

**DISSERTAÇÃO DE MESTRADO**

**Propriedades Dinâmicas de Redes de Neurônios com  
Estímulos Externos e Diluição Assimétrica**

**GERALDO ALVES SOBRAL JÚNIOR**

**UNIVERSIDADE FEDERAL DE ALAGOAS  
CENTRO DE CIÊNCIAS EXATAS E NATURAIS  
DEPARTAMENTO DE FÍSICA**

**UNIVERSIDADE FEDERAL DE ALAGOAS  
DEPARTAMENTO DE FÍSICA**

**Propriedades Dinâmicas de Redes de Neurônios com  
Estímulos Externos e Diluição Assimétrica**

Geraldo Alves Sobral Júnior

Dissertação apresentada ao Departamento de Física da Universidade Federal de Alagoas como parte dos requisitos para obtenção do título de Mestre em Física.

**Orientador: Prof. Dr. Crisógono Rodrigues da Silva**

Maceió-Brasil  
Dezembro de 2005

# **Propriedades Dinâmicas de Redes de Neurônios com Estímulos Externos e Diluição Assimétrica**

Tese de mestrado apresentada ao Departamento de Física da Universidade Federal de Alagoas como parte dos créditos para a obtenção do título de Mestre em Física por Geraldo Alves Sobral Júnior.

Membros da banca examinadora:

---

Sérgio Galvão Coutinho (UFPE)

---

Marcelo Leite Lyra (UFAL)

---

Crisógono Rodrigues da Silva (Orientador)

---

---

# Agradecimentos

Agradeço eternamente à minha mãe pelos exemplos de humildade e respeito que sempre permearam cada pequeno gesto desta grande mulher. Sua presença sempre exercerá grande influência sobre mim.

Agradeço ainda a Deus, por ter me permitido vislumbrar quão pequena é nossa ciência perante as manifestações mais “simplórias” de sua presença.

Voltando ao campo terreno, gostaria de destacar a grande importância da orientação do professor Crisógono, sua paciência, disponibilidade e bom humor foram fundamentais na confecção desta dissertação. Obrigado Crisógono.

Agradeço ainda à minha companheira Sandileide, que também esteve presente em cada passo desta conquista.

Não poderia faltar espaço aqui para lembrar dos grandes amigos que, de uma forma ou outra, fizeram-se presentes nesta pequena peregrinação. Desta forma, iniciarei pelos amigos de longa data Marcelo e Ronaldo que sempre se mostraram preocupados com minha sanidade mental, propiciando, vez ou outra, bons momentos de descontração, regados a muita cerveja. Posso dizer que as lembranças destes momentos muitas vezes serviram de “propulsor” para este trabalho.

Ainda há a “grande família”, com os quais sempre me sinto em casa: Paula, Alex, Taciana, Antônio e Ana Paula. Obrigado pelo carinho.

Aproveito para lembrar dos amigos Alessandro e Wellington, cada um, com estilos bem distintos, soube me estender a mão, firmando amizades que pretendo cultivar ao longo dos anos.

Como destaque (não poderia ser diferente), quero lembrar dos colegas Nádja e Amadeu. A amizade de vocês fortaleceu minha convicção de que no nordeste todos se entendem. Um grande abraço.

Agradeço também ao suporte financeiro da CAPES e FAPEAL, sem o qual eu não teria meios para realizar este trabalho.

# Resumo

Nesta tese são estudadas, analiticamente e mediante simulações numéricas, as propriedades dinâmicas de reconhecimento e armazenamento de padrões do modelo de Hopfield com diluição assimétrica e estímulo externo.

O primeiro capítulo é dedicado a um resumo muito simplificado do funcionamento biológico do cérebro e da evolução das redes neurais artificiais.

Na primeira parte do capítulo 2, introduzimos o modelo de Hopfield em sua versão original e apresentamos seus principais resultados. Na segunda parte, mostramos um desenvolvimento termodinâmico proposto por Amit et al., o qual generaliza o modelo de Hopfield, e uma versão da dinâmica do modelo com diluição e assimetria nas sinapses, proposta por Derrida et al.

No capítulo 3, introduzimos um estímulo externo que depende da superposição  $m(t)$  com o padrão estimulado e estudamos o modelo de Hopfield na versão ultra diluída ( $C \ll \ln N$ ), onde  $C$  é a conectividade média por neurônio e  $N$  o número de neurônios da rede. Duas situações são analisadas: quando o estímulo externo atua sobre o padrão de inicialização e quando o estímulo externo atua sobre um padrão distinto do padrão de inicialização. Em ambos os casos, obtemos o diagrama de fase completo.

Finalmente, no capítulo 4, concluímos resumindo os resultados obtidos na tese e indicando possíveis extensões deste estudo.

# Abstract

In this present thesis the dynamical properties of the Hopfield model with asymmetrical dilution and external stimulus are studied, both analytically and through numerical simulations.

The first chapter is devoted to a very simplified summary of the biological brain functioning and the evolution of the artificial neural networks.

In the first part of chapter 2, we introduce the original version of the Hopfield model and the main results. The second part is devoted to the thermodynamic development proposed by Amit et al. which generalizes the Hopfield model, as well as to a dynamical model including dilution and asymmetry in the synapses, proposed by Derrida et al.

In the chapter 3, we introduce an external stimulus that depends on the overlapping  $m(t)$  of the stimulated pattern and we study the Hopfield model in an ultra-diluted version ( $C \ll \ln N$ ), where  $C$  is the average neuron connectivity and  $N$  is the number of neurons in the network. Two situations are analyzed: When the external stimulus acts on the initialization pattern and when the external stimulus acts on a pattern different from the initialization one. In both cases, we obtain the complete phase diagram.

Finally, in the chapter 4, we conclude summarizing the results obtained in the thesis and indicating possible extensions of this study.

# Índice

<b>Agradecimentos</b> .....	<b>iv</b>
<b>Resumo</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vi</b>
<b>Capítulo 1</b> .....	<b>11</b>
<b>Rede de Neurônios</b> .....	<b>11</b>
1.1 - Introdução .....	11
1.2 - O Neurônio Biológico .....	12
1.3 - Redes neurais artificiais (RNAs) .....	17
1.4 - Neurônio Matemático de McCulloch-Pitts .....	19
1.5 - Tipos de funções de ativação .....	20
1.5.1 - Função de limiar: .....	20
1.5.2 - Função sigmóide: .....	21
1.6 - Arquiteturas de redes neurais .....	21
1.6.1 - Perceptrons ou arquitetura tipo <i>feed-forward</i> .....	21
1.6.2 - Redes recorrentes ou arquitetura tipo <i>feedback</i> .....	23
1.7 - O Modelo de Little: Processos Aleatórios .....	24
<b>Capítulo 2</b> .....	<b>27</b>
<b>Modelo de Hopfield</b> .....	<b>27</b>
2.1 - Introdução .....	27
2.2 - A Natureza da Memória Associativa .....	30
2.3 - Implementação da Dinâmica (padrões estáveis) .....	32
2.3.1 - Rede com um único padrão .....	33
2.3.2 - Rede com muitos padrões .....	33
2.4 - Capacidade de armazenamento da rede de Hopfield .....	34
2.5 - Generalização do Modelo de Hopfield .....	37
2.5.1 - Termodinâmica do Modelo de Hopfield .....	37
2.5.2 - Dinâmica do modelo de Hopfield Diluído e Assimétrico .....	40
<b>Capítulo 3</b> .....	<b>48</b>
<b>Efeito do Estímulo Externo no Reconhecimento de Padrões no Modelo DGZ</b> .....	<b>48</b>
3.1 - Introdução .....	48
3.2 - Padrão estimulado igual ao padrão de inicialização .....	50
3.2.1 - Número Finito de Padrões .....	54
3.2.2 - Número Infinito de Padrões .....	55
3.3 - Padrão estimulado distinto do padrão de inicialização .....	59
3.3.1 - Número Finito de Memórias .....	60
3.3.2 - Número Infinito de Memórias .....	62

3.4 – Simulação numérica .....	67
<b>Capítulo 4.....</b>	<b>72</b>
<b>Conclusões.....</b>	<b>72</b>
<b>Apêndice A.....</b>	<b>76</b>
<b>Cálculo da Equação Dinâmica do Modelo DGZ.....</b>	<b>76</b>
A.1 – Relação de recorrência para a superposição $m(t)$ .....	76
A.1.1 – Limite de conectividade infinita.....	79
<b>Apêndice B.....</b>	<b>81</b>
<b>Equação de recorrência para o modelo DGZ com o estímulo externo .....</b>	<b>81</b>
B.1 – Padrão estimulado paralelo ao padrão de inicialização .....	81
B.2 - Padrão estimulado diferente do padrão de inicialização .....	82
B.2.1 – Superposição do sistema com o padrão inicialização $n$ .....	82
B.2.2 – Superposição do sistema com o padrão estimulado $d$ .....	84
<b>Apêndice C.....</b>	<b>86</b>
<b>Teorema do Limite Central.....</b>	<b>86</b>
<b>Apêndice D.....</b>	<b>87</b>
<b>Automédia.....</b>	<b>87</b>
<b>Referências.....</b>	<b>88</b>

# Índice de Figuras

Figura 1.1: Desenho esquemático de uma rede de neurônios.....	13
Figura 1.2: Desenho de um neurônio típico e sua comunicação biológica.....	15
Figura 1.3: Estrutura química simplificada de um neurônio.....	16
Figura 1.4: Estrutura do neurônio de McCulloch-Pitts. ....	19
Figura 1.5: Gráfico da função de limiar. ....	20
Figura 1.6: Gráfico da função sigmóide.....	21
Figura 1.7: Desenho esquemático de um perceptron de camada única. ....	22
Figura 1.8: Desenho esquemático de um perceptron de múltiplas camadas .....	23
Figura 2.1 Representação esquemática do espaço de configuração.....	30
Figura 2.2: Diagrama de fase $T$ versus $a$ para o modelo de Hopfield generalizado [16]. $T_G$ é a temperatura de transição para os estados de vidro de spin, $T_M$ é a temperatura em que surgem os estados ferromagnéticos, $T_C$ é a temperatura em que estes estados tornam-se mínimos globais, passando por uma transição de primeira ordem. ....	39
Figura 2.3: Diagrama de fase $T$ versus $a$ para o modelo DGZ, obtido da equação (2.33).....	45
Figura 2.4: Superposição $m$ versus $a$ . A linha cheia foi obtida a partir da equação (2.34), enquanto que os símbolos referem-se a simulações numéricas para diferentes tamanhos de rede e conectividade finita $C=20$ [27]. ....	46
Figura 3.1: Superposição $m$ como função de $T$ e $h$ para $a = 0$ com $m_o \approx 1$ .....	54
Figura 3.2: Superposição $m$ como função de $a$ e $h$ para $T = 0$ com $m_o \approx 1$ . ....	56
Figura 3.3: Diagrama de fase $a_c$ x $h$ para $T = 0$ . A linha crítica separa a fase de reconhecimento R ( $m \neq 0$ ) da fase paramagnética P ( $m = 0$ ). ....	57
Figura 3.4: Diagrama de fase $T \times a \times h$ para o modelo DGZ com estímulo externo...58	
Figura 3.5: $m \times T \times h$ para $a = 0$ , $m_n^o = 0,99$ e $m_d^o = 0,01$ (Equações 3.21 e 3.22).(a) Superposição com o padrão de inicialização $n$ ; (b) Superposição com o padrão estimulado $d$ .....	61

<b>Figura 3.6:</b> $m \times a \times h$ para $T = 0$ , $m_n^o = 0,99$ e $m_d^o = 0,01$ . (a) Superposição com o padrão de inicialização $n$ ; (b) Superposição com o padrão estimulado $d$ .	63
<b>Figura 3.7:</b> Diagrama de fase $\alpha \times h$ para $T = 0$ . A linha crítica contínua separa a fase paramagnética (região clara) da fase de reconhecimento dos padrões $n$ e $d$ (região escura). Na região de reconhecimento a linha tracejada indica a transição descontínua.	64
<b>Figura 3.8:</b> $m \times a \times T$ para $m_n^o = 0,99$ e $m_d^o = 0,01$ e diferentes valores de $h$ : (a) $h = 0,0$ , (b) $h = 0,2$ e (c) $h = 0,4$ .	65
<b>Figura 3.9:</b> Diagrama de fase $T \times a \times h$ para o modelo DGZ com estímulo externo. Note que para $h = 1$ o comportamento linear na transição aparece somente para valores de $a/T < 1$	66
<b>Figura 3.10:</b> $m \times a$ para $T = 0$ , $h = 0,2$ e $m_n^o \approx 0,99$ e $m_d^o \approx 0,01$ . A linha cheia corresponde à solução analítica e os símbolos aos resultados das simulações para $N = 10000$ (círculos) $N = 20000$ (quadrados) e $N = 40000$ (triângulos) nos casos: a) $C = 20$ , b) $C = 40$ e c) $C = 80$ .	69
<b>Figura 3.11:</b> $m \times a$ para $T = 0$ , $h = 0,2$ e $m_n^o \approx 0,99$ e $m_d^o \approx 0,01$ . As linha pontilhadas correspondem à solução analítica e os símbolos aos resultados das simulações para $N = 10000$ (círculos) $N = 20000$ (quadrados) e $N = 40000$ (triângulos) nos casos: a) $C = 20$ , b) $C = 80$ .	70

# Capítulo 1

## Rede de Neurônios

### 1.1 - Introdução

O sistema nervoso é uma rede de comunicação eletroquímica interna do corpo. Suas principais partes são: o cérebro, a medula espinhal e os nervos. O cérebro e a medula espinhal formam o sistema nervoso central (SNC) — centro de controle e coordenação do corpo. Bilhões de longos neurônios, a maioria agrupada em nervos, formam o sistema nervoso periférico, transmitindo impulsos nervosos entre o SNC e as demais regiões do corpo.

O cérebro humano, em particular, é um exemplo de um sistema altamente intrincado e com alto grau de complexidade. Contém cerca de 10 bilhões de células nervosas, denominadas neurônios, cada um constituindo-se de uma unidade complexa e independente de processamento que pode se interligar a, aproximadamente,  $10^4$  outros neurônios. A comunicação entre as células nervosas se dá através de ligações eletroquímicas denominadas sinapses, que formam uma rede de comunicação densamente interligada. O cérebro é o maior órgão do sistema nervoso central, funciona como o centro deste sistema e controla todas as atividades do corpo, voluntárias e involuntárias. É também responsável por todas as operações conhecidas como “funções superiores do cérebro”, como por exemplo: pensamento, memória, aprendizagem, emoções, etc.

O estudo do cérebro tem se tornado, nos últimos anos, um campo de interesse para diferentes áreas do conhecimento científico, como as neurociências, engenharias, matemática, química, física, ciências da computação, entre outras. Apesar de se ter muitos

dados resultantes de pesquisas desenvolvidas neste último século, muito ainda é desconhecido com relação ao seu funcionamento biológico. As motivações da comunidade científica nesta área são muito amplas, os interessados vão desde aqueles que desejam entender o verdadeiro funcionamento do cérebro biológico, como, por exemplo, sua anatomia e fisiologia, àqueles cujo interesse é usar as informações derivadas dessas pesquisas, para desenvolver máquinas capazes de simular o comportamento lógico do cérebro, sem a necessidade de utilizar todos os dados disponíveis. Estes últimos são responsáveis pelas áreas de inteligência artificial e teorias de computação neural: modelos de neurônios, dinâmica neural e teorias de aprendizagem (engenharias, física, química, ciências da computação, etc.).

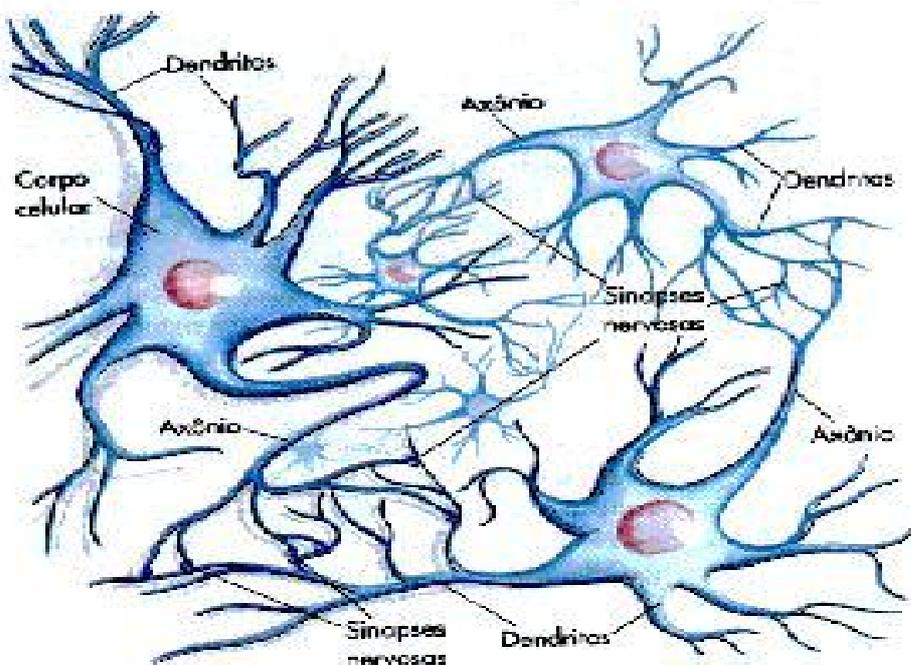
A física, e em particular a física estatística, utilizando ferramentas desenvolvidas ao longo dos anos e muitas vezes as inovando, familiarizada com comportamentos emergentes sob o nome de transições de fase, busca entender o comportamento coletivo que surge das interações mútuas entre os muitos neurônios. Existem numerosas evidências, obtidas a partir de estudos neurológicos, que algumas das funções superiores do cérebro (dentre elas a capacidade de memorizar de forma associativa, aprender e classificar) não surgem somente da especialização de suas partes, mas também, são produtos do trabalho coletivo de um número muito grande de células, que interagem de maneira simples, transportando informação (sinais elétricos). Se o comportamento global do sistema nervoso dependesse dos detalhes mais refinados da estrutura dos neurônios, é bem verdade que a física contribuiria pouco para o seu entendimento, mas se certos níveis de complexidade resultarem das interações entre eles, a física pode fornecer um ponto de vista valioso para a sua compreensão.

## 1.2 - O Neurônio Biológico

A descrição da biologia do neurônio que será apresentada aqui é uma visão muito simplificada do neurônio real. A neurobiologia é, obviamente, um tema muito mais complexo que a imagem aqui esboçada. Nosso principal interesse é descrever de forma sucinta a funcionalidade básica de um neurônio biológico, ressaltando as características passíveis de uma implementação computacional e que servirão de inspiração para os modelos teóricos abordados.

A figura 1.1 mostra o desenho esquemático de uma rede de neurônios e seus elementos básicos: os neurônios e suas conexões sinápticas. Os principais componentes de uma célula nervosa típica são: o **corpo celular** ou **soma**, que contém o núcleo, e é responsável por manter os processos de integração dos estímulos que a ele chegam; os **dendritos**, terminais receptivos, que são ramificações em torno do corpo celular por onde chegam os impulsos nervosos; e o **axônio**, terminal emitente, fibra longa que nasce no corpo celular e se ramifica na outra extremidade e que tem como principal função conduzir os impulsos gerados no soma.

De maneira simples, podemos imaginar uma rede de neurônios baseada nas divisões canônicas do seu elemento principal: a parte por onde chegam os impulsos elétricos (dendritos), a parte de processamento destes impulsos (corpo celular) e a parte de transmissão (axônio).



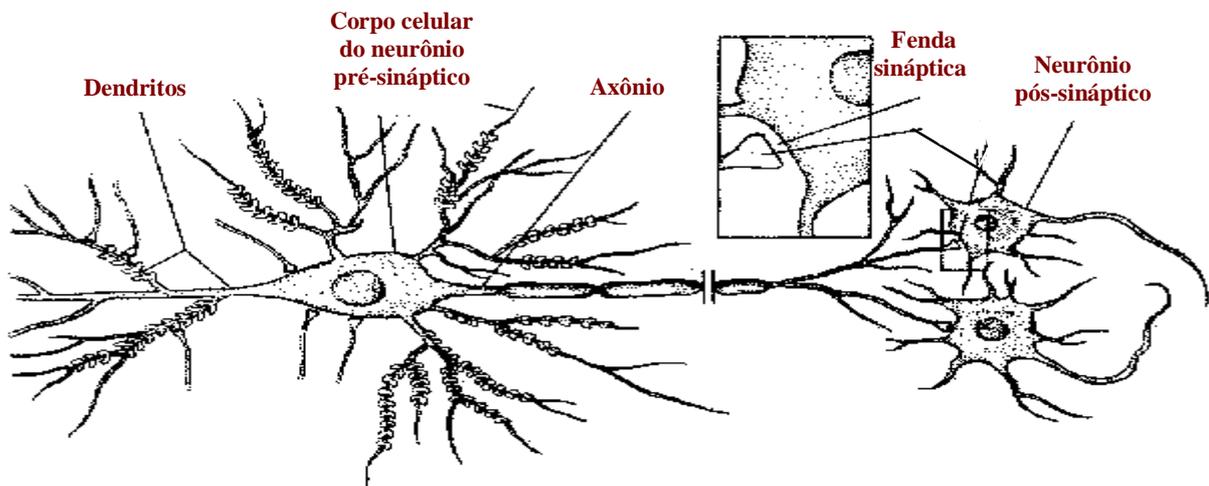
**Figura 1.1:** Desenho esquemático de uma rede de neurônios.

De acordo com sua função específica na rede neural, a morfologia do neurônio é bastante variada, sofrendo alterações de forma e tamanho do corpo celular, do número e da distribuição espacial dos dendritos, bem como do comprimento do axônio. O volume também é extremamente variável, havendo células nervosas desde quatro microns de diâmetro até outras que ultrapassam os cem microns.

De acordo com a disposição dos prolongamentos dos dendritos e axônios com relação ao corpo celular, os neurônios podem ser agrupados em três classes: *unipolares*, *bipolares* e *multipolares*. Os neurônios *unipolares* são os mais simples, eles não apresentam dendritos, simplesmente recebem e integram os impulsos no corpo celular e, caso disparem, transmitem o impulso nervoso pelo axônio. No sistema nervoso, neurônios unipolares controlam a secreção de glândulas exócrinas. Os neurônios *bipolares* apresentam apenas um dendrito e um axônio, originando pólos opostos no corpo celular, o impulso nervoso é recebido pelo dendrito e atravessa o corpo celular rumo ao axônio. As células nervosas sensitivas primárias, representadas por células da mucosa olfativa e pelas células da retina (cones e bastonetes), são exemplos de neurônios bipolares. Os *multipolares* apresentam uma estrutura complexa de distribuição dos dendritos (que podem ser milhares) em torno do corpo celular, e têm um único axônio. Os neurônios *multipolares* são os mais comuns no cérebro e na medula espinhal [1,2].

Quanto à funcionalidade, as células nervosas ainda podem ser classificadas em neurônios *sensitivos*, *motores* e *interneurônios*: os sensitivos captam as informações no meio externo e levam-nas ao sistema nervoso central; os neurônios *motores* ou *eferentes* conduzem o impulso nervoso ao órgão efetuator que, nos mamíferos, é um músculo ou uma glândula e os *interneurônios* ou neurônios de *associação*, fazem a associação dos neurônios entre si [3].

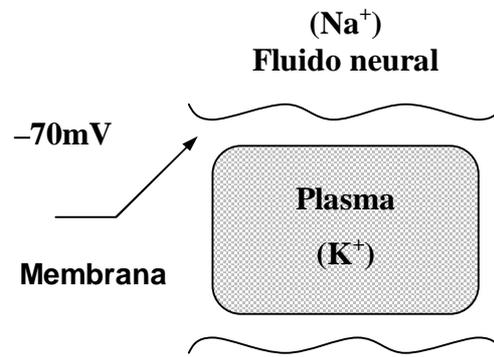
Na figura 1.2 mostramos um neurônio biológico típico, seus principais componentes e a porta de comunicação entre o neurônio pré-sináptico e o neurônio pós-sináptico, denominada junção ou fenda sináptica. Um neurônio, ao ser estimulado, gera um impulso nervoso que será transmitido à rede neural da qual ele faz parte, esse impulso elétrico originado no corpo celular, caracterizado por um pico de despolarização da membrana plasmática do neurônio, propaga-se pelo axônio até alcançar a junção sináptica. Na sinapse, como resposta ao impulso elétrico recebido, as vesículas do neurônio pré-sináptico liberam neurotransmissores que se difundem através da fenda sináptica e agem sobre os receptores do neurônio pós-sináptico. De acordo com os neurotransmissores liberados, o estímulo que chega ao neurônio pós-sináptico será excitatório ou inibitório, tornando este neurônio mais, ou menos, susceptível a emitir um sinal em resposta ao impulso recebido, dessa forma, essencialmente, os neurônios processam informação.



**Figura 1.2:** Desenho de um neurônio típico e sua comunicação biológica.

O plasma intracelular presente no interior das células nervosas e o fluido no qual os neurônios estão imersos contêm diferentes espécies e concentrações de íons. Os principais íons encontrados são o sódio ( $\text{Na}^+$ ) e o potássio ( $\text{K}^+$ ) que são elementos essenciais envolvidos na formação do impulso nervoso. Como mostra a figura 1.3, os íons de potássio estão predominantemente concentrados no interior da célula, dispersos no plasma celular, enquanto que os íons de sódio apresentam-se em maior concentração no lado externo da membrana plasmática. No estado de repouso (inatividade), há uma diferença de potencial de cerca de  $-70\text{mV}$  entre o meio extracelular e o plasma, devido a uma deficiência de íons positivos no interior da célula. Esta diferença de potencial tem origem na membrana celular que se apresenta impenetrável aos íons de sódio [4].

A ação do neurotransmissor é mudar a permeabilidade da membrana plasmática a certos tipos de íons. A mudança de permeabilidade tanto pode excitar como pode inibir o neurônio pós-sináptico, dependendo da concentração intra e extracelular de um determinado íon. Um fluxo de íons-sódio ( $\text{Na}^+$ ) para dentro da célula tende a *despolarizar* o potencial da membrana; este efeito é excitatório. Se íons-potássio ( $\text{K}^+$ ) saem do interior da célula, a membrana será *hiperpolarizada*; este efeito é inibitório. Esta flutuação do potencial da membrana com relação ao potencial de repouso do neurônio é conhecida como *Potencial Pós-Sináptico (PPS)*. Deste modo, o efeito da atividade do neurônio pré-sináptico é excitar ou inibir a polarização da membrana do neurônio pós-sináptico.



**Figura 1.3:** Estrutura química simplificada de um neurônio.

Como apenas o neurônio pré-sináptico libera neurotransmissores, as sinapses são polarizadas, ou seja, a informação é transmitida em um único sentido. No caso da excitação, resultante do efeito conjunto dos PPS que chegam ao corpo celular do neurônio pós-sináptico, ser maior que um certo limiar próprio do neurônio, gera-se um novo impulso que será transmitido pelo axônio. Caso contrário (menor que o limiar) nenhum sinal é produzido. Estes impulsos são chamados de *potenciais de ação* (PA) e constituem a única forma dos neurônios expressarem-se, variando apenas a frequência e a intensidade dos potenciais de ação. Desta forma, podemos representar a atividade da junção corpo celular – axônio como um processo binário: *tudo ou nada*.

Sabe-se hoje que nem todas as sinapses participam igualmente dos processos que ocorrem no cérebro. Pelo contrário, sinapses pouco excitadas são de alguma forma reprimidas, enquanto que as sinapses que participam de forma persistente na transmissão de informação, sofrem alguma mudança metabólica de modo que sua eficiência é aumentada. Costuma-se dizer que as sinapses mais utilizadas são mais *eficientes*, e têm maior “peso” no momento da integração dos impulsos, criando “caminhos” (circuitos) preferenciais para a informação. Esta hipótese, conhecida como hipótese de Hebb (Donald Hebb), é a base das novas teorias conexionistas sobre o funcionamento do cérebro e sua relação com a atividade mental [5].

### 1.3 - Redes neurais artificiais (RNAs)

As pesquisas na área de redes neurais artificiais foram motivadas inicialmente pelo desejo de construir máquinas capazes de realizar funções similares às do cérebro humano, visto que o cérebro processa informação de maneira bem mais eficiente que os computadores digitais. Pesquisadores começaram a olhar para a anatomia e fisiologia dos neurônios como uma possível fonte de inspiração para criar máquinas inteligentes. Neurologistas e outros pesquisadores dos processos mentais também viram nas redes neurais a possibilidade de aplicar suas teorias sobre o funcionamento cerebral e sobre processos como o aprendizado e classificação de padrões.

As RNAs são modelos matemáticos simplificados de sistemas neurais verdadeiros. O objetivo básico de uma rede neural é modelar algumas das propriedades conhecidas sob o título de *funções superiores do cérebro*, constituindo uma estrutura que procura mimetizá-lo. Assim como o cérebro, as RNAs extraem suas características mais importantes do comportamento coletivo de milhares de componentes básicos: os neurônios. Normalmente são implementadas utilizando-se componentes eletrônicos, ou através de programação em computadores.

Recorreremos a Haykin [6] para definir mais precisamente uma rede neural artificial:

*Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro em dois aspectos:*

- 1. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.*
- 2. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.*

As redes neurais artificiais modernas surgiram com o trabalho pioneiro de McCulloch e Pitts (1943), com um modelo inspirado na fisiologia dos neurônios. O neurônio formal é uma espécie de válvula lógica com dois possíveis estados internos: ativo (1) e repouso (0). Durante o processamento, cada neurônio recebe um número pequeno de

entradas provenientes das saídas de outros neurônios, as entradas são somadas e o estado do neurônio é determinado pelo sinal resultante com respeito a um certo limiar. Se o sinal for maior que o limiar, o neurônio é ativado; caso contrário, entrará em estado de repouso.

McCulloch e Pitts acreditavam que com um número suficientemente grande de unidades processadoras trabalhando em regime de cooperação e com pesos sinápticos ajustáveis, seria possível construir computadores de propósito universal, que poderiam realizar a computação de qualquer função calculável.

Frank Rosenblatt (1962) utilizaria posteriormente o neurônio formal de McCulloch e Pitts para implementar sua rede *perceptron*. O perceptron gerou bastante expectativa na comunidade científica, quando Rosenblatt apresentou seu *teorema da convergência do perceptron* e um *algoritmo de aprendizado*, que tornavam os perceptrons capazes de resolver qualquer problema de classificação de entradas separáveis linearmente.

No final da década, Minsky e Papert (1969) em seu livro *Perceptrons*, teceram sérias críticas à capacidade de processamento dos perceptrons. Uma das críticas apresentadas ficou conhecida como o problema do XOR (operador binário que realiza um *ou exclusivo* em lógica booleana). Eles demonstraram que o perceptron de camada simples é incapaz de realizar uma operação XOR, ou seja, dadas duas entradas binárias, deve-se produzir 1, se *apenas uma* das entradas estiver ligada, e produzir 0 em caso contrário.

Este problema pode ser facilmente resolvido empregando-se um perceptron multicamada, mas neste caso o teorema da convergência não pode ser utilizado. Na época em que este trabalho foi publicado não se conhecia um algoritmo de aprendizagem para os perceptrons multicamadas. Esta limitação e o impacto das críticas de Minsky e Papert fizeram com que a pesquisa em redes neurais artificiais permanecesse latente na década de 1970 [7].

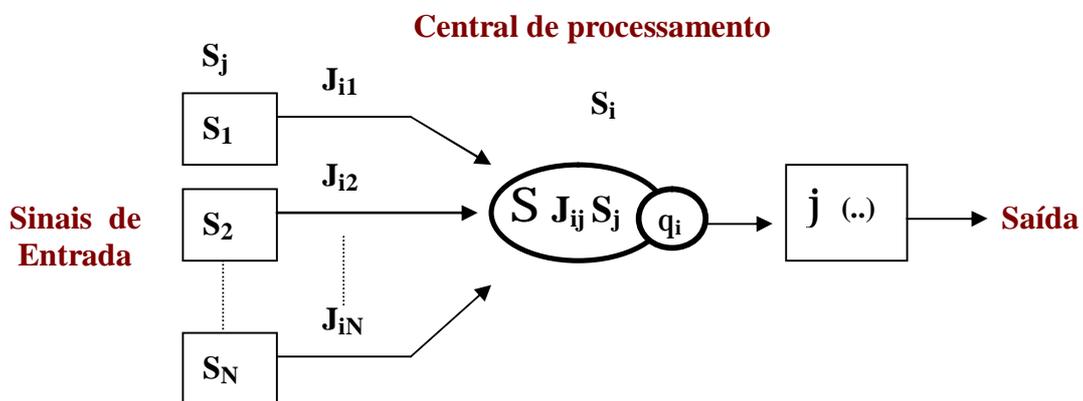
As RNAs tornaram a atrair a atenção de pesquisadores em 1982 com o artigo publicado pelo neurobiologista John Hopfield [8], que propunha um modelo para tratar o processo de memorização do cérebro. Neste trabalho, seguindo as idéias de Little [9], se estabelecia uma estreita correlação entre modelos de redes com atratores e modelos magnéticos de Ising utilizados na física estatística. Esta nova abordagem despertou um particular interesse dos físicos pela modelagem neural, consolidando uma nova linha de pesquisa.

Ainda na década de 1980, o interesse pelos perceptrons multicamadas foi reavivado com a divulgação, por Rumelhart et al.[10], do algoritmo “*backpropagation*” que resolvia

o problema da escolha dos pesos sinápticos para as camadas intermediárias dos perceptrons multicamadas.

#### 1.4 - Neurônio Matemático de McCulloch-Pitts

A figura 1.4 representa a estrutura de um neurônio formal no modelo de McCulloch-Pitts, esta estrutura lógica permanece até hoje como a base da teoria de computação neural. O modelo procura enfatizar os aspectos computacionais presentes no neurônio real e, portanto, define o neurônio formal de maneira extremamente simples do ponto de vista biológico.



**Figura 1.4:** Estrutura do neurônio de McCulloch-Pitts.

No modelo, um neurônio  $j$  pode estar em um dos dois estados possíveis:  $S_j = 1$  para o estado ativo e  $S_j = 0$  para o estado inativo. A influência do neurônio  $j$  sobre o neurônio  $i$  é representada pelo peso sináptico  $J_{ij}$ , onde o índice  $i$  refere-se ao neurônio pós-sináptico, e o índice  $j$  indica qual sinapse do neurônio  $i$  transmitirá o impulso. O peso sináptico  $J_{ij}$  vale 1 se a sinapse for excitatório e  $-1$  se for inibitória.

Para uma rede de  $N + 1$  neurônios, teremos então  $N$  neurônios  $S_j$  ( $j = 1, 2, \dots, N$ ) influenciando o neurônio  $S_i$ . O potencial de ação ou campo local induzido  $h_i$  sobre o neurônio  $i$  pode ser assim definido:

$$h_i = \sum_{j=1}^N J_{ij} S_j . \quad (1.1)$$

A equação (1.2) define o estado final do neurônio  $i$ , determinado pelo valor do campo local induzido  $h_i$  com respeito a um certo limiar  $q_i$ : se o campo local for maior que esse limiar, o neurônio é ativado ( $S_i = 1$ ); caso contrário, ele permanecerá inativo ( $S_i = 0$ ). A função de ativação  $j(\mathbf{K})$  tem o papel de limitar a resposta do neurônio a valores contidos no intervalo  $[0,1]$ .

$$S_i = j(\sum_{j=1}^N J_{ij} S_j - q_i) . \quad (1.2)$$

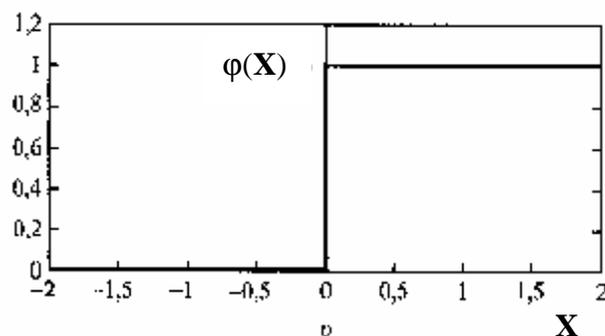
## 1.5 - Tipos de funções de ativação

Neste momento é conveniente descrever os tipos mais comuns de funções de ativação.

### 1.5.1 - Função de limiar:

$$j(X) = \begin{cases} 1, & \text{se } X \geq 0 \\ 0, & \text{se } X < 0 \end{cases}$$

A função de limiar, também chamada de função degrau, determina que se a variável  $X$  for não-negativa, o neurônio disparará com valor 1, caso contrário, a saída apresentará o valor 0. O gráfico da função de limiar está representado a seguir.



**Figura 1.5:** Gráfico da função de limiar.

### 1.5.2 - Função sigmóide:

A função sigmóide é a mais utilizada atualmente como função de ativação nas redes neurais pelo fato de interpolar suavemente os estados 0 e 1. O parâmetro  $a$  está associado à inclinação da função sigmóide como mostrado na figura 1.6. Um exemplo bastante comum de função sigmóide é a função tangente hiperbólica que interpola suavemente os estados -1 e 1.

$$j(X) = \frac{1}{1 + \exp(-aX)}$$

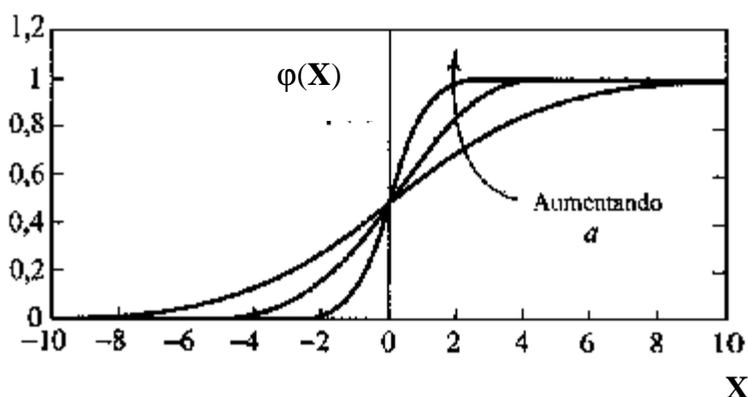


Figura 1.6: Gráfico da função sigmóide.

## 1.6 - Arquiteturas de redes neurais

A arquitetura de uma rede neural, ou seja, a maneira como seus neurônios estão conectados definindo grupos, e como estes grupos ou camadas se comunicam, influenciam profundamente o desempenho, as capacidades específicas que a rede pode desempenhar e até mesmo o algoritmo utilizado para treinar a rede [6].

A seguir destacamos os tipos de arquitetura mais importantes para o nosso estudo.

### 1.6.1 - Perceptrons ou arquitetura tipo *feed-forward*

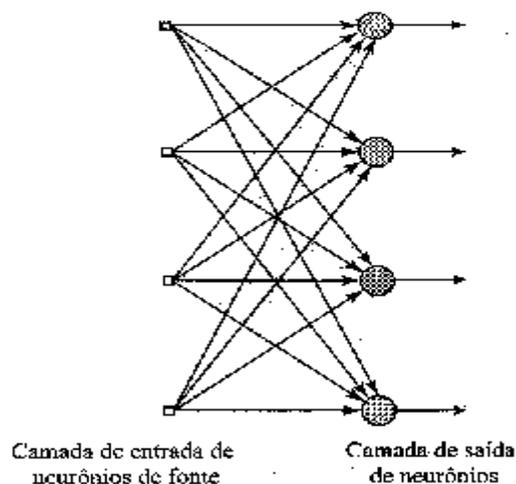
Os perceptrons foram concebidos por Frank Rosenblatt (1962), constituindo um dos primeiros modelos de redes neurais idealizado. A estrutura de um perceptron procura mimetizar algumas características dos neurônios biológicos, por exemplo, à semelhança

dos neurônios, as conexões dos perceptrons apenas permitem a passagem da informação em um único sentido. São por isso classificadas como redes alimentadas adiante (*feed-forward*). Os perceptrons podem ser organizados em redes com uma única camada ou em várias camadas, sendo classificadas em redes de camada única não-recorrente e redes de múltiplas camadas não-recorrentes, respectivamente, como descrito nos dois tópicos seguintes.

- **Rede de camada única não-recorrente**

Nas redes neurais de camada única, a camada de entrada tem o papel de receber os estímulos externos e convertê-los em impulsos que serão transmitidos ao restante da rede, neste nível, a informação de entrada ainda não sofreu nenhum processamento por parte da rede. Utilizando um conceito da teoria de grafos orientados, os neurônios nesse ponto são chamados também de *nós de fonte*, pois, como dissemos, eles apenas repassam a informação de entrada.

No caso mais simples, a rede de camada única não-recorrente compõe-se de uma camada de entrada que alimenta uma camada de saída, onde há o processamento da informação e seu fluxo em apenas um sentido, ou seja, sob nenhuma condição deverão existir conexões entre os neurônios da mesma camada e jamais a camada de saída poderá influenciar a camada de entrada. A figura 1.7 mostra uma rede neural com esta arquitetura.

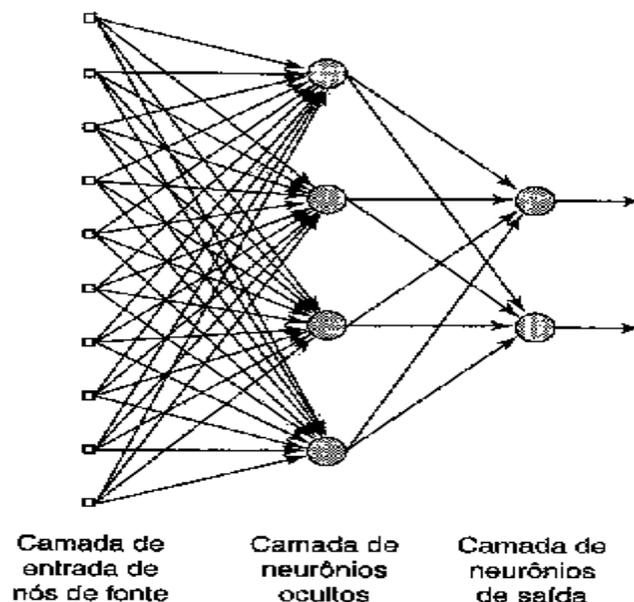


**Figura 1.7:** Desenho esquemático de um perceptron de camada única.

- **Rede de múltiplas camadas não-recorrente**

Na rede de múltiplas camadas, entre os nós de fonte (camada de entrada) e os nós de saída, há outras camadas responsáveis por processamento intermediário da informação, estas camadas convencionalmente são chamadas de camadas ocultas (*hidden*). Não são feitas restrições quanto ao número de camadas ocultas, sendo que estas camadas extras aumentam a capacidade de processamento da rede e o número elevado de conexões sinápticas torna a rede mais tolerante a falhas [6].

Neste modelo de arquitetura de rede os sinais da camada de entrada são transferidos à camada seguinte (oculta), após processamento, os sinais são encaminhados à camada seguinte, e assim por diante até que a camada de saída seja alcançada. A figura 1.8 mostra um desenho deste tipo de rede:



**Figura 1.8:** Desenho esquemático de um perceptron de múltiplas camadas

### 1.6.2 - Redes recorrentes ou arquitetura tipo *feedback*

As redes recorrentes são assim chamadas porque os neurônios podem estabelecer sinapses com uma camada anterior à sua, de forma que os impulsos gerados por estes neurônios irão realimentar a rede e, possivelmente, suas próprias entradas. Isto

corresponderia à troca de informação entre neurônios de uma mesma região do cérebro. No caso em que a saída de um neurônio sirva de entrada para ele mesmo, temos um laço de *auto-realimentação*. Os laços de realimentação também exercem profunda influência sobre a capacidade de aprendizagem da rede e seu desempenho, potencializando estas características [6].

Se considerarmos uma evolução temporal discreta,  $t = 0, 1, 2, \dots$ , podemos escrever a equação (1.2) como uma equação dinâmica,

$$S_i(t+1) = j [\sum_j J_{ij} S_j(t) - q_i] , \quad (1.3)$$

que transforma a entrada do sistema  $\{S_i(t)\}, i = 1, 2, \dots, N$ , no tempo  $t$ , em estados de saída  $\{S_i(t+1)\}$  no instante  $(t+1)$ . A equação acima define uma arquitetura neural do tipo *feedback*, pois a entrada é realimentada pela saída a cada passo de tempo. As redes recorrentes ainda podem ser *totalmente conectadas*, quando cada neurônio projeta-se sobre *todos* os outros neurônios da rede, ou *parcialmente conectadas*, caso em que algumas sinapses estão ausentes.

Independentemente da arquitetura da rede, os sinais obtidos na camada de saída constituem a resposta da rede ao padrão de entrada fornecido pelos nós de fonte (neurônios de entrada).

## 1.7 - O Modelo de Little: Processos Aleatórios

O modelo de Little [9] foi de grande importância no desenvolvimento da dinâmica de redes neurais. Deve-se a ele a semelhança entre os sistemas magnéticos de spins de Ising e o modelo proposto por McCulloch e Pitts. Nesta analogia, os estados dos neurônios são redefinidos pela relação  $S_i = 2S'_i - 1$ , de forma que os valores  $S_i = \{0, 1\}$  passam a ser  $S_i = \{-1, 1\}$  correspondendo aos casos inativo e ativo, respectivamente. Desta forma, o neurônio simplificado e as conexões sinápticas equivaleriam, na parte magnética, aos spins e aos acoplamentos magnéticos respectivamente:  $S_i = +1$  (spin para cima) e  $S_i = -1$  (spin para baixo). Já as conexões sinápticas,  $J_{ij}$ , correspondem às interações físicas entre os spins.

Outra contribuição importante foi a introdução de ruído no processo de comunicação entre os neurônios. São muitas as fontes de ruído envolvidas na transmissão de sinais entre neurônios reais, como por exemplo, a atividade dos neurotransmissores liberados na junção sináptica, flutuações no tempo de resposta de cada neurônio, variações nos limiares de disparo e outros processos não identificados e difíceis de descrever devido a complexidade das redes neurais reais. Uma vez que todos estes processos envolvendo aleatoriedade se refletirão no campo local do neurônio, uma maneira simples de representar este efeito (já que cada neurônio recebe informações de um número muito grande de outros neurônios) é supor o campo local  $h_i$  como uma variável aleatória cuja distribuição de probabilidade é dada pela função gaussiana

$$P_r(h) = \frac{1}{\sqrt{2ps^2}} \exp\left[-\frac{(h-h_i)^2}{2s^2}\right], \quad (1.4)$$

onde  $s$  é a largura da distribuição associada às diferentes fontes de aleatoriedade (ruído) e  $h_i$  é o campo local médio sobre o  $i$ -ésimo neurônio dado pela equação (1.1). Deste modo, redes neurais determinísticas, cujos neurônios obedeciam à regra dinâmica (1.3), tornaram-se redes probabilísticas, onde a nova regra de atualização fornece a probabilidade dos estados  $S_i(t+1)$  assumirem valores  $\pm 1$ :

$$S_i(t+1) = \begin{cases} +1, & \text{com probabilidade } P_r[h_i(t)] \\ -1, & \text{com probabilidade } 1 - P_r[h_i(t)] \end{cases}, \quad (1.5)$$

onde  $P_r[h_i(t)]$  é a distribuição de probabilidade que pode ser expressa em termos da função tangente hiperbólica na forma:

$$P_r[h_i(t)] = \frac{1}{2} [1 + \tanh(b_o h_i(t))]. \quad (1.6)$$

A equação acima (com as modificações apropriadas no significado das variáveis) é idêntica à equação que descreve a dinâmica de um modelo magnético (tipo Ising) em contato com um banho térmico à temperatura finita  $T_o$ . Aqui, o parâmetro  $b_o = 1/T_o$ , associado ao efeito da temperatura nos modelos magnéticos, equivale ao parâmetro de

ruído estocástico no modelo de redes neurais. Além destas características, a dinâmica de atualização dos estados no modelo de Little é síncrona (paralela).

Resumindo, neste capítulo apresentamos uma breve descrição do funcionamento da célula nervosa, um rápido resumo da evolução do estudo das redes neurais, detalhes das arquiteturas e de suas conexões, como também incorporamos a figura do neurônio típico. Apesar desta descrição reunir algumas informações importantes, ignoramos muitos detalhes do funcionamento da rede neural, quer seja pelo desconhecimento das especificidades ou mesmo devido à dificuldade na implementação do modelo. Apesar da complexidade das redes neurais reais, a partir das descrições anteriores, foi possível relacionar um conjunto de características básicas de modo a estruturar a modelagem das redes de neurônios. No próximo capítulo, vamos utilizar essas propriedades para estudar o modelo de Hopfield.

## Capítulo 2

### Modelo de Hopfield

#### 2.1 - Introdução

Redes neurais e sistemas magnéticos compartilham o fato de serem sistemas com um número muito grande de unidades simples que interagem entre si e com potenciais externos, dando origem a fenômenos coletivos de extrema complexidade como, por exemplo, transições de fase sofisticadas, atratores caóticos, etc. Nos últimos anos, com o aprimoramento e o desenvolvimento de novas ferramentas matemáticas e computacionais, a mecânica estatística tem dado contribuições inquestionáveis para o desenvolvimento da ciência, particularmente para o estudo da termodinâmica de equilíbrio e para o tratamento de modelos dinâmicos altamente complexos.

O cérebro humano é um sistema que apresenta um comportamento bastante complexo e o estudo das funções superiores do cérebro tem se tornado um campo de pesquisa para diferentes áreas do conhecimento. Neste contexto, as ferramentas da mecânica estatística têm sido extremamente úteis quando o interesse dos pesquisadores é entender o comportamento coletivo emergente das interações mútuas entre os inúmeros neurônios.

Um modelo de rede neural muito importante, do ponto de vista físico, foi introduzido pelo neurobiologista John Hopfield em 1982 [8] para modelar o processo de memorização no cérebro. O modelo de Hopfield, como é hoje conhecido, destacou-se originalmente ao associar o processo de memorização de uma rede neural ao problema de minimização da função energia da rede. Numa analogia com a física de sistemas magnéticos clássicos desordenados [11,12], o estado de ativação de cada neurônio

equivalaria aos estados “up” e “down” de um spin. O neurônio formal de McCulloch-Pitts é o elemento básico do modelo.

A rede de Hopfield difundiu-se amplamente como um modelo de *memória endereçável por conteúdo* (reconhecimento associativo), ou seja, o acesso às memórias é guiado não pela posição que a memória ocupa no sistema, mas sim pela própria informação armazenada. Uma vez conhecidos os padrões armazenados na rede, o problema consiste em determinar os pesos sinápticos que serão capazes de reproduzir estas memórias como pontos fixos atratores, pontos de mínimo da função energia.

Para ter êxito como um modelo de reconhecimento de padrões por associação, certas limitações, difíceis de serem justificadas de um ponto de vista biológico, foram impostas sobre a rede na sua versão original:

- i) As conexões sinápticas entre os neurônios são simétricas, ou seja,  $J_{ij} = J_{ji}$ , para qualquer  $i$  e  $j$ ;
- ii) Cada neurônio tem uma função de ativação não-linear;
- iii) Sua arquitetura é do tipo totalmente conectada: Cada neurônio conecta-se a todos os outros neurônios da rede, podendo receber ou emitir sinais ininterruptamente, sem que haja auto-realimentação (a saída de um neurônio não pode alimentar sua própria entrada);
- iv) As memórias são variáveis aleatórias independentes que assumem valores  $\pm 1$  com igual probabilidade (não é possível armazenar nenhum padrão com estrutura própria).

Em particular, a última exigência implica em redes com uma atividade média de 50%. Entretanto, sabe-se que nas regiões do cérebro responsáveis pelos processos de reconhecimento associativo, em média, apenas 4% dos neurônios estão simultaneamente ativos [13].

Outras limitações do modelo original dizem respeito à ausência de características como: ruído estocástico, diluição e assimetria nas conexões sinápticas, diferentes escalas de tempo presentes nas atividades neuronais (períodos refratários) e outros ingredientes importantes na biologia das redes neurais reais. Mesmo assim, o modelo de Hopfield foi fundamental como propulsor na busca de modelos mais realistas.

Existem duas fases distintas de processamento para uma rede de Hopfield: *armazenamento e recuperação*. Na fase de armazenamento, a matriz de pesos sinápticos do

sistema é calculada utilizando-se as memórias previamente armazenadas; é nessa etapa que as memórias são codificadas como mínimos de energia para a rede\*. Na fase de recuperação, um vetor de prova é atribuído à rede como seu estado inicial, previamente este vetor foi iniciado com uma versão incompleta de uma das memórias armazenadas. Sob a ação da dinâmica da rede, o sistema percorre o espaço de estados até encontrar a configuração de mínimo de energia mais próxima do padrão informado.

O processo de atualização da rede é, segundo a regra dinâmica para arquiteturas conexionista dada pela equação (1.2). Conhecida também como *redes neurais atratoras* a arquitetura conexionista destaca-se pela sua importância na modelagem dos processos relacionados à memorização. No caso, o estado do neurônio é atualizado se o campo local induzido é maior que um certo limiar. Caso contrário, o estado do neurônio permanece inalterado. Quando cessam as mudanças nos estados dos neurônios, dizemos que a rede atingiu um ponto fixo (atrator da dinâmica).

Neste ponto, deparamo-nos com o seguinte problema: Que método de atualização devemos empregar durante o processo dinâmico do sistema? Considerando-se que a arquitetura da rede é recorrente, o método empregado na atualização das unidades é importante e pode afetar o processamento [14]. Há dois métodos de atualização da rede: atualização *síncrona* (todas as unidades são atualizadas simultaneamente a cada passo de tempo) e atualização *assíncrona* (a cada passo de tempo uma unidade é escolhida aleatoriamente para ser atualizada). No modelo original de Hopfield foi utilizada a dinâmica assíncrona (seqüencial) e no modelo de Little a dinâmica síncrona (paralela).

Uma característica interessante do modelo de Hopfield é a autocorreção, se um neurônio dispara quando deveria permanecer inativo, ou fica inativo quando deveria disparar, os demais neurônios da rede tentam corrigir este comportamento local errôneo, de forma que a rede ainda tenderá a convergir para uma das memórias armazenadas, não afetando significativamente o comportamento do sistema.

Um inconveniente do modelo é a existência de *estados espúrios (indesejáveis)* que surgem quando aumentamos o número de padrões armazenadas na rede. Estados espúrios são estados estáveis que não correspondem às memórias armazenadas da rede.

---

\* Os detalhes de como os pesos sinápticos são determinados serão apresentados na seção seguinte ao falarmos do *postulado de aprendizagem de Hebb*.

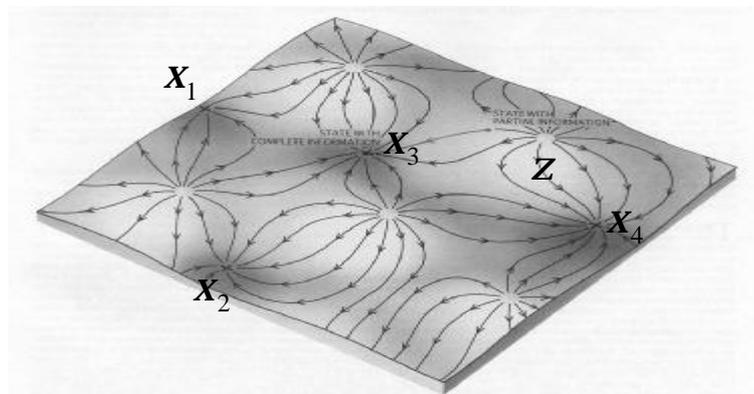
## 2.2 – A Natureza da Memória Associativa

Relembraremos resumidamente alguns dos aspectos do modelo de Hopfield na modelagem dos processos de memória associativa. Vamos considerar uma rede de  $N$  neurônios, cada neurônio modelado por uma variável binária tipo Ising  $S_i$ . Com o objetivo de simplificar algumas expressões da dinâmica do modelo, convencionaremos que os valores binários assumidos pelos neurônios da rede serão  $+1$  (neurônio ativo) e  $-1$  (neurônio inativo).

Usaremos o símbolo  $S_i$  para nos referir ao  $i$ -ésimo neurônio da rede, enquanto que o estado da rede será definido pelo vetor de estado:

$$\vec{S} = [S_1, S_2, \dots, S_N] . \quad (2.1)$$

O espaço de todos os possíveis estados da rede, ou seja, o espaço de configurações, é representado esquematicamente na figura 2.1. Cada padrão (ou memória), representado pela variável  $x$ , corresponde a um atrator dentro deste espaço e as regiões circunvizinhas definem suas bacias de atração. Os picos e vales presentes na superfície representam os níveis energéticos do sistema: cada pico corresponde a um estado energético mais alto e os vales a estados menos energéticos. Uma vez que no modelo de Hopfield as memórias da rede são armazenadas como mínimos de energia do sistema, se a rede for iniciada em um estado próximo de uma memória, como por exemplo, uma cópia incompleta (representada na figura por  $Z$ ), a rede descreverá uma trajetória para o vale mais próximo no espaço de estados, encaminhando-se para a versão completa da memória armazenada. Uma vez que a dinâmica da rede a leve para um dos atratores, ela permanecerá lá indefinidamente.



**Figura 2.1** Representação esquemática do espaço de configuração.

Como vimos, a rede de Hopfield utiliza o neurônio de McCulloch-Pitts, herdando assim seu aspecto “tudo ou nada” característico de sua função de ativação. O processo de atualização é dado segundo a regra determinística:

$$S_i = \begin{cases} +1 & \text{se } h_i \geq 0 \\ -1 & \text{se } h_i < 0 \end{cases}, \quad (2.2)$$

ou, de maneira mais simples, em termos da função sinal ( $sgn$ ),

$$S_i(t+1) = sgn[h_i(t)]. \quad (2.3)$$

O termo  $h_i(t)$  representa o campo local induzido sobre o neurônio  $i$  no tempo  $t$  e é determinado pela soma de todas as contribuições dos demais neurônios da rede. Levando-se em conta as respectivas eficiências sinápticas dos demais neurônios sobre o neurônio  $i$  ( $J_{ij}$ ), e o limiar de ativação  $q_i$  próprio de cada neurônio, o campo local tem a seguinte expressão:

$$h_i(t) = \sum_j^N J_{ij} S_j(t) - q_i. \quad (2.4)$$

A regra de armazenamento no modelo de Hopfield baseia-se no postulado enunciado por Hebb que diz:

*Quando um axônio da célula A está perto o suficiente para excitar uma célula B e participa do seu disparo repetida ou persistentemente, então algum processo de crescimento ou modificação metabólica acontece em uma das células ou em ambas, de tal forma que a eficiência de A como uma das células que dispara B é aumentada.*

Vamos considerar o conjunto de  $p$  configurações que serão armazenadas na rede (chamaremos de padrões ou memórias indistintamente) representadas pela variável  $\vec{x}_i^m = (x_1^m, x_2^m, \dots, x_N^m)$  com  $m = 1, 2, \dots, p$ . A expressão matemática para regra de Hebb, ou regra de Hebb generalizada, como também é conhecida, define o peso sináptico entre o neurônio  $i$  e o neurônio  $j$  como a seguir:

$$J_{ij} = \frac{1}{N} \sum_{m=1}^p \mathbf{X}_i^m \mathbf{X}_j^m, \quad J_{ii} = 0. \quad (2.5)$$

Observe que na definição dos pesos sinápticos fizemos  $J_{ii} = 0$ , justificando o fato de que nas redes de Hopfield não há auto-realimentação. Hopfield mostrou que com esta regra sináptica, a equação (2.3) permitia o armazenamento associativo de memórias, desde que as memórias sejam configurações aleatórias não correlacionadas, ou seja, os  $\mathbf{X}_i^m$  devem ser gerados assumindo valores  $\pm 1$  com igual probabilidade.

Com a mudança dos estados de  $S = \{0, 1\}$  para  $S = \{-1, +1\}$ , o limiar de ativação tem a forma:

$$q_i = - \sum_{j \neq i}^N J_{ij} \quad (2.6)$$

de modo que o campo local (2.4) assumirá a forma mais simples:

$$h_i(t) = \sum_j^N J_{ij} S_j(t). \quad (2.7)$$

Apesar de ser inspirada no princípio de Hebb, a equação (2.5) incorpora características que vão além das presentes no postulado original, e que não são justificadas fisiologicamente, por exemplo, incrementa o peso sináptico positivamente quando nenhum dos neurônios está disparando ( $x_i^m = x_j^m = -1$ ), pode ocorrer também que uma sinapse passe de excitatória para inibitória ou vice versa quando mais padrões são adicionados à rede, o que nunca ocorreria em uma sinapse real.

### 2.3 – Implementação da Dinâmica (padrões estáveis)

Para uma melhor compreensão da dinâmica da rede, demonstraremos que a definição adotada para os pesos sinápticos do sistema mantém algumas propriedades do postulado de Hebb, e, principalmente, que padrões armazenados segundo esta regra tornam-se atratores estáveis da rede. Deste modo, pequenos desvios destes padrões serão corrigidos à medida que a rede evolui.

Como utilizamos uma dinâmica assíncrona de atualização, onde cada unidade é atualizada individualmente, admitiremos que nossa rede  $S$  atingiu um estado de equilíbrio estável (atrator) quando a aplicação da dinâmica, equação (2.3), não causar mais nenhuma mudança em qualquer um dos estados da rede. Trataremos inicialmente do caso mais simples em que a rede tem apenas uma memória, depois abordaremos o caso genérico com  $N$  padrões.

### 2.3.1 - Rede com um único padrão

De acordo com nosso critério de estabilidade, um padrão será estável se os estados microscópicos não sofrerem atualizações quando submetidos à regra (2.3). Vamos considerar o caso em que apenas o padrão  $x_i$  está armazenado na rede. Da condição inicial que  $S_i(t) = x_i$ , é fácil observar que o padrão  $x_i$  é um atrator de ponto fixo da rede, já que os pesos sinápticos obedecem a relação  $J_{ij} \propto x_i x_j$  e  $x_j x_j = 1$ . Da equação (2.3) obtém-se:

$$S_i(t+1) = \text{sgn}\left(\sum_{j \neq i}^N \frac{1}{N} x_i x_j x_j\right) = x_i \quad (\text{para todo } i) . \quad (2.8)$$

Quando a rede é iniciada com a versão apenas parcial da memória, não mais distante que cinquenta por cento do padrão original, os bits “incorretos”, que não correspondem ao padrão inicial, serão eliminados pela maioria dos bits que estão corretos, na determinação do campo local induzido (2.7), e novamente o sistema evolui até a memória  $x_i$ . Esta propriedade indica que a rede apresenta robustez, no sentido que está capacitada a recuperar o padrão completo a partir de seu conhecimento parcial. Quando a rede é inicializada distante mais de cinquenta por cento do padrão a ser reconhecido, o sistema evolui para a antimemória, ou estado reverso,  $-x_i$ , desta forma, para este caso simples há dois atratores, sendo o segundo oposto ao padrão armazenado.

### 2.3.2 - Rede com muitos padrões

Vamos agora considerar o caso em que  $p$  padrões estão armazenados na rede. Consideraremos um desses padrões como condição inicial,  $S_i(t=0) = x_i^n$ . Neste caso, a condição de estabilidade para um padrão particular  $x_n$  assume a forma mais genérica,

$$S_i(t+1) = \text{sgn}[h_i(t)] = x_i^n \quad (\text{para todo } i), \quad (2.9)$$

onde

$$h_i(t) \equiv \sum_j^N J_{ij} S_j(t) = \frac{1}{N} \sum_j \sum_m x_i^m x_j^m x_j^n. \quad (2.10)$$

Separando, na equação (2.10), o termo particular  $m=n$  da soma sobre  $m$ , obtemos

$$h_i(t) = x_i^n + \frac{1}{N} \sum_j^N \sum_{m \neq n}^p x_i^m x_j^m x_j^n. \quad (2.11)$$

O primeiro termo do campo local é chamado de *termo de sinal* (padrão de reconhecimento) e o segundo, *termo de ruído*. Se o valor absoluto do termo de ruído for suficientemente pequeno, o termo de sinal será dominante e o padrão  $x_n$  será estável segundo (2.9). Este é o caso em que o número de padrões armazenados  $p$  é pequeno comparado com o tamanho  $N$  da rede. Assim, para valores pequenos de  $p$ , os padrões armazenados serão todos estáveis. Além do mais, se a rede for iniciada com uma pequena fração de bits trocados, com relação a uma memória armazenada, a rede será corrigida do mesmo modo que no caso simples de uma única memória, pois, novamente, os bits incorretos serão cancelados na soma pela maioria de bits corretos.

## 2.4 - Capacidade de armazenamento da rede de Hopfield

O cérebro humano é limitado com relação ao armazenamento e reconhecimento de informações. No reconhecimento de padrões por associação existe uma quantidade máxima de informação que pode ser retida e que varia de indivíduo para indivíduo, deste modo, quanto maior a quantidade de informação armazenada que pode ser utilizada eficientemente por um indivíduo, maior é sua capacidade de memorização. Para uma quantidade de informação superior à sua capacidade máxima, o cérebro responde de maneira indesejada (erroneamente).

A rede de Hopfield apresenta um comportamento análogo às redes reais no que diz respeito ao reconhecimento de padrões por associação. Como já comentamos anteriormente, um dos problemas presentes na rede é que o número de estados espúrios cresce à medida que o número de padrões armazenados na rede aumenta [15]. Nesta seção,

determinaremos os parâmetros que medem a capacidade de armazenamento da rede e o reconhecimento de uma dada memória.

Podemos fazer uma analogia do armazenamento de memórias em nosso sistema com o problema de interferência sinal-ruído, recorrendo novamente à equação (2.11). O termo de sinal  $x_i^n$ , corresponde à componente desejada do campo local e o termo de ruído sináptico é resultante da superposição de todos os demais padrões armazenados. Chamaremos o termo de ruído de  $d_i^n$ , indicado a seguir:

$$d_i^n = \frac{1}{N} \sum_j^N \sum_{m \neq n}^p x_i^m x_j^m x_j^n . \quad (2.12)$$

Como os  $x_i^s$  são variáveis aleatórias e assumem valores  $\pm 1$  com igual probabilidade, o termo de ruído consiste em uma soma de  $N(p-1)$  termos aleatórios, cada um contribuindo com +1 ou -1 à soma. Desde que os bits em diferentes padrões são descorrelacionados, esta soma pode ser interpretada como uma caminhada aleatória unidimensional de passos unitários. Assim como na caminhada, os passos são dados para a direita ou esquerda com igual probabilidade. No limite de  $N$  muito grande e de acordo com o teorema do limite central (ver apêndice C),  $d_i^n$  é uma variável aleatória gaussiana com média zero e variância dada por:

$$s^2 = (p-1)/N \approx p/N . \quad (2.13)$$

Esta expressão nos mostra que a variância (dispersão) aumenta linearmente com  $p$  e, conseqüentemente, o termo de ruído estático, gerado pela superposição das memórias armazenadas, pode crescer o suficiente para destruir a estabilidade dos padrões armazenados. Já para a componente de sinal  $x_i^n$ , restrita aos valores  $\pm 1$  com igual probabilidade, obtemos média zero e variância unitária. A relação sinal-ruído é definida como o quociente da variância do sinal pela variância do ruído [6,13,16]:

$$r \equiv \frac{1}{(p-1)/N} \cong \frac{N}{p} \quad (\text{Para } p \text{ grande}) . \quad (2.14)$$

Para que as memórias fundamentais sejam estáveis é necessário que a relação sinal-ruído  $r$  seja alta ( $N > p$ ). A recíproca da relação sinal-ruído define um limite aproximado

para a quantidade de memórias  $p$  que podem ser armazenadas na rede em função da quantidade de neurônios  $N$ . Acima deste limite as memórias armazenadas deixam de ser estáveis. Este quantidade é chamada de *parâmetro de carga* ou mais comumente *capacidade de armazenamento da rede*. Da equação (2.13) podemos notar que a capacidade de armazenamento é a própria variância do termo de ruído.

$$a \equiv \frac{p}{N} . \quad (2.15)$$

As redes de Hopfield apresentam um comportamento estável até que o parâmetro de carga atinja o valor crítico  $a_c \cong 0,14$  [8]. Isto indica que numa rede de 1000 neurônios aproximadamente 140 padrões podem ser armazenados e recuperados sem que haja erro. Acima deste valor a rede não reconhecerá nenhum padrão.

Um outro parâmetro importante nas redes de Hopfield, que tem seu análogo na magnetização dos modelos magnéticos, é a superposição entre o estado do sistema e a memória a ser reconhecida  $\mathbf{x}_i^m$ :

$$m_m(t) = \frac{1}{N} \sum_i^N x_i^m S_i(t) . \quad (2.16)$$

Este parâmetro, que assume valores no intervalo  $[-1, +1]$ , mede a proximidade entre o estado do sistema  $S_i(t)$  e a memória que se quer reconhecer  $\mathbf{x}_i^m$ . Por exemplo, se o estado da rede está descorrelacionado completamente com um certo padrão  $\mathbf{x}^n$ , a superposição correspondente  $m_n$  será nula, entretanto, se a correlação com o padrão for completa, teremos para essa superposição particular  $m_n = 1$ . Devido à independência das memórias,  $\langle \mathbf{x}_i^m \mathbf{x}_j^n \rangle = d_{ij} d_{mn}$ , todas as outras superposições tenderão a zero  $m_{m \neq n} \rightarrow 0$ . De qualquer modo, quando o sistema atinge um ponto fixo atrator, os valores das superposições permanecem constantes no tempo.

Os principais resultados originalmente observados por Hopfield foram os seguintes:

- Existe uma faixa de valores de  $a$  para a qual o sistema mantém uma ótima performance na recuperação das memórias. Para  $a$  neste intervalo, os estados estáveis da rede, apesar de modificados pela presença de ruído, continuarão

muito próximos dos padrões originais ( $a < a_c \cong 0,14$ ). Quando  $a$  tornar-se maior que seu valor crítico nenhum padrão poderá ser recuperado;

- As memórias armazenadas e suas inversas são pontos fixos estáveis da dinâmica;
- A transição entre as fases de reconhecimento e não-reconhecimento de padrões é de primeira ordem (descontínua);
- Para valores pequenos de  $\alpha$ , surgem *estados espúrios* que possuem superposição finita com diversos padrões e correspondem a estados de mistura entre as memórias armazenadas.

A analogia entre os modelos magnéticos e a rede de Hopfield, levou Amit, Gutfreund e Sompolinsky em 1985 [17] a propor uma generalização do modelo original. A seguir, analisaremos essa contribuição.

## 2.5 - Generalização do Modelo de Hopfield

### 2.5.1 - Termodinâmica do Modelo de Hopfield

Uma importante contribuição teórica ao modelo de Hopfield deve-se a Amit, Gutfreund e Sompolinsky [17]. Eles propuseram uma generalização do modelo que introduzia ruído na dinâmica do sistema, inspirados nos diferentes elementos estocásticos presentes na transmissão de sinais em sistemas reais. Utilizando um formalismo análogo ao utilizado no estudo de modelos de vidro de spin (método das réplicas) [11], estudaram analiticamente a termodinâmica de equilíbrio do modelo de Hopfield, tendo na temperatura o parâmetro que controla o nível de ruído estocástico.

Nesta analogia com os modelos magnéticos, os estados de equilíbrio, que são mínimos da função energia livre, são governados por um conjunto de equações acopladas obtidas com o método das réplicas simétricas que equivale a uma aproximação de campo médio. A energia livre e as equações que descrevem os estados de equilíbrio do sistema são [17,18]:

$$\begin{aligned}
f = & \frac{a}{2} + \frac{1}{2} \sum_m m_m^2 + \frac{a}{2b} \ln[1 - b(1 - q)] - \frac{1}{2} \frac{aq}{[1 - b(1 - q)]} \\
& + \frac{1}{2} abr(1 - q) - \frac{1}{b} \langle\langle \ln 2 \cosh b(\sum_m m_m x^m + \sqrt{ar}z) \rangle\rangle
\end{aligned} \tag{2.17}$$

e

$$\begin{aligned}
m_n = & \langle\langle x^n \tanh b(\sum_m m_m x^m + \sqrt{ar}z) \rangle\rangle \\
r = & \frac{q}{[1 - b(q - 1)]^2} \\
q = & \langle\langle x^n \tanh^2 b(\sum_m m_m x^m + \sqrt{ar}z) \rangle\rangle,
\end{aligned} \tag{2.18}$$

onde o símbolo  $\langle\langle \dots \rangle\rangle$  significa uma média dupla: uma sobre o conjunto das variáveis  $x$  (padrões armazenado) e a outra sobre o ruído gaussiano  $z$ . Uma derivação heurística simples das propriedades do modelo de Hopfield com acoplamentos simétricos foi desenvolvida por Geszti [19]. Ele obteve um conjunto de equações acopladas, idênticas às equações obtidas por Amit et al., dadas por:

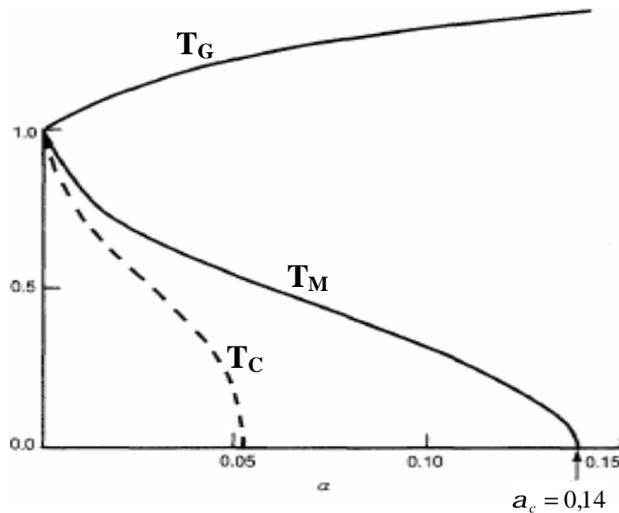
$$\begin{aligned}
m = & \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} \tanh[b(m + \sqrt{ar}z)] dz \\
r = & \frac{q}{[1 - b(q - 1)]^2} \\
q = & \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} \tanh^2[b(m + \sqrt{ar}z)] dz,
\end{aligned} \tag{2.19}$$

onde  $m$  mede a superposição do sistema com o padrão de reconhecimento, o parâmetro  $r$ , conhecido como parâmetro AGS (Amit, Gutfreund, Sompolinsky), está associado com a média do quadrado da superposição do sistema com os outros padrões e  $q$  é o parâmetro Edward-Anderson [12].

Para o caso particular  $T \rightarrow 0$ , o modelo determinista original de Hopfield é recuperado, e o valor crítico encontrado para a capacidade de armazenamento,  $a_c \approx 0,138$ , é muito próximo do obtido por Hopfield através de simulações. Para  $a < a_c$  a rede reconhece eficientemente as memórias armazenadas e tem uma superposição não nula com

um único padrão ( $m \approx 1$ ). Em  $a = a_c$  a superposição com a memória reconhecida é aproximadamente  $m \approx 0,97$  e para  $a > a_c$  as soluções de reconhecimento desaparecem abruptamente e o sistema perde a capacidade de reconhecer as memórias,  $m = 0$  (fase de não reconhecimento) [15].

Na presença do ruído térmico  $T \neq 0$ , apresentamos na figura 2.2 o diagrama de fase do modelo no plano  $T \times a$ .  $T_G$  é a temperatura de transição que separa a fase paramagnética (acima) da fase de vidro de spin (abaixo). As soluções de vidro de spin ( $q \neq 0$ ) existem para todos os valores de  $a$ . O principal resultado é a existência de uma região aproximadamente triangular, abaixo de  $T_M$ , na qual existem soluções de reconhecimento  $m \neq 0$ . Para  $a < a_c$ , os estados de reconhecimento aparecem descontinuamente à medida que  $T$  decresce, ou seja, a transição separando a fase de reconhecimento da fase de não reconhecimento é de primeira ordem. À medida que o ruído térmico aumenta, a descontinuidade na linha de transição  $T_M$  diminui com a diminuição de  $a$  e em  $a = 0$  a transição é de segunda ordem com  $T^* = 1$ . Abaixo de  $T_C$  as soluções de reconhecimento são mínimos globais de energia, enquanto que na região acima de  $T_C$  estas soluções são mínimos locais e os mínimos globais são os estados de vidro de spin [19].



**Figura 2.2:** Diagrama de fase  $T$  versus  $a$  para o modelo de Hopfield generalizado [16].  $T_G$  é a temperatura de transição para os estados de vidro de spin,  $T_M$  é a temperatura em que surgem os estados ferromagnéticos,  $T_C$  é a temperatura em que estes estados tornam-se mínimos globais, passando por uma transição de primeira ordem.

Simulações numéricas à temperatura nula ( $T=0$ ), para redes com mais de 30000 unidades, foram realizadas utilizando a técnica de “multispin coding” [20]. Os resultados obtidos na simulação mostraram uma capacidade de armazenamento maior que a obtida com a aproximação de réplica simétrica  $a_c = 0,143 \pm 0,001$ . Esta divergência no valor do  $a_c$  foi reduzida utilizando o método das réplicas com um passo de quebra de simetria, pelo qual se obteve  $a_c \approx 0.144$  [21].

Os resultados analíticos descritos acima correspondem às propriedades termodinâmicas do modelo de Hopfield quando os estados do sistema são atualizados sequencialmente. Para o processo de atualização em paralelo, introduzido por Little, Peretto [22,23] mostrou que para redes tipo Hopfield a distribuição de Gibbs leva a pequenas modificações no Hamiltoniano efetivo. O modelo de Hopfield com dinâmica paralela foi estudado por Amit et al. [17] para um número de padrões finitos ( $a = 0$ ) e por Fontanari e Köberle [24] para um número extensivo de padrões ( $a \neq 0$ ).

Como vimos, a aproximação de campo médio tem mostrado ser uma ferramenta importante e tem sido aplicada no estudo das redes neurais, particularmente no estudo de redes com conexões simétricas, nas quais a condição de balanço detalhado é satisfeita, permitindo utilizar a termodinâmica de equilíbrio [15,17,24]. Sistemas com conexões assimétricas, diluição nas sinapses e campos locais que dependem do estado do sistema, todos violam esta condição e conseqüentemente não é mais possível utilizar um desenvolvimento termodinâmico. Nestas situações, como alternativa ao tratamento termodinâmico, podemos analisar o desenvolvimento da *dinâmica* do modelo. Dedicaremos a parte final deste capítulo ao estudo da dinâmica do modelo de Hopfield, que servirá de base para o desenvolvimento desta dissertação.

### **2.5.2 Dinâmica do modelo de Hopfield Diluído e Assimétrico**

Normalmente, os processos dinâmicos são bem mais difíceis de se tratar que os processos termodinâmicos. Em geral, um tratamento analítico da dinâmica quase sempre envolve algum tipo de aproximação para descrever a evolução temporal do sistema. Mesmo assim, dificuldades e aproximações muitas vezes são recompensadas pelo fato de que um processo dinâmico proporciona informações válidas não somente para regimes assintóticos, mas também para regimes intermediários e curtos, onde novos efeitos podem ser observados.

Nesta linha, uma outra contribuição importante ao modelo de Hopfield foi desenvolvida por Derrida, Gardner e Zippelius, em 1987 [25]. Eles propuseram a introdução de novos ingredientes biológicos como diluição e assimetria nas conexões sinápticas. A motivação principal deste trabalho deveu-se às inúmeras provas biológicas de que as redes neurais reais não têm todos os seus elementos conectados entre si, ou seja, cada neurônio conecta-se apenas a um número reduzido de outros neurônios mais próximos (da ordem de  $10^4$  neurônios), que é muito inferior ao número total de elementos da rede (da ordem de  $10^{11}$  neurônios). Há dados também que evidenciam, por exemplo, que regiões do neocortex e hipocampo apresentam arquiteturas modulares e elementos com conectividades estruturadas, fatores que estão ausentes em modelos totalmente conectados [3]. Um aspecto importante é que ao introduzir diluição nas sinapses, simplesmente cortando as conexões entre os neurônios, introduz-se também outro ingrediente verificado biologicamente: a assimetria nas conexões sinápticas.

O grande triunfo do modelo proposto por Derrida, Gardner e Zippelius, chamado, de agora em diante, modelo DGZ, está no fato de que ao mesmo tempo em que incorpora dois ingredientes biológicos importantes presentes nas redes neurais reais (diluição e assimetria nas sinapses), ele permite ainda uma solução analítica exata da dinâmica do modelo.

Neste modelo, cada neurônio está conectado, em média, a  $C$  outros neurônios, sendo este grau de conectividade ( $C$ ) muito inferior ao do modelo original de Hopfield ( $N$ ), situação conhecida como extrema diluição ( $C \ll \ln N$ ). Por definição, as conexões sinápticas  $J_{ij}$  entre os neurônios da rede são definidas pela regra Hebbiana:

$$J_{ij} = C_{ij} \sum_{m=1}^p x_i^m x_j^m, \quad (2.20)$$

onde os  $x_i^m$  são variáveis aleatórias independentes, podendo assumir valores  $\pm 1$  com igual probabilidade. A variável  $x_i^m$  representa o estado do  $i$ -ésimo neurônio correspondente ao padrão armazenado de índice  $m$  ( $m = 1, 2, \dots, p$ ). Os  $C_{ij}$  são também variáveis aleatórias independentes que podem tomar valores 0 ou 1, escolhidos segundo a distribuição de probabilidades  $r(C_{ij})$ :

$$r(C_{ij}) = \frac{C}{N} d(C_{ij} - 1) + \left(1 - \frac{C}{N}\right) d(C_{ij}). \quad (2.21)$$

A variável  $C_{ij}$  introduz a assimetria e diluição no modelo. A assimetria surge do fato que, sendo variáveis aleatórias independentes, podemos ter  $C_{ij} \neq C_{ji}$ , enquanto que a diluição ocorre porque agora cada conexão pode existir ( $C_{ij} = 1$ ) com probabilidade  $C/N$ , ou ser eliminada ( $C_{ij} = 0$ ) com probabilidade  $(1 - C/N)$ .

De acordo com o processo de atualização do modelo, cada sítio da rede é atualizado por uma dinâmica estocástica paralela dada pela seguinte regra:

$$S_i(t + \Delta t) = \begin{cases} +1 & \text{com probabilidade } (1 + \exp[-2h_i(t)/T_0])^{-1} \\ -1 & \text{com probabilidade } (1 + \exp[2h_i(t)/T_0])^{-1} \end{cases}, \quad (2.22)$$

que podemos escrever em função da tangente hiperbólica na forma:

$$S_i(t + \Delta t) = \pm 1, \text{ com probabilidade } \Pr(S_i) = \frac{1}{2} \left[ 1 + S_i \tanh\left(\frac{h_i(t)}{T_0}\right) \right]. \quad (2.23)$$

Aqui,  $T_0$  (temperatura, em analogia aos modelos magnéticos) representa a fonte de ruído estocástico do sistema, que difere do ruído estático (devido ao número de padrões armazenados), muito embora, ambos tendam a desestabilizar o reconhecimento dos padrões. O  $h_i(t)$ , que representa o campo local induzido sobre o  $i$ -ésimo neurônio da rede, é definido da maneira convencional:

$$h_i(t) = \sum_j J_{ij} S_j. \quad (2.24)$$

Seguindo estritamente o desenvolvimento elaborado por Derrida, Gardner e Zippelius para o modelo DGZ, vamos considerar a evolução temporal do sistema para o caso em que a configuração inicial no tempo  $t$ ,  $\{S_i(t)\}$ , tem uma superposição não nula com um dos padrões armazenados, padrão  $m = 1$ , e uma superposição nula com os demais  $p - 1$  padrões. A superposição para o padrão  $m = 1$  será então chamada simplesmente  $m(t)$  e definida por

$$m(t+1) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^1 \langle S_i(t+1) \rangle. \quad (2.25)$$

Considerando uma dinâmica paralela, onde todos os neurônios são atualizados simultaneamente segundo a regra (2.23), deduziremos uma relação de recorrência para a superposição  $m(t+1) = f(m(t))$ .

Combinando as equações (2.23) e (2.25), podemos escrever:

$$m(t+1) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^1 \tanh\left(\frac{h_i(t)}{T_0}\right), \quad (2.26)$$

que nos dá a superposição em função do campo local. Convencionando chamar de  $j_1, j_2, \mathbf{K}, j_K$  os  $K$  neurônios conectados ao neurônio  $i$ , tais que  $J_{ij} \neq 0$ , e usando (2.20) e (2.24), podemos expressar o campo local  $h_i(t)$  no argumento da tangente hiperbólica em (2.26) como

$$h_i(t) = \sum_{r=1}^K \sum_{m=1}^p \mathbf{x}_i^m \mathbf{x}_{j_r}^m S_{j_r}(t), \quad (2.27)$$

onde consideramos apenas os termos efetivos dos pesos sinápticos ( $C_{ij} = 1$ ). Separando o termo que desejamos reconhecer (padrão  $m = 1$ ) do somatório acima, temos:

$$h_i(t) = \mathbf{x}_i^1 \left( \sum_{r=1}^K \mathbf{x}_{j_r}^1 S_{j_r}(t) \right) + \sum_{r=1}^K \sum_{m=2}^p \mathbf{x}_i^m \mathbf{x}_{j_r}^m S_{j_r}(t). \quad (2.28)$$

Considerando o limite termodinâmico  $N \rightarrow \infty$ , e a conectividade média da rede  $C$  satisfazendo a condição de *extrema diluição*

$$C \ll \ln N, \quad (2.29)$$

é possível mostrar [26,27] que os  $K$  neurônios  $S_{j_1}(t), S_{j_2}(t), \mathbf{K}, S_{j_K}(t)$ , têm correlações temporais desprezíveis, ou seja, cada termo da série na equação (2.28), para qualquer tempo  $t$ , é independente dos demais (cada neurônio possui sua árvore temporal própria de ancestrais, independente das árvores dos demais). Desta forma, as equações (2.26) e (2.28) podem ser manipuladas para se chegar à seguinte relação de recorrência (ver descrição detalhada no apêndice A):

$$m(t+1) = f(m(t)) = \sum_{k=0}^{\infty} \frac{C^k e^{-C}}{k!} \sum_{n=0}^k \sum_{s=0}^{k(p-1)} \Pr(n, s) \tanh\left(\frac{kp - 2n - 2s}{T_0}\right), \quad (2.30)$$

onde

$$\Pr(n, s) = \frac{1}{2^{kp}} \binom{k(p-1)}{s} \binom{k}{n} (1+m_0)^{k-n} (1-m_0)^n. \quad (2.31)$$

Devido à impossibilidade de calcular os somatórios presentes na expressão (2.30), vamos nos limitar ao caso particular  $C \rightarrow \infty$ . Se garantirmos que a condição (2.29) seja satisfeita, ou seja,  $N \rightarrow \infty$  mais rapidamente que  $C$ , podemos obter a equação dinâmica que determina os valores críticos da temperatura ( $T^*$ ) e da capacidade de armazenamento ( $a_c$ ), a partir dos quais o sistema não reconhece os padrões armazenados.

Neste limite, a expressão (2.28) é uma soma de infinitas variáveis aleatórias independentes que, de acordo com o Teorema do Limite Central, apresenta uma distribuição de probabilidade gaussiana. A média e a variância da série podem ser obtidas facilmente, de modo que a expressão (2.26) pode ser reescrita em forma integral. No apêndice A, mostramos detalhadamente como obter a equação de recorrência para a superposição do estado do sistema com o padrão  $m = 1$ , a qual é dada por:

$$m(t+1) = \int_{-\infty}^{+\infty} Dz \tanh\left(\frac{m(t) + z\sqrt{a}}{T}\right), \quad (2.32)$$

onde definimos  $Dz \equiv \frac{dz e^{-z^2/2}}{\sqrt{2p}}$ , a capacidade de armazenamento para o modelo diluído

$a = p/C$  e a temperatura reduzida  $T = T_0/C$ .

Note que a definição  $a = p/C$  está de acordo com a adotada anteriormente no caso do modelo de Hopfield usual: a conectividade média por neurônio é  $C = N$ . No limite de tempos grandes, o sistema converge para um ponto fixo no qual a relação  $m(t+1) = f(m(t))$  se reduz a  $m^* = f(m^*)$ , e a estabilidade da solução  $m = 0$  pode ser analisada com base no parâmetro  $I = \left. \frac{\partial f(m)}{\partial m} \right|_{m=0}$ . Para  $I < 1$  a solução  $m = 0$  é estável e para  $I > 1$  a solução é instável. O caso  $I = 1$  corresponde à linha crítica que separa a solução  $m = 0$  da solução  $m \neq 0$ . Aplicando a condição  $\lambda=1$  em (2.32) obtemos a linha crítica no plano  $T \times \alpha$ :

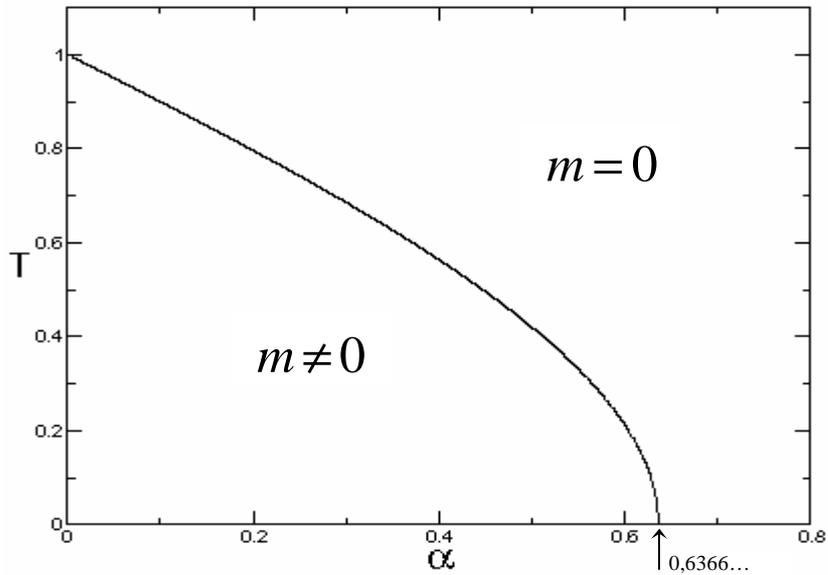
$$T^* = 1 - \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh^2 \left( \frac{z\sqrt{a}}{T^*} \right), \quad (2.33)$$

que pode ser resolvida numericamente.

Na figura 2.3, apresentamos o diagrama de fase  $T$  x  $a$  do modelo ultra diluído. A linha crítica é de segunda ordem, separando a fase de reconhecimento ( $m \neq 0$ ) da fase desordenada ( $m = 0$ ).

Para  $a = 0$ , temos  $T^* = 1$ , e na ausência de ruído térmico ( $T = 0$ ), a equação (2.32) assume a forma mais simples

$$m(t+1) = \text{erf} \left[ \frac{m(t)}{\sqrt{2a}} \right]. \quad (2.34)$$



**Figura 2.3:** Diagrama de fase  $T$  versus  $a$  para o modelo DGZ, obtido da equação (2.33).

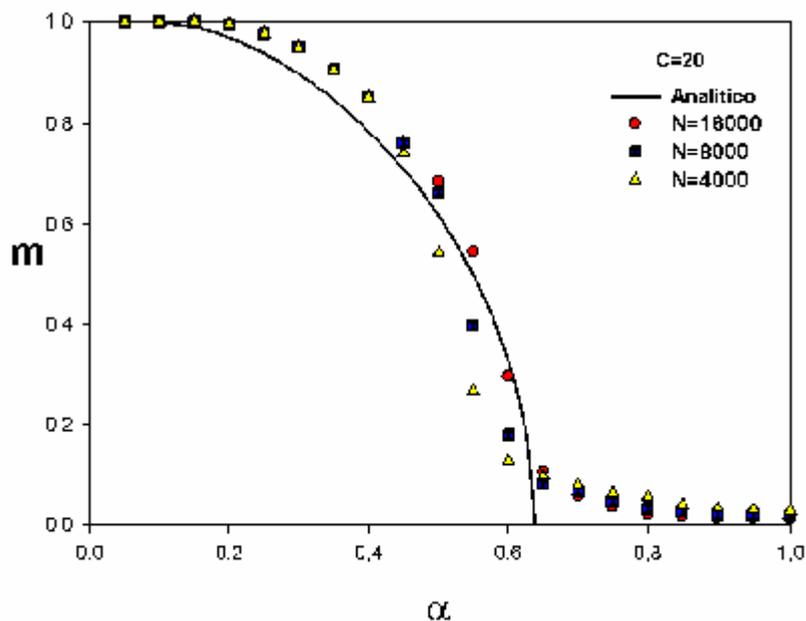
Expandindo a equação (2.34) para valores pequenos de  $m$  (limite  $m \rightarrow 0$ ), obtemos o valor crítico da máxima capacidade de armazenamento da rede  $a_c (T = 0)$ :

$$a_c = \frac{2}{p} = 0,6366\mathbf{K} , \quad (2.35)$$

valor este, muito acima do encontrado para o modelo totalmente conectado de Hopfield ( $a_c \approx 0,14$ ).

A condição de extrema diluição (2.29) pode, a princípio, levar à conclusão de que as simulações numéricas do modelo só obterão dados satisfatórios se o tamanho  $N$  da rede crescer exponencialmente com a conectividade  $C$ . Entretanto, Lemke e Arenzon [28] demonstraram que resultados confiáveis podem ser obtidos, mesmo sem o uso de redes exponencialmente grandes.

Na figura 2.4, mostramos a superposição  $m$  como função de  $a$  para o caso  $T = 0$ , reproduzindo os resultados obtidos por Lemke e Arenzon. A linha sólida corresponde ao resultado analítico, obtido através da equação (2.34), válido para  $N, p$  e  $C \rightarrow \infty$ . Os símbolos vazios são os resultados das simulações numéricas para  $N = 4000, 8000$  e  $16000$  com conectividade finita  $C = 20$ . Como podemos observar, os resultados numéricos apresentam boa concordância com o resultado analítico (conectividade infinita), indicando que o relaxamento da condição de extrema diluição ( $C \ll \ln N$ ), para a condição menos restritiva  $C \ll N$ , não invalida os resultados do modelo.



**Figura 2.4:** Superposição  $m$  versus  $a$ . A linha cheia foi obtida a partir da equação (2.34), enquanto que os símbolos referem-se a simulações numéricas para diferentes tamanhos de rede e conectividade finita  $C=20$  [27].

Em resumo, apresentamos neste capítulo uma revisão dos principais resultados obtidos referentes ao modelo de Hopfield totalmente conectado e ao modelo diluído (DGZ). Apesar do nosso interesse estar voltado para a capacidade de armazenamento e reconhecimento de padrões, existem muitas outras situações nas quais o modelo de Hopfield foi estudado.

Como, por exemplo, no reconhecimento de seqüências temporais [29-33] e classificação de padrões [34-37]. Também foram realizados estudos implementando novas arquiteturas, regras de aprendizado [13], e propondo a incorporação de outros elementos biológicos ao modelo, como os períodos refratários [38-40], estímulos externos [41-43], etc. Este resumo representa somente uma pequena parte dos inúmeros trabalhos desenvolvidos do modelo de Hopfield.

Neste capítulo, descrevemos em parte somente os modelos e técnicas que serão utilizadas no desenvolvimento desta tese. No capítulo seguinte, abordaremos o uso de campos externos como estímulo para a rede, este será o tema principal de nosso trabalho.

## Capítulo 3

# Efeito do Estímulo Externo no Reconhecimento de Padrões no Modelo DGZ

### 3.1 - Introdução

Nos últimos anos muitas características dos neurônios biológicos foram incorporadas aos modelos de redes neurais e, particularmente, ao modelo de Hopfield, com o intuito de torná-los mais realistas ou mesmo visando o surgimento de novas propriedades. Muitos trabalhos foram desenvolvidos tendo como proposta aumentar a capacidade de armazenamento da rede ou alargar as bacias de atração das memórias, tornando as redes mais robustas [44,45].

Neste sentido, um elemento biológico de interesse no estudo das redes neurais são os *estímulos externos* [41-43]. Biologicamente, sempre que um estímulo alcança nossa consciência, seja através de uma imagem, som, ou sensação, um conjunto de neurônios é ativado, estimulando ou inibindo uns aos outros, de modo a formar um determinado estado consciente (ou padrão da rede). Com o fim do estímulo, os neurônios que participaram da formação desse padrão podem ou não retomar a seu estado original. Se o estímulo responsável pela formação de um dado padrão for contínuo (persistente), o conjunto de neurônios ativados tende a fortalecer suas interações, aumentando a eficiência deste estado em resposta ao estímulo.

Diversas propostas têm sido apresentadas levando em conta estímulos externos atuando sobre redes neurais, utilizando técnicas analíticas e/ou numéricas. Uma das vertentes exploradas diz respeito ao uso de campos externos como estímulo para a rede. Amit, Gutfreund e Sompolinsky [15,17], ao resolver a termodinâmica do modelo de Hopfield, analisaram o caso de um *campo magnético estático* conjugado a um ou vários

padrões na ausência de ruído térmico ( $T = 0$ ). O principal resultado obtido, para o caso de apenas um padrão marcado, foi o aumento da capacidade de recuperação deste padrão, fazendo com que ele seja recuperado mesmo quando  $a > 0,14$ . Entretanto, um estímulo sobre vários padrões eleva o ruído interno do sistema, de forma que o crescimento da capacidade de recuperação é insignificante, e cai com o acréscimo de padrões marcados. Um estímulo sobre cinco ou seis padrões já é suficiente para limitar o reconhecimento, mesmo para os padrões marcados.

Engel, Englisch, e Schutte em 1989 [41], estudando as propriedades de reconhecimento de padrões do modelo de Hopfield, propuseram o uso de campos externos correlacionados com o padrão de entrada  $S_i(t=0)$ , ou seja, estímulos paralelos à configuração inicial, o que permite combinar a informação interna à rede (armazenada nas sinapses), com a contida no estímulo de entrada. Resultados analíticos foram comparados com simulações numéricas, demonstrando que campos externos deste tipo permitem a recuperação de informação mesmo quando o número de padrões armazenados excede o valor crítico do modelo de Hopfield. A dependência entre a superposição, o estímulo e a condição inicial  $m_0 = m(t=0)$  foi estudada, revelando uma transição de primeira ordem entre a fase desordenada e a fase de reconhecimento, para valores de  $m_0$  maiores que um certo valor crítico  $m_0^{cri}(a)$ .

O uso de estímulos neurais externos, paralelos à configuração inicial, foi posteriormente estendido para o modelo de Hopfield diluído. Engel et al. [43] obtiveram uma equação dinâmica para a superposição do sistema como função do parâmetro que simula o estímulo externo. Estudando as bacias de atração, constataram o aumento da estabilidade dos estados de memória e uma diminuição dos estados espúrios para uma estreita faixa de valores do estímulo, nesta região o estado de recuperação  $m^* = 1$  pôde ser alcançado mesmo para valores iniciais de superposição muito baixos, como  $m_0 = 0,4$ , isto é, quando o estado inicial do sistema encontrava-se muito afastado da região de atração na ausência de campo. Entretanto, para valores de  $a$  maiores que um certo valor crítico  $a_{max}(m_0)$ , a estabilidade dos estados espúrios cresce com o estímulo, demonstrando que, nesta região, um campo do tipo proposto é incapaz de aumentar a capacidade de recuperação da rede.

Amit, Parisi e Nicolis [42] analisaram o caso em que um campo externo acoplado a um dos padrões é utilizado para estimular a rede. O sistema foi estudado analiticamente

utilizando uma aproximação de campo médio, e as soluções foram comparadas com simulações numéricas. Este modelo difere do caso tratado por Amit, Gutfreund e Sompolinsky [15], pois o potencial externo inclui ruído, que é gerado por uma distribuição de erro. Os resultados obtidos levam à conclusão que o nível de carga de memória no qual a rede pára de atuar como um modelo de reconhecimento associativo, é determinado pela intensidade do estímulo de entrada  $e$ , para todas as formas de distribuição de ruído estudadas, independentemente do valor de  $a$ , há um valor  $h = h_c$ , acima do qual a rede corrige completamente a informação contida no estímulo.

Neste capítulo, que constitui o núcleo desta dissertação, vamos estudar o modelo DGZ quando adicionamos ao campo local um termo que privilegia um determinado padrão da rede. Desde já, vamos distinguir o padrão de inicialização (correlacionado com o estado inicial da rede) do padrão estimulado externamente. Vamos tratar duas situações distintas:

- i) Quando o estímulo externo atua, justamente, sobre o padrão de inicialização;
- ii) Quando o estímulo externo atua sobre um padrão distinto do padrão de inicialização.

No primeiro caso, a convergência do sistema para o padrão de inicialização é fortalecida, pois o campo contém informação idêntica à que está armazenada no padrão estimulado, determinando o curso da dinâmica da rede. Numa situação em que o estado inicial da rede esteja longe do centro da bacia atratora da memória, o sistema evoluirá para este atrator ou ficará significativamente próximo desta memória.

No segundo caso, onde o padrão de inicialização difere do padrão estimulado, o estado inicial da rede influenciará, mas não determinará totalmente, a dinâmica da rede. Nesta situação, a influência do estado inicial da rede será tanto menor quanto maior for a intensidade do campo. Para valores suficientemente intensos do estímulo, a bacia atratora do padrão estimulado poderá dominar completamente o espaço de fase do sistema, o sistema poderá, então, passar de uma bacia atratora (padrão de inicialização) para outra (padrão estimulado).

### 3.2 – Padrão estimulado igual ao padrão de inicialização

Continuaremos estudando o modelo DGZ introduzido no capítulo anterior. Vamos usar uma aproximação de campo médio para analisar as propriedades de armazenamento e reconhecimento quando um estímulo externo atua sobre o padrão de inicialização da rede.

A matriz de pesos sinápticos e a probabilidade de transição dos sítios individuais da rede continuam sendo dados por

$$J_{ij} = C_{ij} \sum_{m=1}^p x_i^m x_j^m, \quad (3.1)$$

e

$$\Pr(S_i) = \frac{1}{2} [1 + S_i \tanh(\beta_o h_i(t))] , \quad (3.2)$$

onde  $\beta_o = 1/T_o$  mede o nível de ruído estocástico da rede.

A introdução de um campo externo agindo sobre o estado da rede, pode ser facilmente introduzida no modelo DGZ com a inclusão de um termo aditivo na expressão do campo local, mantendo-se, em linhas gerais, todos os demais parâmetros do sistema inalterados. Dessa forma, redefiniremos o campo local:

$$h_i(t) = h_i^H(t) + h_{ext}(t), \quad (3.3)$$

onde  $h_i^H(t)$  é o termo usual de Hopfield, dado por

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t), \quad (3.4)$$

e o segundo termo,  $h_{ext}(t)$ , representa a contribuição externa ao campo local no tempo  $t$ , que, em analogia aos modelos de spins, corresponde a um campo externo. Vamos assumir  $h_{ext}(t)$  paralelo ao estado do sistema

$$h_{ext}(t) = h_o S_i(t). \quad (3.5)$$

Um campo externo desse tipo é equivalente a substituir os acoplamentos diagonais  $J_{ii}$  (auto interação), que são nulos no modelo de Hopfield, pela constante  $h_o$ .

Engel et al. [43], acharam que essa escolha ( $J_{ii} = h_o$ ), juntamente com a exigência de extrema diluição ( $C \ll \ln N$ ), fornecia condição necessária e suficiente para garantir a ausência de correlação temporal entre os diferentes neurônios da rede, de modo a assegurar

a forma da equação de recorrência obtida no modelo DGZ,  $m(t+1) = f(m(t))$  para todo tempo  $t$ . Este trabalho foi posteriormente questionado quanto à ausência das correlações temporais, nos casos em que o campo externo tem a forma  $h_o S_i(t)$  ou  $h(t)S_i(t)$ . Nestes casos, a validade das equações fica limitada a tempos extremamente curtos [46].

Um desenvolvimento analítico exato que contorne estes problemas é uma tarefa quase impossível de ser realizada, devido à complexidade das interações e correlações temporais. Contudo, existe um método aproximado para resolver o problema, que, em muitos casos, fornece um resultado satisfatório. No nosso modelo, vamos utilizar uma aproximação de campo médio, de modo a tornar o problema tratável. A aproximação consiste em representar as flutuações do campo externo, na equação (3.3), por seu valor médio  $\langle h_{ext}(t) \rangle = h_o \langle S_i(t) \rangle$  e substituir  $\langle S_i(t) \rangle$  por  $m_v(t)\xi_i^v$  numa associação com os modelos magnéticos [9]. Encontramos, assim, a seguinte expressão para o campo local:

$$h_i(t) = h_i^H(t) + h_o m_v(t) \xi_i^v . \quad (3.6)$$

Aqui,  $h_o$  é o parâmetro que regula a intensidade do estímulo externo,  $\xi_i^v$  é o padrão que será estimulado e  $m_v(t)$  a superposição do estado do sistema com este padrão. Diferentemente dos outros modelos, a inclusão do termo  $m_n$  no campo local confere um comportamento não linear ao sistema, pois na região de não reconhecimento  $m_n \rightarrow 0$  e, portanto, o estímulo pára de atuar. Substituindo  $h_i^H(t)$  dado por (3.4) na equação acima, o campo local assume a forma:

$$h_i(t) = \sum_j J_{ij} S_j(t) + h_o m_n(t) x_i^n . \quad (3.7)$$

Utilizando a definição da matriz sináptica  $J_{ij}$  dada por (3.1) e separando o termo  $v$  dos demais, levando em conta apenas os  $K$  elementos existentes após a diluição ( $C_{ij} = 1$ ), obtemos a seguinte expressão para o campo local:

$$h_i(t) = h_o m_n(t) x_i^n + x_i^n \sum_{r=1}^K x_{j_r}^n S_{j_r}(t) + \sum_{r=1}^K \sum_{m \neq n}^p x_i^m x_{j_r}^m S_{j_r}(t) . \quad (3.8)$$

Adotaremos os mesmos pressupostos admitidos na seção em que tratamos do modelo DGZ, entretanto, dessa vez, observaremos a superposição do sistema com a memória estimulada. Vale lembrar que o padrão estimulado é o mesmo padrão de inicialização da rede. Assim, no tempo  $t=0$ , o estado inicial do sistema terá superposição finita com a memória  $v$  e superposição desprezível com as demais. Realizando a média térmica por meio de (3.2), podemos escrever a equação de recorrência para a superposição com aquela memória:

$$m_n(t+1) = \frac{1}{N} \sum_{i=1}^N x_i^n \tanh\left(\frac{h_i(t)}{T_0}\right), \quad (3.9)$$

onde  $T_0$  representa a fonte de ruído estocástico do sistema (diferentemente do ruído estático associado ao número de padrões da rede). Considerando o limite termodinâmico e usando a propriedade de automédia (ver apêndice D), podemos substituir a média sobre os sítios por uma média sobre as configurações, de forma que:

$$m_n(t+1) = \left\langle \left\langle \tanh\left(\frac{x_i^n h_i(t)}{T_0}\right) \right\rangle \right\rangle, \quad (3.10)$$

onde  $\langle\langle \mathbf{K} \rangle\rangle$  designa a média configuracional sobre as variáveis  $\mathbf{x}$ .

Considerando  $C \rightarrow \infty$  e  $p \rightarrow \infty$ , a média presente em (3.10) pode ser facilmente calculada, bastando usar (3.8) para escrever o termo  $x_i^n h_i(t)$  como uma soma de infinitas variáveis aleatórias, que apresentam, segundo o Teorema do Limite Central, uma distribuição de probabilidades Gaussiana.

No apêndice B, mostramos em detalhe como obter a forma integral para a superposição  $m(t+1)$  com o padrão  $v$ , que é dada por

$$m(t+1) = \int_{-\infty}^{+\infty} Dz \tanh\left(\frac{m(t)(h+1) + z\sqrt{a}}{T}\right), \quad (3.11)$$

onde definimos  $Dz \equiv dz e^{-z^2/2} / \sqrt{2p}$  e  $a = p/C$ . Aqui,  $h = h_0/C$  e  $T = T_0/C$  são definidos como os parâmetros reduzidos (estímulo e ruído térmico, respectivamente). Observe que fazendo  $h=0$  na equação (3.11) recuperamos a expressão obtida por Derrida et

al. [25]. No limite de tempos grandes, o sistema entra em um regime assintótico, de forma que, por simplicidade, adotaremos  $m(t) = m$  para todo  $t$  (atrator ponto fixo).

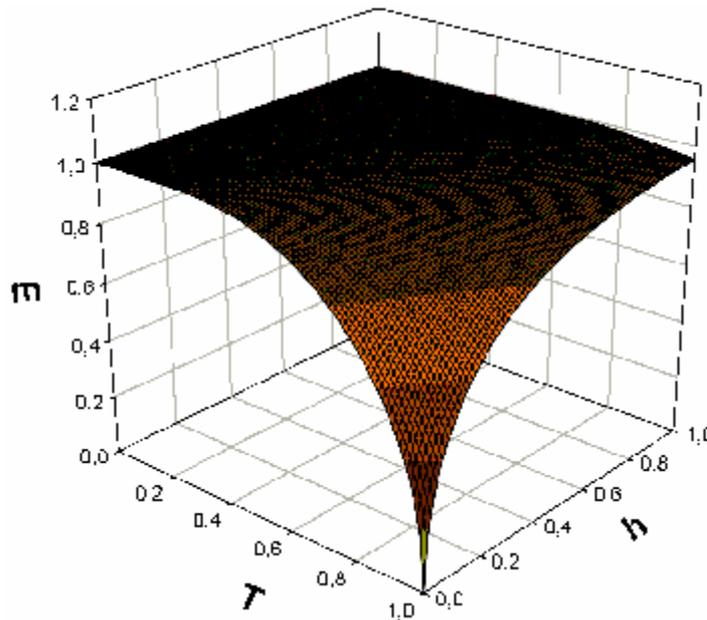
### 3.2.1 - Número Finito de Padrões

Vamos iniciar nossa análise considerando o caso mais simples  $a = 0$  (número finito de padrões). Nesse limite, a equação (3.11) tem a seguinte forma:

$$m = \tanh\left(\frac{m(h+1)}{T}\right) \quad (3.12)$$

Na figura 3.1, mostramos o comportamento da superposição  $m$  em função dos parâmetros  $T$  e  $h$  para  $a = 0$  e configuração inicial próxima do padrão de reconhecimento ( $m_o \equiv m(t=0) \approx 1$ ). Como o padrão de inicialização e o padrão estimulado coincidem, ou seja, armazenam a mesma informação, qualquer valor inicial de  $m_o$ , desde que  $m_o \neq 0$ , não modifica o comportamento assintótico da rede. Essa invariância evidencia o estímulo externo como o parâmetro que determina a evolução dinâmica do sistema.

Para  $h = 0$  recuperamos o modelo DGZ puro ( $T^* = 1$ ). Quando temos poucos padrões, o efeito do estímulo é muito expressivo, na região  $h = 1$  a superposição é grande ( $m \approx 1$ ) para todo o intervalo  $0 \leq T \leq 1$ .



**Figura 3.1:** Superposição  $m$  como função de  $T$  e  $h$  para  $a = 0$  com  $m_o \approx 1$ .

A princípio, no nosso modelo, não há restrições quanto ao valor máximo do parâmetro  $h$ . Para um valor muito intenso do estímulo, o curso da dinâmica do sistema será completamente dominado pelo padrão estimulado. Em quase todas as nossas análises vamos nos restringir à região  $0 \leq h \leq 1$ .

Visto que a transição é contínua, podemos expandir a tangente hiperbólica na equação (3.12), para valores pequenos de  $m$ , obtendo a seguinte expressão para a temperatura crítica:

$$T^* = h + 1, \quad (3.13)$$

que mostra que o estímulo aumenta a estabilidade da memória marcada para reconhecimento. Removendo a ação do campo, fazendo  $h = 0$ , recaímos no caso puro do modelo DGZ:  $T^* = 1$ .

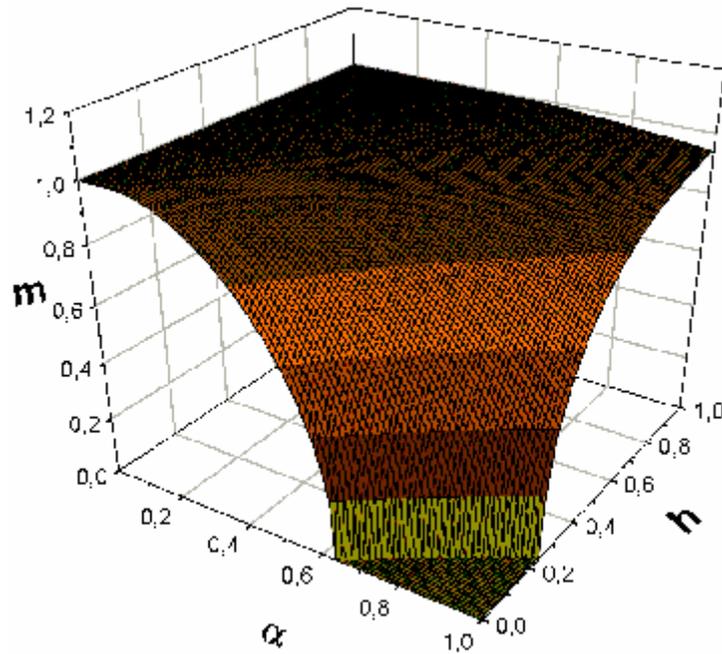
### 3.2.2 - Número Infinito de Padrões

Nesta seção, vamos considerar o caso  $a \neq 0$ . Vamos analisar inicialmente o sistema na ausência de ruído térmico  $T = 0$ . Neste limite, ( $b \rightarrow \infty$ ), a equação (3.11) assume a seguinte forma (ver seção um do apêndice B):

$$m = \operatorname{erf} \left( \frac{m(h+1)}{\sqrt{2a}} \right). \quad (3.14)$$

Na figura 3.2, apresentamos a superposição  $m$  como função de  $\alpha$  e  $h$ , com  $T = 0$ , para uma configuração inicial muito próxima do padrão de reconhecimento (que também é o padrão estimulado). Para  $h = 0$ , recuperamos a linha crítica do modelo DGZ obtida em [25]. Para valores grandes de  $h$  e poucos padrões, o sistema reconhece com superposição  $m = 1$  e decresce muito lentamente mesmo para valores grandes de  $a$  ( $a \approx 1$ ), em  $a = 1$  e  $h \approx 1$  temos  $m \approx 0,94$ .

Como já prevíamos, o parâmetro  $h$  melhora a qualidade do reconhecimento e aumenta consideravelmente a capacidade de armazenamento da rede. Como o estímulo é paralelo ao padrão de inicialização, mesmo quando o sistema parte de uma situação de baixa superposição inicial o estímulo ainda é capaz de corrigi-lo, levando-o para a bacia atratora do padrão estimulado.

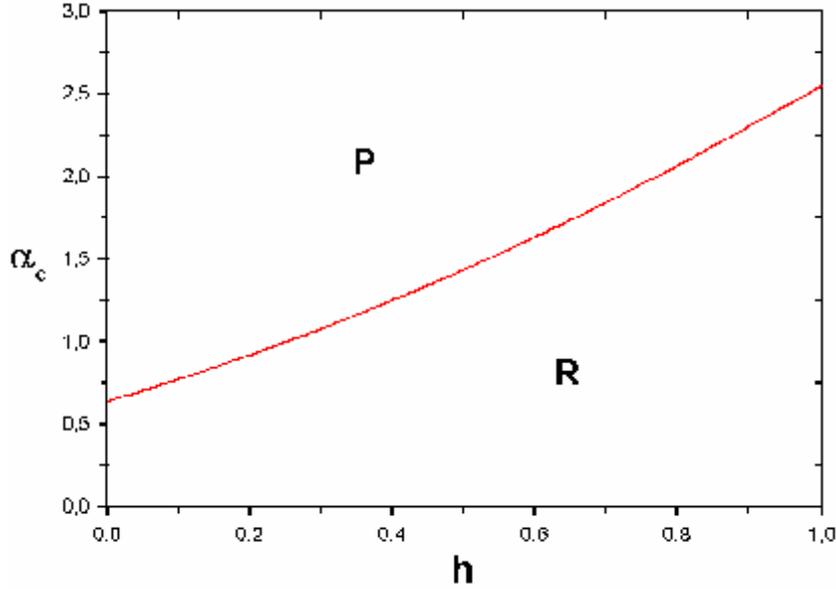


**Figura 3.2:** Superposição  $m$  como função de  $a$  e  $h$  para  $T = 0$  com  $m_0 \approx 1$ .

A linha crítica no plano  $m = 0$  pode ser facilmente obtida pela expansão, para valores pequenos de  $m$ , da função  $erf(\cdot)$  na equação (3.14). A expressão, em função dos parâmetros  $\alpha_c$  e  $h$ , é dado por:

$$\alpha_c = (h+1)^2 \frac{2}{\pi} . \quad (3.15)$$

Na figura 3.3 mostramos o diagrama de fase  $\alpha_c$  versus  $h$  para  $T=0$ . A linha de transição separa a fase de reconhecimento (R) da fase paramagnética (P). Como podemos notar, a capacidade de armazenamento na presença do estímulo cresce de forma quadrática com o parâmetro  $h$ , o que indica que a influência dos demais padrões, que tende a reduzir a capacidade de armazenamento do sistema, é atenuada pela inclusão do estímulo. Para  $h = 0$ , temos a capacidade máxima do modelo DGZ,  $a_c = 0,6366\dots$ , e para  $h = 1$  temos  $a = 2,546\dots$ .



**Figura 3.3:** Diagrama de fase  $a_c \times h$  para  $T = 0$ . A linha crítica separa a fase de reconhecimento **R** ( $m \neq 0$ ) da fase paramagnética **P** ( $m = 0$ ).

Analizamos também o caso  $T \neq 0$ . Partindo da equação (3.11), podemos avaliar a estabilidade da solução  $m = 0$ , a partir do parâmetro  $I = \left. \frac{\partial f(m)}{\partial m} \right|_{m=0}$ . Para valores de  $I < 1$  a solução é estável, enquanto que para valores de  $I > 1$  a solução é instável. Dessa forma, a condição  $\lambda = 1$  define a superfície crítica no espaço dos parâmetros  $T$ ,  $\alpha$  e  $h$ , dada por:

$$T^* = (h+1) - (h+1) \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh^2 \left( \frac{z\sqrt{a}}{T^*} \right). \quad (3.16)$$

A figura 3.4 mostra o diagrama de fase no espaço  $(T, a, h)$ . Observa-se uma transição de segunda ordem para a fase paramagnética ( $m = 0$ ), ao longo da superfície crítica. Abaixo da superfície, o sistema reconhece os padrões armazenados ( $m \neq 0$ ).

Na ausência de estímulo ( $h = 0$ ), obtemos a bem conhecida linha crítica do modelo DGZ, onde  $T^* = 1$  para  $a = 0$ . Entretanto, à medida que mais padrões são armazenados na rede (valores crescentes de  $a$ ), o valor crítico da temperatura diminui até  $T = 0$  para  $a = a_c = \frac{2}{p} = 0,6366\dots$

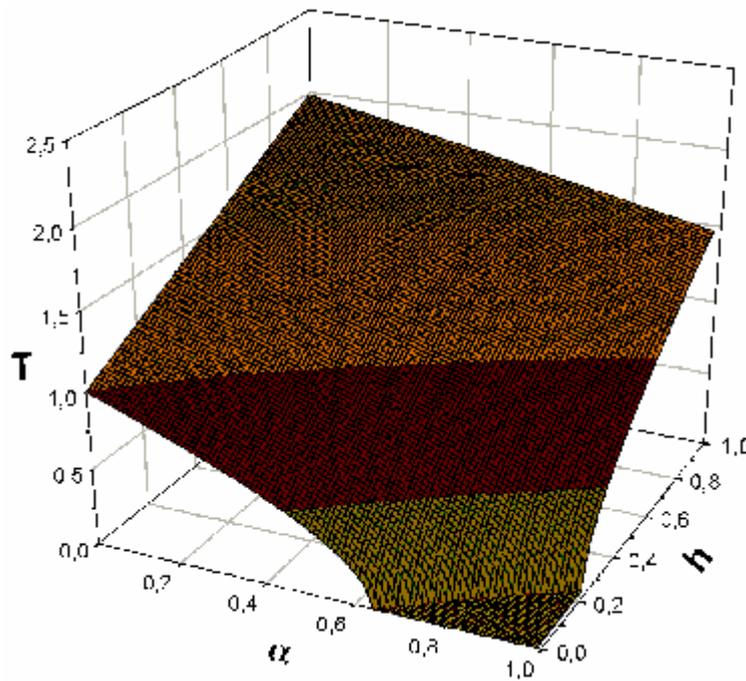
O aumento da intensidade do estímulo exerce influência marcante sobre duas características do sistema: a capacidade de armazenamento e a susceptibilidade ao ruído térmico. Para valores crescentes do estímulo, o sistema apresenta maior robustez com

relação ao ruído térmico, e um número maior de padrões pode ser armazenado sem que a memória estimulada seja desestabilizada.

Existe uma situação bem definida que apresenta comportamento linear na figura 3.4: O crescimento de  $T$  com  $h$  para  $a=0$ , dado pela equação (3.13). Um comportamento semelhante pode ser notado para valores grandes de  $h$  ( $h \approx 1$ ), na região em que  $T > a$ . Para analisar esta situação, vamos expandir o argumento da  $\tanh$  na equação (3.16), obtendo:

$$T \approx (h+1) - (h+1) \left( \frac{a}{T^2} - \frac{2a^2}{T^4} + \frac{5a^4}{3T^6} \dots \right). \quad (3.17)$$

Numa aproximação de primeira ordem, chegamos à equação linear:  $T \approx (h+1) - \frac{a}{(h+1)}$ , como vemos, para  $h=1$  temos que  $T$  decresce de  $T=2$  em  $a=0$ , até  $T \approx 1,5$  em  $a=1$ . Note que essa aproximação só é válida na região em que  $a/T^2 \ll 1$ .



**Figura 3.4:** Diagrama de fase  $T \times a \times h$  para o modelo DGZ com estímulo externo.

Para estímulos muito intensos, como vimos no caso acima, o comportamento assintótico do sistema é completamente determinado pela ação do campo, de forma que o padrão de reconhecimento apresentará a informação contida no estímulo. Um caso mais

interessante ocorre quando o estímulo atua sobre um padrão diferente do padrão de inicialização, esta situação é abordada na seção seguinte.

### 3.3 – Padrão estimulado distinto do padrão de inicialização

Na seção anterior, analisamos o efeito do estímulo externo atuando sobre o padrão de inicialização. Como vimos, a influência do campo, além de intensificar a tendência natural do sistema em convergir para o padrão de reconhecimento, também aumenta a bacia atratora do padrão, tornando a rede mais robusta.

Analisaremos agora as propriedades de reconhecimento e armazenamento de memórias do modelo DGZ para o caso em que o padrão de entrada apresentado à rede corresponde a uma memória diferente daquela que será estimulada. Dessa forma, teremos uma situação de competição, em que, por um lado, a dinâmica da rede tenta levar o sistema para a bacia de atração da memória mais próxima, e por outro, o estímulo tende a guiar o sistema para o padrão estimulado.

Convencionaremos chamar o padrão correlacionado inicialmente com o estado da rede de  $v \{x^n\}$  (padrão de inicialização) e o padrão estimulado externamente de  $\delta \{x^d\}$ .

Para determinar as superposições  $m_n(t)$  e  $m_d(t)$ , vamos substituir a matriz sináptica  $J_{ij}$ , dada por (3.4), na equação (3.7) e separar no somatório do campo local tanto o termo estimulado  $m = d$ , quanto o termo que queremos reconhecer  $m = n$ , deste modo, podemos escrever  $h_i(t)$  na forma seguinte:

$$h_i(t) = x_i^d \left( h m_d(t) + \sum_{r=1}^K x_{j_r}^d S_{j_r}(t) \right) + \sum_{r=1}^K x_i^n x_{j_r}^n S_{j_r}(t) + \sum_{r=1}^K \sum_{\substack{m \neq n \\ m \neq d}}^p x_i^m x_{j_r}^m S_{j_r}(t), \quad (3.18)$$

onde consideramos apenas os  $K$  elementos restantes após a diluição.

Seguindo os mesmos passos da seção anterior, tomando o limite termodinâmico  $N \rightarrow \infty$  e assumindo conectividade infinita  $C, p \rightarrow \infty$ , as equações básicas para analisar o reconhecimento e armazenamento do padrão de inicialização e do padrão estimulado podem ser calculadas. Nas seções B.2.1 e B.2.2 do apêndice B, mostramos em detalhe como obter o conjunto de equações acopladas para as superposições  $m_n(t+1)$  e  $m_d(t+1)$ , dadas por:

$$m_n(t+1) = \int_{-\infty}^{+\infty} Dz \tanh \left( \frac{m_n(t) + z\sqrt{m_d^2(t)(1+h)^2 + a}}{T} \right) \quad (3.19)$$

e

$$m_d(t+1) = \int_{-\infty}^{+\infty} Dz \tanh \left( \frac{m_d(t)(h+1) + z\sqrt{m_n^2(t) + a}}{T} \right). \quad (3.20)$$

Observe que para  $h = 0$  as equações tornam-se idênticas, ou seja, na ausência de campo o sistema não privilegia nenhum dos padrões. Todo o desenvolvimento a seguir terá por base as equações acima.

### 3.3.1 - Número Finito de Memórias

Vamos iniciar considerando o caso  $a = 0$ . Fazendo  $m(t) \rightarrow m$  para o regime assintótico, as equações (3.19) e (3.20) têm as formas simplificadas dadas por:

$$m_n = \int_{-\infty}^{+\infty} Dz \tanh \left( \frac{m_n + zm_d(1+h)}{T} \right) \quad (3.21)$$

e

$$m_d = \int_{-\infty}^{+\infty} Dz \tanh \left( \frac{m_d(h+1) + zm_n}{T} \right). \quad (3.22)$$

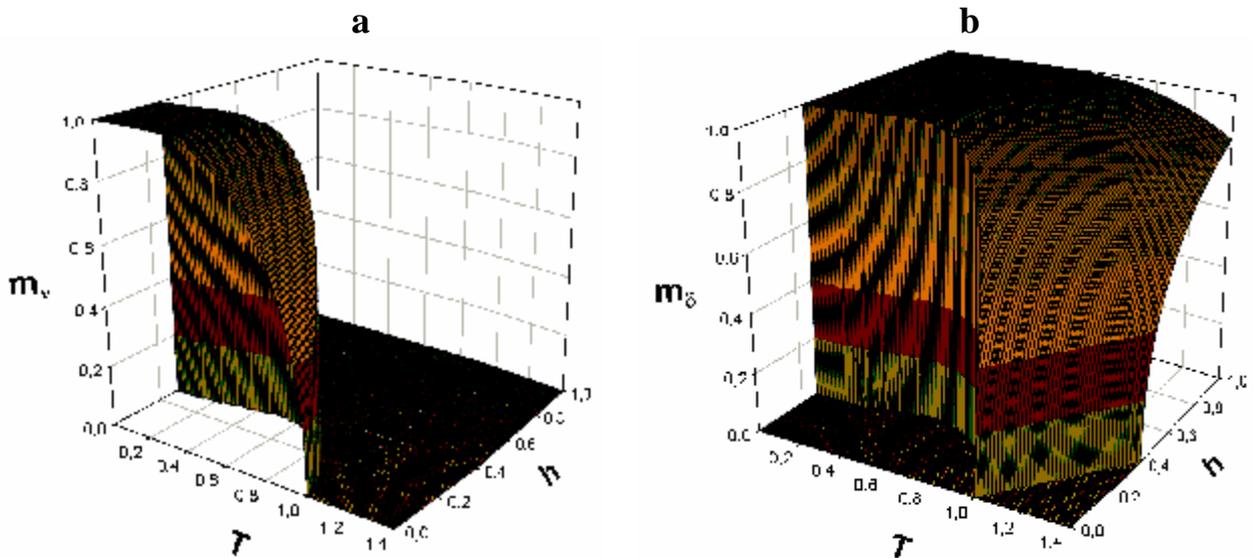
Na figura 3.5 apresentamos as superposições  $m_n$  e  $m_d$  em função dos parâmetros  $T$  e  $h$  para  $a = 0$  e condições iniciais  $m_n^o = 1$  e  $m_d^o = 0,01$ . Para  $h = 0$ , na figura 3.5a, reproduzimos a linha crítica do modelo DGZ onde  $m_n = 0$  em  $T^* = 1$ , com  $m_d = 0$  para todo  $T$ .

Para  $h \neq 0$  existe uma transição de primeira ordem entre a fase de reconhecimento do padrão  $n$  e a fase de reconhecimento do padrão  $d$ , que para  $T = 0$  apresenta um ponto crítico dado por  $h_c \approx 0,279$ . Para  $T < 1$  existe uma pequena região de domínio do padrão de inicialização  $m_n$  que se estende desde  $T = 0$  para  $h_c \approx 0,279$  até  $T = 1$  para  $h = 0$ . Ao

longo da fronteira desta região o sistema passa abruptamente da bacia atratora do padrão  $\nu$  para a bacia do padrão  $\delta$ .

Para valores pequenos de  $h$  ( $h < 0,8$ ), o padrão estimulado é mais instável à medida que  $T$  cresce, de modo que ocorre uma pequena, mas pronunciada, reentrância da região de domínio do padrão  $\nu$  sobre a região do padrão  $\delta$ . Com o crescimento do parâmetro  $T$ , o padrão  $n$  é completamente desestabilizado, independentemente do valor inicial  $m_n^0$ . Para  $T > 1$  o sistema só reconhece o padrão estimulado.

Na figura 3.5b, para  $T < 1$  e aumentando o valor de  $h$ , o padrão estimulado surge descontinuamente ( $m_d \approx 1$ ). Para  $T > 1$  o padrão  $n$  não é mais reconhecido, e a superposição com o padrão  $d$  cai continuamente com  $T$ . Nesta região, para valores pequenos de  $h$ , o sistema não reconhece nenhum dos padrões (fase paramagnética). A linha crítica no plano  $T \times h$ , equação (3.13)  $T^* = h + 1$ , é facilmente obtida pela expansão da equação (3.22) para  $m_q \rightarrow 0$  com  $m_\nu = 0$ .



**Figura 3.5:**  $m \times T \times h$  para  $a = 0$ ,  $m_n^0 = 0,99$  e  $m_d^0 = 0,01$  (Equações 3.21 e 3.22).(a)

Superposição com o padrão de inicialização  $n$ ; (b) Superposição com o padrão estimulado  $d$ .

Verificamos também, através de análises numéricas, que o comportamento do sistema independe da condição inicial  $m_n^0$  (superposição inicial com o padrão de inicialização), desde que  $m_n^0 \neq 0$ , e que para valores diferentes de  $m_d^0$  (superposição inicial

com o padrão estimulado) o sistema apresenta comportamentos quantitativamente diferentes.

### 3.3.2 - Número Infinito de Memórias

Nesta seção vamos considerar o caso  $a \neq 0$ . Vamos analisar inicialmente o sistema na ausência de ruído térmico  $T \rightarrow 0$ . Neste limite ( $b \rightarrow \infty$ ), as equações (3.19) e (3.20) assumem a seguinte forma (ver apêndice B, seções B.2.1 e B.2.2):

$$m_n = \operatorname{erf} \left( \frac{m_n}{\sqrt{2} \sqrt{m_d^2 (1+h)^2 + a}} \right) \quad (3.23)$$

e

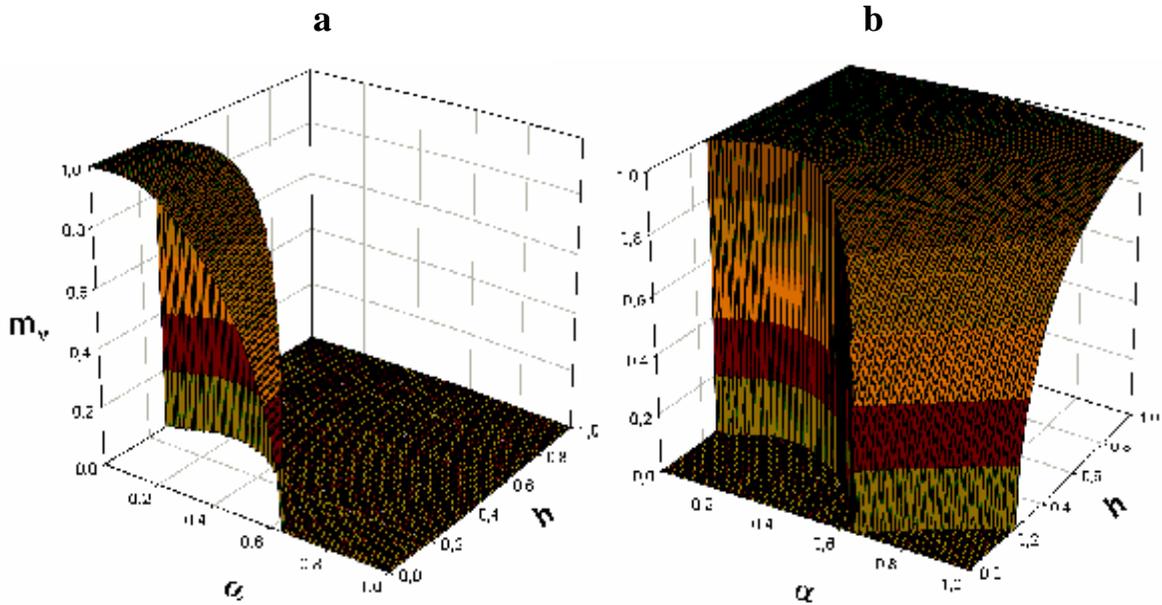
$$m_d = \operatorname{erf} \left( \frac{m_d (h+1)}{\sqrt{2} \sqrt{m_n^2 + a}} \right). \quad (3.24)$$

As superposições  $m_n$  e  $m_d$ , como função de  $\alpha$  e  $h$ , são mostradas na figura 3.6, para  $T = 0$ ,  $m_n^o = 0,99$  e  $m_d^o = 0,01$ .

Note que na figura 3.6b, para  $h = 0$ , temos  $m_\delta = 0$  para qualquer valor do parâmetro  $\alpha$ , enquanto que na figura 3.6a,  $m_n$  tende a zero continuamente e se anula em  $a_c = 2/p \approx 0,6366\dots$ . Como podemos observar, o parâmetro  $\alpha$  desempenha um papel similar ao parâmetro  $T$ . As transições de fase descontínuas e o comportamento reentrante na região de domínio do padrão  $d$  são qualitativamente semelhantes ao caso anterior (figura 3.5).

Na figura 3.6b, para  $\alpha > \alpha_c$ , temos  $m_v = 0$  e a transição entre a fase de reconhecimento do padrão estimulado ( $m_d \neq 0$ ) e a fase desordenada ( $m_\delta = 0$ ) é de segunda ordem. A linha crítica no plano  $\alpha \times h$  pode ser facilmente calculada pela expansão da equação (3.24) para  $m_d \rightarrow 0$  com  $m_v = 0$ , resultando na equação (3.15) dada

$$\text{por } a_c = \left( \frac{2}{p} (h+1)^2 \right).$$

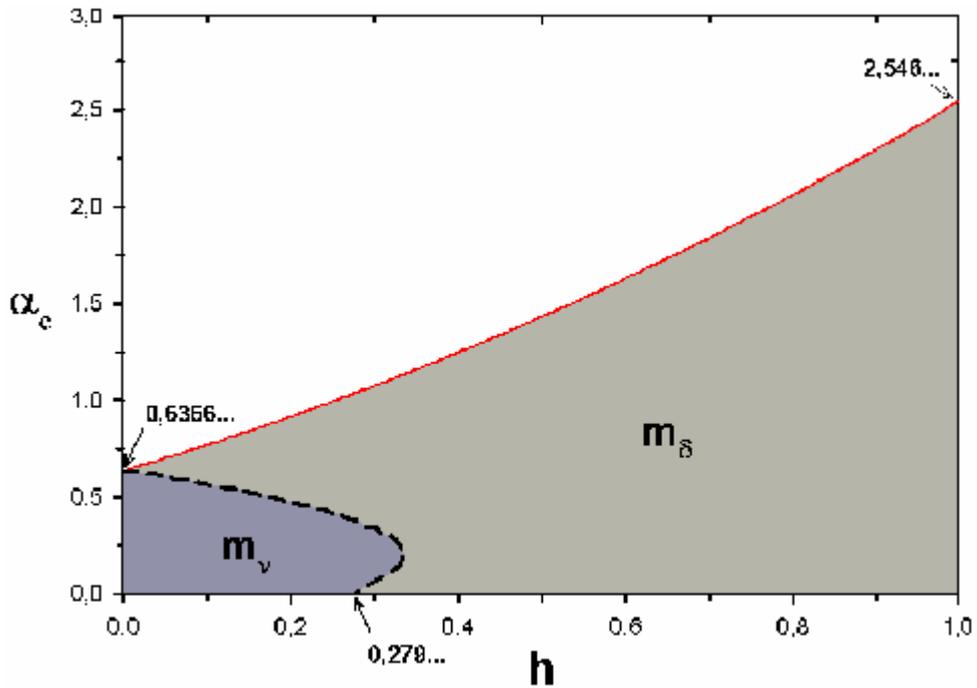


**Figura 3.6:**  $m \times a \times h$  para  $T = 0$ ,  $m_n^o = 0,99$  e  $m_d^o = 0,01$ . (a) Superposição com o padrão de inicialização  $n$ ; (b) Superposição com o padrão estimulado  $d$ .

Na figura 3.7, apresentamos o diagrama de fase no plano  $a \times h$  para  $T = 0$ . Na fase paramagnética (região clara) o sistema não reconhece nenhum dos padrões armazenados. Na região de reconhecimento (escura), o domínio do padrão  $n$  ocorre apenas numa pequena faixa do diagrama de fase e apresenta uma reentrância sobre a grande região de domínio do padrão estimulado.

Este comportamento vem do fato de que, para valores de  $a$  suficientemente pequenos, o termo de sinal do padrão de inicialização é forte, de forma que na região de  $h$  pequeno o sinal domina e o padrão estimulado é desestabilizado primeiro. Com o crescimento de  $a$ , ocorre uma inversão já que o padrão de inicialização é mais sensível ao efeito do ruído estático. A figura 3.7 ainda exhibe os valores extremos da capacidade de armazenamento: Para  $h = 0$  temos  $a_c = 0,6366\dots$  (máxima capacidade do modelo DGZ) e para  $h = 1$  temos  $a = 2,546\dots$

O caso mais geral  $T$  e  $a \neq 0$ , pode ser analisado a partir das equações (3.19) e (3.20) se fixarmos convenientemente o parâmetro  $h$ . A figura 3.8 exhibe as superposições  $m_n$  e  $m_d$  como função de  $\alpha$  e  $T$  para (a)  $h = 0$ , (b)  $h = 0,2$  e (c)  $h = 0,4$ , sujeito às condições iniciais  $m_n^o = 0,99$  e  $m_d^o = 0,01$ .



**Figura 3.7:** Diagrama de fase  $\alpha \times h$  para  $T = 0$ . A linha crítica contínua separa a fase paramagnética (região clara) da fase de reconhecimento dos padrões  $n$  e  $d$  (região escura).

Na região de reconhecimento a linha tracejada indica a transição descontínua.

Na ausência de estímulo externo  $h = 0$  (figura 3.8a), temos  $m_d = 0$  para todo o espaço, enquanto que o padrão  $n$  é desestabilizado (transição de segunda ordem) com o crescimento de  $T$  e  $a$ , apresentando valores assintóticos  $a = 0$  para  $T^* = 1$  e  $T = 0$  para  $a_c = 0,6366\dots$

Na figura 3.8b ( $h = 0,2$ ), aumentando  $a$  e  $T$ , o sistema apresenta uma transição de fase de primeira ordem, passando da fase de reconhecimento do padrão de inicialização ( $m_n \neq 0$ ) para a fase paramagnética ( $m_n = 0$ ). Na linha crítica  $a_c(T)$ , onde o padrão  $n$  deixa de ser reconhecimento, surge abruptamente o padrão  $d$  com superposição constante da ordem de  $m_d \approx 0,87$ . Seguindo, dentro da região de domínio do padrão  $d$  ( $m_n = 0$ ), o sistema passa por uma nova transição de fase na qual o parâmetro de ordem  $m_d$  tende a zero continuamente, caracterizando uma segunda linha crítica  $a_c(T)$  com  $h$  fixo.

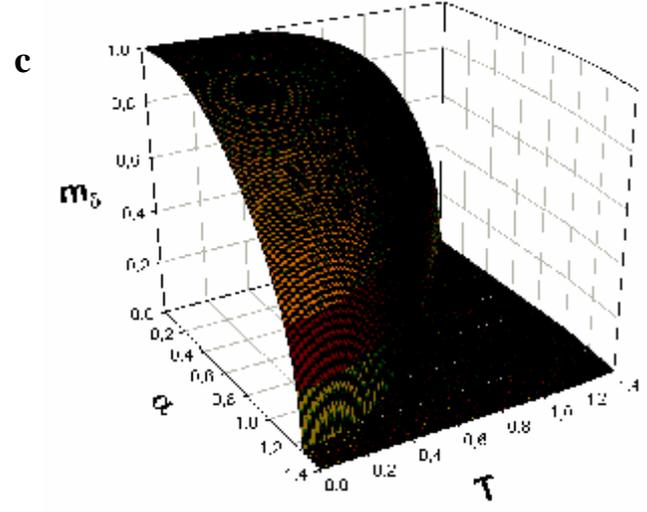
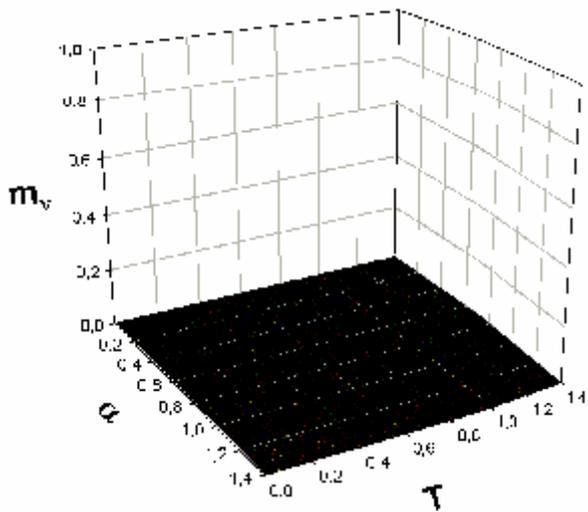
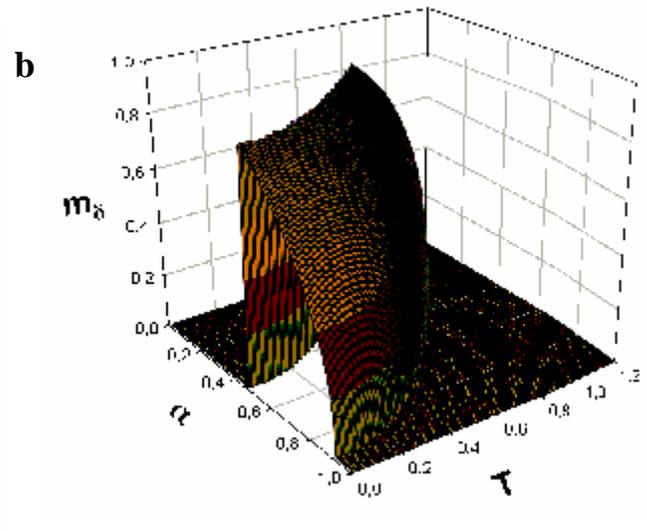
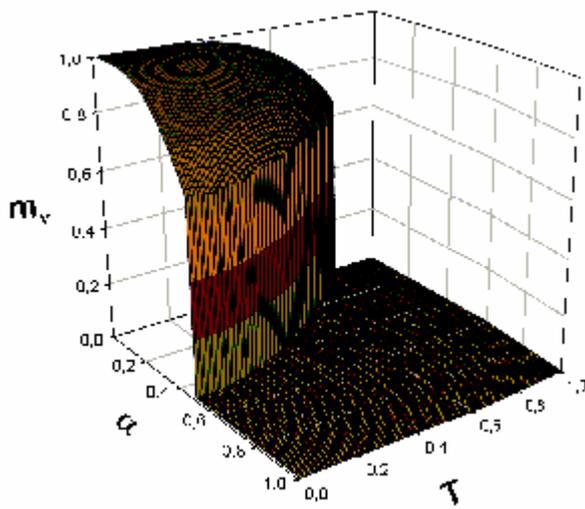
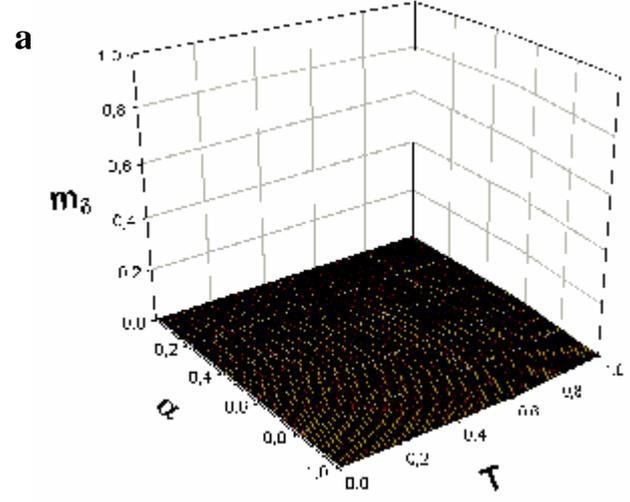
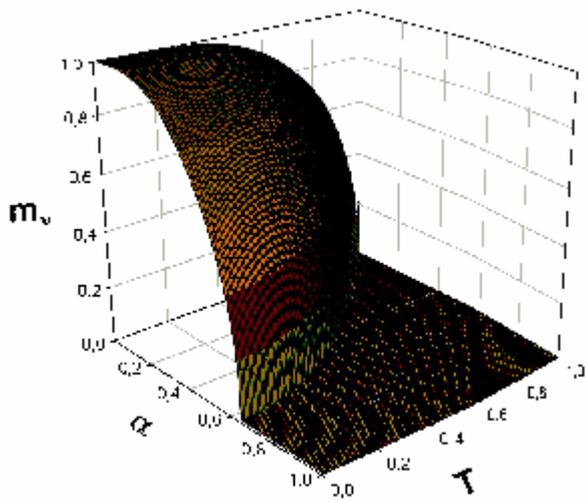


Figura 3.8:  $m \times a \times T$  para  $m_n^o = 0,99$  e  $m_d^o = 0,01$  e diferentes valores de  $h$ : (a)  $h = 0,0$ , (b)  $h = 0,2$  e (c)  $h = 0,4$ .

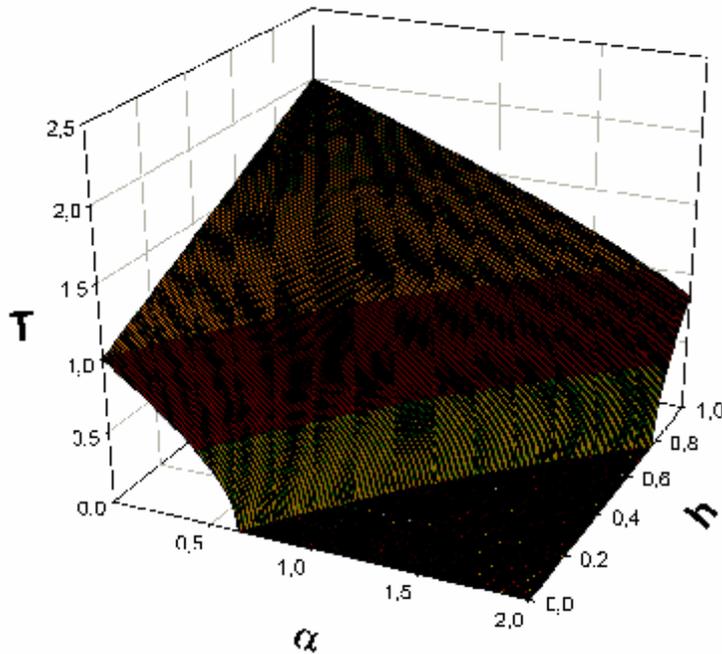
A superfície crítica no espaço dos parâmetros  $T$ ,  $\alpha$  e  $h$  pode ser obtida da equação (3.20), usando a condição  $\left. \frac{\partial f(m_d, m_n = 0)}{\partial m_d} \right|_{m_d=0} = 1$  que determina a estabilidade da solução  $m_d = 0$ , dessa forma:

$$T^* = (h+1) - \frac{(h+1)}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh^2 \left( \frac{z\sqrt{a}}{T^*} \right). \quad (3.25)$$

Os pontos extremos da figura 3.8b podem ser obtidos facilmente. Considerando  $a = 0$ , a equação (3.25) se reduz a  $T^* = h+1$ , que para  $h = 0,2$  resulta em  $T^* = 1,2$ . Em  $T = 0$  temos a linha crítica  $\alpha_c = \frac{2}{\pi}(h+1)^2$  dada por (3.15), tomando  $h = 0,2$  obtemos  $\alpha_c = 0,916\dots$

Para valores grandes de  $h$  (figura 3.8c), o padrão estimulado domina inteiramente a dinâmica do sistema, apresentando uma transição de fase contínua, com valores assintóticos  $\alpha_c = 1,247\dots$  para  $T = 0$  e  $T^* = 1,4$  para  $a = 0$ .

O diagrama de fase no espaço  $T$ ,  $a$  e  $h$ , que corresponde à figura 3.4 em escala modificada, mostra a superfície crítica separando as regiões de reconhecimento dos padrões  $n$  e  $d$  da fase paramagnética.



**Figura 3.9:** Diagrama de fase  $T \times a \times h$  para o modelo DGZ com estímulo externo. Note que para  $h = 1$  o comportamento linear na transição aparece somente para valores de  $a/T < 1$ .

### 3.4 – Simulação numérica

Em geral, as simulações numéricas contribuem para um maior conhecimento do comportamento dinâmico do modelo, além de ser um recurso fundamental para testar os resultados analíticos, já que podem ser considerados como verdadeiros experimentos. Nesta seção apresentaremos os resultados das nossas simulações numéricas para  $T = 0$ , a fim de testar as aproximações analíticas obtidas nas seções anteriores.

Devido à impossibilidade de implementar simulações computacionais em que a condição de extrema diluição  $C \ll \ln N$  seja satisfeita com  $C \gg 1$ , utilizaremos a condição menos restritiva  $C \ll N$  em nossas simulações, alguns trabalhos realizados para conectividade  $C$  pequena e grandes redes, têm mostrado boa concordância com os resultados analíticos [28,37,39].

Para implementar a evolução dinâmica do modelo, o campo local pode ser convenientemente escrito em termos da superposição entre o estado do sistema e os padrões de reconhecimento, de forma a obter a equação dinâmica:

$$S_i(t+1) = \text{Sinal} \left[ x_i^n m_n(t)(h+1) + \frac{1}{C} \sum_{k=1}^C \sum_{m \neq n}^p x_i^m x_k^m S_k(t) \right] \quad (3.26)$$

para o caso em que o estímulo externo atua sobre o padrão de inicialização, e

$$S_i(t+1) = \text{Sinal} \left[ x_i^n m_n(t) + x_i^d m_d(t)(h+1) + \frac{1}{C} \sum_{k=1}^C \sum_{m \neq n}^p x_i^m x_k^m S_k(t) \right], \quad (3.27)$$

quando o estímulo externo atua sobre um padrão distinto do padrão de inicialização. Os termos separados nos somatórios estão associados com o padrão de inicialização  $n$  e com o padrão estimulado  $d$ , enquanto que o termo geral representa o ruído estático. Note que em (3.26) o estímulo  $h$  atua no padrão  $n$  e em (3.27) o estímulo atua no padrão  $d$ , sendo que em ambos os casos a rede é inicializada apresentando superposição máxima com o padrão  $n$ .

Em nossas simulações, o valor médio da superposição do sistema com os padrões  $n$  e  $d$ , após o sistema ter entrado em regime estacionário, foi obtido com base em 50 amostras diferentes, para cada uma destas amostras um novo conjunto de memórias foi gerado aleatoriamente. Utilizando um processo de atualização em paralelo, as simulações numéricas foram realizadas para redes com tamanhos  $N = 10000$ ,  $20000$  e  $40000$  com

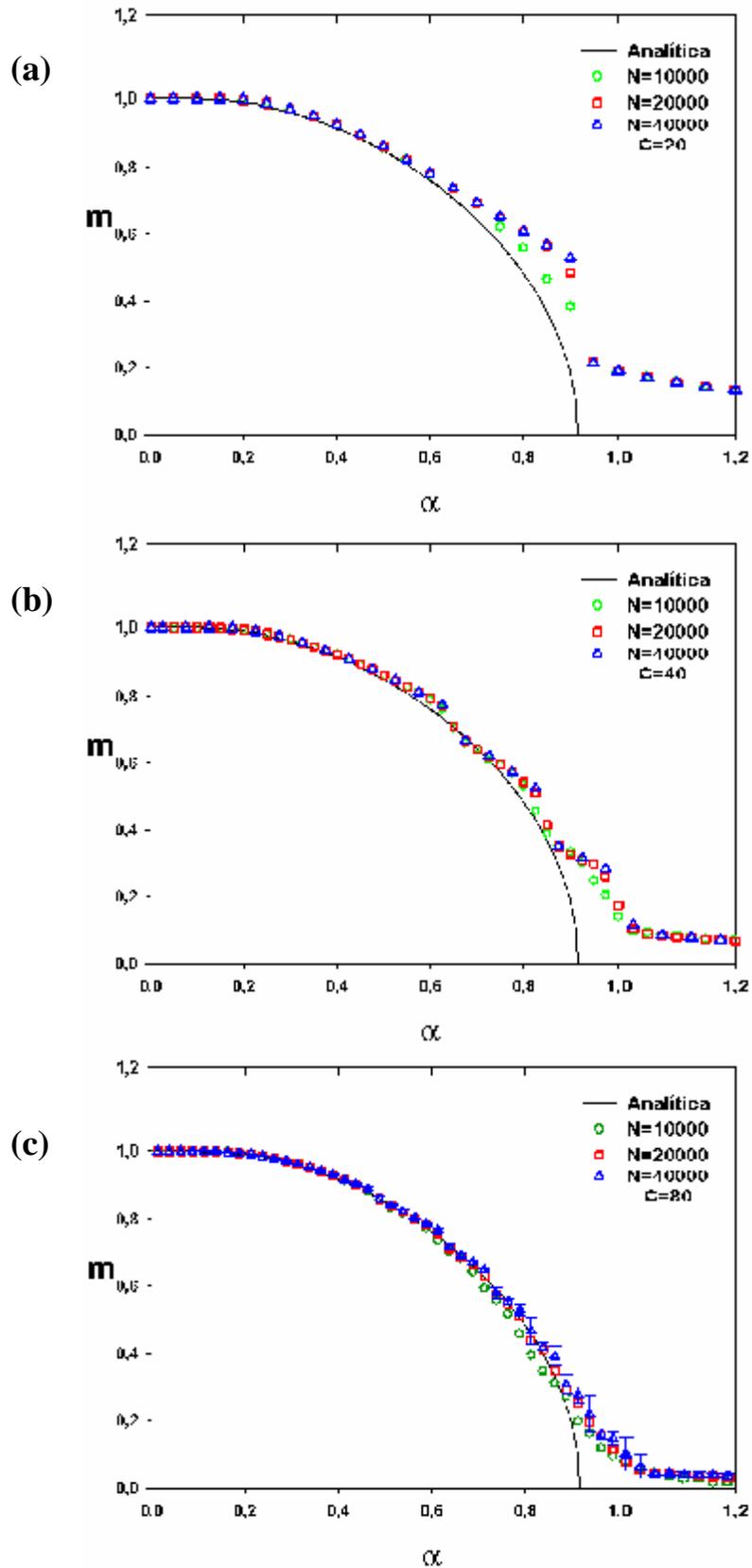
conectividades  $C = 20, 40$  e  $80$ . Em todas as amostras, partimos sempre de uma configuração inicial com superposição  $m_n^o \approx 0,99$  e  $m_d^o \approx 0,01$ .

Na figura 3.10, apresentamos a superposição  $m$  como função de  $a$  para  $T = 0$  e  $h = 0,2$ , no caso em que o padrão estimulado é o mesmo que o de inicialização. As linhas sólidas são as soluções da equação (3.14) válida para  $C \rightarrow \infty$  e os símbolos são os resultados das simulações numéricas. Observe que para valores pequenos de  $a$  as soluções analítica (linha cheia) e numérica (símbolos) coincidem perfeitamente, indicando que a superposição  $m$ , para esta região, parece ser insensível ao tamanho  $N$  e à conectividade  $C$  do sistema. Próximo à região de transição, percebemos que os resultados numéricos apresentam qualitativamente o mesmo comportamento (que se assemelha ao resultado analítico), entretanto é evidente a dependência dos resultados com a conectividade  $C$  e o tamanho da rede  $N$ .

Na figura 3.10a (para  $C=20$ ), à medida que tomamos redes cada vez maiores, percebemos que os resultados da simulação tendem a se afastar da solução analítica. Este comportamento não era esperado. Ao contrário, esperávamos uma concordância cada vez maior entre a solução analítica e a solução computacional, com o crescimento de  $N$ . Este comportamento pode ser melhor entendido se atentarmos que apesar de estarmos indo na direção do limite termodinâmico ( $N$  muito grande), a condição  $C \gg 1$  está sendo negligenciada. Isto fica mais evidente nas simulações numéricas mostradas nas figuras 3.10b e 3.10c, onde usamos conectividades maiores,  $C = 40$  e  $C = 80$  respectivamente.

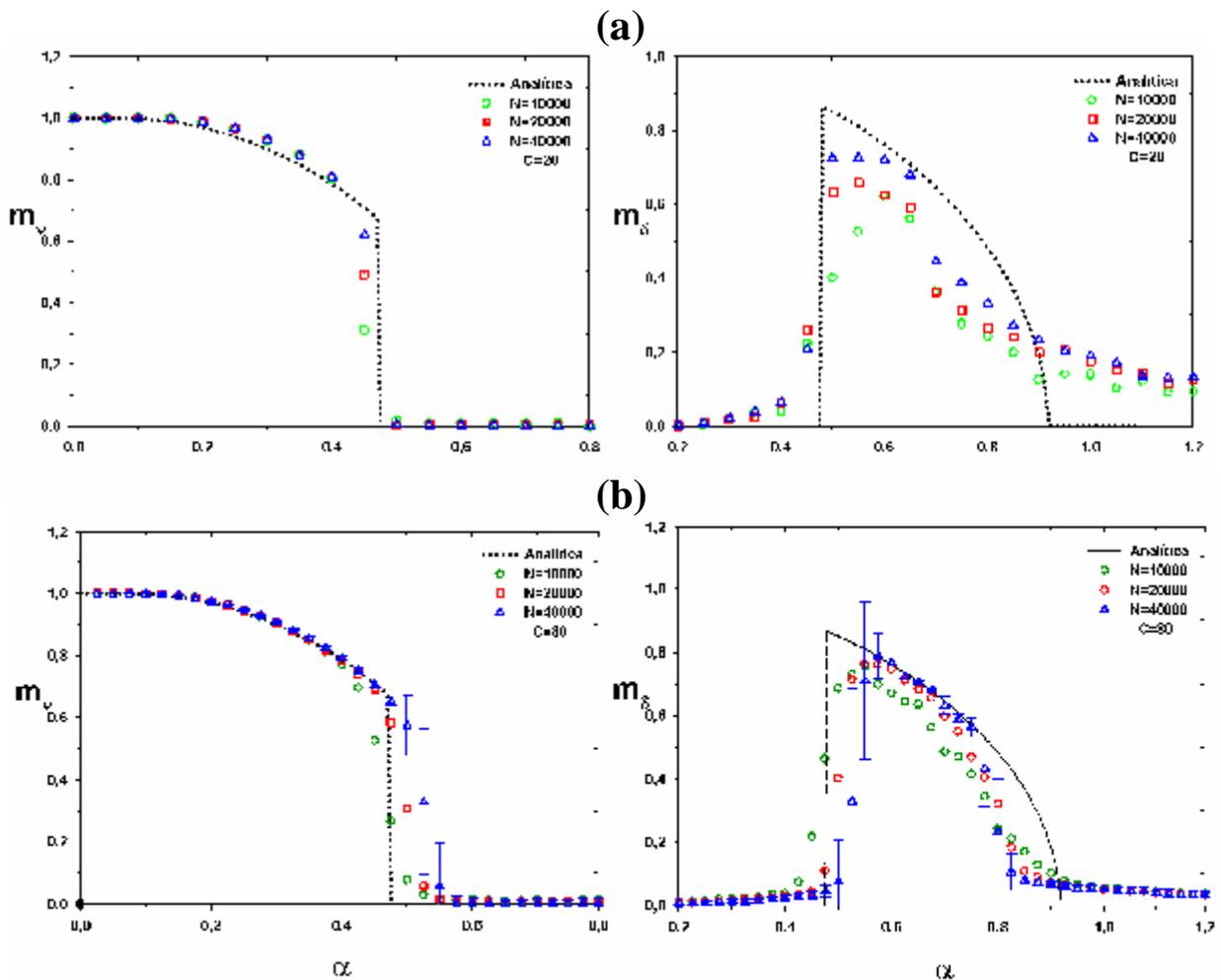
Na figura 3.11 apresentamos, as superposições  $m_v$  e  $m_\delta$  como função de  $a$  para  $T = 0$  e  $h = 0,2$ , no caso em que o padrão estimulado é diferente do padrão de inicialização. Novamente percebemos que para valores pequenos de  $a$  (antes da transição descontínua), as simulações apresentam um comportamento que praticamente independe do tamanho da rede, coincidindo perfeitamente com o resultado analítico. Na região além da transição, as simulações ainda reproduzem, qualitativamente, o comportamento deduzido das equações analíticas.

As barras de erro, exibidas nas figuras 3.10c e 3.11b, correspondem ao maior tamanho da rede  $N = 40000$  e  $C = 80$ .



**Figura 3.10:**  $m \times \alpha$  para  $T = 0$ ,  $h = 0,2$  e  $m_n^o \approx 0,99$  e  $m_d^o \approx 0,01$ . A linha cheia corresponde à solução analítica e os símbolos aos resultados das simulações para  $N = 10000$  (círculos)  $N = 20000$  (quadrados) e  $N = 40000$  (triângulos) nos casos: a)  $C = 20$ , b)  $C = 40$  e c)  $C = 80$ .

Como podemos observar nas figuras 3.11a e 3.11b, os efeitos do tamanho finito da conectividade  $C$  e do tamanho da rede  $N$  exercem uma forte influência no comportamento das superposições do estado do sistema com os padrões de reconhecimento na região de transição de fase. Note que quando  $N$  cresce os resultados da superposição  $m_n$  e  $m_d$  caracterizam bem uma transição de primeira ordem encontrada nos cálculos analíticos e na transição de segunda ordem no parâmetro  $m_d$  o efeito de tamanho finito é fundamental nos resultados. Contudo, acreditamos que exista um valor ótimo para a razão  $C/N$  ( $C \gg 1$  e  $C \ll N$ ) no qual os resultados das simulações e os resultados analíticos apresentem uma excelente concordância.



**Figura 3.11:**  $m \times \alpha$  para  $T = 0$ ,  $h = 0,2$  e  $m_n^o \approx 0,99$  e  $m_d^o \approx 0,01$ . As linha pontilhadas correspondem à solução analítica e os símbolos aos resultados das simulações para  $N = 10000$  (círculos)  $N = 20000$ (quadrados) e  $N = 40000$  (triângulos) nos casos: a)  $C = 20$ , b)  $C = 80$ .

Estes resultados mostram que a condição menos restritiva  $C \ll N$ , na simulação, por si só, não é suficiente para garantir que as propriedades de reconhecimento, para valores grandes do parâmetro  $a$ , reproduzam os resultados do modelo ultra diluído. É necessário também que tenhamos  $C \gg 1$  além de  $C \ll N$ . Um estudo mais específico da dependência da conectividade  $C$  com o tamanho do sistema  $N$ , seria fundamental para entender o papel da condição menos restritiva  $C \ll N$  em comparação ao regime ultra diluído  $C \ll \ln N$ .

# Capítulo 4

## Conclusões

Neste capítulo, apresentamos um resumo dos principais resultados obtidos nesta tese. Em termos gerais, utilizamos as ferramentas da mecânica estatística para estudar as propriedades de reconhecimento e armazenamentos do modelo de Hopfield quando submetido à ação de estímulos externos.

No capítulo 1, descrevemos de maneira simplificada a biologia do neurônio, o surgimento das redes neurais artificiais, a evolução das idéias nesta área de pesquisa, as arquiteturas para conexões dos elementos da rede, os tipos de funções de ativação, e por fim, sua associação com os modelos magnéticos.

No capítulo 2, introduzimos o modelo de Hopfield, discutimos a estabilidade das memórias, destacando como este modelo, inspirado no postulado de aprendizagem de Hebb, define a matriz de pesos sinápticos. Apresentamos ainda os principais resultados obtidos para o modelo, e descrevemos uma generalização termodinâmica proposta por Amit et al. [17], por último, apresentamos uma versão diluída, conhecida como modelo DGZ, proposta por Derrida, Gardner e Zippelius [25].

No capítulo 3, estudamos analiticamente e através de simulações numéricas a dinâmica de uma rede de neurônios com diluição e assimetria nos acoplamentos sinápticos, onde incluímos um campo para modelar o estímulo externo sobre um dado padrão da rede. Utilizando uma aproximação de campo médio que permitiu estudar o comportamento assintótico do sistema, analisamos as propriedades de armazenamento e reconhecimento de padrões. Duas situações foram estudadas: i) quando o estímulo externo atua sobre o padrão de inicialização e ii) quando o estímulo externo atua sobre um padrão distinto do padrão de inicialização. Os resultados obtidos mostram, em ambos os casos, que o parâmetro  $h$

melhora a qualidade do reconhecimento e aumenta consideravelmente a capacidade de armazenamento da rede.

No primeiro caso, obtivemos uma equação dinâmica para a superposição  $m(t)$  do estado do sistema em função dos parâmetros  $T$ ,  $\alpha$  e  $h$ . Este último é o parâmetro que ativa o estímulo externo. Para  $h = 0$ , a equação para a superposição reproduz os resultados do modelo DGZ puro.

Para  $h \neq 0$  e sendo o padrão de inicialização e o padrão estimulado coincidentes, qualquer valor inicial de  $m_o$ , desde que  $m_o \neq 0$ , não modifica o comportamento assintótico da rede. Para  $a = 0$  o efeito do estímulo é muito expressivo e na região  $h = 1$  a superposição é grande ( $m \approx 1$ ) para todo o intervalo  $0 \leq T \leq 1$ . Para  $a \neq 0$  e  $T = 0$ , obtivemos o diagrama de fase e a transição entre a fase de reconhecimento (R) e a fase paramagnética (P) é de segunda ordem como acontece no caso  $h = 0$ . Neste caso, a capacidade de armazenamento cresce de forma quadrática com o parâmetro  $h$  (equação (3.15)), indicando que a influência dos demais padrões é atenuada pela inclusão do estímulo. Para  $h = 0$ , recuperamos a capacidade máxima do modelo DGZ,  $a_c = 0,6366\dots$ , e para  $h = 1$  temos  $a = 2,546\dots$ .

Analizamos o caso  $a$  e  $T \neq 0$  e, estudando a estabilidade da solução  $m = 0$ , obtivemos uma equação analítica para a superfície crítica no espaço dos parâmetros  $T$ ,  $\alpha$  e  $h$ . O diagrama de fase (figura 3.4) mostra que para estímulos muito intensos o comportamento assintótico do sistema é completamente determinado pela ação do campo, de forma que o padrão de reconhecimento apresentará a informação contida no estímulo. Podemos resumir este primeiro caso concluindo que a influência do campo, além de intensificar a tendência natural do sistema em convergir para o padrão de reconhecimento, também aumenta a bacia atratora do padrão, tornando a rede mais robusta.

No segundo caso, o estímulo atua sobre um padrão diferente do padrão de inicialização. Obtivemos duas equações acopladas para a dinâmica do sistema, que descrevem as superposições: com o padrão de inicialização  $m_n(t)$  e com o padrão estimulado  $m_d(t)$  em função dos parâmetros  $T$ ,  $\alpha$  e  $h$ . Para  $h = 0$  as equações tornam-se idênticas e o sistema não privilegia nenhum dos padrões.

Para  $h \neq 0$  a transição entre a fase de reconhecimento do padrão  $n$  e a fase de reconhecimento do padrão  $d$  é sempre de primeira ordem, caracterizando a existência de um ponto crítico dado por  $h_c \approx 0,279$  para  $a = 0$  e  $T = 0$ .

Para  $a = 0$  e  $T < 1$ , existe uma pequena região de domínio do padrão de inicialização  $m_n$  que se estende desde  $T = 0$  para  $h_c \approx 0,279$  até  $T = 1$  para  $h = 0$ . Ao longo da fronteira desta região o sistema passa abruptamente da bacia atratora do padrão  $v$  para a bacia do padrão  $\delta$ . Para  $T > 1$  o padrão  $n$  é completamente desestabilizado e para valores pequenos de  $h$ , ocorre uma transição de fase de segunda ordem, no parâmetro  $m_d$ , para a fase paramagnética.

Obtivemos a fronteira crítica entre a fase de reconhecimento e a fase paramagnética no plano  $a \times h$  para  $T = 0$  (figura 3.7). Na região de reconhecimento, o domínio do padrão  $n$  ocorre apenas numa pequena faixa do diagrama de fase que apresenta uma reentrância sobre a grande região de domínio do padrão  $d$ .

Analisamos o caso  $a$  e  $T \neq 0$ . Para  $h = 0$  (figura 3.8a) o padrão  $n$  é dominante e experimenta uma transição de segunda ordem com o crescimento de  $T$  e  $a$ . Para  $h \neq 0$  e aumentando  $a$  e  $T$ , o sistema apresenta uma transição de fase de primeira ordem, passando da fase de reconhecimento do padrão  $n$  ( $m_n \neq 0$ ) para a fase de reconhecimento do padrão  $d$  com superposição constante da ordem de  $m_d \approx 0,87$ . Seguindo, dentro da região de domínio do padrão  $d$  ( $m_n = 0$ ), o sistema experimenta uma nova transição de fase na qual o parâmetro de ordem  $m_d$  tende a zero continuamente, caracterizando uma segunda linha crítica  $a_c(T)$  com  $h$  fixo ( $h = 0,2$ ). Para valores grandes de  $h$  (figura 3.8c), o padrão estimulado domina inteiramente a dinâmica do sistema, apresentando uma transição de segunda ordem para a fase paramagnética.

O diagrama de fase no espaço dos parâmetros  $T$ ,  $a$  e  $h$  é mostrado na figura 3.9 separando as regiões de reconhecimento da fase paramagnética.

Verificamos também, através de análises numéricas, que o comportamento do sistema independe da condição inicial  $m_n^0$  (superposição inicial com o padrão de inicialização), para  $m_n^0 \neq 0$ , e que para valores diferentes de  $m_d^0$  (superposição inicial com o padrão estimulado) o sistema apresenta comportamentos quantitativamente diferentes.

Simulações numéricas corroboram os resultados analíticos.

Como perspectiva futura, sugerimos a realização de simulações numéricas para redes maiores, visando um estudo mais detalhado sobre a dependência da conectividade  $C$  e o tamanho do sistema  $N$ , para entender o papel da condição menos restritiva  $C \ll N$  em comparação ao regime ultra diluição  $C \ll \ln N$ . Também, um estudo numérico das bacias atratoras, tempo médio de convergência, estados espúrios em torno das memórias, etc. O

estudo desse modelo aplicado ao modelo usual de Hopfield já está sendo realizado utilizando a derivação heurística proposta por Geszti [19].

# Apêndice A

## Cálculo da Equação Dinâmica do Modelo DGZ

### A.1 – Relação de recorrência para a superposição $m(t)$

Detalharemos aqui os cálculos da seção 2.5.2, onde encontramos as expressões para a superposição no modelo DGZ desenvolvido por Derrida, Gardner e Zippelius [25].

Para encontrar a equação de recorrência (2.30), estabeleceremos, primeiro, a relação entre  $m(t = 0)$  e  $m(t = 1)$  e depois generalizaremos para todo tempo  $t$ .

A partir da equação (2.26), temos

$$\begin{aligned} m(t = 1) &= \frac{1}{N} \sum_{i=1}^N \tanh\left(\frac{\mathbf{x}_i^1 h_i(t = 0)}{T_0}\right) \\ &= \left\langle \left\langle \tanh\left(\frac{\mathbf{x}_i^1 h_i(t = 0)}{T_0}\right) \right\rangle \right\rangle, \end{aligned} \quad (\text{A.1})$$

onde incorporamos o termo  $\mathbf{x}_i^1$  ao argumento da tangente hiperbólica e introduzimos a notação  $\langle \langle \dots \rangle \rangle$  para expressar a média sobre as variáveis  $\mathbf{x}$ .

Para obter o termo  $\mathbf{x}_i^1 h_i(t = 0)$  vamos multiplicar a equação (2.28) por  $\mathbf{x}_i^1$ , donde

$$\mathbf{x}_i^1 h_i(t = 0) = \sum_{r=1}^K \mathbf{x}_{j_r}^1 S_{j_r}(t = 0) + \mathbf{x}_i^1 \sum_{r=1}^K \sum_{m=2}^p \mathbf{x}_i^m \mathbf{x}_{j_r}^m S_{j_r}(t = 0). \quad (\text{A.2})$$

Os dois termos acima são somas de  $k$  e  $k(p-1)$  termos respectivamente que podem assumir valores  $\pm 1$  independentemente. Definindo  $n$  como o número de termos negativos na primeira soma e  $s$  como o número de termos negativos na segunda soma, temos:

$$\mathbf{x}_i^1 h_i(t=0) = (k-n) - n + [k(p-1) - s] - s = kp - 2(n+s)$$

onde  $n = 0, 1, \mathbf{K}, k$  e  $s = 0, 1, \mathbf{K}, k(p-1)$ .

Se no tempo  $t = 0$  o sistema está correlacionado com o padrão  $\mathbf{x}^1$ , de forma que

$$\langle \mathbf{x}_{j_r}^1 S_{j_r}(t=0) \rangle = m_0, \quad (\text{A.3})$$

onde  $m_0 = m_{j_r}(t=0)$ , então a probabilidade do produto  $S_{j_r} \mathbf{x}_{j_r}^1$  ser  $+1$  ou  $-1$  pode ser escrita como:

$$\Pr(S_{j_r} \mathbf{x}_{j_r}^1 = \pm 1) = \frac{(1 \pm m_0)}{2}. \quad (\text{A.4})$$

Logo, a probabilidade de haver  $n$  termos negativos na primeira soma é dada por

$$\Pr(n) = \binom{k}{n} \left( \frac{1-m_0}{2} \right)^n \left( \frac{1+m_0}{2} \right)^{k-n} = \binom{k}{n} (1-m_0)^n (1+m_0)^{k-n} \frac{1}{2^k}. \quad (\text{A.5})$$

Para a segunda soma, a configuração inicial não está correlacionada com as memórias restantes ( $m \neq 1$ ), portanto, a probabilidade de ter  $s$  termos negativos na segunda soma é:

$$\Pr(s) = \binom{k(p-1)}{s} \left( \frac{1}{2} \right)^s \left( \frac{1}{2} \right)^{k(p-1)-s} = \binom{k(p-1)}{s} \frac{1}{2^{k(p-1)}}. \quad (\text{A.6})$$

Como em  $t = 0$  as duas somas são independentes, a probabilidade total é obtida pelo produto das probabilidades  $\Pr(n)\Pr(s)$ , dada por:

$$\Pr(n, s) \equiv \frac{1}{2^{kp}} \binom{k(p-1)}{s} \binom{k}{n} (1+m_0)^{k-n} (1-m_0)^n. \quad (\text{A.7})$$

Agora podemos expressar a média em (A.1) na seguinte forma:

$$m(t = 1) = \sum_{k=0}^N \Pr(k) \sum_{n=0}^k \sum_{s=0}^{k(p-1)} \Pr(n, s) \tanh\left(\frac{kp - 2n - 2s}{T_0}\right), \quad (\text{A.8})$$

onde somamos sobre todos os valores possíveis de  $n$  e  $s$ , e  $\Pr(k)$  é a probabilidade do neurônio  $i$  estar conectado a  $k$  outros neurônios após a diluição. Seguindo a distribuição de probabilidades adotada, equação (2.21), obtemos

$$\Pr(k) = \binom{N}{K} \left(\frac{C}{N}\right)^k \left(1 - \frac{C}{N}\right)^{N-k}. \quad (\text{A.9})$$

onde

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}.$$

Tomando o limite  $N \rightarrow \infty$ , com  $\frac{C}{N} \ll 1$ , a distribuição  $\Pr(k)$  assume a forma mais simples de uma distribuição de Poisson:

$$\Pr(k) \approx \frac{C^k e^{-C}}{k!}. \quad (\text{A.10})$$

A distribuição de Poisson, cuja média é  $C$  e dispersão  $\sqrt{C}$ , no limite  $C \rightarrow \infty$  tende a uma função delta de Dirac,  $d(C - k)$ .

A generalização da equação (A.8) para qualquer tempo pode ser feita se pudermos garantir que os sítios  $S_{j_r}(t)$ , na equação do campo local, são descorrelacionados. É possível mostrar [26,27] que esta condição é satisfeita desde que a constante  $C$  seja finita, ou mais precisamente, se

$$C \ll \ln N.$$

Esta condição é conhecida como *ultra-diluição* e garante que:

- Praticamente todos os ciclos de retroalimentação são eliminados;

- Todos os neurônios são independentes para todo tempo  $t$ , pois as árvores de ancestrais (neurônios aferentes), conectando cada neurônio a seu estado inicial no tempo  $t = 0$ , não têm neurônios em comum.

Sendo assim, a expressão (A.8) pode ser escrita para tempos arbitrários, constituindo a equação de recorrência (2.30) apresentada no segundo capítulo:

$$m(t+1) = \sum_{k=0}^{\infty} \frac{C^k e^{-C}}{k!} \sum_{n=0}^k \sum_{s=0}^{k(p-1)} \binom{k}{n} \binom{k(p-1)}{s} \frac{(1+m)^{k-n} (1-m)^n}{2^{kp}} \tanh\left(\frac{kp - 2n - 2s}{T_0}\right) \quad (\text{A.11})$$

### A.1.1 – Limite de conectividade infinita

Devido a impossibilidade de realizar a soma na equação (A.11), vamos nos limitar ao caso particular  $C \rightarrow \infty$  para obter a expressão integral para a superposição  $m(t)$  no modelo DZG, equação (2.32).

No limite  $C$  e  $p \rightarrow \infty$ , e mantendo em mente que  $N \rightarrow \infty$  primeiro, a expressão (A.2) é uma soma de infinitas variáveis aleatórias independentes, na qual podemos aplicar o Teorema do Limite Central. Segundo o teorema, a variável aleatória dada por:

$$X = \sum_{r=1}^K x_{j_r}^1 S_{j_r} + x_i^1 \sum_{r=1}^K \sum_{m=2}^p x_i^m x_{j_r}^m S_{j_r}, \quad (\text{A.12})$$

possui uma distribuição de probabilidades Gaussiana

$$P(X) = \frac{1}{\sqrt{2ps^2}} e^{-(X-\bar{X})^2 / 2s^2}. \quad (\text{A.13})$$

O valor médio e a variância da série podem ser calculados,

$$\langle x_i^1 h_i(t) \rangle = \left\langle \sum_{r=1}^k x_{j_r}^1 S_{j_r}(t) \right\rangle + \left\langle x_i^1 \sum_{r=1}^k \sum_{m=2}^p x_i^m x_{j_r}^m S_{j_r}(t) \right\rangle = \left\langle \sum_{r=1}^k x_{j_r}^1 S_{j_r}(t) \right\rangle = Cm(t), \quad (\text{A.14})$$

onde usamos o fato de que o termo de ruído (segundo termo) apresenta média nula devido a independência dos  $x$ . Para a variância, encontramos de forma semelhante

$$s^2 = \langle (\mathbf{x}_i^1 h_i(t))^2 \rangle - \langle \mathbf{x}_i^1 h_i(t) \rangle^2 = C(p-1) \equiv C^2 \mathbf{a}, \quad (\text{A.15})$$

onde  $\mathbf{a} = (p-1)/C$ . Usando a representação Gaussiana, a equação (A.1) pode ser escrita como:

$$m(t+1) = \frac{1}{\sqrt{2ps^2}} \int_{-\infty}^{+\infty} dx e^{-(x-Cm(t))^2/2s^2} \tanh\left(\frac{x}{T_0}\right), \quad (\text{A.16})$$

onde  $s = C\sqrt{\mathbf{a}}$ . Por fim, fazendo a mudança de variável  $z = \frac{x-Cm(t)}{s}$ , obtemos a equação desejada (2.32):

$$m(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh\left(\frac{m(t) + z\sqrt{\mathbf{a}}}{T}\right), \quad (\text{A.17})$$

onde  $T = b^{-1}$  é a temperatura reduzida  $T = T_0 / C$ .

## Apêndice B

### Equação de recorrência para o modelo DGZ com o estímulo externo

#### B.1 – Padrão estimulado paralelo ao padrão de inicialização

Vamos obter analiticamente a equação de recorrência (3.11) para a superposição do sistema com o padrão estimulado,  $m_n(t+1)$ , no caso  $C \rightarrow \infty$  e  $p \rightarrow \infty$ .

Utilizando a equação (3.8), vamos especificar o termo do campo local que aparece no argumento da tangente hiperbólica em (3.10):

$$\mathbf{x}_i^n h_i(t) = h_o m_n(t) + \sum_{R=1}^k \mathbf{x}_{j_r}^n S_{j_r}(t) + \mathbf{x}_i^n \sum_{r=1}^K \sum_{m \neq n}^p \mathbf{x}_i^m \mathbf{x}_{j_r}^m S_{j_r}(t). \quad (\text{B.1})$$

No limite  $N \rightarrow \infty$ , a equação acima é uma soma de infinitas variáveis aleatórias independentes que segundo o Teorema do Limite Central, apresentam uma distribuição de probabilidades Gaussiana com valor médio dado por:

$$\langle \mathbf{x}_i^n h_i(t) \rangle = h_o m_n(t) + C m_n(t) \quad (\text{B.2})$$

e variância dada por (A.15),  $s^2 = C(p-1) \equiv C^2 a$ . Usando a representação Gaussiana a equação (3.10) pode ser escrita como:

$$m(t+1) = \frac{1}{\sqrt{2ps^2}} \int_{-\infty}^{+\infty} dx e^{-(x - cm(t))^2 / 2s^2} \tanh\left(\frac{hm(t) + x}{T_0}\right). \quad (\text{B.3})$$

Fazendo a mudança de variável  $z = \frac{x - Cm(t)}{S}$ , obtemos a seguinte equação:

$$m(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh\left(\frac{m(t)(h_o + C) + S z}{T_0}\right). \quad (\text{B.4})$$

Definindo  $T = T_0/C$  e  $h = h_o/C$  como os parâmetros reduzidos, temos a expressão desejada para a superposição do sistema com o padrão estimulado, equação (3.11). Esta equação permite estudar as propriedades do modelo DGZ com estímulo externo e é válida para todo tempo  $t$ :

$$m(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh\left(\frac{m(t)(h+1) + z\sqrt{a}}{T}\right). \quad (\text{B.5})$$

No caso limite  $T \rightarrow 0$  ( $b \rightarrow \infty$ ), podemos utilizar a identidade matemática

$$\int \frac{dz}{\sqrt{2p}} e^{-z^2/2} \tanh[b(az+b)] = \text{erf}\left(\frac{b}{\sqrt{2a}}\right), \quad (\text{B.6})$$

onde  $\text{erf}$  é a função erro, definida como  $\text{erf}(x) = \frac{2}{\sqrt{p}} \int_0^x e^{-u^2} du$ . A equação resultante deste

limite, corresponde a equação (3.12) para  $m(t) = m$ , tem a forma:

$$m(t+1) = \text{erf}\left[\frac{m(t)(h+1)}{\sqrt{2a}}\right]. \quad (\text{B.7})$$

## B.2 - Padrão estimulado diferente do padrão de inicialização

### B.2.1 – Superposição do sistema com o padrão inicialização $n$ .

Encontraremos aqui a equação de recorrência para  $m_n(t+1)$  no caso em que o padrão de inicialização da rede e o padrão estimulado são distintos.

Iniciaremos determinando a superposição do sistema com a memória utilizada na inicialização, padrão  $x^n$ . Da equação (3.18), temos:

$$x_i^n h_i(t) = \sum_{r=1}^K x_{j_r}^n S_{j_r}(t) + x_i^n \sum_{r=1}^K \sum_{\substack{m \neq n \\ m \neq d}}^p x_i^m x_{j_r}^m S_{j_r}(t) + x_i^n x_i^d \sum_{r=1}^K x_{j_r}^d S_{j_r}(t) + x_i^n x_i^d h m_d(t). \quad (\text{B.8})$$

No limite  $C$  e  $p \rightarrow \infty$  a expressão (B.8) é uma soma de infinitas variáveis aleatórias independentes na qual podemos aplicar o Teorema do Limite Central. A superposição do estado do sistema com o padrão de inicialização  $n$ , definida em (3.11), é:

$$m_n(t+1) = \left\langle \left\langle \tanh\left(\frac{x_i^n h_i(t)}{T_0}\right) \right\rangle \right\rangle. \quad (\text{B.9})$$

Levando em conta a independência das variáveis  $x$ ,  $\langle x_i^m x_j^n \rangle = d_{ij} d_{mn}$ , o valor médio e a variância do termo no argumento da tangente hiperbólica podem ser facilmente calculados:

$$\langle x_i^n h_i(t) \rangle = C m_n(t) \quad (\text{B.10})$$

e

$$s^2 = Cp + C^2 m_d^2(t)(1 + 2h_o/C + h_o^2/C^2). \quad (\text{B.11})$$

Numa representação Gaussiana a equação (B.9) assume a forma:

$$m_n(t+1) = \frac{1}{\sqrt{2ps^2}} \int_{-\infty}^{+\infty} dx e^{-\frac{(x - C m_n(t))^2}{2s^2}} \tanh\left(\frac{x}{T_0}\right). \quad (\text{B.12})$$

Fazendo a mudança de variável  $z = \frac{x - C m_n(t)}{s}$  e definindo o parâmetro de carga

$a = p/C$  obtemos a seguinte equação para a superposição  $m_n(t+1)$

$$m_n(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} e^{-z^2/2} \tanh\left(\frac{sz + C m_n(t)}{T_0}\right) dz. \quad (\text{B.13})$$

Renormalizando a temperatura  $T_0$  e o parâmetro  $h_o$  ( $T = T_o/C$ ,  $h = h_o/C$ ) dado que  $s = C\sqrt{a + m_d^2(t)(1+h)^2}$ , chegamos à equação de recorrência (3.19):

$$m_n(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} e^{-z^2/2} \tanh\left(\frac{m_n(t) + z\sqrt{m_d^2(t)(1+h)^2 + a}}{T}\right) dz. \quad (\text{B.14})$$

O limite  $T \rightarrow 0$  pode ser obtido utilizando a identidade (B.6), resultando na equação (3.23) para  $m_n(t) = m_n$ :

$$m_n(t+1) = \text{erf}\left(\frac{m_n(t)}{\sqrt{2}\sqrt{m_d^2(t)(1+h)^2 + a}}\right), \quad (\text{B.15})$$

### B.2.2 – Superposição do sistema com o padrão estimulado $d$ .

A equação para o reconhecimento do padrão estimulado é desenvolvida de forma análoga à aplicada na seção anterior.

O valor médio para  $m_d(t+1)$  é dado por:

$$m_d(t+1) = \left\langle \left\langle \tanh\left(\frac{\mathbf{x}_i^d h_i(t)}{T_0}\right) \right\rangle \right\rangle, \quad (\text{B.16})$$

cujo termo  $\mathbf{x}_i^d h_i(t)$  pode ser posto na forma

$$\mathbf{x}_i^d h_i(t) = \sum_{r=1}^K \mathbf{x}_{j_r}^d S_{j_r}(t) + \mathbf{x}_i^d \sum_{r=1}^K \sum_{\substack{m \neq n \\ m \neq d}}^p \mathbf{x}_i^m \mathbf{x}_{j_r}^m S_{j_r}(t) + \mathbf{x}_i^d \sum_{r=1}^K \mathbf{x}_i^n \mathbf{x}_{j_r}^n S_{j_r}(t) + h_o m_d(t). \quad (\text{B.17})$$

No limite termodinâmico a expressão em (B.17) é uma soma de infinitas variáveis aleatórias independentes com média e a variância dada pelas equações:

$$\left\langle \xi_i^\delta h_i(t) \right\rangle = h_o m_\delta(t) + C m_\delta(t) \quad (\text{B.18})$$

e

$$s^2 = Cp + m_n^2(t)C^2. \quad (\text{B.19})$$

Tomando o limite  $C, p \rightarrow \infty$ , renormalizando os parâmetros  $T_0$  e  $h_0$ , e após algumas simplificações podemos escrever a equação (B.16) na forma integral, chegando à equação desejada (3.20):

$$m_d(t+1) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} \tanh \left[ \frac{m_d(t)(h+1) + z\sqrt{m_n^2(t) + a}}{T} \right], \quad (\text{B.20})$$

onde  $T = b^{-1}$  é a temperatura reduzida  $T = T_0 / C$  e  $h = h_0 / C$ .

No limite  $T \rightarrow 0$  e novamente recorrendo à identidade (B.6), a equação (B.20) assume a seguinte forma (equação (3.24) para  $m_d(t) = m_d$ ):

$$m_d(t+1) = \text{erf} \left( \frac{m_d(t)(h+1)}{\sqrt{2}\sqrt{m_n^2(t) + a}} \right). \quad (\text{B.21})$$

# Apêndice C

## Teorema do Limite Central

Nos limitaremos aqui a citar o teorema do limite central, um tratamento mais detalhado, bem como sua demonstração, podem ser encontrados em qualquer bom livro de estatística, como, por exemplo, a referência [47].

Considere uma seqüência infinita de variáveis aleatórias independentes  $x_1, x_2, \mathbf{K}$  com densidades de probabilidade  $P_1(x_1), P_2(x_2), \mathbf{K}$ . Se  $X$  é a média dos primeiros  $N$  termos da série e  $s$  é a incerteza média quadrática (isto é,  $s^2 = (\Delta x_1^2 + \mathbf{K} + \Delta x_N^2)/N$ ), então, para valores grande de  $N$ , a distribuição  $P_X(x)$  tende a uma distribuição Gaussiana, na forma:

$$P_X(x) = \left( \frac{N}{2\pi s^2} \right)^{1/2} e^{-Nx^2/2s^2} .$$

# Apêndice D

## Automédia

A propriedade de automédia nos permite escrever a média presente na expressão (A.1), realizada sobre os sítios, como uma média realizada sobre as configurações dos padrões  $\mathbf{x}^m$ . Para mais detalhes sobre a propriedade de automédia, veja a referência [48].

As memórias  $\mathbf{x}^m$  relacionadas com um certo neurônio, podem assumir  $2^p$  configurações distintas. Como estas memórias são geradas aleatoriamente, cada configuração distinta deve ocorrer com a mesma probabilidade  $1/2^p$ . Dessa forma, a probabilidade de que uma dada configuração apareça  $n$  vezes é dada por:

$$\Pr(N(\mathbf{x}) = n) = \binom{N}{n} \left(\frac{1}{2^p}\right)^n \left(1 - \frac{1}{2^p}\right)^{N-n}. \quad (\text{D.1})$$

O valor médio de  $n$  é

$$\langle n \rangle = \frac{N}{2^p}, \quad (\text{D.2})$$

enquanto que a variância é:

$$s^2(n) = \langle n^2 \rangle - \langle n \rangle^2 = \frac{N}{2^p} \left(1 - \frac{1}{2^p}\right). \quad (\text{D.3})$$

Para  $N \gg 2^p$  isto demonstra que a dispersão em torno do valor médio torna-se cada vez menor, de forma que para memórias descorrelacionadas podemos escrever:

$$\frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i) \approx \frac{1}{2^p} \sum_{\mathbf{x}} f(\mathbf{x}) \equiv \langle\langle f(\mathbf{x}) \rangle\rangle, \quad (\text{D.4})$$

onde introduzimos a notação  $\langle\langle \rangle\rangle$  para designar a média calculada sobre as configurações das variáveis  $\mathbf{x}^m$ .

## Referências

- [01] Barbizet, J., *Manual de Neuropsicologia*, 1. ed., Artes Médicas, Porto Alegre, 1985.
- [02] Martin, J. H., *Neuroanatomia*, 2. ed., Artes Médicas, Porto Alegre, 1998.
- [03] Contu, P., *Anatomia funcional do sistema nervoso*, URGS, Porto Alegre, 1972.
- [04] Gupta, M. M.; Jin, L. e Homma, N., *Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory*, 1.ed., John Wiley, New York, 2003.
- [05] Hebb, D. O., *The Organization of Behavior: A Neuropsychological Theory*, Wiley, New York, 1949.
- [06] Haykin, S., *Redes neurais: princípios e prática*, 2. ed, Bookman, Porto Alegre, 2001.
- [07] Rich, E. e Kevin, K., *Inteligência Artificial*, 2. ed., Makron Books, São Paulo, 1993.
- [08] Hopfield, J. J., *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [09] Little, W. A., *Math. Biosc.* **19**, 101 (1974); **39**, 281 (1978).
- [10] Rumelhart, D. E., Hinton, G. E. e Williams, R. J., *Nature*: **323**, 533 (1986).
- [11] Sherrington, D. e Kirkpatrick, S., *Phys. Rev. Lett.* **35**, 1792 (1975).
- [12] Edward, S. F. e Anderson, P. W., *J. Phys. F* **5**, 965 (1975).
- [13] Amit, D. J., *Modeling brain function: the world of attractor neural networks*, 1. ed., Cambridge University Press, 1989.
- [14] Davey, N., Hunt, S.P., Adams, R.G., *Neurocomputing* **62**, 459 (2004).
- [15] Amit, D. J., Gutfreund, H., e Sompolinsky, H., *Ann. Phys. N. Y.*, **173**, 30 (1987); *Phys. Rev. Lett.*, **55**, 1530(1985).

- [16] Hertz, J., Krogh, A. e Palmer, R. G., *Introduction to the theory of neural computation*, Addison-Wesley, Santa Fe, 1991.
- [17] Amit, D. J., Gutfreund, H., e Sompolinsky, H., *Phys. Rev. A*, **32**, 1007 (1985).
- [18] Huang, K., *Statistical Mechanics*, New York, Wiley, 1987.
- [19] Geszti, T., *Physical Models of Neural Networks*, World Scientific, 1990.
- [20] Oliveira, P. M. C., *Computing Boolean Statistical Models*, World Scientific Publishing, Singapore (1991).
- [21] Crisanti, A., Amit, D. J. e Gutfreund H., *Europhys Lett.* **2**, 337 (1986).
- [22] Peretto, P., *An Introduction to the modeling of Neural Network*, Cambridge University Press, Cambridge, 1994.
- [23] Peretto, P., *Biol. Cybern.* **50**, 51 (1984).
- [24] Fontanari, J. F. e Köberle, R., *J. Phys. France* **49**, 13 (1988).
- [25] Derrida, B., Gardner, E. e Zippelius A., *Europhys. Lett.* **4**, 167 (1987).
- [26] Derrida B. e Weisbuch G., *J. Phys. (France)* **47**, 1297 (1986).
- [27] Hilhorst H. J. e Nijmijer M., *J. Phys. (France)* **48**, 185 (1987).
- [28] Arenzon, J.J. e Lemke N., *J. Phys A: Math. Gen.*, **27**, 5161 (1994).
- [29] Gutfreund ,H. e Mezard, M., *Phys. Rev. Lett.* **61**, 235 (1988).
- [30] Sompolinsky, H. e Kanter, I., *Phys. Rev. Lett.* **57**, 2861 (1986).
- [31] Buhmann, J. e Schulten, K., *Europhys. Lett.* **4**, 1205 (1987).
- [32] Horn, D., *Physica A* **200**, 594 (1993); Horn, D. e Usher, M., *Phys. Rev. A.* **40**, 1036 (1989).
- [33] Zertuche, F., López-Pena, R. e Waelbroech, H., *J. Phys. A: Math. Gen.* **27**, 5879 (1994).
- [34] Fontanari, J. F., *J. Phys. I (France)* **51**, 2412 (1990).

- [35] Miranda, E. N., *J. Phys. I (France)* **1**, 999 (1991).
- [36] Stariolo, D. A. e Tamarit, F. A., *Phys. Rev. A* **43**, 5249 (1992).
- [37] Silva, C. R., Tamarit, F. A., Lemke, N., Arenzon, J. J. e Curado, E. M. F., *J. Phys. A: Math. Gen.* **28**, 1593 (1995).
- [38] Silva, C. R., Tamarit, F. e Curado, E. M. F. *Int. J. Mod. Phys. C:* **7**, 43 (1996).
- [39] Silva, C. R., Tamarit, F. e Curado, E. M. F. *Phys. Rev. E:* **55**, 3320 (1996).
- [40] Tamarit, F. A., Stariolo, D. A. e Cannas, S. A. *Phys. Rev. E:* **53**, 5146 (1996).
- [41] Engel, A., Englisch, H., e Schutte, A., *Europhys. Lett.* **8**, 393 (1989).
- [42] Amit D. J., Parisi G. e Nicolis S., *Network* **1**, 75 (1990).
- [43] Engel, A., Bouten, M., Komoda, A., e Serneels, R., *Phys. Rev. A* **42**, 4998 (1990).
- [44] Storkey, A. J., Valabregue, R., *Neural Networks* **12**, 869 (1999).
- [45] Singh, M. P., *Phys. Rev. E* **64**, 051912 (2001).
- [46] Englisch, H., Xiao, Y., e Yao, K., *Phys. Rev. A* **44**, 1382 (1991).
- [47] Garrod, C., *Statistical mechanics and thermodynamics*, Oxford University Press, New York, 1995.
- [48] Lautrud, B., *Lecture notes on the theory of the Hopfield model*, The Niels Bohr Institute, 1988.
- [49] Stein, D. L., *Spin Glasses*, *Scientific American*, 1989.
- [50] Tank, D. W.; Hopfield, J. J., *Collective Computation in Neuronlike Circuits*, *Scientific American*, 1987.