes gerações são armazenadas em diferentes redes, uma rede para cada geração. Desta forma, se damos como estímulo a uma das redes que armazenam ancestrais um padrão descendente, a rede reconhecerá as semelhanças e, portanto, reconhecerá a classe antes do que os indivíduos. Outra possibilidade explorada por Tamarit e Curado [54,74] é armazenar padrões correlacionados aos pares. Este modelo também apresenta três regimes: reconhecimento dos indivíduos do par, reconhecimento do par, embora não dos indivíduos (classe), e um regime “paramagnético” no qual o sistema não reconhece nem a classe nem o indivíduo. Finalmente, Virasoro estudou, em uma série de trabalhos [75, 40,76], modelos hierárquicos que podem ter relevância no estudo de doenças cerebrais, como *agnosia* e *prosopagnosia*, quando um certo número das sinapses são eliminadas (diluição).

Todos os modelos discutidos (exceto o da ref. [74] no qual os padrões são estatisticamente correlacionados) apresentam, como ponto em comum, a necessidade de se modificar a regra de aprendizado de Hopfield, com o objetivo de armazenar padrões hierarquicamente organizados. Os ancestrais, ou padrões representativos das classes ou categorias, são explicitamente armazenados na rede. Desta forma a capacidade de categorizar não é espontânea nestes modelos, ela é imposta como condição na prescrição da regra de aprendizado. Neste ponto fazemo-nos a seguinte pergunta: Será possível fazer um modelo que seja capaz de criar um estado representativo de um conjunto de padrões similares, sem que esse estado apareça explicitamente na regra de aprendizado? Este é basicamente o problema de *generalização* em redes de neurônios com atratores, ao qual vamos dedicar a próxima seção .

4.3 Generalização numa Rede de Neurônios Analógica

Vimos na seção (4.1), que a capacidade de generalização numa rede tipo perceptron está relacionada à capacidade de inferir uma regra a partir de um conjunto de exemplos da mesma. Em redes que apresentam atratores, como as redes analógicas estudadas nos capítulos anteriores, as possibilidades de generalização têm sido estudadas fundamentalmente no que diz respeito às possibilidades de criar *categorias*, de agrupar os padrões em diferentes *classes*. Na seção anterior, vimos como modificando a prescrição para as sinapses, pode-se obter uma rede de neurônios com memórias organizadas hierarquicamente em categorias. No contexto de redes de neurônios com atratores, então , a capacidade de generalização referir-se-á à capacidade do sistema em criar um conceito, que seja representativo de um conjunto de padrões . Em outras palavras, iremos dizer que a rede generaliza se é capaz de inferir as características comuns aos padrões , independentemente dos detalhes que cada um deles possa apresentar.

Nesta linha, J.F.Fontanari [78] estudou a capacidade de generalização num modelo de Hopfield a temperatura zero, com uma organização hierárquica das memórias em dois níveis. Utilizando a técnica de réplicas, demonstrou que o sistema consegue generalizar, desde que os conjuntos de memórias do nível mais baixo da hierarquia apresentem um

número mínimo de padrões . A técnica de réplicas prediz uma transição descontínua de uma fase de reconhecimento de indivíduos, ou uma fase vidro de spin dependendo do valor de α , a uma fase de generalização . Na fase de generalização , quando o número de exemplos cresce o erro de generalização decai exponencialmente. E.N.Miranda [79] fez um estudo numérico do modelo anterior, confirmando as previsões gerais do trabalho analítico de Fontanari; entretanto as simulações por ele feitas com redes de até 512 neurônios detectaram uma transição contínua, e um decaimento do erro de generalização como uma potência do número de exemplos. Num trabalho recente, mediante simulações com redes de até 8192 neurônios, M.C.Branchtein e J.J.Arenzon [80] estudaram o mesmo modelo, utilizando como condição inicial não só um conceito (como no trabalho de Miranda), como também fizeram simulações partindo dos exemplos. Os dois casos apresentam curvas de generalização diferentes, pelo menos quando $\alpha \rightarrow 0$. Partindo de um exemplo, eles recuperam a transição descontínua prevista pela técnica de réplicas. No caso da condição inicial ser um conceito, a transição é contínua. Na fase de generalização , as curvas coincidem para ambos casos, e o erro de generalização decai com uma lei exponencial do número de exemplos, confirmando novamente os resultados analíticos de Fontanari.

Dando continuidade ao estudo de diversas propriedades em redes de neurônios analógicas, vamos apresentar uma análise das possibilidades de generalização neste tipo de redes [81].

4.3.1 O Modelo

O modelo é , no que diz respeito a sua dinâmica, o mesmo que já foi estudado nos capítulos 2 e 3. A mudança está na regra de aprendizado, onde agora voltamos à regra mais simples possível, a regra de Hopfield. Só para lembrar, a dinâmica é definida por um conjunto de mapas acoplados

$$x_i(t+1) = \tanh(gh_i(t)), \quad i = 1, \dots, N \quad (4.24)$$

sendo g o parâmetro de ganho da função de transferência, e

$$h_i(t) = \sum_{j=1}^N J_{ij} x_j(t) \quad (4.25)$$

representa o potencial pós-sináptico, ou campo local, no neurônio i -ésimo no tempo t . Neste caso, J_{ij} , será a matriz simétrica de Hopfield (1.9),

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad J_{ii} = 0 \quad (4.26)$$

onde $\{\xi_i^\mu = \pm 1\}, \mu = 1, \dots, p$, representam os padrões armazenados.

Devemos gerar uma rede com uma arquitetura adequada à tarefa de generalização. Para isso, primeiramente geramos uma configuração aleatória $\{\eta_i\}$, com $\eta_i = \pm 1$ com igual probabilidade. Esta configuração apresentará as características comuns a um conjunto de s padrões. As componentes destes s padrões serão variáveis estatisticamente independentes, escolhidas de acordo com a seguinte distribuição de probabilidades:

$$P(\xi_i^\mu) = \frac{1}{2} (1 + \eta_i b) \delta(\xi_i^\mu - 1) + \frac{1}{2} (1 - \eta_i b) \delta(\xi_i^\mu + 1) \quad (4.27)$$

onde $b \in [0, 1]$. Vamos considerar o caso no qual s é finito, com $p = \alpha N$. As componentes dos restantes $(p - s)$ padrões também são variáveis estatisticamente independentes, tomando os valores ± 1 com igual probabilidade. Com esta escolha para os padrões, o sistema apresenta a estrutura hierárquica mais simples possível compatível com o armazenamento de um número de padrões proporcional a N (ver fig. 4.3). Podemos pensar na configuração $\{\eta\}$, como representando um *conceito*. É importante notar que este conceito *não está explicitamente representado na regra de aprendizado*, diferentemente do que acontecia com os modelos hierárquicos para estudar memória associativa. No entanto, o conceito é introduzido na rede por meio de s exemplos do mesmo. O parâmetro b mede a proximidade, a semelhança, dos exemplos com o conceito que os representa. Vamos considerar o processo de aprendizado como uma estratégia de treinamento, ou seja, mediante a apresentação de um certo número de exemplos, vamos procurar o número mínimo necessário para que a rede tenha o conceito como um atrator. A medida que os exemplos são apresentados, a rede deverá modificar sua estrutura de atratores, de forma a gerar um atrator particular para o conceito.

De acordo com a estratégia acima proposta, vamos nos interessar pela superposição, a tempos longos, entre o estado da rede e o conceito que ela deverá gerar.

$$m_c(t) = \frac{1}{N} \sum_{i=1}^N \eta_i x_i(t) \quad (4.28)$$

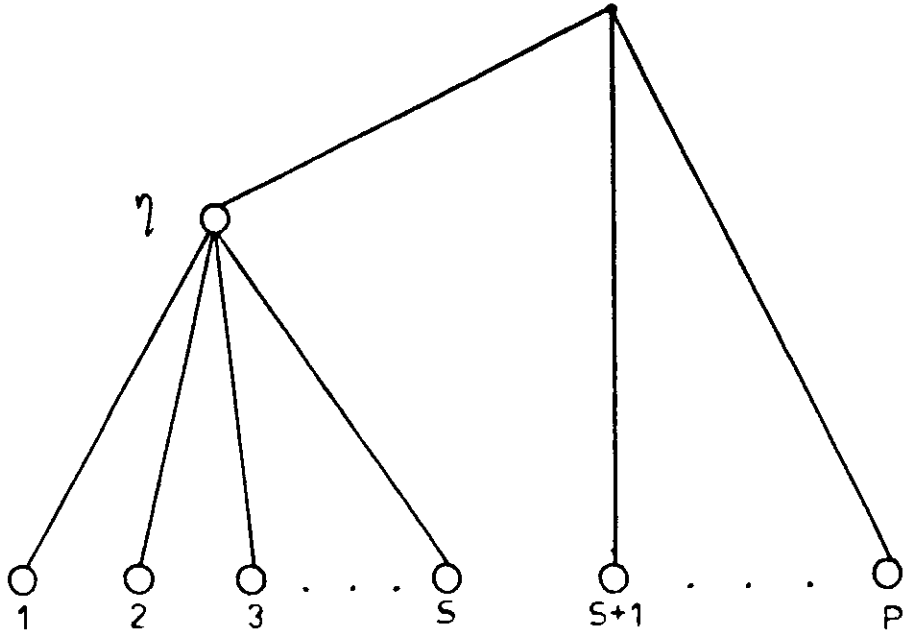


Figura 4.3: Organização hierárquica das memórias na rede analógica

Neste contexto, vamos definir o *erro de generalização*, ϵ_g , como o número de sítios diferentes entre o conceito e o estado da rede para tempos longos (distância de Hamming). Conseqüentemente, ϵ_g está relacionado ao valor de m_c nos pontos fixos dos mapas (4.24) mediante

$$\epsilon_g = \frac{1 - m_c}{2} \quad (4.29)$$

4.3.2 Análise Estatística e Equações dos Pontos Fixos

Vamos utilizar a técnica introduzida no capítulo 2 na análise de uma rede analógica com diluição simétrica. Primeiramente, vamos considerar as superposições entre cada uma das memórias e um ponto fixo das eqs.(4.24)

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu x_i^* \quad \mu = 1, \dots, p \quad (4.30)$$

Procuramos um conjunto de equações auto-consistentes para estas superposições. Em virtude das eqs.(4.24), (4.26) e (4.30), podemos escrever as equações para o ponto fixo

$$x_i^* = \tanh\left(g \sum_{\mu=1}^p \xi_i^\mu m^\mu\right), \quad i = 1, \dots, N \quad (4.31)$$

Vamos supor que as superposições são quantidades auto-mediadas, ou seja, no limite $N \rightarrow \infty$, uma realização particular das variáveis aleatórias “congeladas” é representa-

tiva do comportamento médio das mesmas. Conseqüentemente, podemos substituir a soma no índice i pelo valor médio

$$m^\mu = \langle\langle \xi^\mu \tanh(g \sum_{\mu=1}^p \xi^\mu m^\mu) \rangle\rangle, \quad \mu = 1, \dots, p \quad (4.32)$$

$\langle\langle \dots \rangle\rangle$ representa uma média nas variáveis $\{\xi_i^\mu\}$ e $\{\eta_i\}$, nessa ordem.

Já que estamos interessados em reconhecer o conceito (não as memórias individuais), e ele apresenta a mesma superposição com todos os exemplos, consideraremos soluções que tenham a mesma superposição de $O(1)$ com os s exemplos, e uma superposição de $O(N^{-1/2})$, com os restantes $(p - s)$ padrões. Em linguagem mais técnica, estamos interessados em *soluções simétricas* da forma

$$sm_s = \langle\langle z_s \tanh \left[g(m_s z_s + \sum_{\mu>s} \xi^\mu m^\mu) \right] \rangle\rangle \quad (4.33)$$

onde $z_s = \sum_{\mu=1}^s \xi^\mu$. Vamos fazer uma aproximação na qual, para tomar a média nas primeiras s variáveis ξ^μ , z_s é considerada uma variável gaussiana. O valor médio de z_s será $sb\eta$, e a variância $\sigma_1^2 = s(1 - b^2)$. Notar que até aqui não foi tomada a média na variável η . Nos limites $p \rightarrow \infty$, $N \rightarrow \infty$, com $\alpha = p/N$ finito, pode-se demonstrar [16] que o segundo termo do argumento da função de transferência representa um ruído gaussiano, com média zero e variância $\sigma_2^2 = \alpha(g - C)/g(1 - C)^2$. C vem definida pela equação

$$C = \langle\langle g \operatorname{sech}^2 \left[g(m_s z_s + \sum_{\mu>s} \xi^\mu m^\mu) \right] \rangle\rangle \quad (4.34)$$

Tomando as médias no ruído gaussiano, z_s e η , nessa ordem, obtemos um conjunto de equações não-lineares acopladas

$$m_s = \frac{1}{2} \int Dz \int Dh \{ [b + \sigma_1 z/s] \tanh(\Theta^+) - [b - \sigma_1 z/s] \tanh(\Theta^-) \} \quad (4.35)$$

$$C = \frac{g}{2} \int Dz \int Dh [\operatorname{sech}^2(\Theta^+) + \operatorname{sech}^2(\Theta^-)] \quad (4.36)$$

com $\Theta^\pm = m_s(z\sigma_1 \pm sb) + \sigma_2 h$, e

$$Dz = \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2), \quad Dh = \frac{dh}{\sqrt{2\pi}} \exp(-h^2/2) \quad (4.37)$$

Fazendo uma análise similar para a superposição com o conceito, obtemos

$$m_c = \frac{1}{2} \int Dz \int Dh [\tanh(\Theta^+) - \tanh(\Theta^-)] \quad (4.38)$$

Nas próximas seções estudaremos numericamente as soluções das equações para m_s , C e m_c .

4.3.3 O limite $g \rightarrow \infty$

No limite $g \rightarrow \infty$ a \tanh vira uma função sinal, e as variáveis dinâmicas tornam-se discretas. Neste limite as equações para m_s , m_c e C se simplificam

$$m_s = \sqrt{\frac{2}{\pi}} \frac{m_s(1-b^2)}{\Delta} \exp\left[-\frac{(m_s sb)^2}{2\Delta^2}\right] + b \operatorname{erf}\left[\frac{m_s sb}{\sqrt{2}\Delta}\right] \quad (4.39)$$

$$C = \sqrt{\frac{2}{\pi\Delta^2}} \exp\left[-\frac{(m_s sb)^2}{2\Delta^2}\right] \quad (4.40)$$

$$m_c = \operatorname{erf}\left[\frac{m_s sb}{\sqrt{2}\Delta}\right] \quad (4.41)$$

com $\Delta^2 = s(1-b^2)m_s^2 + \alpha/(1-C)^2$.

Na fig. 4.4 podemos ver o erro de generalização ϵ_g em função do número de exemplos s , para diferentes valores de α e para $b = 0.2$. Ao aumentar o valor de α , aumenta o ruído gerado pelos padrões com superposição nula e, conseqüentemente, o sistema deve experimentar um número cada vez maior de exemplos para começar a generalizar. Para cada valor particular de α , existe um valor crítico de exemplos s_c , necessário para que a rede comece a reconhecer a categoria de forma contínua. Este resultado é diverso do que acontece no modelo de Hopfield, no qual a transição é descontínua [78]. Para $s \gg 1$, $m_s \rightarrow b$, $m_c \rightarrow 1$ e ϵ_g decai exponencialmente

$$\epsilon_g \approx \sqrt{\frac{\gamma_1}{2\pi}} \exp\left[-\frac{1}{2\gamma_1}\right] \quad (4.42)$$

com $\gamma_1 = (1-b^2)/(sb^2)$. Utilizando o fato de que, perto da transição, $m_s \ll 1$, expandimos as equações para m_s e C , obtendo uma relação entre os valores críticos α_c , b_c e s_c

$$\alpha_c = \frac{2}{\pi} b^4 (s_c - 1)^2 \quad (4.43)$$

Este resultado está de acordo com resultados prévios obtidos por Fontanari e Köberle [93] e Erichsen e Theumann [94]. Na fig. 4.5 apresentamos, para diferentes valores de α_c , as curvas críticas, acima das quais aparecem soluções não nulas para a superposição

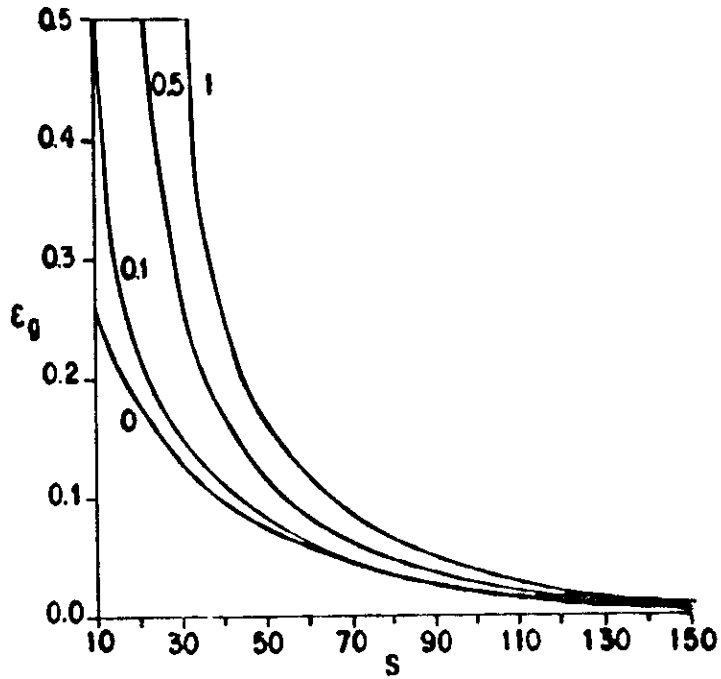


Figura 4.4: Erro de generalização ϵ_g vs. número de exemplos s no limite $g \rightarrow \infty$, para $b=0.2$ e $\alpha = 0, 0.1, 0.5 \in 1$

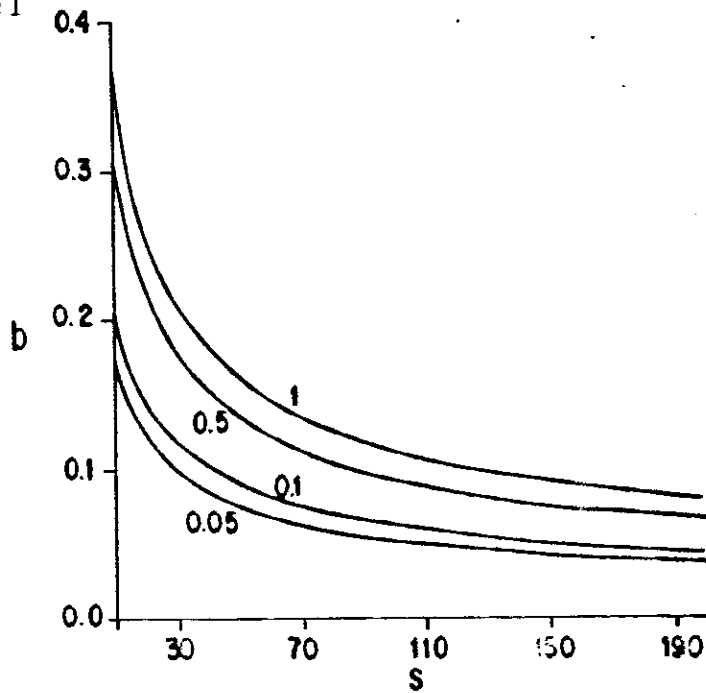


Figura 4.5: As curvas críticas $b(s)$ no limite $g \rightarrow \infty$, para $\alpha_c = 0.05, 0.1, 0.5 \in 1$. O sistema começa a generalizar acima da correspondente curva.

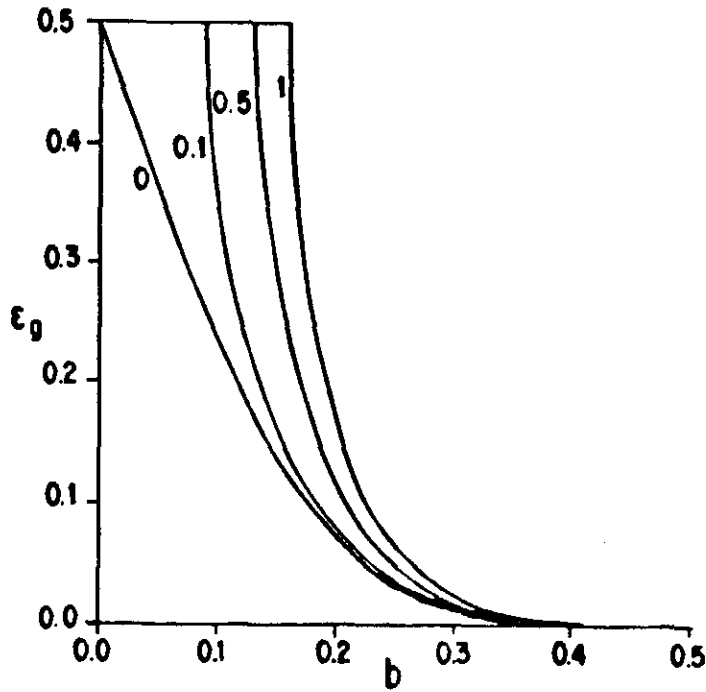


Figura 4.6: Erro de generalização ϵ_g vs. b no limite $g \rightarrow \infty$, para $s=50$ e $\alpha = 0, 0.1, 0.5 e 1$.

m_s . É interessante notar que α_c diverge com s_c , apesar dele não estar relacionado com a capacidade de armazenamento crítica do modelo. Abaixo destas curvas, o sistema apresenta soluções de reconhecimento, ou vidro de spin, dependendo do valor de α . Um estudo sistemático destas soluções ainda não foi feito.

Se exigimos que o sistema não cometa erros na representação do conceito, apresentando um número de exemplos fixos, então necessariamente se devem cumprir as condições $b \approx 1$ e $\alpha \approx 0$. O erro de generalização apresenta a mesma forma funcional da eq.(4.42)

$$\epsilon_g \approx \sqrt{\frac{\gamma_2}{2\pi}} \exp\left[-\frac{1}{2\gamma_2}\right] \quad (4.44)$$

com $\gamma_2 = \frac{\alpha + sb^2(1-b^2)}{s^2b^4}$. Na fig. 4.6 pode se ver ϵ_g como função de b para $s=50$ e para diferentes valores de α .

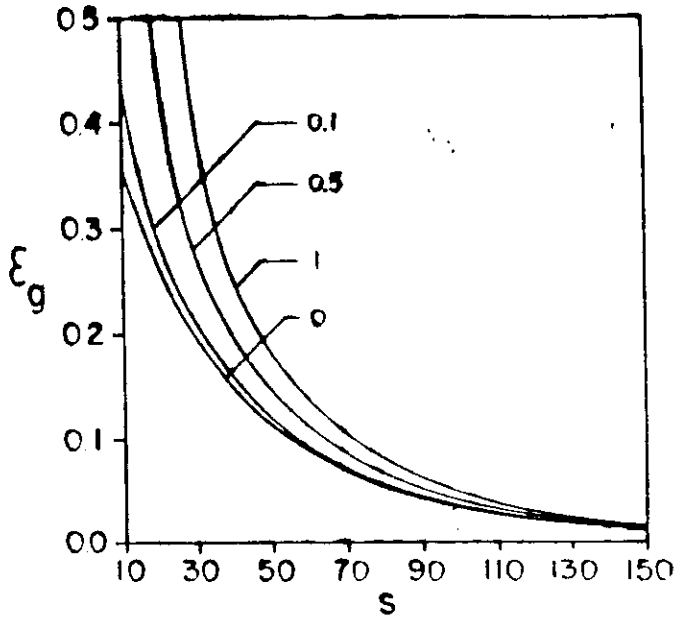


Figura 4.7: Erro de generalização ϵ_g vs. número de exemplos s , para $g=1$, $b=0.2$ e $\alpha = 0, 0.1, 0.5 e 1$.

4.3.4 Resultados para g finito

Na fig. 4.7 temos a resposta do modelo através do erro de generalização em função do número de exemplos quando $g=1$, $b=0.2$ e diferentes valores de α . Da comparação destes resultados com os obtidos para $g \rightarrow \infty$, notamos uma melhora na capacidade de generalizar da rede: o sistema necessita um número menor de exemplos para começar a generalizar. Esta melhora pode-se observar claramente na fig. 4.8, na qual comparam-se as curvas críticas $b_c(s_c)$, para $\alpha = 0.1$ e dois valores de g . Finalmente, na fig. 4.9 comparamos ϵ_g vs s , para $g = 1$ e $g \rightarrow \infty$, sendo $b = 0.1$ e $\alpha = 0.1$. Notamos que, se bem que para $g = 1$ a transição acontece para um número de exemplos menor do que para $g \rightarrow \infty$, neste último caso o sistema funciona melhor se requerermos um erro pequeno ($s \gg 1$).

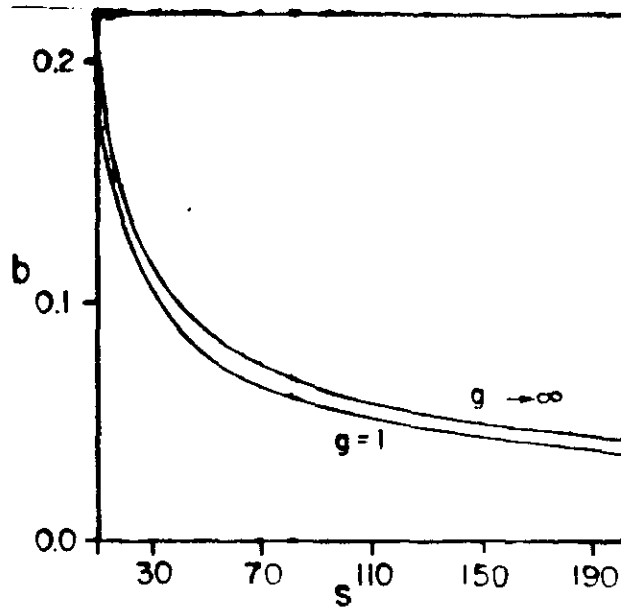


Figura 4.8: As curvas críticas $b(s)$, acima das quais o sistema generaliza, para $g=1$ e $g \rightarrow \infty$, com $\alpha = 0.1$.

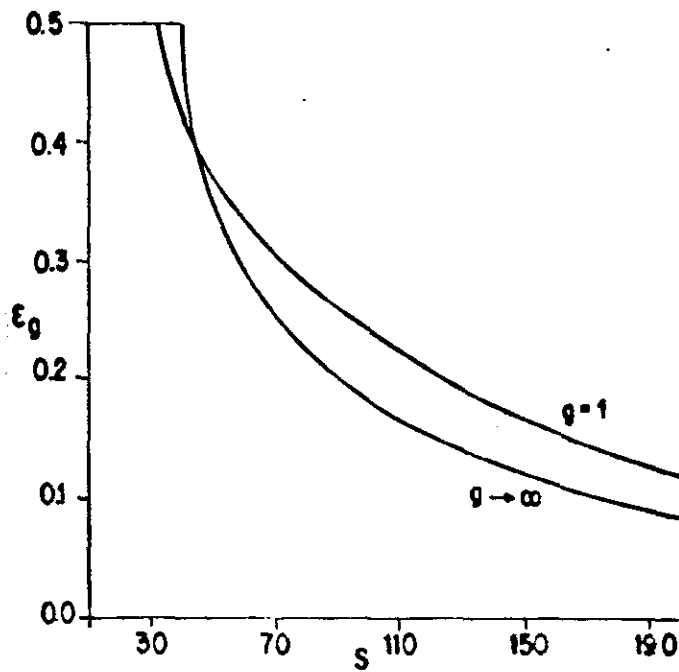


Figura 4.9: Erro de generalização ϵ_g vs. número de exemplos s , para $g = 1$ e $g \rightarrow \infty$, sendo $b = 0.1$ e $\alpha = 0.1$.

4.3.5 Conclusões

Neste trabalho foi demonstrado que uma rede de neurônios analógica é capaz de classificar uma dada quantidade de estímulos (entradas), de acordo com a sua proximidade. Ou seja, a rede é capaz de generalizar sempre que um mínimo número de exemplos da classe lhe seja apresentado. Se compararmos com o limite discreto do modelo $g \rightarrow \infty$, valores finitos do ganho provocam uma pequena melhoria na capacidade de generalização. Isto quer dizer que um menor número de exemplos da classe é necessário para que o sistema comece a generalizar. Para um futuro trabalho fica uma análise detalhada das diferentes fases destes modelos. Também seria interessante estudar a performance, no que diz respeito a generalização, de modelos com estrutura hierárquica mais complexa. Finalmente, ainda não tem sido considerada a possibilidade de estudar estes modelos por meio de uma análise no espaço das interações sinápticas, como foi feito para os modelos tipo perceptron. Acharmos que uma análise nessa linha seria muito esclarecedora no que diz respeito as capacidades de generalização das redes de neurônios com atratores.

4.4 Discussão e Perspectivas

No presente capítulo temos analisado o problema do aprendizado e generalização em redes de neurônios, considerando de forma diferente os problemas segundo a arquitetura específica. Fizemos uma revisão do formalismo desenvolvido para arquiteturas tipo perceptron, e depois introduzimos e analisamos um modelo explícito, no caso de redes com atratores. As duas formas de abordagem do problema apresentam diferenças claras. O formalismo mecânico-estatístico desenvolvido para arquiteturas tipo perceptron é mais geral e poderoso que o utilizado nas redes com organização hierárquica das memórias. Este fato talvez seja consequência do caráter um tanto restrito da função de generalização em redes com atratores. Nestas redes, como já foi apontado, generalização é sinônimo de categorização ou classificação. A arquitetura especial das redes tipo perceptron as torna adequadas para realizar diferentes tipos de cálculos, que nós chamamos de “regras”. Segundo a opinião de D.Amit [11], as redes tipo perceptron são particularmente adaptadas para realizar “cálculos espontâneos”, e como sistemas

de entrada-saída. No entanto as redes com atratores estariam melhor adaptadas como sistemas internos, capazes de realizar funções superiores. As duas arquiteturas são , neste contexto, complementares, e poderiam formar parte de uma estrutura maior.

Um ponto de vista um pouco diferente é apresentado por G.Parisi num trabalho recente [95]. Ele apresenta um formalismo estatístico no qual as variáveis são as diferentes arquiteturas possíveis, e o objetivo é estudar, em função da arquitetura, a propensão de uma rede para generalizar uma certa regra. Esta “propensão ” é quantificada definindo uma “entropia” de forma adequada. Se duas arquiteturas diferentes apresentam uma propensão equivalente , a definição de uma medida da distância entre as arquiteturas permite distinguir a performance das mesmas sobre diferentes conjuntos de regras. O formalismo utilizado está baseado na distribuição estatística de Boltzmann-Gibbs. No entanto, é possível considerar tais problemas sob o ponto de vista de outros formalismos estatísticos baseados na teoria da informação . Um destes formalismos, introduzido por C.Tsallis [96], generaliza o conceito de entropia e contém a estatística de Boltzmann-Gibbs como caso particular. Estas estatísticas poderiam ser mais adequadas do que a de Boltzmann-Gibbs como estimativas da propensão de generalização das regras [97,98]. Já entrando numa fase mais especulativa, Parisi se pergunta como definir a inteligência de uma arquitetura. Uma possibilidade consiste em definir uma arquitetura que será a mais inteligente. A inteligência de qualquer outra viria dada pela distância à anterior. Para definir “arquitetura mais inteligente”, o autor especula que conceitos, tais como *complexidade algorítmica* e *profundeza lógica* (logical depth) de uma regra, são grandezas importantes. A complexidade algorítmica de uma regra é definida como o comprimento do menor código que consegue calcular a regra, por exemplo, o menor número de sinapses necessárias num perceptron para realizar certo cálculo. A profundeza lógica é o tempo de cálculo, o número de operações , necessário para calcular a saída para todas as possíveis entradas, numa máquina que utiliza o menor código possível. Em outras palavras, o número mínimo de bits necessários para identificar uma regra representa a sua complexidade algorítmica. O tempo necessário para utilizar esta informação é a sua profundeza lógica. Estes dois conceitos são , de certa forma, complementares. Muitas regras com complexidade algorítmica pequena têm profundeza

lógica muito longa. Por exemplo, o programa para calcular os primeiros 10^{10} dígitos de π pode ser relativamente curto, no entanto o tempo de cálculo pode ser muito longo. Uma generalização inteligente para uma máquina seria escolher a regra mais simples compatível com os dados que se querem analisar.

Estas especulações sobre o problema de generalização em redes de neurônios deixam-nos com a sensação de que o problema pode ser ainda muito explorado, e sem dúvida futuros trabalhos irão esclarecer o papel da generalização em relação ao comportamento inteligente.

Capítulo 5

Sistemas Dinâmicos Complexos: Mapas Unidimensionais e Automata Celulares

Os sistemas dinâmicos podem apresentar uma grande variedade de comportamentos nas suas evoluções temporais. Este comportamento depende de muitas condições, algumas gerais, outras particulares de cada sistema. Por exemplo, uma condição geral é a classificação dos sistemas dinâmicos em *conservativos* e *dissipativos*. Sem tentarmos ser rigorosos, os sistemas conservativos preservam, em sua evolução dinâmica, o volume no espaço de fases. Se diz que “conservam” o espaço de fases. Em termos energéticos, eles conservam a energia (por exemplo, o oscilador harmônico simples). Os dissipativos evoluem num subespaço do espaço de fases inicial, as vezes um simples ponto, e não conservam a energia (por exemplo um oscilador amortecido). Uma outra grande classificação é entre sistemas *lineares* e *não-lineares*. Os sistemas lineares são

aqueles cuja evolução dinâmica é governada por forças lineares (oscilador harmônico). Diferentemente, os sistemas não-lineares apresentam não-linearidades nas equações que definem sua dinâmica (oscilador anarmônico). Os sistemas lineares apresentam uma fenomenologia mais simples do que os não-lineares e, conseqüentemente, as suas propriedades são muito mais conhecidas do que as dos sistemas não-lineares. Nos capítulos anteriores vimos exemplos de uma família particular de sistemas dinâmicos não-lineares e dissipativos: as redes de neurônios. Vimos como, para tempos de evolução longos, a dinâmica pode ser muito simples: pontos fixos ou ciclos limite. Porém, as vezes, sob certas condições particulares, os sistemas podem apresentar uma dinâmica muito mais complexa do que as anteriores. É o caso das dinâmicas caóticas. A denominação *sistemas complexos* faz referência ao conjunto de sistemas dinâmicos que podem apresentar, para certas condições específicas, dinâmicas não triviais como, por exemplo, caos. Uma introdução sistemática aos sistemas dinâmicos está fora dos objetivos da presente tese. O leitor interessado pode encontrar excelentes introduções aos sistemas dinâmicos nas referências [99], e aos sistemas caóticos, em particular, nas [100,101].

No presente capítulo vamos apresentar um estudo da interação entre dois sistemas dinâmicos complexos: um mapa unidimensional e um autômato celular. Os mapas unidimensionais são, surpreendentemente, os exemplos mais simples de sistemas complexos [102]. São definidos pela iteração (tempo discreto) de uma variável x que, geralmente, toma valores reais num certo intervalo,

$$x_{t+1} = F(x_t, r) \quad (5.1)$$

r é um parâmetro que pode representar uma grande variedade de condições do sistema, por exemplo, temperatura num sistema físico. O exemplo mais conhecido de mapa unidimensional é a *parábola logística*, cujas características básicas serão descritas na próxima seção. Um exemplo muito mais elaborado de sistema dinâmico complexo é representado pelos *automata celulares*, sistemas com muitos graus de liberdade nos quais o tempo, o espaço, e cada grau de liberdade são discretizados. Na seção 5.2 resumimos as principais características dos automata celulares. Estudos completos podem se encontrar nas referências [103,104]. Na seção 5.3 introduzimos e estudamos a dinâmica

que resulta da interação entre a parábola logística e um autômato celular determinista.

5.1 A parábola logística

A parábola logística é um sistema dinâmico não-linear, dissipativo, em tempo discreto. O valor da variável x_t no tempo t é mapeado ao novo valor x_{t+1} no tempo $t + 1$, de acordo com a função não-linear

$$x_{t+1} = rx_t(1 - x_t) \quad (5.2)$$

O parâmetro de controle r satisfaz $0 < r \leq 4$, para os quais x_t fica restrito ao intervalo $0 \leq x_t \leq 1$ se assim for x_0 . Para uma condição inicial x_0 no intervalo $(0, 1)$, estamos interessados no comportamento do mapa 5.2 depois de passadas um número muito grande de iterações, especificamente, no limite $t \rightarrow \infty$.

A dinâmica do mapa 5.2 depende fortemente do valor do parâmetro r . Para $0 < r < 1$, x_t vai assintoticamente (para $t \rightarrow \infty$) para 0, independentemente da condição inicial. Ou seja, após o transiente, todos os pontos no intervalo $(0, 1)$ são atraídos para o ponto fixo x^* em $x = 0$. Formalmente

$$\lim_{t \rightarrow \infty} x_t = x^* = 0. \quad (5.3)$$

Uma simples análise de estabilidade linear mostra que o zero é estável no intervalo $(0, 1)$. Para $1 < r < 3$ encontramos um ponto fixo atrator que depende do valor de r . Podemos calculá-lo facilmente já que, num ponto fixo, $x_t = x_{t+1} = x^*$. Substituindo na eq.(5.2) obtemos que

$$\lim_{t \rightarrow \infty} x_t = x^*(r) = 1 - \frac{1}{r} \quad (5.4)$$

De acordo com a relação anterior, quando variamos r entre 1 e 3, o valor do ponto fixo x^* vai desde 0 até $2/3$. É importante notar que no intervalo $(1, 3)$ o ponto fixo anterior é estável. O zero continua a ser um ponto fixo, porém, neste intervalo é instável. A mudança de estabilidade em $r = 1$ é chamada *bifurcação*. Uma bifurcação mais interessante acontece em $r = 3$. Agora, em lugar de se aproximar de um ponto fixo, o comportamento assintótico do mapa oscila entre dois valores: o sistema apresenta um

ciclo limite atrator de período 2. Ao continuar aumentando r novas bifurcações acontecem. Para valores de $r=3.5$ e maiores, o sistema bifurca em ciclos limite atratores de período 4, 8, 16, 32, ..., 2^n . Todas estas bifurcações acontecem até um valor de $r_c \approx 3.57$. Acima de r_c , o conjunto atrator, para muitos valores de r , não apresenta periodicidade alguma. O conjunto consiste de uma seqüência de pontos x_t que nunca se repetem. Para estes valores de r , a simples parábola logística apresenta *caos determinista*, e o conjunto atrator (muito mais complexo que um ponto fixo, ou um ciclo limite) é conhecido como *atrator estranho*.

Este complexo comportamento aperiódico motivou o nome “*caos*”. No entanto, uma evidencia analítica de que o atrator é mesmo caótico pode se obter através da medida da sensibilidade as condições iniciais. Um atrator estranho apresenta grande sensibilidade as condições iniciais; dois pontos arbitrariamente próximos num tempo t não permanecem próximos a medida que o tempo transcorre, eles se separam exponencialmente no tempo. Tecnicamente, a sensibilidade as condições iniciais é determinada pelo valor do *expoente de Liapunov*, λ , definido como

$$\lambda(r) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \ln |F'(r, x_t)| \quad (5.5)$$

Com esta definição, um valor de $\lambda > 0$ implica que pontos inicialmente próximos se separam exponencialmente. Desta forma, às regiões caóticas correspondem valores de $\lambda > 0$, e às regiões periódicas $\lambda < 0$.

Uma outra questão interessante destes sistemas não-lineares é a forma pela qual tais sistemas passam de um movimento periódico para um movimento caótico. Na parábola logística, isto acontece por meio de uma cascata de bifurcações com duplicações de período, ou seja, por meio de uma sucessão de ciclos limite, com períodos que se incrementam como 2^n . Feigenbaum [105] demonstrou que os valores do parâmetro r , para os quais acontecem as bifurcações, convergem geometricamente. Se definimos

$$\delta_n = \frac{r_n - r_{n-1}}{r_{n+1} - r_n}, \quad (5.6)$$

ele obteve

$$\lim_{n \rightarrow \infty} \delta_n \equiv \delta = 4.669... \quad (5.7)$$

Uma observação de Feigenbaum ainda mais importante foi que δ não depende dos detalhes da parábola logística; qualquer função que possua um máximo “genérico” (derivada segunda não nula) terá a mesma δ . Desta forma δ é uma grandeza *universal*.

Finalmente, é importante ressaltar que a cascata com duplicação de período não é o único caminho ao caos. Tem sido encontradas outras formas, tais como *quasiperiodicidade* [106], *intermitência* [107], e o *gap road* [108].

5.2 Automata Celulares

Os automata celulares (AC) são sistemas dinâmicos formados por N unidades em interação. Como foi apontado acima, nos automata celulares, o espaço, o tempo e o número de estados por unidade são *discretos*. Cada unidade discreta do espaço (conhecida como “célula”) possui um certo número k de entradas, e uma regra que determina o seu estado no intervalo de tempo seguinte em função das entradas. Todas as células atualizam seus estados simultaneamente. Cada uma das N células pode tomar r estados diferentes, $\sigma_i = 0, \dots, r-1$; $i = 1, \dots, N$. Então um autômato celular determinista fica definido pela regra

$$\sigma_i(t+1) = f_i(\sigma_j(t); j = 1, \dots, k), \quad (5.8)$$

onde as funções f_i tomam os valores $0, \dots, r-1$. Muitas generalizações desta definição podem ser feitas [110], por exemplo podemos definir um autômato probabilístico, incluindo uma dependência num número aleatório z

$$\sigma_i(t+1) = f_i(\sigma_j(t); j = 1, \dots, k; z) \quad (5.9)$$

Como podemos notar, a definição de automata celulares é muito geral e, como consequência, eles apresentam um comportamento dinâmico muito rico.

Para k entradas por célula, temos um total de r^k configurações de entrada, e para cada uma destas configurações associamos um dos r possíveis estados de saída. Então temos um total de r^{r^k} regras. Na tabela 5.1 apresentamos dois exemplos de regras com $k=3$ e $r=2$, as regras 90 e 22 decompostas em forma binária. Podemos ver que a regra

regra	111	110	101	100	011	010	001	000
90	0	1	0	1	1	0	1	0
22	0	0	0	1	0	1	1	0

Tabela 5.1: As regras C=90 e C=22 decompostas em forma binária.

90 dá 1 sempre que as duas células vizinhas estejam em estados diferentes, e a regra 22 dá 1 sempre que somente uma das três entradas seja 1. As regras também podem apresentar simetrias. Por exemplo, invertendo todos os 0's a 1's, e todos os 1's a 0's nas entradas e saídas da tabela 5.1, a regra 90 vira a regra 165, e a regra 22 vira a regra 151, no entanto a regra 51 é invariante frente a esta simetria. A consideração das simetrias simplifica a análise das regras.

Wolfram propôs uma classificação dos automata celulares de acordo com os seus comportamentos para tempos longos. A fig.5.1 mostra os diagramas espaço-temporais de alguns automata celulares com $k=5$ e $r=2$, e configuração inicial aleatória. De acordo com Wolfram, podemos observar quatro classes:

- Classe 1 (regras 4 e 16 na fig.5.1): apresenta um estado final homogêneo.
- Classe 2 (regras 8 e 24): produz estruturas periódicas separadas espacialmente.
- Classe 3 (regras 2,6,10,12,14,18 e 22): produz padrões desordenados que apresentam buracos de diferentes tamanhos.
- Classe 4 (regra 20): produz estruturas complexas localizadas.

As classes com comportamento mais interessante, são claramente, as 3 e 4. À classe 3 pode ser associado um comportamento caótico. Também é possível observar invariância de escala nos padrões produzidos pelas regras da classe 3. No limite de tempo infinito, esta invariância produz estruturas *auto-similares*, típicas dos sistemas fractais. Uma estrutura é auto-similar quando vistas de regiões com diferentes ampliações (e a mesma resolução) são indistinguíveis. A classe 4 apresenta comportamentos ainda mais complexos, com estruturas localizadas no espaço e no tempo, que dependem muito fortemente das condições iniciais. Alguns pesquisadores conjecturam que estes automata podem funcionar como computadores universais.

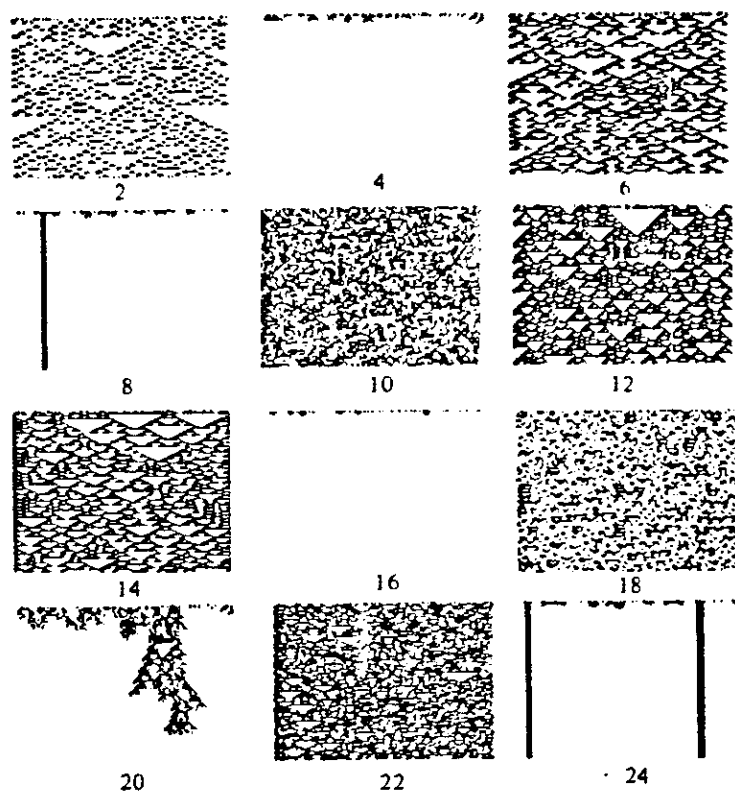


Figura 5.1: Diagramas espaço-temporais para algumas regras com $k=5$ e $r=2$. No sentido vertical, a evolução temporal.

A classificação anterior é pouco precisa quantitativamente. No entanto, foram introduzidas diversas grandezas para estudar o comportamento dos automata celulares, por exemplo, diferentes entropias, e mesmo uma grandeza similar ao expoente de Liapunov λ definido para sistemas dinâmicos contínuos. Uma possível definição de λ em automata celulares [109], consiste em considerar duas configurações iniciais idênticas, exceto em uma única célula. Deixar evoluir os sistemas, e graficar as células diferentes para cada tempo. λ resulta ser a inversa da inclinação do contorno da figura produzida. Fisicamente, λ é a velocidade com a qual o dano inicial se propaga, ou seja, uma medida da sensibilidade às condições iniciais. A idéia de propagação de danos em sistemas discretos tem sido generalizada e amplamente aplicada ao estudo de diferentes sistemas, como automata celulares probabilísticos [111] e sistemas magnéticos [54].

5.3 Interação entre a Parábola Logística e um Autômato Celular

As vezes, é possível descrever a dinâmica de um sistema físico particular, tomando como ponto de partida um modelo conhecido, e acrescentando flutuações na dinâmica. As flutuações podem refletir a influência de outros sistemas sobre o considerado, ou simplesmente representar o efeito de algum parâmetro externo. Nesta seção vamos ver a influência que produz um autômato celular determinista na dinâmica da parábola logística. Surpreendentemente, sendo ambas dinâmicas deterministas, o efeito pode ser visto como uma forma de ruído não trivial.

5.3.1 A Dinâmica da Parábola Logística com Perturbações

A natureza das flutuações provocadas num sistema dinâmico depende do acoplamento entre o sistema físico modelado e a fonte de perturbações. A dinâmica de muitos sistemas reais pode ser satisfatoriamente modelada considerando perturbações na parábola logística. Crutchfield et al. [112] estudaram o efeito de adicionar um ruído gaussiano na forma:

$$x_{t+1} = F(r, x_t) + p_t, \quad (5.10)$$

onde p_t é uma variável aleatória, com distribuição gaussiana, média zero e variância σ . σ representa o nível de ruído introduzido no sistema. O resultado qualitativo mais importante é que, a medida que o nível de ruído aumenta, as órbitas periódicas tornam-se bandas, semelhantes a atratores caóticos. As flutuações afetam a estabilidade local dos atratores. Como efeito das perturbações, a distinção entre órbitas periódicas e atratores caóticos deixa de ser simples, como no caso puramente determinista. Todos os atratores passam a encher intervalos de uma largura que depende do nível de ruído. Pode-se dizer que o efeito das flutuações é tomar a média de um conjunto de atratores deterministas que correspondem a um certo intervalo nos valores do parâmetro. Neste sentido, existe uma equivalência entre a perturbação de uma órbita e a perturbação do parâmetro. Então, uma forma equivalente da eq.(5.10) é

$$x_{t+1} = F((r + q_t), x_t) \quad (5.11)$$

Rössler et. al. [113] estudaram a dinâmica da parábola logística com parâmetro dependente do tempo, quando o mesmo pode tomar somente dois valores: $r_t = A, B$, em seqüências periódicas ou aleatórias. Eles encontraram que o sistema pode entrar em caos, para valores de A e B inferiores ao limiar $r_c \approx 3.569$. Por exemplo para $A=3.36$ e $B=2.25$ o expoente de Liapunov é positivo, indicando a presença de caos na dinâmica. Os autores denominam este fenômeno de “early chaos”. Também é possível o fenômeno contrario, valores de (A,B) que apresentam um λ negativo, e nos quais A ou B são maiores que r_c . A explicação do fenômeno está baseada em manter o sistema num transiente permanente, introduzindo perturbações com a fase correta. Também foi estudado o caso de uma modulação periódica do parâmetro r [114]. Quando a freqüência da modulação está perto da freqüência própria da parábola logística é observado novamente o fenômeno do “early chaos”.

5.3.2 A Parábola Logística Controlada por um Autômato Celular Determinista: uma classe de ruído não trivial

Nesta seção vamos apresentar um estudo da dinâmica da parábola logística, com valores do parâmetro dependentes do tempo:

$$x_{t+1} = r_t x_t (1 - x_t) \quad (5.12)$$

com a seqüência $\{r_t\}$ controlada por um autômato celular (AC) determinista [115]. Consideramos

$$r_t = a + b \sigma_i(t), \quad (5.13)$$

com a e b constantes reais. $\sigma_i(t)$ representa o valor (0 ou 1) que corresponde ao sítio i -ésimo do AC no tempo t . Com esta prescrição, r_t pode tomar somente dois valores, a ou $a+b$. O caso particular $b=0$ corresponde à parábola logística usual, eq.(5.2). Esta classe de modelos pode representar sistemas reais, nos quais a evolução dinâmica local é controlada por fenômenos cooperativos globais (ver refs. [113,114] e referências nesses trabalhos).

Especificamente, nós consideramos o AC determinista definido pela regra 90, segundo a classificação de Wolfram (ver tabela 5.1). Nesta regra, o estado de uma célula particular é determinado simplesmente pela soma módulo 2 das células vizinhas no intervalo de tempo anterior. Sendo a regra determinista, ela gera, para cada célula do AC, uma seqüência determinista de zeros e uns.

É importante notar que esta modulação do parâmetro r_t é diferente das introduzidas por Rössler et. al. [113]. Nos dois casos estudados por esses autores, a modulação é invariante por *translação temporal* (translação arbitraria no caso aleatório, e translação num múltiplo do período no caso periódico). No nosso caso, a modulação é invariante por *escala temporal* (como foi apontado na seção 5.2, a regra 90 gera uma estrutura fractal).

Consideramos um AC com N células, e condições de contorno periódicas. Na configuração inicial, todas as células estão no estado 0, exceto uma que está no estado 1. Para a parábola logística fixamos $x_0 = 0.2$. Foram estudados os atratores do sistema

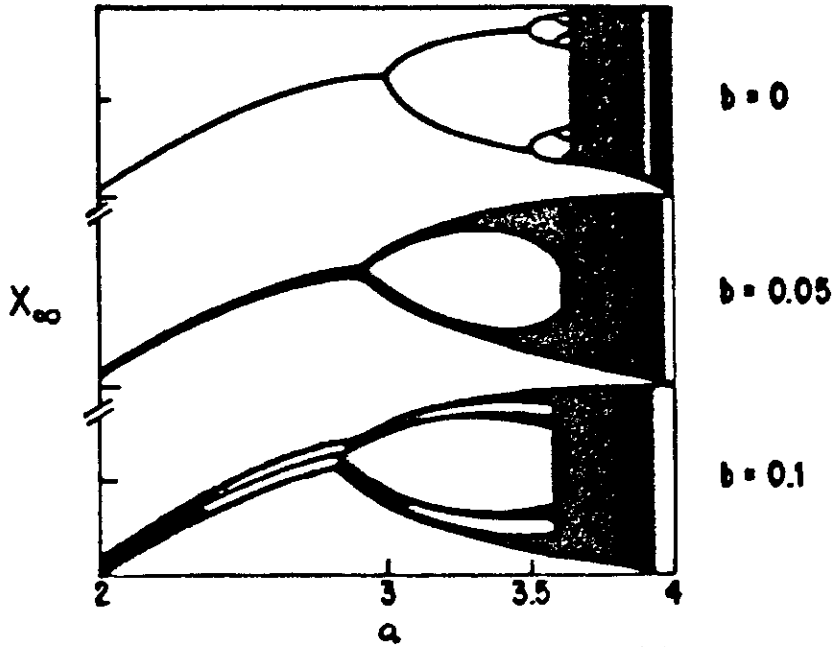


Figura 5.2: Representação esquemática dos atratores das eqs.(5.12) e (5.13), para três níveis de ruído diferentes

definido pelas eqs.(5.12) e (5.13), analisando os diagramas espaço-temporais, e as distribuições de freqüências $D(x)$ no espaço de fases ($\int_0^1 D(x)dx = 1$). Não observamos diferenças relevantes para N sendo par ou ímpar. A distribuição assintótica $D(x)$ é a mesma para todas as células (exceto para algumas com alta simetria, como a que esteve no estado 1 na condição inicial). Trabalhamos no limite termodinâmico, no sentido de N ser suficientemente grande para não mais influenciar $D(x)$ ($N=50$ foi suficiente para todos os casos estudados). Foram descartados os primeiros 5000 passos de tempo para eliminar efeitos de transientes: o relacionado ao tempo necessário para o sinal atingir todas as células do AC (da ordem de N), e o necessário para a evolução temporal global atingir o estado assintótico. Para determinar satisfatoriamente $D(x)$ consideramos as 10000 iterações posteriores.

Vamos apresentar os resultados, seguindo a influencia de b nos atratores (ver fig.5.2). Para $b = 0$ recuperamos o caminho ao caos via duplicação de período, que foi descrito na seção 5.1. A estrutura de bifurcações começa a apagar-se ao aumentar o valor

de b , por exemplo, para $b = 0.05$ somente a primeira bifurcação é claramente distinguível. Neste intervalo de valores de b , pode-se intuir o resultado, que é basicamente o obtido por Crutchfield et.al. [112] introduzindo um ruído branco. Podemos notar que a primeira bifurcação (assim como o caos) é antecipada, no sentido que acontece para $a \approx 2.9$, que é inferior ao valor ($a = 3$) correspondente para $b = 0$. Esta propensão a antecipar a entrada no caos é intuitiva, e se acentua ao aumentar o valor de b . Na fig.5.3(a) apresentamos um exemplo típico ($a = 2.95$). Consideremos agora o caso $b = 0.1$ da fig.5.2. Neste exemplo é que observamos um tipo de ruído não trivial, já que apresenta uma estrutura interna a priori não esperada: um atrator com *múltiplas* bandas, no lugar da banda *simples* gerada por um ruído branco, ou observada no caso $b = 0.05$. Na fig.5.4(b) podemos ver um exemplo típico ($a = 2.38$), com um ruído que apresenta três bandas. O diagrama espaço-temporal mostra que, toda vez que se apresenta uma seqüência de zeros no AC, o sistema converge rapidamente para o atrator de $(a, b) = (2.38, 0)$, que é um ponto fixo. Quando o AC produz uma seqüência alternada de zeros e uns, o sistema é mantido num atrator ciclo dois. Como nunca aparece uma seqüência de uns, o sistema nunca atinge o atrator de $(a, b) = (2.48, 0)$, que também é um ponto fixo. Esta ausência de simetria entre zeros e uns é devida ao fato de termos inicializado o AC com uma única célula com valor um (se tivéssemos começado com uma única célula com valor zero, a situação complementar seria observada). Se inicializássemos o AC com uma configuração aleatória, esperaríamos que esta assimetria desaparecesse. É interessante notar que a convergência para o atrator com $b = 0$ já estava presente na fig.5.3(a) sempre que uma seqüência de zeros do AC aparecesse, porém com uma velocidade de convergência muito menor, devido à proximidade da bifurcação ($a = 2.95$). Outro caso típico é apresentado na fig.5.5(c). De forma similar ao que acontece nas figs.5.3(a) e 5.4(b), sempre que se apresenta uma seqüência de zeros no AC o sistema converge ao atrator correspondente a $b = 0$ (neste caso, o ciclo dois correspondente a $a = 3.14$). Sempre que alternam zeros e uns, o sistema se aproxima de um atrator diferente (porém, também um ciclo dois).

O comportamento dinâmico do sistema parábola logística-AC pode ser descrito satisfatoriamente como a alternância entre dois tipos de transientes: um deles tentando

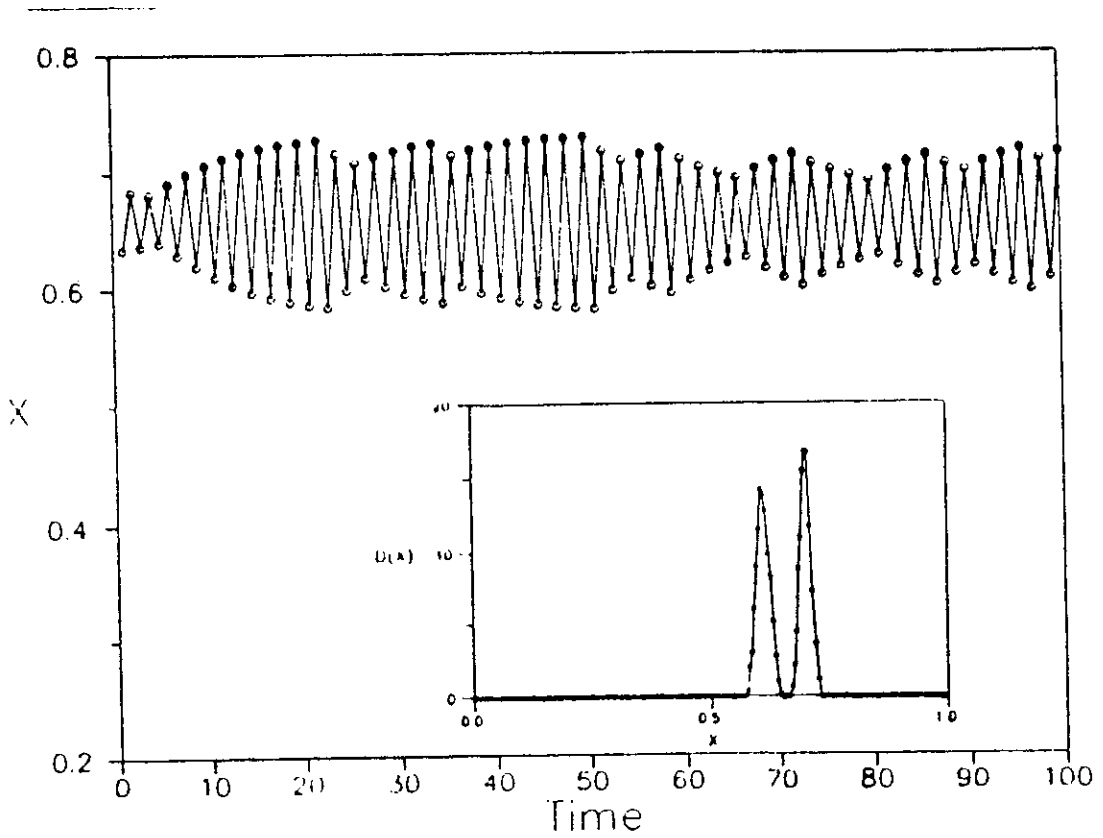


Figura 5.3: Diagramas espaço-temporais dos atratores das eqs.(5.12) e (5.13) para as primeiras 100 iterações depois do transiente (5000 iterações); círculos cheios correspondem a $\sigma_i(t) = 1$, e os vazios a $\sigma_i(t) = 0$. No detalhe, as correspondentes distribuições de frequências, para 10000 iterações após o transiente (o intervalo $[0,1]$ foi particionado em 2000 subintervalos iguais). (a)(a,b)=(2.95,0.05).

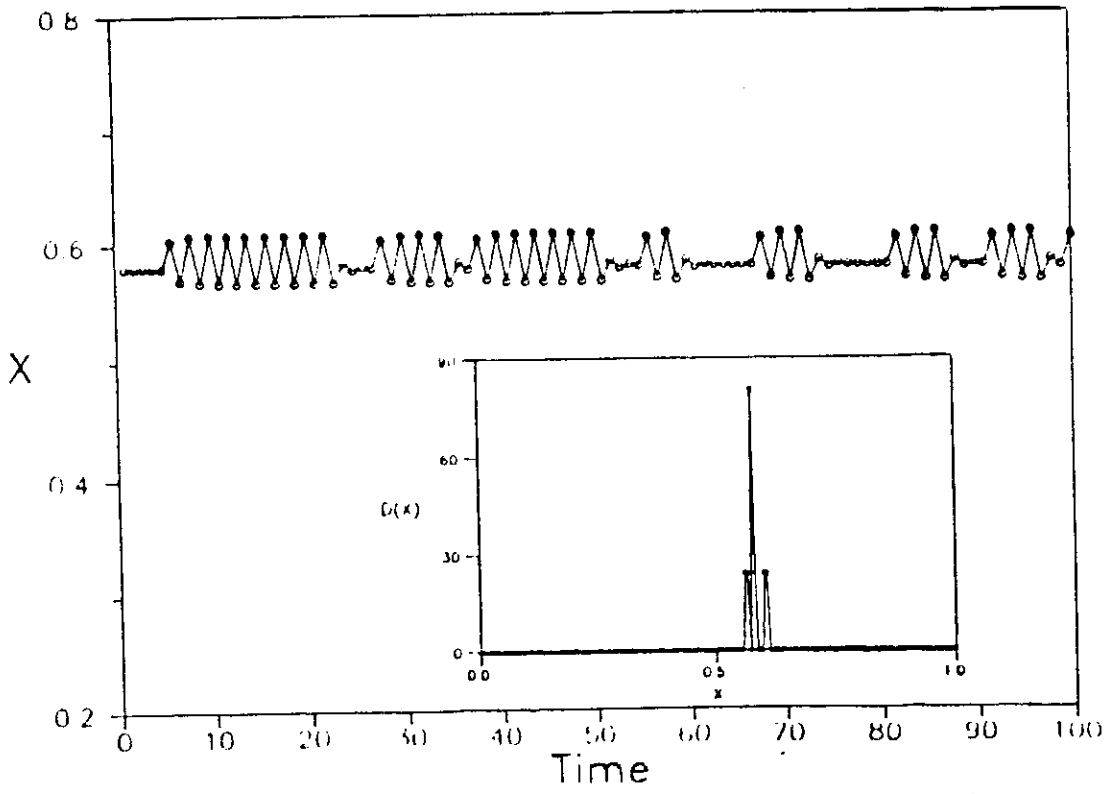


Figura 5.4: $(b)(a,b)=(2.38,0.1)$.

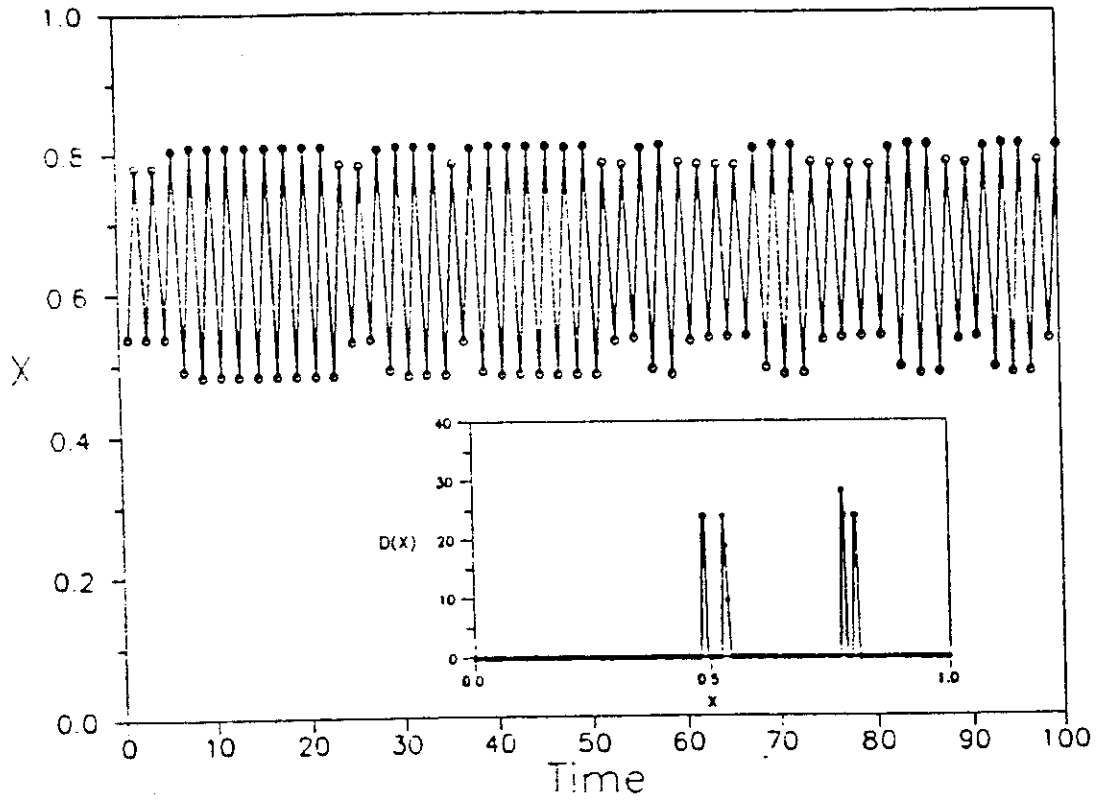


Figura 5.5: $(c)(a,b)=(3.14,0.1)$.

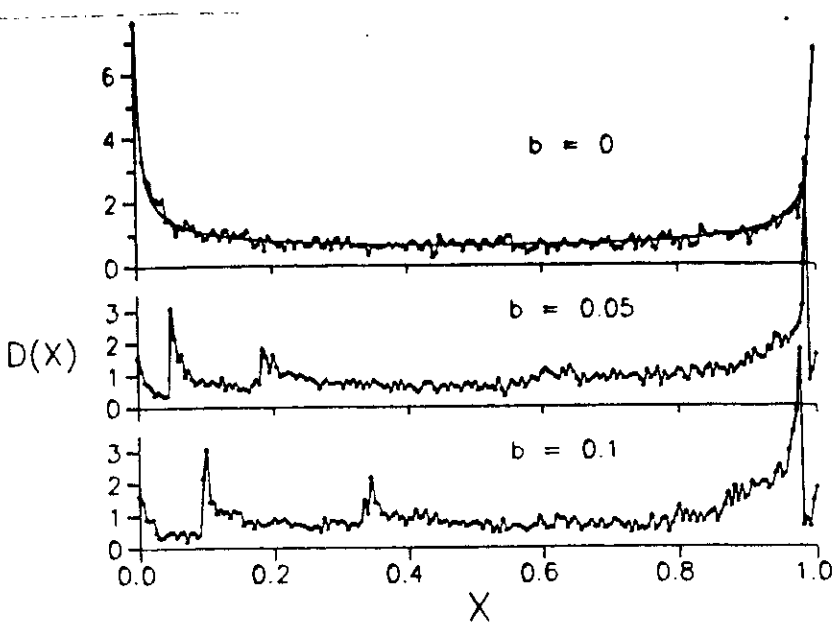


Figura 5.6: Distribuição de freqüências para três casos típicos que satisfazem $a + b = 4$. Superposto ao caso $b=0$, a distribuição exata $1/\pi\sqrt{x(1-x)}$.

convergir para o atrator correspondente a $b = 0$ (quando se apresenta uma seqüência de zeros no AC), o outro dirigido a um atrator ciclo dois (induzido pela alternancia de zeros e uns na dinâmica do AC). Os atratores entre os quais alterna a dinâmica dependem das particularidades do AC, que determina uma seqüência particular de zeros e uns. Para outros AC's, diferentes da regra 90, é de se esperar outras alternancias, porém esperamos o *mesmo fenômeno básico*.

Completando a descrição da dinâmica do sistema, notamos que para $a + b \approx 4$ o atrator finito é abandonado (a parábola logística pura abandona o atrator finito exatamente para $r = 4$). A densidade $D(x)$ é apresentada na fig.5.6, para os valores $b = 0, 0.05$ e 0.1 . Notamos que o ruído gerado pelo AC ($b \neq 0$) produz picos na densidade, não presentes no caso $b = 0$.

5.4 Conclusões

No presente capítulo foram considerados dois tipos de sistemas dinâmicos de naturezas completamente diferentes. Por um lado, a parábola logística, exemplo clássico de mapa unidimensional, modelo descrito pela evolução em tempo discreto de uma variável real.

Por outro, um autômato celular determinista, sistema com muitas variáveis, discretas, em interação . Os dois tipos de sistemas apresentam dinâmicas complexas. Embora apresentem diferenças básicas, ambas dinâmicas podem ser caracterizadas usando, basicamente, as mesmas ferramentas: expoentes de Liapunov, distribuições de frequências, diagramas espaço-temporais, entre outras. Utilizando estes dois últimos métodos, estudamos a dinâmica que resulta da interação entre os dois tipos de sistemas. Introduzimos um modelo que apresenta características que não estão presentes nas dinâmicas dos sistemas componentes. A dinâmica do modelo apresentado pode ser entendida, basicamente, como produzida pela alternância entre transientes induzidos pelas seqüências geradas pelo AC. Desta forma, o sistema permanece num estado de transiente perpetuo. Este fato é responsável pela geração de um tipo de ruído não trivial. Este ruído está caracterizado pela presença de bandas *múltiplas*. Para melhorar a compreensão do comportamento destes sistemas, seria interessante fazer um estudo dos expoentes de Liapunov λ , que permitiriam caracterizar melhor a região caótica. Igualmente, seria importante estudar a influência detalhada das condições iniciais do AC, da parábola logística, bem como do fato de considerar outros AC's diferentes da regra 90.

Capítulo 6

Conclusões

Neste capítulo fazemos um resumo dos principais resultados e conclusões da tese. O capítulo 2 foi dedicado a estudar como são afetadas as funções de armazenamento e reconhecimento de padrões em redes de neurônios analógicos quando é cortada uma certa fração das sinapses da rede. Nos casos estudados a rede permanece densamente conectada depois da diluição. O caso de diluição extrema não foi considerado nesta tese [53]. Primeiramente consideramos o caso de *diluição simétrica* aleatória de sinapses. Partimos da regra de aprendizado de Hopfield e, para poder fazer um estudo analítico, aproveitamos o fato da equivalência entre diluir as sinapses e acrescentar um termo de ruído na regra de aprendizado [41]. Por meio de uma análise estatística das superposições calculamos as curvas que dão a variação da superposição (m) e da capacidade de armazenamento crítico (α_c) em função do nível de diluição (d), e para diversos valores do parâmetro de ganho (g) (ver figs.2.1 e 2.2). Achamos que a performance do

modelo melhora quando g aumenta. No caso ótimo ($g \rightarrow \infty$), uma diluição de até 80% das sinapses permite reconhecer padrões com uma superposição maior que 90%. Para níveis de diluição entre 0 e 40%, a capacidade de armazenamento cai de 0.14 (quando o modelo se reduz ao de Hopfield) até 0.1 respectivamente. Para $g=2$, com uma diluição de 80% a rede ainda reconhece com uma superposição maior que 0.8. O α_c cai de 0.09 até 0.05 para diluições de 0 e 40% respectivamente. Estes dados numéricos demonstram que a rede analógica é muito robusta, tanto com respeito à diluição severa de sinapses, quanto à diferentes valores do ganho g .

Numa segunda parte incorporamos um elemento realista no modelo: assimetria. Devido às dificuldades para se tratar analiticamente modelos assimétricos, recorremos a simulações por computador. Encontramos uma fenomenologia muito mais rica do que no caso de diluição simétrica [55]. Para um valor do parâmetro $\alpha = 0.01$, a dinâmica assintótica só apresenta duas fases: uma fase origem, na qual todos os neurônios vão para o valor zero para valores pequenos do ganho g , e uma fase de reconhecimento para valores grandes de g . Quando a diluição cresce, a fase de reconhecimento torna-se estável para valores grandes de g (ver fig.2.3). Para um valor de $\alpha = 0.06$ aparecem duas novas fases no modelo (ver fig.2.4). Para valores moderados da diluição ($d=0.1-0.7$) observamos uma fase vidro de spin entre as fases origem e reconhecimento. Este resultado não concorda com o resultado de alguns trabalhos [56] nos quais se suspeitava que uma pequena quantidade de assimetria seria suficiente para suprimir a fase vidro de spin, indesejável num modelo de memória. Como sempre, a fase de reconhecimento é estável para valores grandes do ganho. Para um nível de diluição $d=0.9$ (o maior alcançado nas simulações) só estão presentes duas fases: uma fase origem para valores pequenos do ganho e uma fase caótica para valores de $g \geq 3$. As propriedades assintóticas do modelo mudam com o tamanho da rede simulada. Para valores grandes da diluição e redes de 50 e 100 neurônios existem ciclos limite que separam fases definidas por pontos fixos estáveis e caos. A janela de ciclos limite vai colapsando a medida que o tamanho da rede aumenta. Já para uma rede de 480 neurônios (o maior tamanho que utilizamos) não há presença de ciclos limite. A existencia de uma fase caótica certamente é indesejável do ponto de vista do reconhecimento de padrões; no entanto dinâmicas caóticas têm

sido detectadas em experiências com neurônios reais embora ainda não seja claro o seu significado no que diz respeito às funções cerebrais [60]. Certamente é um dos pontos de interesse para investigações futuras.

Em geral, tanto para diluição simétrica como assimétrica, os modelos continuam a funcionar muito bem como memórias associativas ainda depois de sofrer o corte de uma fração considerável das sinapses. A pequena diminuição nos valores das capacidades críticas de armazenamento (α_c) dos modelos diluídos com respeito aos dos completamente conectados é compensada pela grande economia no número de sinapses necessárias. Este fato é importante na hora de fabricar redes neuronais artificiais, o espaço físico que deve ocupar a rede é, na prática, muito limitado. Para futuros trabalhos resta fazer um estudo sistemático da fase caótica nas redes assimétricas para melhor caracterizá-la. Um estudo analítico do modelo talvez possa ser feito tomando como base o trabalho de Tirozzi e Tsodyks [58] para o caso de diluição extrema. Seria interessante também fazer um estudo da capacidade de armazenamento crítica do modelo já que, como vimos, nos dois valores de α simulados, os diagramas de fases são bem diferentes e apresentam diferentes fases.

No capítulo 3 estudamos as propriedades de reconhecimento em uma variedade de modelos caracterizados por diferentes regras de aprendizado Hebbianas, simétricas e assimétricas, e por armazenar padrões com níveis arbitrários de atividade dos neurônios [68]. As regras de aprendizado consideradas estão caracterizadas por 4 parâmetros (A,B,C e D) que permitem uma grande variedade de condições nas atividades dos neurônios pré e pós-sinápticos. Primeiramente fazemos um estudo analítico da estabilidade dos padrões, em função do grau de assimetria (γ) e atividade (a), por meio de uma análise sinal-ruído. A análise evidencia a necessidade de se introduzir dois parâmetros adicionais para otimizar o sinal, um campo externo uniforme (U) e um parâmetro (b) que define os possíveis valores da atividade dos neurônios. Como resultado da análise comprovamos que o valor máximo para a capacidade de armazenamento, $\alpha_c = \infty$, se obtém para uma regra simétrica ($\gamma = 1$) [65] no limite de baixa atividade ($|a| \rightarrow 1$). Esta solução é pouco robusta frente à assimetria (ver fig.3.1). Para valores de $|a| < 1$ as regras assimétricas são gradativamente mais robustas. A capacidade de armazenamento cai a zero para

regras antissimétricas ($\gamma = -1$) para qualquer valor da atividade e converge de forma não uniforme ao valor 1 no limite de $a = 0$ como no modelo de Hopfield ($\gamma = 1$, $a=0$). Cientes que a análise sinal-ruído só permite uma primeira aproximação à performance de um modelo, fizemos simulações numa rede de neurônios analógica para dois valores do ganho $g=100$ e $g=2.5$. Os resultados para $a=0.2$ confirmam as predições gerais da análise sinal-ruído.

Considerando padrões com atividade é importante conhecer a quantidade de informação armazenada nos padrões ; este estudo é objetivo de um futuro trabalho. Também seria interessante considerar termos que misturem memórias na matriz de sinapses para estudar o possível armazenamento de seqüências temporais nos modelos [11].

O capítulo 4 está dedicado ao estudo da capacidade de generalização em redes neuronais analógicas [81]. A capacidade de generalizar é evidenciada quando o sistema é capaz de aprender uma certa regra através de exemplos. Nós demonstramos, por meio de uma análise estatística do modelo, que ele é capaz de criar um atrator representativo das características comuns a uma série de padrões armazenados. O ponto interessante é que este atrator não é armazenado explicitamente na regra de aprendizado, ele é gerado pelo sistema como consequência da superposição existente entre um certo número de padrões (que sim são armazenados na regra de aprendizado), numa rede com uma estrutura hierárquica simples. Esperamos que esta capacidade de generalizar, ou classificar padrões de acordo com a sua semelhança, continue a estar presente em modelos com estrutura hierárquica mais complexa. Achamos que o sistema começa a generalizar para um valor crítico de exemplos do conceito a ser aprendido. A transição na rede analógica é contínua, diferentemente do que acontece no modelo discreto [78] no qual a transição para a fase de generalização é descontínua. Notamos que ao aumentar o valor do parâmetro $\alpha = p/N$, aumenta o ruído gerado pelos padrões que não são exemplos do conceito e então são necessários mais exemplos para o sistema começar a generalizar. Para um número de exemplos $s \gg 1$ o erro de generalização decai exponencialmente. O sistema apresenta uma leve melhoria na capacidade de generalização para valores finitos do ganho g com respeito ao valor $g \rightarrow \infty$ (ver figs.4.4 e 4.7), ou seja, para valores finitos

de g é necessário apresentar ao sistema um número menor de exemplos para que ele comece a generalizar do que para $g \rightarrow \infty$.

Em futuros trabalhos seria interessante estudar as diferentes fases deste modelo, assim como o tamanho das bacias atratoras na fase de generalização. As extensões deste trabalho podem ser muitas, já que pouco se sabe sobre as possibilidades de generalização em redes com atratores. Até hoje o esforço tem se concentrado no estudo do aprendizado e generalização em redes multicamadas, sendo o exemplo mais simples o perceptron. Seria muito interessante estudar estes problemas em redes com atratores introduzindo um formalismo mecânico-estatístico no espaço das interações. Um ponto de partida poderia ser a introdução de uma entropia adequada para tais problemas como é sugerido na ref.[95]. Uma possível candidata é a forma entrópica introduzida por Tsallis [96] que generaliza a forma clássica de Shannon. Essa forma generalizada para a entropia poderia ter como consequência a necessidade de se apresentar um número reduzido de exemplos para que o sistema generalize. Já que a estatística do aprendizado em redes de neurônios não tem que obedecer necessariamente a estatística de Boltzman-Gibbs, possíveis generalizações como a proposta podem ser importantes.

No capítulo 5 estudamos a dinâmica que surge da interação entre dois sistemas dinâmicos muito diferentes: um mapa unidimensional (a parábola logística) e um autômato celular determinista (AC) [115]. Interações entre sistemas dinâmicos diversos podem ser de utilidade na descrição de fenômenos complexos. Acharmos um fenômeno interessante que não está presente nas dinâmicas dos sistemas componentes. O autômato celular, mesmo sendo um sistema determinista, perturba a dinâmica da parábola logística e gera um atrator com ruído. A diferença com o que acontece quando se adiciona uma fonte de ruído branco é que os atratores apresentam bandas *múltiplas*, em lugar de bandas simples como as estudadas na ref.[112]. As seqüências de zeros e uns produzidas pelo AC mantém o sistema num estado de transiente permanente, o qual pode ser visualizado como um ruído não trivial. O cálculo dos expoentes de Liapunov para diversos valores dos parâmetros permitiria caracterizar melhor os atratores deste modelo.

Bibliografia

- [1] W.S.Mc Culloch e W.A.Pitts, A Logical Calculus of the Ideas Immanents in Neural Nets, *Bull.Math.Biophys.* **5**,115(1943).
- [2] W.A.Little, The Existence of Persistent States in the Brain, *Math.Biosci.* **19**, 101(1974).
- [3] D.A.Stariolo, Ising Spin Glasses with Multispin Interactions, *Physica*, **A106**, 622(1990).
- [4] J.J.Hopfield, Neural Networks and Physical Systems with Emergent Selective Computational Abilities, *Proc.Natl.Acad.Sci. USA*, **79**, 2554(1982).
- [5] B.Katz, *Nerve, Muscule and Synapse*, (Mc-Graw Hill, NY, 1966).
- [6] E.R.Kandel, *Cellular Basis of Behaviour*, (W.H. Freeman, San Francisco, 1976).
- [7] C.Eccles, *The Understanding of the Brain*, (Mc. Graw Hill, NY, 1973).
- [8] M.Abeles, *Local Cortical Circuits*, (Springer- Verlag, Berlin, 1982).

- [9] J.J.Hopfield, Neurons with Graded Response have Collective Computational Properties like those of Two-State Neurons, *Proc.Natl.Acad.Sci. USA*, **81**, 3088(1984).
- [10] M.A.Cohen e S.Grossberg, Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks, *IEEE Transactions on Systems, Man and Cybernetics*, **13**, 815(1983).
- [11] D.J.Amit, *Modelling Brain Function*, (Cambridge University Press, 1989).
- [12] D.J.Amit, H.Gutfreund e H.Sompolinsky, Spin-glass Models of Neural Networks, *Phys.Rev.A***32**, 1007(1985).
- [13] D.J.Amit, H.Gutfreund e H.Sompolinsky, Storing Infinite Number of Patterns in a Spin-Glass Model of Neural Networks, *Phys.Rev.Lett.***55**, 1530(1985).
- [14] D.J.Amit, H.Gutfreund e H.Sompolinsky, Statistical Mechanics of Neural Networks Near Saturation, *Ann.Phys.*, **173**, 30(1987).
- [15] D.J.Amit, H.Gutfreund e H.Sompolinsky, Information Storage in Neural Networks with Low Levels of Activity, *Phys.Rev. A***35**, 2293(1987).
- [16] C.M.Marcus, F.R.Waugh e R.M.Westervelt, Associative Memory in an Analog Iterated-Map Neural Network, *Phys.Rev.A***41**, 3355(1990).
- [17] C.M.Marcus e R.M.Westervelt, Dynamics of Iterated Map Neural Networks, *Phys.Rev.A***40**, 501(1989).
- [18] C.M.Marcus e R.M.Westervelt, Stability of Analog Neural Networks with Delay, *Phys.Rev.A***39**, 347(1989).
- [19] C.M.Marcus e R.M.Westervelt, Stability and Convergence of Analog Neural Networks with Multiple-Time-Step Parallel Dynamics, *Phys.Rev. A***42**, 2410(1990).
- [20] A.V.M.Herz, Global Analysis of Parallel Analog Networks with Retarded Feedback, *Phys.Rev.A***44**, 1415(1991).

- [21] T.Fukai e M.Shiino, Large Suppression of Spurious States in Neural Networks of Nonlinear Analog Neurons, *Phys.Rev.A* **42**, 7459(1990).
- [22] M.Minsky, *The Society of Mind* (Heinemann, London, 1985).
- [23] P.Churchland, *The Neurophylosophy of Mind* (MIT Press, Cambridge, Mass., 1969).
- [24] D.O.Hebb, *The Organization of Behaviour: A Neurophysiological Theory* (Wiley, New York, 1949).
- [25] G.Palm, *Neural Assemblies: an Alternative Approach to Artificial Intelligence*, (Springer-Verlag, Berlin, 1982).
- [26] P.Peretto, On Learning Rules and Memory Storage Abilities of Asymmetrical Neural Networks, *J.de Physique*, **49**, 711(1988).
- [27] W.B.Levy, Associative Changes at the Synapse: LTP in the Hippocampus, in *Synaptic Modification, Neuron Selectivity and Nervous System Organization*, eds.W.B.Levy, J.A.Anderson e S.Lehmkuhle, (Lawrence Erlbaum Associates, London, 1985).
- [28] T.Kohonen, *Self Organization and Associative Memories*, (Springer -Verlag, Berlin, 1984).
- [29] L.Personnaz, I.Guyon e G.Dreyfus, Information Storage and Retrieval in Spin-Glass Like Neural Networks, *J.Physique Lett.*, **46**, 359(1985).
- [30] I.Kanter e H.Sompolinsky, Associative Recall of Memory without Errors, *Phys.Rev.A* **35**, 380(1986).
- [31] M.Blatt e E.Vergini, Neural Networks: A Local Learning Prescription for Arbitrary Correlated Patterns, *Phys.Rev.Lett.* **66**, 1793(1991).
- [32] M.Mezard, J.P.Nadal e G.Toulouse, Solvable Models of Working Memories, *J.Physique*, **47**, 1457(1986).

- [33] J.C.Eccles, *The Physiology of Synapses*, (Springer-Verlag, Berlin, 1964).
- [34] S.Shinomoto, A Cognitive Associative Memory, *Biol.Cybern.*, **57**, 197(1987).
- [35] A.Treves e D.J.Amit, Low Firing Rates: an Effective Hamiltonian for Excitatory Neurons, *J.Phys.A***22**, 2205(1989).
- [36] W.Krauth e M.Opper, Critical Storage Capacity of the $J = \pm 1$ Neural Network, *J.Phys.A* **22**, L519(1989).
- [37] B.M.Forrest, Content-addressability and Learning in Neural Networks, *J.Phys.A* **21**, 245(1988).
- [38] M.Minsky e S.Papert, *Perceptrons* (MIT Press, Cambridge, Mass., 1969).
- [39] E.Gardner, The Phase Space of Interactions in Neural Network Models, *J.Phys.A* **21**, 257(1988).
- [40] M.A.Virasoro, Disordered Models of the Brain, em *Proceedings of the Conference on New Developments in Hardware and Software for Computational Physics*, Eds. A.M.J.Ferrero-R.P.J.Perazzo e S.L.Reich, North Holland, Amsterdan, **93**(1989).
- [41] H.Sompolinsky, The Theory of Neural Networks: The Hebb Rule and Beyond, em *Proceedings of the Heidelberg Colloquium on Glassy Dynamics*, Eds. L.van Hemmen e I.Morgenstern, Springer-Verlag, Heidelberg, (1987).
- [42] C.Tsallis e S.Redner, *Phys.Rev.***B28**, 6603(1983).
- [43] B.Derrida,E.Gardner e A.Zippelius, An Exactly Solvable Asymmetric Neural Network Model, *Europhys. Lett.*,**4**, 167(1987).
- [44] P.Spitzner e W.Kinzel, Hopfield Network with Directed Bonds, *Z.Phys.*,**B74**, 539(1989).
- [45] A.Canning e E.Gardner, Partially Connected Models of Neural Networks, *J.Phys.A* **21**, 3275(1988).

- [46] A.J.Noest, Domains in Neural Networks with Restricted-Range Interactions, *Phys.Rev.Lett.***63**,1739(1989).
- [47] A.Treves e D.J.Amit, The Number of Metastable States in Asymmetrically Diluted Neural Networks, *J.Phys.***A21**, 3155(1988).
- [48] M.Bouten,A.Engel,A.Komoda e R.Serneels, Quenched versus Annealed Dilution in Neural Networks, *J.Phys.***A23**, 4643(1990).
- [49] A.J.Noest, Associative Memory in Sparse Phasor Neural Networks, *Europhys. Lett.*,**6**, 469(1988).
- [50] K.E.Kürten, Adaptive Architectures for Hebbian Network Models, *J.Phys. I France*,**2**, 615(1992).
- [51] H.Sompolinsky, Neural Networks with Nonlinear Sinapses and a Static Noise, *Phys.Rev.***A34**, 2571(1986).
- [52] G.Toulouse,S.Dehaene e J.P.Changeux, Spin Glass Model of Learning by Selection, *Proc.Natl.Acad.Sci. USA*,**83**, 1695(1986).
- [53] F.A.Tamarit,D.A.Stariolo e E.M.F.Curado, Diluted Neural Network Model with Higher Order Interactions, *Phys.Rev.***A43**, 12(1991).
- [54] F.A.Tamarit, Tese de Doutorado, Centro Brasileiro de Pesquisas Físicas, Ríó de Janeiro, 1992.
- [55] D.A.Stariolo, Dynamics of Neural Networks with Continuous Variables, *Phys.Lett.***A152**, 349(1991).
- [56] J.A.Hertz, G.Grinstein e S.A.Solla, Irreversible Spin Glasses and Neural Networks, em *Preceedings of the Heidelberg Colloquium on Glassy Dynamics*, Eds. L.van Hemmen e I.Morgenstern, Springer-Verlag, Heidelberg, (1987).
- [57] D.Amit, The Properties of Models of Simple Neural Networks, *Proceedings of the Heidelberg Colloquium on Glassy Dynamics*, Eds.L. van Hemmen e I.Morgenstern, Springer-Verlag, Heidelberg, (1987).

- [58] B.Tirozzi e M.Tsodyks, Chaos in Highly Diluted Neural Networks, *Europhys.Lett.*, **14**, 727(1991).
- [59] H.Sompolinsky, A.Crisanti e H.J.Sommers, Chaos in Random Neural Networks, *Phys.Rev. Lett.*, **61**, 259(1988).
- [60] K.Aihara, T.Takabe e M.Toyoda, Chaotic Neural Networks, *Phys.Lett.A* **144**, 333(1990).
- [61] A.L.Hodgkin e A.F.Huxley, A quantitative description of current and its application to conduction and excitation in nerve, *J.Physiol.(London)*, **117**, 500(1952).
- [62] M.Abelles, *Studies in Brain Function*, (Springer-Verlag, New York, 1982).
- [63] J.Buhmann, R.Divko e K.Schulten, Associative Memory with High Information Content, *Phys.Rev.A* **39**, 2689(1989).
- [64] M.V.Tsodyks e M.V.Feigel'man, The Enhanced Storage Capacity in Neural Networks with Low Activity Level, *Europhys.Lett.*, **6**, 101(1988).
- [65] C.J.Perez Vicente e D.Amit, Optimized Network for Sparsely Coded Patterns, *J.Phys.A* **22**, 559(1989).
- [66] M.B.Levy, Synaptic Modification, Neuron Selectivity and Nervous System Organization, Eds. M.W.Levy, J.A.Anderson e S.Lehmkuhle (Lawrence Erlbaum Ass., London, 1985), p.5.
- [67] J.P.Rauschecker e W.J.Singer, *J.Physiol.*, **310**, 215(1981).
- [68] D.A.Stariolo e C.Tsallis, Optimal Hebbian Learning Rules and the Role of Asymmetry, submetido a *Europhys.Lett.*
- [69] D.Sherrington e S.Krikpatrick, Solvable Model of a Spin Glass, *Phys.Rev.Lett.*, **35**, 1792(1975).
- [70] M.Mézard, G.Parisi e M.A.Virasoro, *Spin Glass Theory and Beyond*, (World Scientific, Singapore, 1987).

- [71] N.Parga e M.A.Virasoro, The Ultrametric Organization of Memories in a Neural Network, *J.Physique*, **47**, 1857(1986).
- [72] M.V.Feigel'man e L.B.Ioffe, The Augmented Models of Associative Memory: Asymmetric Interaction and Hierarchy of Patterns, *Int.Journ.of Mod.Phys.*,**B1**, 51(1987).
- [73] H.Gutfreund, Neural Networks with Hierarchically Correlated Patterns, *Phys.Rev.***A37**, 570(1988).
- [74] F.A.Tamarit e E.M.F.Curado, Pair Correlated Patterns in Hopfield Model of Neural Networks, *J.Stat.Phys.*, **62**, 473(1991).
- [75] M.A.Virasoro, The Effect of Synapsis Destruction on Categorization in Neural Networks, *Europhys.Lett.*,**7**, 293(1988).
- [76] M.A.Virasoro, Maximum Entropy Principle and Brain Models, *Neural Networks and Spin Glasses*, Eds. W.K.Theumann e R.Köberle, (World Scientific, Singapore, 1989).
- [77] F.Rosenblatt, *Principles of Neurodynamics*, (Spartan, New York, 1962).
- [78] J.F.Fontanari, Generalization in a Hopfield Network, *J.Physique*, **51**, 2421(1990).
- [79] E.N.Miranda, Generalization in the Hopfield Model: Numerical Results, *J.Phys.I France*, **1**, 999(1991).
- [80] M.C.Branchtein e J.J.Arenzon, Categorization in the Hopfield Model, a ser publicado no *J.Phys.I France*, novembro 1992.
- [81] D.A.Stariolo e F.A.Tamarit, Generalization on an Analog Neural Network, *Phys.Rev.***A46**, 5249(1992).
- [82] P.Carnevali e S.Patarnello, Exhaustive Thermodynamical Analysis of Boolean Learning Networks, *Europhys.Lett.*,**4**, 1199(1987).

- [83] G.Györgyi e N.Tishby, *Neural Networks and Spin Glasses*, Eds. W.K.Theumann e R.Köberle, (World Scientific, Singapore, 1989).
- [84] P.Del Giudice, S.Franz e M.A.Virasoro, Perceptron Beyond the limit of Capacity, *J.Physique*, **50**, 121(1989).
- [85] E.Gardner e B.Derrida, Three Unfinished Works on the Optimal Storage Capacity of Neural Networks, *J.Phys.A* **22**, 1983(1989).
- [86] D.Hansel e H.Sompolinsky, Learning from Examples in a Single- Layer Neural Network, *Europhys.Lett.*, **11**, 687(1990).
- [87] H.Sompolinsky, N.Tishby e H.S.Seung, Learning from Examples in Large Neural Networks, *Phys.Rev.Lett.*, **65**, 1683(1990).
- [88] G.Györgyi, Inference of a Rule by a Neural Network with Thermal Noise, *Phys.Rev.Lett.*, **64**, 2957(1990).
- [89] G.Györgyi, First-Order Transition to Perfect Generalization in a Neural Network with Binary Sinapses, *Phys.Rev.A* **41**, 7097(1990).
- [90] E.Levin, N.Tishby e S.Solla, *Proc. IEEE*, **78**, 1568(1990).
- [91] R.Meir e J.F.Fontanari, Learning From Examples in Weight Constrained Neural Networks, 1991 preprint.
- [92] H.S.Seung, H.Sompolinsky e N.Tishby, Statistical Mechanics of Learning from Examples, *Phys.Rev.A* **45**, 6056(1992).
- [93] J.F.Fontanari e R.Köberle, Storing Correlated Patterns in Hopfield's Model, *J.Phys.A* **21**, 2477(1988).
- [94] R.Erichsen Jr. e W.K.Theumann, Mixture States and Storage with Correlated Patterns in Hopfield's Model, preprint IF-UFRGS (1990).
- [95] G.Parisi, On the Classification of Learning Machines, preprint ROM2F-90/37.

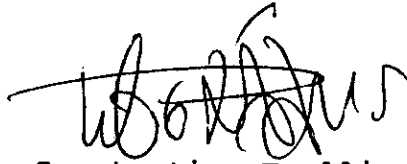
- [96] C.Tsallis, *J.Stat.Phys.***52**, 479(1988).
- [97] E.M.F.Curado e C.Tsallis, *J.Phys.***A 24**, L69(1991).
- [98] Silver e C.Tsallis, Generalized Entropy as a Basis for a Nonlinear Schroedinger Equation, preprint CBPF1993.
- [99] J.Guckenheimer e P.Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, (Springer, New York, 1983).
- [100] H.G.Schuster, *Deterministic Chaos: an Introduction*, (Weinheim: Physik Verlag, 1984).
- [101] P.Bergé , Y.Pomeau e C.Vidal, *L'Ordre dans le Chaos*, Hermann et Cie. Editeurs, Paris, 1984).
- [102] P.Collet e J.P.Eckmann, *Iterated Maps on the Interval as Dynamical Systems*, (Birkhauser, Cambridge, MA, 1980).
- [103] S.Wolfram, *Theory and Applications of Cellular Automata*, (World Scientific, Singapore, 1986).
- [104] Artigos em *Physica*, **10D**, vol.1, (1984).
- [105] M.J.Feigenbaum, Universal Behavior in Nonlinear Systems, *Los Alamos Science* **1**, 4(1980).
- [106] D.Rend, S.Ostlund, J.Sethna e E.D.Siggia, Universal Transition from Quasiperiodicity to Chaos in Dissipative Systems, *Phys.Rev.Lett.*, **49**, 132(1982).
- [107] Y.Pomeau e P.Manneville, Intermittent Transition to Turbulence in Dissipative Dynamical Systems, *Commun.Math.Phys.*, **74**, 189(1980).
- [108] M.C.de Sousa Vieira, E.Lazo e C.Tsallis, New Road to Chaos, *Phys.Rev.***A35**, 945(1987).
- [109] H.J.Herrmann, Cellular Automata, *Nonlinear Phenomena in Complex Systems*, ed. A.N.Proto, (North-Holland, 1989).

- [110] C.Tsallis, em "Condensed Matter Theories", **8**, ed. L.Blum e F.B.Malik (Plenum,1993).
- [111] M.L.Martins, H.F.V.de Resende, C.Tsallis e A.C.N.de Magalhães, Evidence for a New Phase in the Domany-Kinzel Cellular Automaton, *Phys.Rev.Lett.* **66**, 2045(1991).
- [112] J.P.Crutchfield, J.D.Farmer e B.A.Huberman, Fluctuations and Simple Chaotic Dynamics, *Phys.Reports* **92**, 45(1982).
- [113] J.Rössler, M.Kiwi, B.Hess e M.Markus, Modulated Nonlinear Processes and a Novel Mechanism to Induce Chaos, *Phys.Rev.* **A39**, 5954(1989).
- [114] J.Rössler, E.Lazo e M.Kiwi, Effect of a Resonant Periodic Perturbation on Nonlinear Processes, *Proceedings of the Second Latinamerican Workshop on Nonlinear Phenomena*, eds. R.Benguria, P.Cordero e B.Nachtergale, (North-Holland, 1991).
- [115] D.A.Stariolo e C.Tsallis, Cellular-Automaton-Controlled Logistic Map: an Unusual Type of Noise, *Int.Jour.Mod.Phys.C*, **3**, 547(1992).

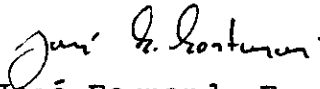
"REDES DE NEURÔNIOS E OUTROS SISTEMAS DINÂMICOS NÃO-LINEARES"

DANIEL ADRIAN STARIOLO

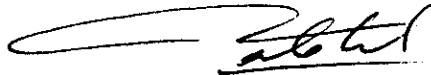
Tese de Doutorado apresentada no Centro Brasileiro de Pesquisas Físicas do Conselho Nacional de Desenvolvimento Científico e Tecnológico, fazendo parte da banca examinadora, os seguintes professores:



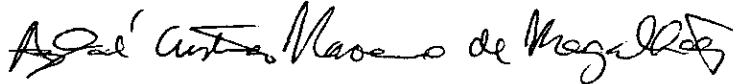
Constantino Tsallis - Presidente



José Fernando Fontanari



Paulo Murilo Castro de Oliveira



Aglaé Cristina Navarro de Magalhães



Jorge Silvio Helman

Rio de Janeiro, 05 de março de 1993