

TESE DE
DOUTORADO

Propriedades Dinâmicas de Redes de
Neurônios com Períodos Refratários e
Diluição Assimétrica

CRISÓGONO RODRIGUES DA SILVA

CENTRO BRASILEIRO DE PESQUISAS FÍSICAS
RIO DE JANEIRO, MARÇO DE 1997

Dedico esta tese aos meus filhos

Rafael e Rodrigo.

AGRADEÇO

A Evaldo M. F. Curado e Francisco A. Tamarit pela amizade e orientação.

Aos meus filhos Rafael Delgado e Rodrigo Delgado pela compreensão das muitas horas de carinho e brincadeiras que deixamos de compartilhar.

A Cristina, por todos esses anos de carinho, dedicação e compreensão.

Aos meus pais, Cristovão e Martina, pelo carinho e apoio ao longo de muitos anos.

A Constantino Tsallis e Aglaé Magalhães pela amizade e apoio durante estes anos de convivência.

A Ladário da Silva, pela amizade, sua constante ajuda e colaboração.

A Ney Lemke e Jeferson Arenzon pela amizade e colaboração.

A José Arnaldo, André Maurício, Daniel Stariolo, Andrés Papa, Isaías Oliveira, Sergio Cannas, Carmen Ochi, Fortunato Menezes, Julio Ortiz, Celia Anteneodo, Fabricio Toscano, Raul Vallejos e Sergio Curilef, pela amizade, discussões e os bons momentos compartilhados.

A Vanda, Almério e Myriam pela amizade e ajuda constante.

Aos colegas do Departamento de Física da UFAL.

Ao Núcleo de Computação de Alto Desempenho (NACAD) da COPPE–UFRJ pelo uso do CRAY, com o qual obtive parte dos resultados numéricos.

A UFAL, CAPES e CNPq pelo suporte financeiro

Abstract

In this present thesis some aspects of the dynamical properties of neural network models with refractory periods and asymmetrical dilution are studied, both analytically and through numerical simulations

The first part of chapter 1 is devoted to a summary of the biological brain functioning and the second one to the main results of the Hopfield model.

In chapters 2 e 3 we introduce a threshold dependent on the local activity of the neuron at previous time, which mimics the effect of the refractory periods, and we study the Hopfield model regarding both the fully connected model and the ultra diluted version respectively. In the later one, contrary to the fully connected model, for large values of the threshold the system can always retrieve stored patterns. In both cases we obtain the complete phase diagram

In chapter 4 we study numerically the nature of the retrieval attractors in an asymmetrically diluted Hopfield model $C \ll N$ (C is the average connectivity by neuron and N the number of neuron of the network), through the damage propagation technique, and we compare the results with those obtained analytically with the ultra diluted version ($C \ll \ln N$). An external field h that breaks the dynamical transition presented by the damage is introduced, allowing us to obtain, with better precision, the chaotic-fixed point transition in the recognition phase, in which the susceptibility associated to the damage tends to diverge.

In chapter 5 we study the generalization capacity of the ultra diluted Hopfield model. We showed that in this case the network classify better than in the fully connected case, i.e., the network generalizes with a smaller number of examples.

Índice

Folha de rosto	i
Dedicatória	ii
Agradecimentos	iii
Resumo em port.	iv
Resumo em ing.	v
Índice	v
Lista de figuras	ix
Lista de tabelas	xiii
Introdução	1
1 Modelos de Rede de Neurônios	10
1.1 Descrição da Biologia do Neurônio	10
1.1.1 O neurônio biológico	10
1.1.2 Comunicação biológica entre os neurônios	12
1.1.3 Escalas de tempo biológico	15
1.1.4 Aprendizado biológico	16
1.2 Da Biologia à Modelagem das Redes Neurais	17
1.2.1 Um modelo protótipo	17
1.2.2 Refinamento do modelo: processos aleatórios	20
1.3 Memórias Associativas: O Modelo de Hopfield	22
1.3.1 A natureza da memória associativa e a analogia física	22

1.4	Modelo de Hopfield generalizado	27
1.4.1	Termodinâmica do modelo de Hopfield	27
1.4.2	Dinâmica do modelo de Hopfield	32
2	Influência dos Períodos Refratários no Modelo de Hopfield	38
2.1	Os períodos refratários	39
2.2	O modelo	42
2.3	Equações de ponto fixo	44
2.3.1	O caso determinista $T = 0$	48
2.3.2	O caso $T \neq 0$	50
2.4	Simulações Numéricas	52
2.5	Conclusões	57
3	Rede de Neurônios com Diluição Assimétrica e Períodos Refratários	59
3.1	O Modelo	60
3.2	A Dinâmica	62
3.2.1	Número finito de memórias	66
3.2.2	Número infinito de memórias	68
3.3	Simulação Numérica	71
3.4	Conclusões	76
4	Propagação de Danos em Uma Rede de Neurônios Diluída e Assimétrica	77
4.1	Propagação de Danos	77
4.2	O Modelo	82
4.3	Método e Resultados Numéricos	83
4.4	Conclusões	90
5	Generalização em uma Rede de Neurônios com Diluição Assimétrica	93
5.1	O Modelo	95
5.2	As Equações de Recorrência	96
5.3	Limite de Conectividade Infinita	99

5.3.1	Número finito de conceitos	99
5.3.2	Número extensivo de conceito	101
5.4	Simulações Numéricas	105
5.5	Conclusões	107
	Conclusões	109
A	- Apêndice A	115
A.1	Reconhecimento de Padrões	115
A.1.1	Limite de conectividade infinita	118
A.2	Generalização	119
A.2.1	Limite de conectividade infinita	121
A.2.2	Número extensivo de conceitos	122
	Referências	126

Lista de Figuras

1.1	Desenho esquemático de um neurônio biológico.	11
1.2	Estrutura sináptica do neurônio.	12
1.3	O potencial de repouso e as diferentes concentrações iônicas. Os tamanhos relativos dos símbolos das espécies iônicas representam aproximadamente a concentração relativa de cada espécie dentro e fora da célula.	13
1.4	Representação esquemática de uma superfície de energia. Os mínimos correspondem as memórias armazenadas.	24
1.5	Diagrama de fase $T-\alpha$. A área sombreada corresponde a região de reconhecimento da rede (A e B). Fora da área sombreada existem a fase de vidro de spin (região C) e a fase paramagnética (região D).	30
1.6	A superposição m versus α a $T = 0$ na região de quebra de simetria de réplica. Os círculos cheios são os resultados da simulação para $N \simeq 34000$ neurônios. A linha cheia corresponde a solução das equações de campo médio com aproximação de réplica simétrica e a linha tracejada corresponde a solução analítica com um passo de quebra de simetria.	31
1.7	Diagrama de fase $T-\alpha$ do modelo de Hopfield ultra diluído.	36
1.8	Superposição $m(t)$ vs. α . Comparação entre três diferentes condições do modelo de Hopfield diluído assimetricamente. Linha cheia: resultados analíticos para N, p e $C \rightarrow \infty$ e $C \ll \ln N$; círculos cheios: $N \rightarrow \infty$, conectividade finita ($C = 20$); símbolos vazios: simulação numérica para $N = 4000, 8000$ e 16000 e conectividade finita ($C = 20$).	36

2.1	Gráfico m vs. α para $T = 0$ e diferentes valores de Δ . Em $\alpha_c(\Delta)$ o sistema sofre uma transição descontínua da fase de reconhecimento para uma fase de não reconhecimento.	50
2.2	Diagrama de fase T vs. α para $\Delta = 0.0, 0.2, 0.4$ e 0.6 . Abaixo das linhas críticas o sistema reconhece com ponto fixo e a transição para a fase de não reconhecimento é de primeira ordem.	51
2.3	Linha crítica $T = f(\Delta)$ para $\alpha = 0$. $(T^*, \Delta^*) \simeq (0.463, 0.610)$ é um ponto tricrítico. Para $\Delta < \Delta^*$ a transição é de segunda ordem (linha cheia) enquanto que para $\Delta > \Delta^*$ a transição é descontínua (linha tracejada). . .	52
2.4	Diagrama de fase numérico (Δ, α) para $T = 0$ e $N = 3200$ neurônios, mostrando as regiões com atratores PF (abaixo dos círculos cheios), periódicos (entre as duas formações de triângulos) e caóticos (acima dos círculos vazios). A linha obtida da simulação (linha tracejada) e a analítica (linha cheia), separam a fase de reconhecimento (abaixo) da fase de não reconhecimento (acima).	54
2.5	A superposição m e a frequência na qual cada tipo de atrator aparece como função de Δ para $\alpha = 0.01$: a) dinâmica paralela; b) dinâmica sequencial. A linha cheia corresponde a solução analítica.	55
2.6	A superposição m e a frequência na qual cada tipo de atrator aparece como função de Δ para $\alpha = 0.04$: a) dinâmica paralela; b) dinâmica sequencial. A linhas cheia corresponde a solução analítica.	56
2.7	Atividade média vs. Δ para $\alpha = 0.01$ e $\alpha = 0.04$	57
3.1	Diagrama de fase Δ vs. T para $\alpha = 0$	66
3.2	m como função do Δ para $\alpha = 0$ e diferentes valores de T (0.0, 0.1, 0.4 e 0.5).	67
3.3	Digrama de fase Δ vs. α para $T = 0$. As linhas críticas separando a fase de recuperação da fase paramagnética correspondem a $l = 4, 6, 8$ e ∞ . . .	69
3.4	A superposição m como função de α e Δ a $T = 0$ com $m_0 = 1$	70

3.5	Superfície crítica separando a fase de reconhecimento (abaixo) da fase paramagnética (acima) no espaço dos parâmetros (T, α, Δ)	71
3.6	Linha crítica separando a fase de reconhecimento da fase paramagnética no plano Δ vs. α a $T = 0$. A linha cheia corresponde a solução analítica para $l \rightarrow \infty$ e os círculos cheios correspondem aos resultados da simulação numérica com $C = 80$ e $N = 80000$	73
3.7	m vs. α para $\Delta = 0.2$, $m_0 = 1$ e $T = 0$. A linha cheia corresponde à solução analítica e os círculos aos resultados das simulações para $N = 80000$ com $C = 40$ (círculos vazios) e $C = 80$ (círculos cheios).	74
3.8	m vs. α para $\Delta = 0.7$, $C = 80$ e $T = 0$. As linhas cheias correspondem à solução analítica e os símbolos aos resultados das simulações: os círculos e os quadrados vazios correspondem a $N = 40000$ e $N = 80000$ respectivamente, com $m_0 = 0.6$. Os círculos e os quadrados cheios correspondem a $N = 40000$ e $N = 80000$ respectivamente, com $m_0 = 1$	74
3.9	$ m(t) $ vs. t para $C = 80$, $N = 80000$, $\Delta = 1.4$ e $\alpha = 0.0375$ (círculos), 0.05 (quadrados), 0.1 (triângulos) and 0.125 (lozângulos).	75
4.1	A distância de Hamming dh e a superposição m em função do α para $C = 40$ e $N = 40000$ (círculos vazios) e 80000 (círculos cheios). As linhas sólidas correspondem as soluções analíticas no regime de ultra diluição. . .	85
4.2	dh e m em função do α para $N = 80000$ e para $C = 10$ (círculos vazios), $C = 20$ (círculos cheios) e $C = 40$ (triângulos). As linhas sólidas correspondem as soluções analíticas no regime de ultra diluição.	86
4.3	A frequência com a qual aparecem os diferentes tipos de atratores em função de α para $N = 80000$ e diferentes valores de C : 20 (no topo), 40 (no meio) e 80 (na base).	87

4.4	dh e m em função de α para $N = 80000$, $C = 40$ e $h = 0.0$ (círculos vazios), 0.001 (círculos cheios) e 0.01 (triângulos). As linhas sólidas correspondem ao regime ultra diluído. Note que para $h > 0$ o dh é sempre diferente de zero (detalhe).	89
4.5	A susceptibilidade de dano χ_{dh} vs. α para $C = 40$, $h = 0.001$ e $N = 20000$ (círculos vazios), 40000 (círculos cheios) e 80000 (triângulos).	90
4.6	O χ_{dh} em função do α para $h = 0.001$, $N = 80000$ e conectividade $C = 20$ (círculos vazios), 40 (círculos cheios) e 80 (triângulos).	91
5.1	O erro de generalização ε em função do número de exemplos s para $\alpha = 0$, $b = 0.2$ e $T = 0.0, 0.5, 1.0$ e 2.0	100
5.2	Diagrama de fase T vs. s para $\alpha = 0$ para dois valores do parâmetro de correlação: $b = 0.2$ e 0.4 . As linhas separando as fases de generalização e não generalização correspondem a ε menor que 5%.	101
5.3	O erro de generalização ε em função do número de exemplos s para $T = 0$, $b = 0.2$ e $\alpha/\alpha_0 = 0.01, 0.05, 0.1, 0.12, 0.1479$ e 0.2	102
5.4	O erro de generalização ε em função do número de exemplos s para $b = 0.3$, $\alpha/\alpha_0 = 0.3$ e $T = 0.04, 0.4, 1.2$ e 2.0	104
5.5	Linhas críticas T vs. α/α_0 separando a fase de generalização (G) da fase de não generalização (NG) para 50 exemplos, $b = 0.3$ e $b = 0.4$. As linhas correspondem a um erro $\varepsilon = 0.05$	105
5.6	O erro de generalização ε vs. o número de exemplos s para $b = 0.2$ e $\alpha/\alpha_0 = 0.12$. A linha sólida corresponde as soluções analíticas. Para as simulações usamos $p = 2$, $C = 26$ e $N = 10000$ (círculos vazios) e 20000 (quadrados cheios).	106
5.7	O erro de generalização ε vs. o número de exemplos s para $b = 0.2$ e $\alpha/\alpha_0 = 0.2$. A linha sólida corresponde as soluções analíticas. Para as simulações usamos $p = 2$, $C = 16$ e $N = 10000$ (círculos vazios) e 20000 (quadrados cheios).	107

Lista de Tabelas

5.1	O erro de generalização ε para $b = 0.4$, $\alpha = 0.028$ e diferentes valores de s , para ambos os modelos; Hopfield e Hopfield diluído.	104
-----	---	-----

Introdução

A física estatística, em seu constante esforço para explicar o comportamento macroscópico da matéria a partir das leis microscópicas da física, vem se destacando ao longo das últimas décadas pelo surgimento e desenvolvimento de novos modelos e teorias, das quais tem sido possível explicar uma grande variedade de fenômenos observados na natureza. A física da matéria condensada e, em particular, a teoria dos fenômenos críticos, são talvez os exemplos que melhor ilustram (longe de serem os únicos) a valiosa contribuição realizada pela mecânica estatística à física moderna. Estas contribuições, por sua vez, transcendem constantemente o plano meramente acadêmico, tornando a mecânica estatística uma ferramenta fundamental no desenvolvimento de novas tecnologias. Como exemplo, podemos citar o desenvolvimento de novos materiais, com suas propriedades físicas e químicas cada vez mais específicas, desenvolvimento este que seria impossível sem o profundo conhecimento prévio obtido pela física estatística. Por outro lado, são numerosas as aplicações que outras áreas do conhecimento (engenharia, economia, meteorologia, entre tantas outras) têm implementado graças ao desenvolvimento da física de sistemas não lineares. Conceitos que há alguns anos pareciam meras especulações teóricas, hoje se mostram não somente importantes mas fundamentais. Problemas tão abstratos como a difusão em meios fractais são utilizados diariamente em tarefas tão concretas como estimar a capacidade de uma bacia petrolífera.

Começar sempre com modelos simples e introduzir ingredientes microscópicos mais e mais realistas parece ser a principal receita seguida (com êxito) ao longo dos anos. E neste processo, as vezes lento e tortuoso, a física estatística (em particular a teoria dos fenômenos críticos) mostrou também que a explicação de muitos comportamentos comple-

xos observados na natureza não reside necessariamente na complexidade dos componentes microscópicos em questão. Pelo contrário, unidades teóricas simples capazes de captar aqueles ingredientes relevantes na descrição dos fenômenos analisados, quando interagem entre si, ainda que de forma simples, podem dar origem a fenômenos difíceis de compreender a primeira vista, como por exemplo uma transição de fase. Neste contexto, o modelo de Ising é talvez o paradigma do modelo estatístico: apesar de sua extrema simplicidade, foi fundamental para a compreensão da teoria moderna de transições de fase.

A fim de compreender os mais variados e complicados sistemas, a mecânica estatística foi incorporando diferentes ferramentas matemáticas, as quais lhe permitiram abordar problemas nos quais, nem as unidades microscópicas, nem as interações entre elas, tinham que ser tão simples como no modelo de Ising. E quando estas ferramentas matemáticas se mostraram insuficientes, foi então necessário recorrer às simulações numéricas para o entendimento destes sistemas. O desenvolvimento de computadores cada vez mais poderosos tem sido fundamental nesta área. Hoje é possível simular a dinâmica de populações com bilhões de indivíduos, sistemas imunológicos com 10^9 anticorpos (próximo do número observado em alguns animais) e redes neuronais com mais de 10^5 neurônios.

Entre todos estes sistemas, os vidros de spin, por incluir aleatoriedade e frustração nas interações efetivas entre os momentos magnéticos, representaram talvez o maior desafio dos últimos anos. Assim como o modelo de Ising é o modelo paradigmático no estudo dos fenômenos críticos, a teoria dos vidros de spin foi um dos marcos iniciais do que hoje se conhece como estudo de sistema complexos, entendendo por estes, sistemas constituídos por muitas unidades simples que interagem de forma em geral simples, resultando num comportamento macroscópico altamente não trivial.

Existem muitos outros exemplos de interesse e é evidente que os sistemas complexos não são exclusividade da física. Unidades simples e interações não muito complexas por um lado, aleatoriedade e frustração por outro, são facilmente encontrados em muitos outros sistemas, que até pouco tempo estavam fora das principais linhas de pesquisa em física: o sistema imunológico dos vertebrados, dinâmica de populações, o problema do tráfego em uma grande cidade, o comportamento da bolsa de valores de um mercado

emergente, são apenas alguns exemplos de sistemas que cada vez mais despertam interesse na física estatística. O desafio consiste agora em explicar, não somente a partir da física microscópica, mas também, a partir dos conhecimentos obtidos por estas outras disciplinas científicas (biologia, sociologia, engenharia, economia), porque estes sistemas apresentam comportamentos cooperativos extremamente complexos (a auto-regulação no sistema imunológico, o desaparecimento súbito de populações aparentemente não ameaçadas de extinção, os congestionamentos em avenidas, as oscilações bruscas e a sensibilidade aos agentes externos das bolsas de valores).

O estudo de sistemas complexos consiste em modelar estes fenômenos e aplicar de forma sistemática as técnicas desenvolvidas ao longo de muitos anos pela física estatística e pela teoria da matéria condensada, tanto analíticas quanto numéricas. Nesta modelagem, é fundamental distinguir dentre os ingredientes do sistema aqueles que são os essenciais e que tornam o modelo tratável, sem se afastar do sistema real.

Nosso cérebro, em particular, é um exemplo de sistema extremamente complexo. Existe hoje numerosas evidências, obtidas a partir de estudos neurológicos, das quais é possível concluir que as chamadas funções superiores do cérebro (dentre elas a capacidade de memorizar de forma associativa, aprender, classificar e generalizar) não surgem da especialização das suas partes mas, pelo contrário, são produtos do trabalho coletivo de um número muito grande de células, que interagem de maneira simples transportando informações (sinais elétricos).

O estudo do cérebro tem se tornado nas últimas décadas um campo de pesquisa para diferentes áreas do conhecimento. Apesar do objetivo comum, os interesses da comunidade científica nesta área são em gerais direcionados para o que ela acredita ser importante e de interesse dentro das metodologias de suas disciplinas específicas. Estes interesses incluem desde aqueles cuja meta é entender sua anatomia, fisiologia e funcionamento, originalmente restrito as neurociências, a aqueles cujo intuito é usar as informações derivadas dessas pesquisas para desenvolver novas máquinas artificiais (computadores): máquinas capazes de reproduzir diferentes (mas não necessariamente simultaneamente) funções cerebrais. Os pesquisadores desta área começaram a usar essas máquinas para tentar simular o

comportamento lógico do cérebro. A matemática, ciências da computação e a engenharia compartilham desses interesses.

Existe também outros cujo interesse é entender o comportamento coletivo emergente das interações mútuas entre os muitos elementos simples (neurônios.). Exemplos desses comportamentos emergentes são bastante familiares para os físicos sob o nome de transição de fase (os momentos magnéticos dos spins alinham-se numa determinada direção e sentido quando a temperatura decresce, um sistema de muitos elétrons comporta-se como um supercondutor, etc.). Ferromagnetismo e supercondutividade são fenômenos coletivos que só podem ser entendidos usando as ferramentas apropriadas da mecânica estatística.

Neste contexto, a física estatística, utilizando e muitas vezes inovando as suas ferramentas desenvolvidas ao longo dos anos, ingressou em definitivo no estudo das funções do cérebro, consolidando esta nova linha de pesquisa conhecida como "Rede de Neurônios". A seguir, citaremos alguns dos trabalhos pioneiros, ao longo dos últimos 50 anos, nos quais tentaremos abstrair as características que tornaram possível o desenvolvimento desta área.

A primeira modelagem de neurônios foi projetada em 1943 por um neurobiologista Warren McCulloch e um matemático Walter Pitts. O neurônio formal, era uma espécie de válvula lógica com dois possíveis estados internos; ativo (1) ou repouso (0). Cada neurônio tem um número pequeno de entradas resultante 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 é maior que o limiar o neurônio é ativado; caso contrário, permanece no repouso.

Estas idéias foram salvas da possível obscuridade, quando os primeiros computadores foram sendo construídos. O físico Jonh Von Neumann em seus estudos sobre as possibilidades lógicas das máquinas inteligentes, se referiu as redes de McCulloch e Pitts como sendo não robustas (contrário às redes de neurônios reais). Ele sugeriu que o trabalho feito por um neurônio poderia ser compartilhado por um conjunto de neurônios cooperativos: robustez surge com a redundância. Pesquisadores começaram a olhar para a anatomia e fisiologia dos neurônios como uma possível fonte de inspiração para criar máquinas inteligentes.

Em 1949, o neurobiologista Donald Hebb (no livro *"A organização do comportamento"*), chama a atenção para a conexão entre a psicologia e a fisiologia, evidenciando que um caminho neuronal é reforçado cada vez que é usado. A "regra de Hebb", como é conhecida e usada nos dias atuais, forneceu a base fisiológica para o estudo dessas redes.

Em 1954, Gragg e Temperley, adotaram o conceito de neurônios cooperativos de Von Neumann, mas deram a essa idéia um novo significado. Eles consideraram que a atividade de uma rede de neurônio resulta do comportamento coletivo (similar ao início do ferromagnetismo, por exemplo). Esta talvez seja a origem dos últimos desenvolvimentos na teoria de redes neuronais.

No final dos anos 50, um novo dispositivo denominado de "perceptron" foi inventado por Frank Rosenblatt. Seu intento era de ilustrar algumas das propriedades dos sistemas inteligentes em geral (classificar, aprender), sem levar em conta certas condições especiais e desconhecidas, que são válidas para organismos biológicos concretos. Ele acreditava que a conectividade que se desenvolvia nas redes biológicas tinha um elevado percentual de aleatoriedade e não como nas redes de McCulloch e Pitts, nas quais se empregava lógica simbólica para analisar estruturas bastante idealizadas. Em 1969, surgiu o livro *"Perceptrons: uma introdução a geometria computacional"*, escrito por Minsky e Papert, que para algumas pessoas significou o final dos modelos de rede de neurônios. Neste livro, fazia-se uma dura crítica a esses dispositivos, em termos de suas capacidades e limitações. Apesar de ser ainda objeto de discussão nos dias atuais, muitos dos problemas que foram abordados com os perceptrons, seguem sendo válidos e serviram para implementar novos modelos como os perceptrons com multicamadas, modelos estes não afetados pelas críticas de Minsky e Papert. Nestes modelos, a informação flui numa única direção sem que haja realimentação.

Na década de 70, Little [1] foi um passo além no desenvolvimento da teoria de redes neuronais. Ele introduziu ruído nas conexões entre os neurônios. Desta forma, redes neuronais tornaram-se redes de autômatos probabilísticos.

Um dos modelos mais importante nesta área, do ponto de vista físico, foi introduzido pelo neurobiologista John Hopfield em 1982 [2] para modelar o processo de memorização

do cérebro. O *modelo de Hopfield*, como é hoje conhecido, destacou-se originalmente por supor que certas funções do cérebro podiam ser modeladas considerando o próprio cérebro como um sistema desordenado. Sua estreita relação com a física de sistemas magnéticos clássicos desordenados, onde o neurônio simplificado e as conexões sinápticas equivaleriam, na parte magnética, aos spins e aos acoplamentos magnéticos respectivamente, e a estrutura de multivales do espaço de fase encontrada em modelos de campo médio de vidro de spin [3, 4] a qual sugere uma forma adequada de modelar o processo de memorização (cada um dos mínimos de energia correspondendo a uma dada memória da rede), permitiu o uso das ferramentas matemáticas desenvolvidas pela mecânica estatística no estudo desses modelos. Assim, o problema de memorização resume-se a encontrar uma forma de garantir que as memórias que se deseja armazenar coincidam ou estejam próximas desses mínimos.

Apesar de possuir mecanismos difíceis de serem justificados do ponto de vista biológico (todos os neurônios são interconectados, podendo receber ou transmitir sinais elétricos ininterruptamente, as conexões sinápticas entre os neurônios são simétricas, entre outras), o modelo de Hopfield na sua versão original (dinâmica determinística) foi fundamental como ponto de partida na busca de modelos mais realistas. Desde a proposta original, novos ingredientes vão aos poucos sendo incorporados ao modelo a partir das informações e conhecimentos das outras áreas e a medida que se compreendem os mecanismos do processo de resposta das funções modeladas. Uma importante contribuição teórica deve-se a Amit, Gutfreund e Sarnpolinsky [5, 6]. Eles introduziram ruído no processo de transmissão das informações entre os neurônios e resolveram a *mecânica estatística* do modelo de Hopfield. Nesta analogia com os modelos magnéticos, o sistema (agora descrito por uma função Hamiltoniana e tendo na temperatura o parâmetro que controla o nível de ruído) pôde ser estudado analiticamente.

O avanço nesta linha foi marcante, nos últimos anos, com as crescentes contribuições na busca de uma maior aproximação aos sistemas biológico. Extensões e generalizações dos modelos originais, aparecem como possíveis mecanismos capazes de revelar novas propriedades, tais como reconhecimento e armazenamento de sequências temporais e ca-

tegorização (generalização).

Ingredientes como diluição (em média cada neurônio está conectado a $10^3 \sim 10^4$ outros neurônios) e assimetria nas conexões sinápticas (os acoplamentos entre os neurônios não são simétricos), são exemplos de implementações que tornaram o estudo dessas redes mais estimulante como objeto de pesquisa: por um lado modelos mais realistas podem mostrar novos comportamentos até então não revelados e, por outro, a complexidade quanto a sua tratabilidade é visto como um desafio no desenvolvimento de técnicas mais sofisticadas.

Introduzir assimetria nas conexões sinápticas, não mais permite definir uma função energia e conseqüentemente utilizar um formalismo termodinâmico. Derrida, Gardner e Zippellius (1987) [7] introduziram, de maneira aleatória, diluição e assimetria nas conexões sinápticas do modelo de Hopfield e estudaram a dinâmica.

Outra implementação importante não levada em conta no modelo original de Hopfield é o tempo de recomposição do neurônio após ter disparado um sinal elétrico. Estes dispositivos, denominados *Período Refratário Absoluto* (PRA) e *Período Refratário Relativo* (PRR) serão tratados em boa parte desta tese.

O trabalho desenvolvido nesta tese tem como objetivo introduzir e analisar os modelos de redes de neurônios, quanto a sua capacidade de armazenamento e reconhecimento de padrões e categorização (reconhecer classes), quando novos ingredientes biológicos são introduzidos na dinâmica. Este estudo está voltado em grande parte para um tratamento dinâmico da rede, que em geral, pode prover de informações com mais detalhes a respeito das suas propriedades. No capítulo 1 apresentamos numa primeira parte, de forma muito limitada, a biologia de um neurônio, suas complexas conexões com outros neurônios, seus tempos de respostas quanto à transmissão de sinais elétricos e o processo de memorização e reconhecimento. Numa segunda parte, analisamos o modelo de Hopfield [2] e sua generalização (introduzindo estocasticidade) [5, 6], apresentando seus principais resultados obtidos a partir de um desenvolvimento termodinâmico, e por fim, apresentamos um desenvolvimento dinâmico exato, onde novos ingredientes biológicos são incorporados ao modelo. Paralelamente a estes resultados analíticos, apresentamos os resultados obtidos por simulações numéricas.

Nos capítulos seguintes, apresentamos as contribuições originais desta tese. Nos capítulos 2 e 3 analisamos o efeito, sobre as propriedades de reconhecimento e armazenamento do modelo de Hopfield, ao introduzir os períodos refratários na dinâmica da rede. Nesta análise, os períodos refratários são introduzidos por meio de um limiar de ativação que depende do estado do mesmo neurônio em tempos diferentes. No capítulo 2, estudamos analiticamente, usando um desenvolvimento de campo médio, e através de simulação numérica a dinâmica do modelo de Hopfield não diluído. Apesar de simples, a aproximação de campo médio tem mostrado ser uma ferramenta importante no estudo destes modelos, reproduzindo as propriedades obtidas mediante análise termodinâmica. No estudo numérico obtivemos o diagrama de fase completo do modelo e observamos que o espaço dos parâmetros apresenta um rico cenário com, além dos atratores ponto fixo usualmente presentes no modelo padrão de Hopfield, atratores periódicos e caóticos. Este resultado é bastante interessante do ponto de vista biológico, já que atratores ponto fixo são altamente improváveis nas redes de neurônios reais. No capítulo 3 apresentamos um estudo da dinâmica do modelo de Hopfield ultra diluído com esses períodos refratários. Desenvolvemos um método aproximado que permite estudar o comportamento de longo tempo do sistema e obtivemos o diagrama de fase completo do modelo. Estes resultados foram comparados com os obtidos através da simulação numérica. É importante destacar que neste caso, estamos modelando simultaneamente três ingredientes biológicos importantes.

No capítulo 4 realizamos um estudo numérico do modelo de Hopfield ultra diluído, sujeito a condição menos restritiva introduzida por Arenzon [22] (válidas para a capacidade de armazenamento e reconhecimento de padrões), para analisar a natureza dos atratores de reconhecimento. Para isto, usamos a técnica de propagação de dano (distância de Hamming ou dano dh) entre duas configurações, inicialmente próximas uma da outra e tendo ambas uma projeção finita com um dos padrões. A análise do comportamento assintótico do dano caracteriza a natureza dinâmica destes atratores. Aplicamos também um novo campo externo h , introduzido recentemente na literatura [59] no contexto de autômato celular, que faz o papel do campo conjugado do dano e calculamos a susceptibilidade associada ao dano dh .

No capítulo 5 passamos a estudar as possibilidades de generalização do modelo apresentado no capítulo anterior. Usamos uma rede diluída assimetricamente com uma estrutura hierárquica simples das memórias, e investigamos a habilidade da rede para criar um conceito representativo de uma série de exemplos comuns apresentados a rede no processo de aprendizado. A dinâmica do modelo é resolvida exatamente e os resultados são comparados com os obtidos nas simulações numéricas. Por último, no capítulo 6, resumimos os resultados apresentados na tese e indicamos possíveis extensões destes estudos.

Capítulo 1

Modelos de Rede de Neurônios

Antes de descrevermos os modelos de redes de neurônios é importante chamar atenção para dois pontos importantes: em primeiro lugar a descrição da biologia de um neurônio aqui apresentada é uma visão muito simplificada. A neurobiologia é, obviamente, um tema muito mais complexo que a imagem aqui esboçada e para os leitores interessados neste tema, as referências [8, 9, 10] fornecem uma descrição completa e atualizada. Em segundo lugar, embora existam muitas analogias entre os conceitos básicos das neurociências e as redes neuronais, não devemos necessariamente imaginar que esses sistemas modelam sistemas nervosos reais. O nosso principal interesse é desenvolver dispositivos capazes de *reproduzir* somente algumas das funções superiores do sistema nervoso.

Vamos começar este capítulo de revisão descrevendo de maneira simplificada a biologia dos neurônios e das redes neuronais reais para, inspirados no seu funcionamento, passar a descrever alguns dos principais modelos matemáticos desenvolvidos nos últimos anos.

1.1 Descrição da Biologia do Neurônio

1.1.1 O neurônio biológico

Resumiremos aqui somente as características que serão essenciais para a construção dos modelos de redes neuronais. Esses elementos básicos são os neurônios e suas conexões sinápticas. Apesar de existir uma grande variedade de tipos de neurônios no sistema

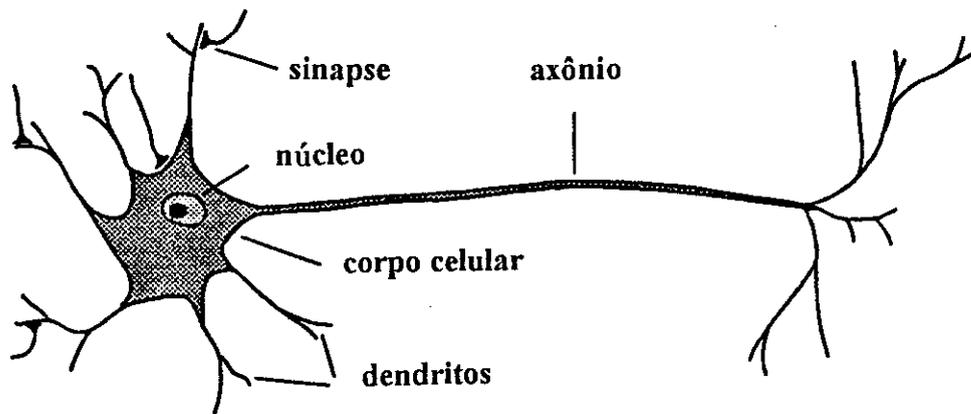


Figura 1.1: Desenho esquemático de um neurônio biológico.

nervoso — variedades no tamanho, na estrutura e na função — para nosso propósito assumiremos que todos são idênticos. Ou seja, somente estamos interessados nas propriedades comuns a todos eles e referidas ao processo de transmissão e processamento de informação. Na figura 1.1 apresentamos as componentes principais de uma célula nervosa típica: o **corpo celular** ou **soma**, que contém o núcleo, é o responsável em manter os processos metabólicos da célula e a integração dos estímulos que a ele chegam; os **dendritos**, que são ramificações do corpo celular especializadas em receber os estímulos que chegam de células sensoriais ou de outros neurônios e o **axônio**, que geralmente é uma fibra longa que nasce no corpo celular e se ramifica na outra extremidade para conectar-se com outros neurônios. Sua principal função é conduzir a informação, após ser integrada na junção corpo celular-axônio, até as extremidades da árvore axonal. De maneira simples, podemos imaginar uma rede de neurônios baseada nas divisões canônicas do seu elemento principal: a parte por onde chega os estímulos (**dendritos**), a parte de processamento (**corpo celular**) e a parte de transmissão (**axônio**).

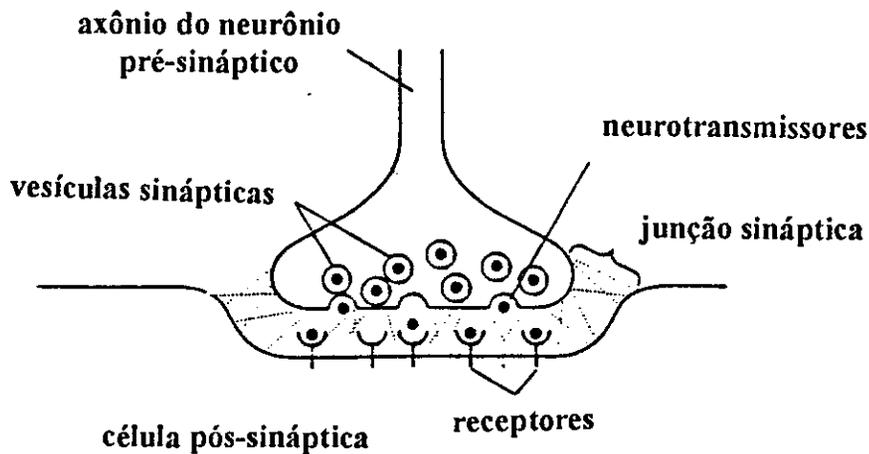


Figura 1.2: Estrutura sináptica do neurônio.

1.1.2 Comunicação biológica entre os neurônios

Em média, cada neurônio está ligado a $10^3 \sim 10^4$ outros neurônios numa intrincada rede de conexões. O contato onde ocorre a comunicação entre eles é denominado de *junção sináptica* ou *sinapse*. Estes contatos estão localizados entre as extremidades da árvore axonal do neurônio pré-sináptico e os dendritos e corpo celular do neurônio pós-sináptico. As atividades que se desenvolvem na membrana pré-sináptica tem como resultado a liberação de moléculas protéicas chamadas neuro-trasmissores. Essas moléculas difundem-se através da junção sináptica e se ligam aos canais receptores da membrana pós-sináptica alterando sua permeabilidade (ver Figura 1.2).

A membrana que recobre o neurônio separa o plasma intracelular do fluido intersticial que está fora da célula. Ela desempenha um papel importante no processo de transmissão dos sinais elétricos. Na ausência desses sinais, que chegam através dos canais receptores, mantem-se uma diferença de potencial entre o fluído intracelular e o fluído extracelular denominado *potencial de repouso*. Esta diferença de potencial deve-se à existência de diferentes concentrações de íons (aos quais a membrana celular é permeável) em ambos lados da membrana, principalmente sódio (Na^+) e potássio (K^+).

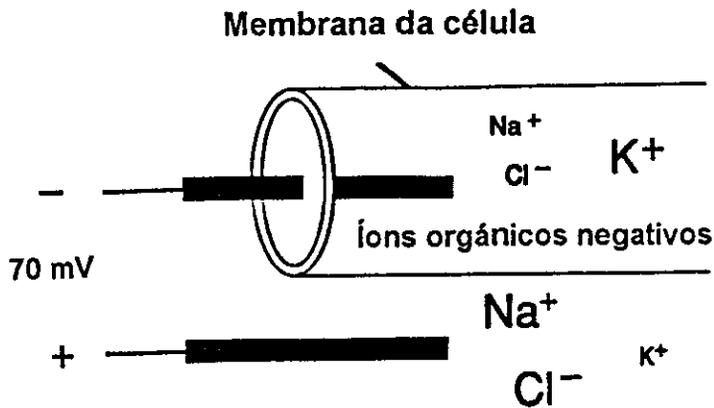


Figura 1.3: O potencial de repouso e as diferentes concentrações iônicas. Os tamanhos relativos dos símbolos das espécies iônicas representam aproximadamente a concentração relativa de cada espécie dentro e fora da célula.

Um dos mecanismos responsável pela existência desta diferença de potencial envolve os chamados íons orgânicos. Estes, ao contrário de todos os outros, não podem se difundir através da membrana por causa do seu maior tamanho. Dado que não podem então sair da célula, sua carga negativa dificulta a entrada de íons de cloro (Cl^-), resultando numa maior concentração de íons de cloro fora da célula. Outro dispositivo importante é conhecido como "bomba sódio-potássio" e é o responsável pela maior concentração de potássio dentro da célula e de sódio fora dela. Existem na membrana canais que são seletivamente mais permeáveis para íons de potássio que de sódio. O gradiente químico de potássio tende a fazer com que os íons K^+ saiam da célula por difusão, enquanto que a forte atração dos íons orgânicos tendem a mantê-los no seu interior. O resultado dessas forças opostas é que se alcança um equilíbrio no qual há mais íons de sódio e cloro fora da célula e mais íons orgânicos e de potássio dentro do corpo celular. Este equilíbrio resulta num potencial de repouso de aproximadamente -70mV , sendo mais negativo o fluído intracelular. A figura 1.3 mostra a diferença de potencial resultante das diferentes concentrações de íons que se estabelece no interior e exterior da célula.

Quando os canais receptores da membrana pós-sináptica são ativados, a ação química que se produz altera a permeabilidade da membrana para certos íons. Um fluxo de íons positivos Na^+ para dentro da célula tende a despolarizar o potencial da membrana; este efeito é *excitatório*. Se íons positivos K^+ saem do interior da célula, a membrana será hiperpolarizada; este efeito é *inibitório*. Deste modo, o efeito da atividade do neurônio pré-sináptico é excitar ou inibir a polarização da membrana do neurônio pós-sináptico. Esta flutuação do potencial da membrana com relação ao potencial de repouso é conhecida como *Potencial Pós-Sináptico* (PPS).

O efeito conjunto dos PPS que chegam ao corpo celular do neurônio pós-sináptico são somados na junção do corpo celular com o axônio e, se o resultado da integração for maior que um certo limiar, gera-se um sinal elétrico ou *Potencial de Ação* (PA). Caso contrário, nenhum sinal é produzido. Desta forma, diz-se que a atividade da junção corpo celular-axônio é um processo tudo ou nada: no primeiro caso o sinal elétrico é transmitido ao longo do axônio sendo replicado nas suas ramificações, mantendo-se quase estável quanto a forma e a amplitude de propagação. No outro caso, não há propagação do sinal ao longo do axônio.

Um ingrediente importante no funcionamento dos sistemas nervosos é a natureza aleatória dos processos relacionados à transmissão de sinais elétricos, principalmente nas junções sinápticas. Três importantes fontes de aleatoriedade associadas aos neurotransmissores são:

- o número e o tamanho dos neuro-transmissores químicos são diferentes;
- somente parte dos neuro-transmissores liberados alcançam os canais receptores da membrana pós-sináptica;
- é possível que mesmo na ausência de estímulos haja liberação de neuro-transmissores.

Como veremos nas próximas seções, existem também outras fontes de aleatoriedade. Os mecanismos responsáveis por estes processos são em muitos casos ainda pouco compreendidos ou mesmo desconhecidos pela biologia.

1.1.3 Escalas de tempo biológico

Quando um sinal elétrico se propaga ao longo do axônio, bloqueia a possibilidade de transmissão de um segundo sinal durante um certo intervalo de tempo, limitando assim a taxa de disparo dos neurônios. A frequência observada em neurônios reais varia entre 40 e 150 disparos por segundo (no cortex), dependendo da área considerada [11]. Outro dado importante na transmissão do sinal ao longo do axônio é o seu caráter unidirecional, ou seja, o sinal gerado na junção corpo celular–axônio afasta-se do mesmo e as partes que ficam para trás levam um certo tempo para se recuperarem.

Pelo menos duas escalas de tempo podem ser identificadas na comunicação entre neurônios biológicos: o ciclo temporal ($\approx 2ms$) é o tempo desde a emissão de um sinal elétrico no neurônio pré-sináptico até a criação de um PA no neurônio pós-sináptico [11]. Este ciclo inclui o tempo de transmissão do sinal elétrico ao longo do axônio, o “retardo sináptico” (que compreende a liberação e a difusão dos neuro-transmissores na separação sináptica), a criação do PPS e a formação do PA. A outra escala corresponde ao intervalo de tempo entre dois disparos consecutivos por um único neurônio. Este intervalo pode durar alguns milissegundos [11]. Após a emissão de um sinal, a membrana da célula necessita de um tempo adicional para restaurar o potencial de repouso e emitir um novo sinal. Num primeiro período, conhecido como *Período Refratário Absoluto* (PRA) e de aproximadamente $2ms$, o neurônio não pode emitir outro sinal, independente dos PPS que a ele chegam. Durante o PRA a membrana da célula encontra-se inicialmente despolarizada e abrem-se os canais de potássio (K^+). Os íons de potássio fluem para fora da célula repolarizando-a abaixo do valor de repouso ($\approx -80mV$) [12]. Seguindo este curto período, o neurônio entra em um novo regime que pode durar, em alguns casos, mais de $7ms$ [12]. Este é o *Período Refratário Relativo* (PRR). Durante o PRR o neurônio recupera parcialmente sua capacidade de emitir um sinal elétrico, só que o PPS requerido para ativar a abertura dos canais de sódio (íons de sódio penetram na célula e o potencial da membrana cresce positivamente) é maior que o limiar de repouso. Resumindo, o tempo de resposta de um neurônio pode ser separado em duas partes: o ciclo temporal do neurônio

e um limiar temporal de disparo devido aos períodos refratários.

1.1.4 Aprendizado biológico

Sabe-se hoje que os sistemas neuronais biológicos não são totalmente pré-programados geneticamente. Pelo contrário, o processo de aprendizagem que ocorre ao longo do tempo modifica de alguma forma a rede neuronal para incluir novas informações. Uma resposta elegante sobre como aprendemos, foi sugerida pelo neurofisiologista Donald Hebb [13] em 1949 no livro *Organização do Comportamento*; e é a base das novas teorias *conexionistas* sobre o funcionamento do cérebro e sua relação com a atividade mental. A idéia principal foi expressada pelo próprio Hebb da seguinte forma:

Quando o axônio de uma célula A está suficientemente próximo para excitar uma célula B e toma parte no seu disparo de forma persistente, algum processo de crescimento ou alguma mudança metabólica ocorre em uma ou ambas as células, de tal modo que a eficiência de A, como uma das células que dispara B, é aumentada.

Desde que a conexão entre os neurônios ocorre através das sinapses, é razoável supor, se a hipótese de Hebb for verdadeira, que qualquer mudança durante o aprendizado deveria ser produzida nelas. Ele acreditava, por exemplo, que a área da junção sináptica devia aumentar. Teorias mais recentes afirmam que o responsável é o crescimento da velocidade com que os neuro-transmissores são liberados da membrana pré-sináptica. Hoje sabe-se também que os chamados mecanismos de potenciação de longo termo permitem modificar a química das conexões sinápticas e são responsáveis pelo processo de memorização de curto tempo.

Apesar de não se conhecer totalmente quais os mecanismos biológicos responsáveis por estas modificações sinápticas, este mecanismo de aprendizagem aparece de uma forma ou de outra em muitos dos modelos de redes neuronais que existem na atualidade, seja em modelos de aprendizagem, memorização, generalização e outras funções cerebrais.

Neste rápido resumo temos ignorado muitos detalhes do funcionamento das células nervosas e da arquitetura das suas conexões. Também criamos a imagem simplificada de um neurônio típico. Este não é o tratamento ideal, mas é o razoável dentro do contexto da física teórica. Se o comportamento global do sistema nervoso dependesse dos detalhes finos da estrutura dos neurônios, é bem verdade que a física contribuirá muito pouco para 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. Seria por demais simples, pensar que comportamentos tão altamente complexos como os observados nos sistemas nervosos biológicos derivassem simplesmente da redundância do número de células e do grau de conectividade entre elas. Seguramente tanto a arquitetura das conexões como a especificidade dos diferentes tipos de neurônios são fundamentais para entender como o cérebro trabalha. Embora não haja modelos que reúnam como um todo esses fundamentos, seja por desconhecimento das especificidades ou devido a complexidade na sua implementação como modelo, grande parte das pesquisas atuais (e em particular o trabalho desta tese) caminham nesta direção.

1.2 Da Biologia à Modelagem das Redes Neurais

1.2.1 Um modelo protótipo

A partir das descrições anteriores, podemos relacionar um conjunto de propriedades básicas para estruturar a modelagem das redes de neurônios. Estas propriedades são:

- A transmissão da informação entre os neurônios se dá através de sinais elétricos e cada neurônio pode ser considerado como um autômato binário que pode tomar os valores $+1$ e -1 representando os estados *ativo* e *inativo* respectivamente.
- A junção sináptica pode favorecer a polarização (se for excitatória) ou a despolarização (se for inibitória).

- A resposta do neurônio aos estímulos é função da integração dos PPS que a ele chegam.
- O processo de transmissão de informação entre neurônios está submetido a diferentes fontes de ruído.
- A existência de um ciclo temporal e os períodos refratários, estabelecem uma escala de tempo padrão para a dinâmica do neurônio e um limiar temporal de disparo, respectivamente.
- A resposta do sistema como um todo a um novo estímulo reflete-se na estrutura das conexões sinápticas

Estes ingredientes simplificadores tomados da biologia do neurônio, como veremos nos próximos capítulos, são suficiente para modelar algumas das funções observadas em sistemas biológicos.

Uma rede de neurônios será então para nós um sistema de N unidades (neurônios), cada uma representada por uma variável dinâmica S_i que pode tomar os valores $+1$ ou -1 . O vetor $\vec{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))$ representa então o estado da rede num dado tempo t . Vamos por enquanto considerar só dois neurônios: o neurônio j que emite o sinal (pré-sináptico) e o neurônio i que recebe o sinal elétrico (pós-sináptico). A contribuição do neurônio j ao PPS do neurônio i pode ser escrita como:

$$V_{ij}(t) = \frac{V_{ij}}{2}(1 + S_j(t)) . \quad (1.1)$$

Levando em conta às possíveis contribuições de todos os outros neurônios da rede, o PPS total pode ser expresso da seguinte forma:

$$V_i(t) = \sum_{j \neq i}^N V_{ij}(t) = \sum_{j \neq i}^N J_{ij}(1 + S_j(t)) , \quad (1.2)$$

onde $J_{ij} = V_{ij}/2$. O J_{ij} mede a eficiência sináptica entre os neurônios ($J_{ij} > 0$ representa uma sinapse excitatória e $J_{ij} < 0$ representa um sinapse inibitória). Se dois neurônios

não estão conectados, então $J_{ij} = 0$. A emissão de um sinal pelo neurônio pós-sináptico i dependerá do PPS total (1.2) ser maior que um certo limiar próprio θ_i . Uma maneira simples de representar a dinâmica determinística é supor que a evolução temporal do sistema obedece a seguinte regra:

$$S_i(t + \Delta t) = \text{Sinal}(V_i(t) - \theta_i) = \text{Sinal}(h_i(t)), \quad (1.3)$$

onde $h_i(t)$ é denominado usualmente o campo local do i -ésimo neurônio da rede no tempo t (em analogia com modelos magnéticos):

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t) + \sum_{j \neq i}^N J_{ij} - \theta_i. \quad (1.4)$$

Se o argumento da função Sinal (1.3) for positivo ($h_i(t) > 0$), no tempo $t + \Delta t$ o neurônio i estará ativo e emitirá um sinal elétrico. Caso contrário, ($h_i(t) < 0$) nenhum sinal será emitido. O valor de Δt caracteriza então o ciclo temporal do neurônio (que vamos supor igual para todos eles).

Os modelos neuronais geralmente levam em conta dois tipos de regras dinâmicas: *paralela* ou *sequencial*. Na dinâmica paralela todos os neurônios são atualizados simultaneamente (podemos escolher então $\Delta t = 1$ como a unidade temporal de atualização), enquanto que na dinâmica sequencial escolhe-se um único neurônio por vez para ser atualizado ($\Delta t = 1/N$). A atualização em paralelo, além de ser mais conveniente para realizar grandes simulações, é também aparentemente mais realista do ponto de vista biológico. Os rápidos processos de resposta do cérebro às funções vitais relacionadas aos órgãos do sentido por exemplo, só são possíveis graças a bilhões de neurônios trabalhando simultaneamente.

Dada uma função cerebral que se deseja modelar, o problema agora consiste em definir:

1. uma dada arquitetura da rede e
2. um conjunto de acoplamentos sinápticos $\{J_{ij}\}$,

de forma tal que a dinâmica introduzida (1.3) permita associar configurações iniciais (entradas) com atratores do sistema (saídas).

1.2.2 Refinamento do modelo: processos aleatórios

Um elemento importante que deve ser levado em conta na modelagem das redes neurais é à aleatoriedade envolvida na transmissão de sinais entre os neurônios. Além dos descritos anteriormente, relacionados aos neuro-transmissores, outros processos aleatórios ocorrem como por exemplo, flutuações no tempo de resposta de cada neurônio, variações nos limiares de disparo e outros. Não é simples identificá-los ou mesmo descrevê-los adequadamente. Afortunadamente isto é simplificado pelo fato que todos eles serão refletidos no campo local de cada neurônio. Dado que cada um recebe sinais de um número grande de neurônios, uma maneira simples de representar esse efeito é supor que o campo local é uma variável aleatória cuja distribuição de probabilidade é descrita por uma função gaussiana,

$$P_r(h) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(h-h_i)^2}{2\sigma^2}\right], \quad (1.5)$$

onde σ é a largura da distribuição associada às diferentes fontes de aleatoriedade (ruído) e h_i é o campo local médio do i -ésimo neurônio dado por (1.4). A probabilidade para o neurônio i disparar um sinal elétrico é:

$$P_r(S_i = 1) \equiv P_r(h > 0) = \int_0^\infty dh P_r(h). \quad (1.6)$$

Substituindo (1.5) em (1.6) e introduzindo a função erro, definida como

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_0^x dt \exp(-t^2), \quad (1.7)$$

a probabilidade torna-se:

$$P_r(S_i = 1) \equiv P_r(h > 0) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{h_i}{\sqrt{2}\sigma}\right) \right]. \quad (1.8)$$

De maneira análoga,

$$P_r(S_i = -1) \equiv P_r(h < 0) = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{h_i}{\sqrt{2}\sigma} \right) \right]. \quad (1.9)$$

Dado que a função erro é ímpar, podemos representar a distribuição de probabilidade na forma:

$$P_r(S_i) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{S_i h_i}{\sqrt{2}\sigma} \right) \right]. \quad (1.10)$$

A expressão acima pode ser aproximada com uma margem de erro da ordem $\approx 1\%$, por

$$P_r(S_i) = \frac{1}{2} \left[1 + \tanh \left(\frac{S_i h_i}{\sqrt{2}\sigma} \right) \right] = \frac{1}{2} [1 + \tanh(\beta S_i h_i)]. \quad (1.11)$$

onde β é o parâmetro de controle do ruído.

Uma expressão idêntica à equação acima (com modificações apropriadas no significado das variáveis), descreve a dinâmica de um spin tipo Ising em contato com um reservatório térmico à temperatura finita $T = \beta^{-1}$. Para fazer um paralelo entre o nosso modelo protótipo aleatório de rede neuronal e os modelos magnéticos, basta associar o estado do neurônio com o momento magnético, o campo local e as conexões sinápticas com o campo efetivo e os acoplamentos magnéticos respectivamente e por último, o parâmetro de ruído com a temperatura.

Dado que é necessário modelar funções sinápticas excitatórias e inibitórias, é preciso incluir acoplamentos positivos e negativos. Isto, junto à complexidade da arquitetura de conexões produz como resultado, geralmente, o surgimento de *frustração* entre os acoplamentos. Por outro lado, como estamos interessados nas propriedades estatísticas destes modelos, imaginamos que a matriz de acoplamento J_{ij} é definida por uma dada distribuição de probabilidade para os seus elementos, que serão *variáveis aleatórias*. A analogia entre o modelo protótipo de rede neuronal e o modelo (também protótipo) de Ising para sistemas magnéticos, somado aos dois elementos acima descritos, *frustração* e *aleatoriedade* põe em evidência a estreita relação entre os modelos de vidro de spin tipo Ising e as redes neuronais.

No caso de vidros de spin o problema pode ser descrito da seguinte forma: dada uma topologia da rede, a sua dimensão e um conjunto de acoplamentos magnéticos, desejamos conhecer o comportamento assintótico ($t \rightarrow \infty$) do sistema. No caso de redes neuronais o problema é exatamente o inverso. Em geral desejamos associar a cada configuração inicial (ou entrada do sistema) um dada configuração final (ou saída do sistema). Nos perguntamos então qual é a arquitetura de conexões e quais os valores da matriz de acoplamentos sinápticos que permitem implementar a associação desejada.

Dentre as diferentes arquiteturas, duas se destacaram pelas suas importâncias com relação as funções modeladas. A primeira é uma extensão natural do neurônio de McCulloch e Pitts e sua estrutura é formada por uma ou mais camadas de neurônios, nas quais a informação flui numa única direção preferencial. Este tipo de arquitetura é conhecida em geral como *perceptron*. Ele tem sido muito utilizado na modelagem do processo de aprendizagem (ou seja, como dispositivo capaz de generalizar uma dada regra de associação entre entradas e saídas a partir de um dado conjunto de exemplos) e de classificação. A outra arquitetura é a chamada *redes de neurônios atratoras* (*attractor neural networks*) e tem sido mais utilizada na modelagem dos processos ligados à memorização. Nesta tese vamos usar somente esta última arquitetura e possíveis variações de modo a torná-la mais realista do ponto de vista biológico.

1.3 Memórias Associativas: O Modelo de Hopfield

1.3.1 A natureza da memória associativa e a analogia física

Relembrar alguma coisa consiste em associar um conceito com alguma apreensão sensorial. Por exemplo, a foto de uma pessoa conhecida ou de um lugar visitado faz relembrar uma série de acontecimentos ou recobrar imagens memorizadas ao longo do tempo. Este paradigma pode ser entendido supondo que coleções básicas de dados, ordenados e interligados de alguma maneira, são armazenados na forma de padrões (memórias) no cérebro. No exemplo acima, existe uma coleção de memórias associadas à pessoa e outra associada

com o lugar visitado. Uma outra maneira de colocar o problema da memória associativa é apresentar por exemplo, uma letra parcialmente alterada ou a caricatura de um personagem conhecido. Num processo rápido, todo o padrão é recobrado ou associado a ele: no caso da letra, sua arquitetura completa (forma) e, da caricatura, a fisionomia real do personagem. Note que neste caso, somente parte do padrão foi usado como apreensão e todo ele é restabelecido.

Os computadores convencionais podem realizar estas funções de uma maneira muito limitada. Um software típico para tal é conhecido usualmente como "banco de dados". A apreensão sensorial é normalmente uma palavra, sequência de palavras ou números. Por exemplo, o catálogo de uma livraria é um conjunto de dados que são armazenados por autores, títulos, data de edição etc. Para se ter sucesso na pesquisa desejada os dados de entrada têm que ser totalmente especificados e completos. Se a informação estiver fragmentada ou mesmo uma letra permutada, não existe mecanismos para corrigir e recuperar a informação correta.

O modo pelo qual as redes de neurônios podem corrigir e recuperar uma informação fragmentada ou imperfeita, surgiu ao associar o processo de memorização ao problema de minimizar dinamicamente uma função de muitos mínimos, e associar estes mínimos com as memórias armazenadas. Esta analogia permite, em alguns casos, visualizar a operação da rede como a dinâmica de um sistema físico cujo comportamento é descrito em termos de uma função energia. Vamos imaginar as memórias como sendo os pontos de mínimos de uma superfície de energia formada de vales e barreiras (ver figura 1.4). Cada um dos mínimos (M_1 , M_2 e M_3) corresponde a uma dada memória da rede e a largura entre as barreiras de cada vale representa a "bacia de atração" da memória correspondente. Uma informação parcial ou fragmentada (ponto P da figura 1.4) cuja apreensão sensorial está relacionada com a memória M_1 , deslizará sobre a superfície até atingir, depois de um certo tempo, esta memória. Deste modo, apreensões sensoriais (estímulos) localizados na bacia atratora de uma dada memória reconhecerão a mesma memória. A este fato dá-se o nome de *reconhecimento associativo*.

A esquematização metafórica da figura 1.4 não tem que ser tão simples assim. Uma

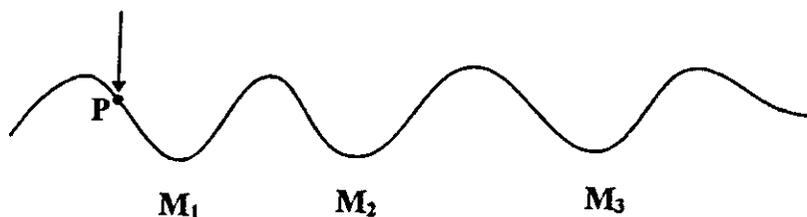


Figura 1.4: Representação esquemática de uma superfície de energia. Os mínimos correspondem as memórias armazenadas.

representação mais completa teria que envolver, além dos mínimos de energia M_1 , M_2 e M_3 associados às memórias, também mínimos espúrios que permitissem ao sistema não associar corretamente (responder por exemplo que não sabe a resposta correta ou até se confundir entre diferentes respostas).

O grande mérito de John Hopfield foi encontrar uma arquitetura muito simples e um conjunto de acoplamentos sinápticos $\{J_{ij}\}$ que lhe permitiu definir o problema de memorização em termos da minimização de uma função tipo energia.

A arquitetura por ele idealizada supõe que cada neurônio recebe e envia sinais de todos e para todos os outros neurônios. Ainda mais, o sinal entre dois neurônios quaisquer é o mesmo em ambas as direções ($J_{ij} = J_{ji}$) e o processo de atualização era sequencial.

A regra de armazenamento ou prescrição sináptica baseia-se no princípio enunciado por Hebb. Vamos considerar um conjunto de p configurações que desejamos armazenar (chamaremos a estas configurações de *padrões* ou *memórias* indistintamente): $\vec{\xi}_i^\mu = (\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu)$ com $\mu = 1, 2, \dots, p$. Então o acoplamento sináptico entre os

neurônios i e j será dado por

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad J_{ii} = 0. \quad (1.12)$$

Hopfield mostrou que com esta regra sináptica, a equação (1.3) permitia o armazenamento associativo de memórias desde que:

- as memórias sejam configurações aleatórias não correlacionadas, ou seja, escolhidas com uma distribuição de probabilidade da forma:

$$P(\{\xi_i^\mu\}) = \prod_{i=1}^N \prod_{\mu=1}^p P_r(\xi_i^\mu) \quad (1.13)$$

onde os ξ_i^μ podem tomar os valores ± 1 com igual probabilidade, ou seja:

$$P_r(\xi_i^\mu) = \frac{1}{2} \delta(\xi_i^\mu - 1) + \frac{1}{2} \delta(\xi_i^\mu + 1); \quad (1.14)$$

- o limiar de ativação θ_i tenha a forma

$$\theta_i = \sum_{j \neq i}^N J_{ij}, \quad (1.15)$$

com o qual o campo local ganha a forma simplificada

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t). \quad (1.16)$$

Notemos que esta regra é aditiva e local no sentido de que ao acrescentar uma nova memória todos os acoplamentos sinápticos são alterados em:

$$\Delta J_{ij} = \frac{1}{N} \xi_i^{\mu+1} \xi_j^{\mu+1} \quad (1.17)$$

e esta modificação só depende dos valores das memórias nos dois neurônios envolvidos na sinapse.

Usando o fato de que os acoplamentos são simétricos $J_{ij} = J_{ji}$ e que $J_{ii} = 0$ (não forma sinapse com ele mesmo), pode-se mostrar que a função energia, definida como

$$H = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j, \quad (1.18)$$

é uma função de Liapunov para a dinâmica do modelo. Na próxima seção veremos como isto permite desenvolver um formalismo termodinâmico do problema.

Os atratores do sistema podem ser obtidos monitorando a evolução temporal das superposições entre o estado do sistema e as memórias armazenadas, definidas como:

$$m_\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i(t) \quad \text{com } \mu = 1, 2, \dots, p. \quad (1.19)$$

Esta quantidade mede a proximidade do estado do sistema com a memória ξ_i^μ . Se o atrator coincide com uma dada memória, por exemplo a ν -ésima, então $m_\nu = 1$ e todas as outras serão próximas de zero ($\mathcal{O}(N^{-1/2})$) por serem as memórias não correlacionadas. Quanto mais próximo da memória está o atrator, melhor será o reconhecimento. No limite em que o atrator está descorrelacionado de todas as memórias, todas as superposições serão nulas e o sistema não consegue memorizar.

Uma das propriedades que mede a qualidade da rede é sua capacidade máxima de armazenamento α_c definida como o máximo valor da constante de proporcionalidade entre o número de memórias e o número de conexões por neurônio $\alpha = p/(N-1) \approx p/N$ para o qual o sistema ainda reconhece.

Os principais resultados observados originalmente por J. Hopfield foram os seguintes:

- os padrões armazenados e suas inversa são pontos fixos da dinâmica;
- as bacias atratoras dos padrões armazenados são grandes, ou seja, ainda que começando com uma configuração cuja superposição inicial com a memória é muito pequena, o sistema reconhece;
- existem muitos atratores espúrios;

- para uma rede com 100 neurônios, o número máximo de padrões armazenados e recuperados foi em torno de 13 (com muito erro no reconhecimento quando α é próximo ao valor máximo), ou seja $\alpha_c \simeq 0.13$.

A pesar do seu grande valor como modelo protótipo na modelagem neuronal o modelo de Hopfield tem sérias limitações, algumas das quais são:

- os acoplamentos são simétricos ao contrário do que acontece nos sistemas biológicos;
- os neurônios são totalmente conectados, sendo que em sistemas reais a conectividade é alta mais muito menor que o número de neurônios;
- a capacidade máxima de armazenamento é relativamente baixa;
- existem estados espúrios que atrapalham o mecanismo de reconhecimento.

1.4 Modelo de Hopfield generalizado

1.4.1 Termodinâmica do modelo de Hopfield

Em 1985 D. Amit, H. Gutfreund e H. Sompolsky [5] propuseram uma generalização do modelo de Hopfield a fim de incluir nele os diferentes elementos estocásticos envolvidos na transmissão de sinais. A dinâmica determinista (1.3) foi substituída por uma dinâmica estocástica análoga à dinâmica Monte Carlo banho térmico comumente utilizada na simulação numérica de modelos tipo Ising. Segundo esta regra, a cada passo de tempo um spin é atualizado com a seguinte probabilidade:

$$S_i(t+1) = \begin{cases} +1 & \text{com probabilidade} & \frac{1}{2}[1 + \tanh \beta h_i(t)] \\ -1 & \text{com probabilidade} & \frac{1}{2}[1 - \tanh \beta h_i(t)] \end{cases} \quad (1.20)$$

Dado que H definido em (1.18) é uma função de Liapunov, é possível extrair o comportamento assintótico do sistema a partir de um cálculo termodinâmico. Se β^{-1} é interpretado

como uma medida do ruído do sistema, pode-se imaginar a rede como em contato com um reservatório a temperatura fixa $(\beta \times k_B)^{-1}$. Eles utilizaram um formalismo análogo ao utilizado no estudo de modelos campo médio de vidros de spin tipo Ising e resolveram a termodinâmica do modelo.

Quando o número de padrões armazenados cresce linearmente com o tamanho da rede ($p = \alpha N$) a análise requer o uso do método das réplicas [3]. O estado de equilíbrio é governado pelos mínimos da função energia livre que surgem das seguintes equações ponto de sela:

$$\begin{aligned}
 f(\beta) &= \frac{\alpha}{2} + \frac{1}{2} \sum_{\mu} (m_{\mu})^2 + \frac{\alpha}{2\beta} \left[\ln(1 - \beta - \beta q) - \frac{\beta q}{1 - \beta - \beta q} \right] \\
 &\quad + \frac{1}{2} \alpha \beta r (1 - q) - \frac{1}{\beta} \langle \langle \ln 2 \cosh \beta \left(\sqrt{\alpha r} z + \sum_{\mu} m_{\mu} \xi^{\mu} \right) \rangle \rangle \\
 m_{\mu} &= \langle \langle \xi^{\mu} \tanh \beta \left(\sqrt{\alpha r} z + \sum_{\mu} m_{\mu} \xi^{\mu} \right) \rangle \rangle \\
 q &= \langle \langle \tanh^2 \beta \left(\sqrt{\alpha r} z + \sum_{\mu} m_{\mu} \xi^{\mu} \right) \rangle \rangle \\
 r &= \frac{q}{(1 - \beta - \beta q)^2}.
 \end{aligned} \tag{1.21}$$

onde $\langle \langle \dots \rangle \rangle$ corresponde a média sobre o conjunto dos ξ^{μ} (padrões condensados) e sobre o ruído gaussiano z . q é o parâmetro Edwards-Anderson do modelo [4] e r é o parâmetro AGS (Amit, Gutfreund, Sampaolinsky) [6] identificado como a média do quadrado da superposição do sistema com os padrões não condensados.

Para o caso particular $\alpha = 0$ (ou seja para p finito e no limite termodinâmico $N \rightarrow \infty$) eles mostraram [5] que:

- acima de $T_c = 1$ os únicos estados termodinamicamente estáveis têm superposição nula com os padrões ($m = 0$) (fase paramagnética);
- em $T_c = 1$ o sistema sofre uma transição de segunda ordem da fase desordenada para a fase de reconhecimento (só uma memória e reconhecida);

- no intervalo $0 \leq T \leq 1$ as soluções de reconhecimento são as de menor energia. Apesar de existirem outras soluções, o sistema não apresenta nenhuma outra transição de fase termodinâmica;
- para $T > 0.46$ somente as soluções de reconhecimento e suas inversas são estáveis. Apesar do sistema reconhecer as memórias com muito erro ($m \neq 1$) o ruído térmico impede o surgimento dos estados espúrios;
- para $T < 0.46$ surgem estados meta-estáveis simétricos (espúrios), ou seja, que tem a mesma superposição com um sub-conjunto das memórias e superposição nula com os restantes. Os estados simétricos com um número ímpar de padrões são estáveis enquanto que com um número par são sempre instáveis. Quando T decresce a zero o número de estados espúrios cresce;
- as bacias de atração das memórias são grandes. Um estado inicial com superposição maior que $\mathcal{O}(1/\sqrt{N})$ com um único padrão e superposição da ordem $1/\sqrt{N}$ com os demais, flui rapidamente para este padrão;
- no limite $T \rightarrow 0$ o modelo determinista original de Hopfield é reobtido.

O principal resultado obtido para o caso mais genérico $\alpha \neq 0$ e na aproximação de réplica simétrica é que, para $T \leq 1$, existe um valor crítico $\alpha_c(T)$ para a capacidade de armazenamento da rede. Em particular, para $T = 0$ eles acharam $\alpha_c \simeq 0.138$ (próximo do valor achado por Hopfield numericamente). Para $\alpha < \alpha_c$ a rede sempre reconhece as memórias armazenadas e a superposição com uma das memórias é próxima de 1. Em α_c , o erro no reconhecimento é de aproximadamente 1.5% e m passa descontinuamente de um valor $m \simeq 0.97$ para 0 ($\alpha > \alpha_c$).

Na figura 1.5 apresentamos o diagrama de fase T - α . A área sombreada corresponde as regiões de reconhecimento do sistema ($m \neq 0$, regiões A e B). Fora da área sombreada não existe reconhecimento ($m = 0$, regiões C e D). As regiões C e D correspondem as fases de vidro de spin ($m = 0$ e $q \neq 0$) e paramagnética ($m = 0$ e $q = 0$) respectivamente. As soluções vidro de spin ($q \neq 0$) existem para todos os valores de α (regiões A, B e C). Na

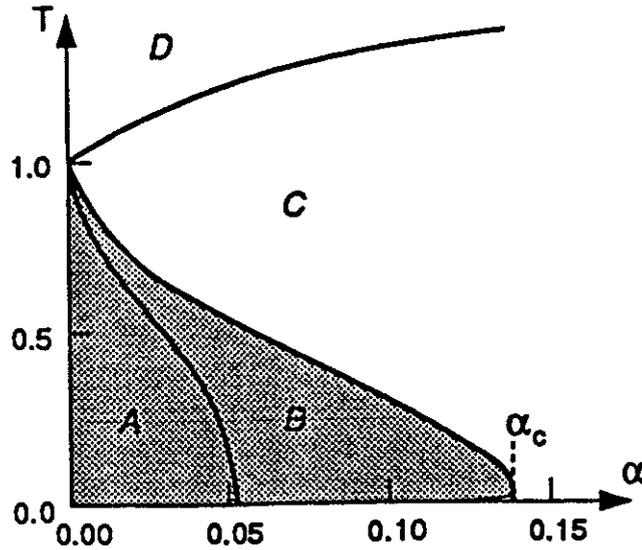


Figura 1.5: Diagrama de fase T - α . A área sombreada corresponde a região de reconhecimento da rede (A e B). Fora da área sombreada existem a fase de vidro de spin (região C) e a fase paramagnética (região D).

região A as soluções de reconhecimento são os mínimos globais da energia enquanto que em B, esses mínimos correspondem às soluções vidro de spin. A transição na linha crítica, ao longo da fronteira da área sombreada, ocorre descontinuamente e essa descontinuidade torna-se mais suave a medida que o ruído térmico cresce. Para $\alpha = 0$ a transição é de segunda ordem a $T_c = 1$. Aumentando o ruído térmico as soluções de reconhecimento e vidro de spin são desestabilizadas e a linha de transição para a fase paramagnética é de segunda ordem. Para pequenos valores de α e T existem também as soluções simétricas.

Amit et al. [6] realizaram simulações numéricas Monte Carlo a temperatura nula para redes até 3000 neurônios. Eles obtiveram uma estimativa para a capacidade de armazenamento da rede $\alpha_c = 0.145 \pm 0.01$. A pequena diferença no valor de α obtido do tratamento analítico ($\alpha \simeq 0.138$), pode ser atribuído a restrição da aproximação de réplica simétrica. Crisanti et al. [14], usando uma generalização do método de Parisi [15], mostraram que com um passo de quebra de simetria, para $T = 0$, existe uma região $0.138 < \alpha < 0.144$ na qual os estados de equilíbrio termodinâmico são estados de reconhecimento. Além da concordância com os resultados numéricos [6], a quebra de simetria de réplica melhora o

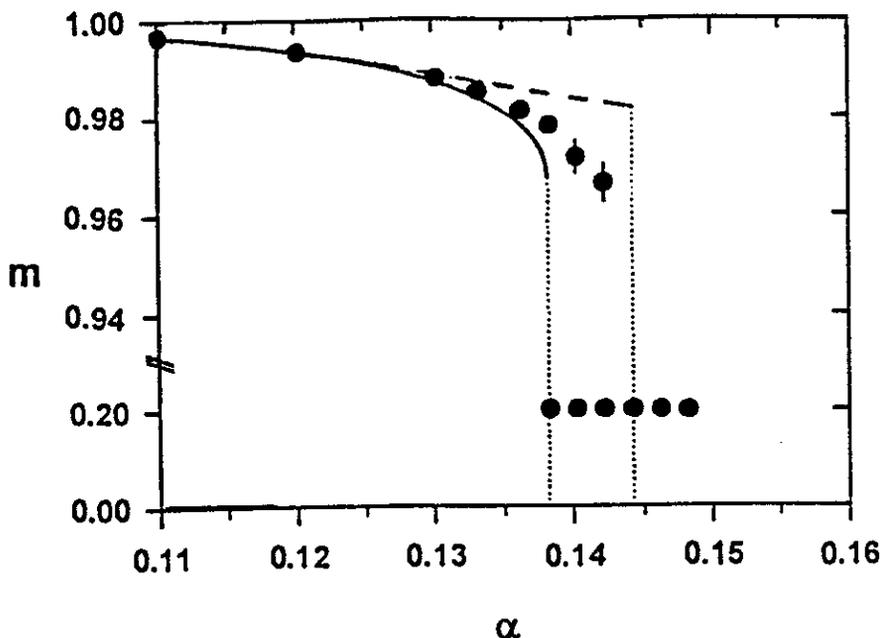


Figura 1.6: A superposição m versus α a $T = 0$ na região de quebra de simetria de réplica. Os círculos cheios são os resultados da simulação para $N \simeq 34000$ neurônios. A linha cheia corresponde a solução das equações de campo médio com aproximação de réplica simétrica e a linha tracejada corresponde a solução analítica com um passo de quebra de simetria.

reconhecimento: em α_c o erro no reconhecimento decresce de 1.5% para 0.9%. Kohring [16] usando uma técnica mais eficiente, introduzida por Pena e Oliveira [17], realizou simulações numéricas na região de quebra de simetria de réplica para redes de tamanho maior que 30000 neurônios (ver figura 1.6). Ele encontrou que o parâmetro de ordem m nesta região tem dois valores estáveis distintos, não previsto pela teoria de quebra de simetria: um com superposição $m \simeq 1$ e outro com $m \simeq 0.2$. A capacidade máxima de armazenamento obtida na simulação foi $\alpha_c = 0.143 \pm 0.001$.

A figura 1.6 mostra os resultados a $T = 0$ da simulação para $N \simeq 34000$ (círculos cheios) e das soluções analíticas obtidas pelo método de réplicas: a linha cheia corresponde a solução para réplica simétrica [6] e a linha tracejada corresponde a solução com um passo de quebra de simetria [14].

Os resultados analíticos descritos acima correspondem as propriedades termodinâmicas do modelo de Hopfield quando os estados do sistema são atualizados sequencialmente. Para os processos de atualização em paralelo, introduzido por Little [1], Peretto [18] mostrou

que para redes do 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. [5] para um número finito de padrões ($\alpha = 0$) e por Fontanari e Köberle [19] para um número extensivo de padrões ($\alpha \neq 0$). Em ambos os casos as propriedades de equilíbrio são similares as do modelo com atualização sequencial, com exceção do último, onde outras particularidades estão presentes como por exemplo, a existência de atratores periódicos de ciclo dois.

1.4.2 Dinâmica do modelo de Hopfield

O processo dinâmico no reconhecimento de memórias é, em geral, mais difícil de se tratar que as propriedades de equilíbrio. A princípio, um tratamento rigoroso necessitaria introduzir um número muito grande de parâmetros para descrever a evolução temporal do sistema. Isto implica que num tratamento analítico da dinâmica, quase sempre envolve algumas aproximações. Mesmo assim, a investigação dinâmica é sempre importante, pois oferece informações válidas não somente para regimes assintóticos, mas também para tempos curtos e intermediários, onde novos efeitos podem ser observados.

Um modelo particular, para o qual a dinâmica pôde ser resolvida exatamente, é o caso do modelo de Hopfield com diluição e assimetria nas sinapses. Este problema foi estudado por B. Derrida, E. Gardner e A. Zippelius [7] em 1987. Vamos passar então a descrevê-lo.

Vamos supor que a configuração inicial do sistema, no tempo $t + 1$, tem superposição não nula com uma única memória, por exemplo a primeira ($\mu = 1$), que chamaremos de $m(t + 1)$ e superposição nula com as restantes ($\mu \neq 1$)

$$m_{\mu}(t + 1) = \begin{cases} m(t + 1) & \text{se } \mu = 1 \\ \mathcal{O}(1/\sqrt{N}) & \text{se } \mu \neq 1, \end{cases} \quad (1.22)$$

onde a superposição $m_\mu(t+1)$ está definida por:

$$m_\mu(t+1) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i(t+1) \rangle . \quad (1.23)$$

A média $\langle \dots \rangle$ é sobre o ruído térmico a temperatura T_0 e sobre as condições iniciais. A dinâmica do sistema é governada por um processo Monte Carlo banho térmico com atualização paralela, isto é, a cada passo de tempo todos os estados da rede são atualizados simultaneamente segundo a regra:

$$P_r(S_i) = \frac{1}{2} \left[1 + \tanh\left(\frac{S_i h_i}{T_0}\right) \right] . \quad (1.24)$$

O campo local sobre o i -ésimo neurônio pode ser escrito na forma:

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t) \quad (1.25)$$

e os acoplamentos sinápticos J_{ij} são dados pela expressão:

$$J_{ij} = C_{ij} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu . \quad (1.26)$$

Os ξ_i^μ 's são variáveis aleatórias independentes que valem ± 1 com igual probabilidade (1.14). O ξ_i^μ representa o estado do i -ésimo neurônio no μ -ésimo padrão armazenado ($\mu = 1, \dots, p$). Os C_{ij} são variáveis aleatórias independentes que representam as diluições nos acoplamentos e são escolhidas com a seguinte distribuição de probabilidade:

$$P_r(C_{ij}) = \frac{C}{N} \delta(C_{ij} - 1) + \left(1 - \frac{C}{N}\right) \delta(C_{ij}) . \quad (1.27)$$

Dado que os C_{ij} pode valer somente 1 e 0 representando a existência ou não da conexão entre os neurônios e que os C_{ij} e C_{ji} são independentes, pode acontecer que ambos sejam diferentes. Isto quer dizer que a própria diluição introduz a assimetria nas sinapses. O C é a conectividade média por neurônio da rede. Estes dois ingredientes, *diluição* e

assimetria, são importantes do ponto de vista de redes de neurônios reais, e não estão presentes no modelo original de Hopfield.

Substituindo a equação (1.26) em (1.25), e estabelecendo que após a diluição somente k neurônios estarão conectados ao neurônio i , representados por $\{S_{j_1}, S_{j_2}, \dots, S_{j_k}\}$ ($r = 1, 2, \dots, k$), podemos reescrever o campo local na forma:

$$h_i(t) = \sum_{j_r \neq i}^k \sum_{\mu=1}^p \xi_i^\mu \xi_{j_r}^\mu S_{j_r}(t). \quad (1.28)$$

Separando na soma sobre μ o termo $\mu = 1$ (padrão a ser reconhecido) dos demais, obtemos:

$$h_i(t) = \sum_{j_r \neq i}^k \xi_i^1 \xi_{j_r}^1 S_{j_r}(t) + R_{ij}^\mu(t), \quad (1.29)$$

onde R_{ij}^μ , que corresponde as memórias não correlacionadas com o estado do sistema, é dado pela seguinte expressão:

$$R_{ij}^\mu(t) = \sum_{j_r \neq i}^k \sum_{\mu>1}^p \xi_i^\mu \xi_{j_r}^\mu S_{j_r}(t). \quad (1.30)$$

Como queremos calcular a superposição do estado do sistema com a primeira memória, de (1.23) e (1.24) obtemos a seguinte expressão para o $m(t+1)$:

$$m(t+1) = \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh\left(\frac{h_i(t)}{T_0}\right), \quad (1.31)$$

onde $h_i(t)$ está dado acima.

É possível provar [20, 21] que se C satisfaz a condição $C \ll \ln N$, onde N é o número de neurônios, as correlações temporais entre os diferentes neurônios da rede são desprezíveis e a desordem congelada (quenched) pode ser tratada como uma desordem recozida (annealed). Deste modo, no limite $N \rightarrow \infty$ não existe circuito neuronal fechado e, para qualquer tempo t dois neurônios são independentes com probabilidade 1. Ou seja, cada neurônio possui uma árvore temporal própria de ancestrais, independentemente das

árvores dos demais. A condição $C \ll \ln N$ é conhecida como *extrema diluição* e o modelo como *modelo de Hopfield ultra diluído*.

A descrição acima significa que para qualquer tempo t todos os $S_{j_r}(t)$ são variáveis aleatórias independentes, e que o termo R_{ij}^μ , no limite termodinâmico, pode ser representado por uma distribuição gaussiana com média nula e variância σ^2 . No limite C e $p \rightarrow \infty$ com $\alpha \equiv p/C = \text{constante}$, a dinâmica do modelo ultra diluído é exata. Note que a definição $\alpha \equiv p/C$ está de acordo com a adotada anteriormente já que C é a conectividade média por neurônio: no caso não diluído $C = N$ (cada neurônio está conectado a todos os outros).

A equação de recorrência para a superposição do estado do sistema com o padrão $\mu = 1$ (ver descrição detalhada no apêndice A) é:

$$m(t+1) = \int Dz \tanh \beta \left(m(t) + z\sqrt{\alpha} \right), \quad (1.32)$$

onde $T = \beta^{-1}$ é a temperatura generalizada $T = T_0/C$ e

$$Dz = \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}}.$$

Na figura 1.7 apresentamos o diagrama de fase $T-\alpha$ do modelo. Na linha crítica, a transição da fase de reconhecimento ($m \neq 0$) para a fase paramagnética ($m = 0$) é de segunda ordem e na ausência de ruído ($T = 0$) a capacidade de armazenamento da rede $\alpha_c(T = 0) = 2/\pi \simeq 0.637$, é bem superior ao modelo de Hopfield [2].

Apesar da condição de extrema diluição ser pouco realista e impossível de implementar, seja em um dispositivo ou em uma simulação, estudos numéricos a $T = 0$ realizados por Arenzon e Lemke [22] mostraram que estas equações são válidas, no que se refere às propriedades de reconhecimento, sob a condição menos restritiva $C \ll N$. A figura 1.8 mostra a superposição $m(t)$ versus α a $T = 0$ para três diferentes condições do sistema [22]: a linha cheia corresponde ao caso de extrema diluição ($C \ll \ln N$) com N , C e $p \rightarrow \infty$ (equação 1.32), os círculos cheios correspondem a solução analítica para $N \rightarrow \infty$

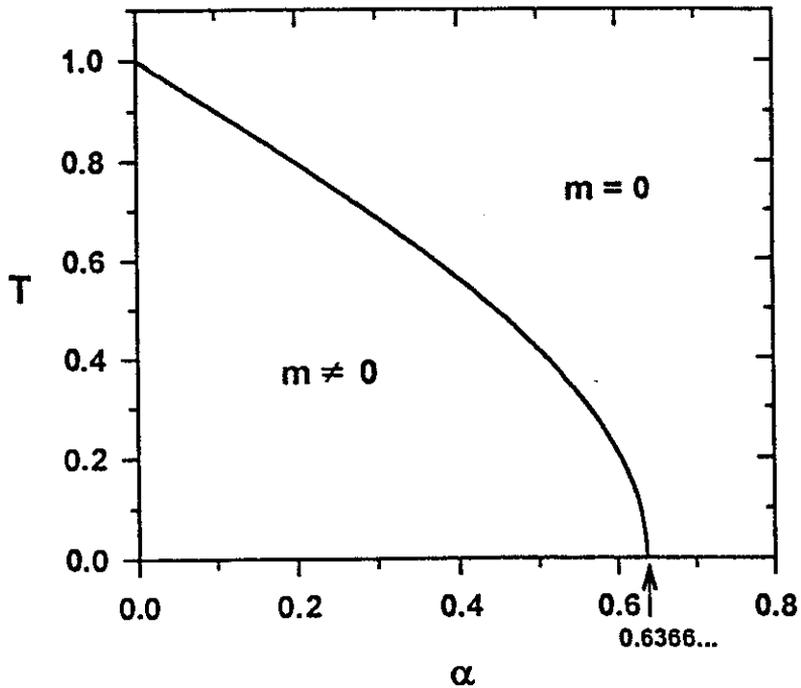


Figura 1.7: Diagrama de fase T - α do modelo de Hopfield ultra diluído.

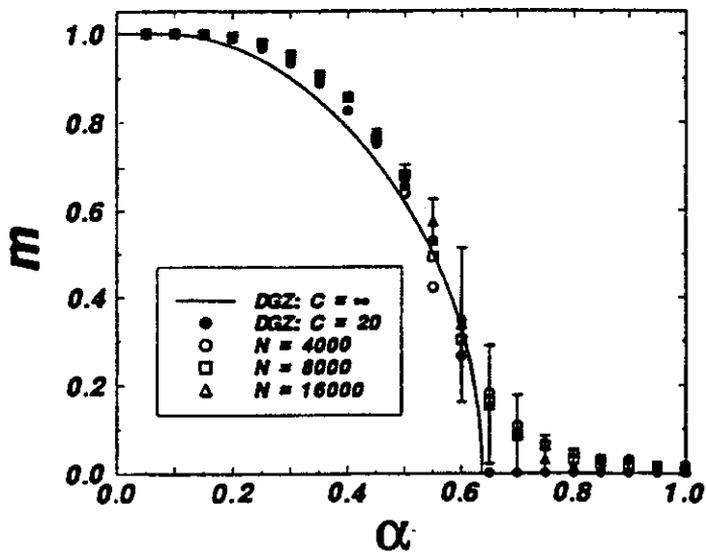


Figura 1.8: Superposição $m(t)$ vs. α . Comparação entre três diferentes condições do modelo de Hopfield diluído assimetricamente. Linha cheia: resultados analíticos para N , p e $C \rightarrow \infty$ e $C \ll \ln N$; círculos cheios: $N \rightarrow \infty$, conectividade finita ($C = 20$); símbolos vazios: simulação numérica para $N = 4000$, 8000 e 16000 e conectividade finita ($C = 20$).

e $C = 20$ (finito) e os símbolos vazios são o resultado da simulação para $C = 20$ e $N = 4000, 8000$ e 16000 . O relaxamento da condição de extrema diluição ($C \ll \ln N$) para ($C \ll N$) parece estar de acordo com os resultados da simulação .

Resumindo, neste capítulo apresentamos os mecanismos básicos necessários para modelar as redes neuronais, bem como os principais resultados do modelo de Hopfield, relacionados com o nosso trabalho, tanto termodinâmicos como dinâmicos. Apesar do interesse maior estar centrado na capacidade de armazenamento e reconhecimento de padrões, outras funções importantes como reconhecimento de sequências temporais [23, 24, 25, 26, 27] e categorização [28, 29, 30, 31], tem sido estudadas no modelo de Hopfield. Esta última, em particular, será tratada no capítulo 5. O desafio maior, na tentativa de superar suas limitações como modelo de redes neuronais, parece estar no desenvolvimento de novas técnicas ou mesmo na implementação das já existentes, quando novos ingredientes biológicos são incorporados ao modelo. Nos capítulos 2 e 3 abordaremos esta questão.

Uma grande quantidade de trabalhos e outros métodos dinâmicos não foram aqui citados mas, suas contribuições foram importantes para complementar estes estudos. Aqui descrevemos em parte somente os modelos e as técnicas que serão usadas no desenvolvimento desta tese.

Capítulo 2

Influência dos Períodos Refratários no Modelo de Hopfield

Como foi visto no primeiro capítulo, existe uma forte analogia entre o modelo de Hopfield e modelos estatísticos de materiais magnéticos. E esta analogia torna-se mais clara ainda quando se introduz nos modelos neuronais uma dinâmica estocástica para simular a aleatoriedade envolvida nos processos biológicos. Neste capítulo vamos explorar esta analogia a fim de poder aplicar alguns dos métodos da física estatística de materiais magnéticos (em particular modelos tipo Ising de vidros de spin) ao modelo de Hopfield com períodos refratários.

Nesta analogia, um neurônio ativo $S_i = +1$ corresponde, num modelo magnético tipo Ising, a um spin orientado paralelo à direção do campo magnético, e um neurônio inativo $S_i = -1$ corresponde a um spin orientado antiparelado a esta direção. Modelos tipo Ising são muito usados [32], não somente para modelar materiais magnéticos de spin 1/2, mais também para muitos sistemas físicos os quais podem ser descritos por meio de variáveis dinâmicas binárias (modelos de gás de rede, entre outros). Embora em muitos casos seja uma descrição idealizada, é bem sabido que ela é suficiente para fornecer resultados qualitativos corretos de muitos fenômenos cooperativos observados na natureza.

2.1 Os períodos refratários

Durante os últimos anos muitas características dos neurônios reais foram incorporadas aos modelos neuronais (e em particular ao modelo de Hopfield), na tentativa de torná-los mais realistas do ponto de vista biológico e visando entender as consequências destes ingredientes na resposta da rede. Elementos tais como ruído, diluição, assimetria, interações multi-sinápticas, entre outros, tem sido vastamente estudados na literatura [5, 6, 7, 23, 24, 25, 33, 34, 35].

Um elemento importante na modelagem das redes neuronais (já mencionados no capítulo anterior) são os *períodos refratários*. O seu efeito mais evidente nas redes biológicas é limitar a taxa de disparos dos neurônios. Nos últimos anos uma grande quantidade de trabalhos foram feitos levando em conta as diferentes escalas de tempo biológicas, e em particular, alguns modelaram os períodos refratários utilizando diferentes técnicas tanto analíticas como numéricas.

Buhmann e Schulten [36] investigaram um modelo de rede neuronal com uma dinâmica de tempo contínuo, onde diferentes escalas de tempo biológicas (tais como tempo de relaxação do potencial da membrana do neurônio, períodos refratários absoluto e relativo e escalas de tempo para controlar as mudanças sinápticas, entre outras), foram levadas em conta. Neste modelo, as variáveis dinâmicas são o potencial da membrana, as correntes axonais e as eficácias sinápticas. Estas últimas obedecem uma regra Hebbiana (mas não a de Hopfield) dinâmica, com intensidade que muda para tempos pequenos (da ordem de dezenas de milisegundos). A dinâmica do potencial da membrana é representada por um conjunto de equações diferenciais não lineares acopladas, que envolve a relaxação da membrana para um potencial de repouso, bem como o crescimento ou decréscimo do potencial devido às interações sinápticas entre os neurônios. Eles demonstraram que quando o ruído é introduzido na dinâmica do potencial da membrana, este não destrói a habilidade da rede para o reconhecimento e para a reconstrução de padrões associativamente. Em comparação com o modelo determinístico, a rede com ruído pode reconhecer com baixas frequências de entrada e com uma taxa média de disparo baixa.

Amit et al. [37] (seguindo a linha de trabalho de Buhmann [38]) desenvolveram um modelo que leva em conta uma grande quantidade de ingredientes biológicos. Neste modelo é possível distinguir entre neurônios excitatórios e inibitórios, correspondendo, respectivamente, ao armazenamento associativo e ao controle da taxa de disparos dos neurônios. Os diferentes elementos biológicos envolvidos são:

- dinâmica contínua para o potencial da membrana;
- diferenciação entre as operações excitatórias e inibitórias;
- períodos refratários absoluto e relativo, este último devido a hiperpolarização dos pulsos pós-sinápticos;
- retardamentos nas transmissões de pulsos uniformes ou aleatórios;
- inibição por hiperpolarização;
- estímulos curtos e persistentes, representados como correntes sinápticas dentro de uma fração dos neurônios de um padrão;
- ruído, o qual modela o decaimento gradual do potencial da membrana.

Com relação ao reconhecimento associativo de padrões, onde a matriz sináptica é dada pela regra de Hebb (1.12) alguns modelos dinâmicos foram propostos para um número finito de padrões. M. Y. Choi [39] estudou um modelo que leva em conta explicitamente a existência de várias escalas de tempo tais como os períodos refratários, tempo de duração do potencial de ação e o retardamento da propagação do sinal, sem que o tempo seja discretizado. A dinâmica não é nem totalmente paralela nem sequencial, o que pode ser mais realista. A evolução temporal da rede é descrita por duas equações diferenciais acopladas correspondendo a atividade e a superposição do estado da rede com os padrões. Como exemplo, ele investigou o modelo dinâmico com uma regra tipo Hebb, e obteve características similares aos modelos de Little [1] e Hopfield [2]. Em adição, ele encontrou um rico diagrama de fases, o qual, para $\alpha = 0$ apresenta um ponto tricrítico, como acontece no nosso modelo, segundo veremos nas próximas seções.

Outros trabalhos, considerando a regra de Hebb e os períodos refratários, foram estudados por Aihara et al. [40] (redes neuronais caóticas) e Gerstner e van Hemmen [41]. Neste último, eles direcionaram suas investigações para ver o que acontece se os modelos de neurônios formais (tipo Hopfield) fossem substituídos por redes de neurônios onde a codificação da informação é feita pela presença ou ausência de um pulso. O comportamento coletivo do sistema para uma rede totalmente conectada, é descrito por um conjunto de equações dinâmicas, as quais são exatas no limite de rede infinita e para um número finito de padrões armazenados. Eles mostraram que, em um estado de recuperação estacionário, a rede comporta-se similarmente ao modelo do neurônio formal e, no caso de um estado de recuperação oscilatório o ruído e as constantes de tempo internas dos neurônios tornam-se importantes e determinam o comportamento da rede.

Um modelo bastante interessante foi proposto recentemente por Tamarit et al. [42], onde eles estudaram uma rede com dinâmica estocástica, no qual levam em conta os efeitos dos períodos refratários (absoluto e relativo). A dinâmica é resolvida analiticamente para o modelo de Hopfield ultra diluído na aproximação de campo médio. Para simular os períodos refratários na dinâmica da rede, eles definiram um neurônio com três estados $\{S_i = 0, \pm 1\}$. A temperatura nula, um neurônio que está ativo no tempo t ($S_i(t) = 1$), no tempo seguinte $t+1$ estará no estado zero ($S_i(t+1) = 0$). No tempo posterior $t+2$ este pode estar no estado $S_i(t+2) = -1$ ou no estado ativo novamente, mas neste caso com um grande limiar de ativação. Este último processo simula o período refratário relativo. Eles encontraram, no regime de reconhecimento, baixa atividade dos neurônios e atratores ponto fixo, ciclos periódicos e caos.

Trabalhos envolvendo limiares de ativação, que de certa forma reduzem a taxa de disparos dos neurônios, foram também estudados no contexto de reconhecimento de padrões [26, 43, 44]. A seguir vamos introduzir e estudar um modelo de rede neuronal onde os períodos refratários são levando em conta na dinâmica, através de uma análise de campo médio, em analogia aos modelos magnéticos [45]

2.2 O modelo

Uma maneira simples de introduzir os períodos refratários na dinâmica do modelo de Hopfield, é por meio de um limiar de ativação dependente do tempo e que atua como um campo externo que depende da história do neurônio. Este limiar atua somente após o neurônio ter disparado um sinal e o resultado tem que ser desfavorecer a atividade do neurônio no próximo (ou próximos) instantes de tempo.

Como vimos, diferentes limiares de ativação já tinham sido introduzidos no modelo de Hopfield [26, 43], simulando o efeito de fadiga observado nos neurônios reais. A primeira consequência importante de incluir estes limiares é que a dinâmica não mais satisfaz a condição de balanço detalhado, sendo impossível então desenvolver uma análise termodinâmica como a vista no capítulo 1 para o modelo de Hopfield. Em segundo lugar, mesmo na ausência de aleatoriedade (ou seja, com dinâmicas deterministas), os atratores ponto fixo são desestabilizados dando lugar a comportamentos oscilatórios ou até caóticos, os quais são biologicamente muito mais realistas [11, 12]. Em particular, estes atratores tem merecido muito interesse no estudo de armazenamento de sequências temporais, dado que agora o sistema pode, eventualmente, percorrer diferentes regiões do espaço de fase.

O nosso modelo, embora simples, possui uma grande riqueza nos seus comportamentos dinâmicos observados. Usando a aproximação de campo médio analisaremos analiticamente sua capacidade de reconhecimento e, finalmente, o compararemos com resultados obtidos através de simulações numéricas, as quais nos fornecem informação sobre a natureza dinâmica do sistema.

Lembremos que o modelo de Hopfield consta de uma rede de N neurônios binários, cada um modelado por uma variável tipo Ising S_i , a qual pode tomar valores $\{-1, +1\}$, representando os estados ativo e passivo respectivamente. Para levar em conta o efeito do período refratário sobre o i -ésimo neurônio, adicionamos um limiar que depende do tempo mas, somente através do estado do próprio neurônio, ou seja, que é uma função da variável $S_i(t)$. Assim, o potencial pós-sináptico (ou campo local) no tempo t é dado

agora por:

$$h_i(t) = h_i^H(t) - \frac{\Delta}{2}(1 + S_i(t)) \quad (2.1)$$

onde h_i^H é o PPS usual de Hopfield:

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

J_{ij} é a matriz sináptica de Hopfield, cujos elementos tem a forma:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu . \quad (2.3)$$

Os ξ_i^μ são variáveis aleatórias independentes que valem ± 1 com igual probabilidade e $\{\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu\}$ $\mu = 1, 2, \dots, p$ representa o μ -ésimo padrão armazenado da rede (note que somente consideramos memórias não correlacionadas). A dinâmica da rede é governada por um processo Monte Carlo banho térmico, dado por:

$$P(S_i(t+1) = \pm 1) = \frac{1}{2} [1 \pm \tanh(\beta h_i(t))] . \quad (2.4)$$

Lembremos que o parâmetro $T = 1/\beta$ mede o nível de ruído da rede e que no limite $T \rightarrow 0$ recuperamos a dinâmica determinística:

$$S_i(t+1) = \text{Sinal} \left[h_i^H(t) - \frac{\Delta}{2}(1 + S_i(t)) \right] . \quad (2.5)$$

Desta expressão podemos facilmente entender o efeito deste campo adicional: se o neurônio i dispara um sinal no tempo t ($S_i(t) = +1$), ele necessitará de uma contribuição extra Δ ao PPS para disparar novamente. Por outro lado, se ele não disparou no tempo t ($S_i(t) = -1$), então somente o termo usual de Hopfield contribuirá para o PPS. Observe que neste modelo não há distinção entre os períodos refratários absoluto e relativo (é na verdade considerado um valor médio de ambos) e não se inclui o efeito de fadiga (longa história no tempo), pois depende somente do instante de tempo anterior.

A habilidade de reconhecimento da rede no tempo t é caracterizada pelo cálculo do vetor de superposição \vec{m} entre o estado do sistema $\{S_i(t)\}$ e os padrões armazenados, e cujas componentes são dadas por:

$$m_\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i(t) \rangle. \quad (2.6)$$

Aqui $\langle \dots \rangle$ denota a média térmica a temperatura T . Dizemos que o sistema reconhece um dos padrões sempre que ele evolui no tempo para um atrator caracterizado por um ou mais (dado que agora podemos ter ciclos ou caos) valores não nulos de uma *única* componente do vetor \vec{m} . Os parâmetros relevantes no nosso modelo são Δ , T e $\alpha = p/N$.

2.3 Equações de ponto fixo

Quando Δ é igual a 0 a capacidade crítica de armazenamento pode ser obtida analiticamente para ambas as dinâmicas: paralela (Little) [1] e sequencial (Hopfield) [2]. Neste caso, a condição de balanço detalhado é satisfeita e as distribuições de equilíbrio são bem conhecidas, permitindo derivar o comportamento de longo tempo do sistema a partir da termodinâmica de equilíbrio. Mas, como ocorre quando introduzimos assimetria nas sinapses, a presença de um termo que dependa do estado do sistema no seu campo local (2.1) viola a condição de balanço detalhado e, nestas circunstâncias, as ferramentas da termodinâmica não podem mais ser usadas. Como consequência da inclusão deste termo na dinâmica da rede, verificamos que (ver seção (2.4)), além dos atratores estáveis ponto fixo e ciclo dois comumente encontrados nos modelos de Hopfield e Little, o sistema pode também apresentar ciclos de período maior que dois e mesmo trajetórias caóticas.

A aproximação de campo médio tem mostrado ser (como acontece usualmente na física dos fenômenos críticos) uma ferramenta importante no estudo das redes neuronais. Ela tem sido aplicada ao estudo de redes com conexões simétricas, reproduzindo os valores obtidos mediante a análise termodinâmica (quando esta é possível) [27, 46]. Mas também, tem sido usada para redes com conexões assimétricas, onde, como no nosso caso, não é

possível usar a termodinâmica. Em particular, a derivação heurística das equações segue as idéias usadas por Geszti [47] e Peretto [48], este último no estudo de modelos Hebbianos assimétricos com dinâmica estocástica paralela.

Deste modo, suponhamos que o estado do sistema tem superposição macroscópica diferente de zero com o primeiro padrão $\mu = 1$ (adotaremos $m_1 = m$) e superposição microscopicamente nula ($m_\mu = \mathcal{O}(1/\sqrt{N})$) com todos os outros ($\mu \neq 1$). Realizando a média térmica usando (2.4) e utilizando a aproximação de campo médio [47] na qual o campo local h_i é substituído pelo seu valor médio, podemos escrever m como:

$$\begin{aligned} m &= \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh(\beta \langle h_i \rangle) \\ &= \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh \left[\beta \left(\sum_{\mu}^p \xi_i^\mu \frac{1}{N} \sum_{j \neq i}^N \xi_j^\mu \langle S_j \rangle - \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right]. \end{aligned} \quad (2.7)$$

Separando na soma sobre μ o termo $\mu = 1$, podemos reescrever a equação para o m na forma:

$$m = \frac{1}{N} \sum_{i=1}^N \tanh \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right], \quad (2.8)$$

onde m_μ é o valor assintótico do $m_\mu(t)$ (2.6) no atrator ponto fixo. Note que o termo $\xi_i^1 = \pm 1$ foi posto para fora do argumento da \tanh . Substituído $\langle S_i \rangle$ por $m \xi_i^1$ [47] na equação acima encontramos para a seguinte expressão para a superposição:

$$m = \frac{1}{N} \sum_{i=1}^N \tanh \left[\beta \left(m(1 - \Delta/2) + \xi_i^1 \sum_{\mu \neq 1}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} \right) \right]. \quad (2.9)$$

Como pretendemos armazenar um número extensivo de padrões $p = \alpha N$, não podemos desprezar os efeitos dos $p - 1$ padrões restantes m_μ , presentes no segundo termo do argumento da função \tanh . Vamos assumir que os pequenos m_μ , $\mu \neq 1$, são variáveis aleatórias independentes com média zero e variância dada pela soma das variâncias de cada termo. Note que $\xi_i^1 \xi_i^\mu$ são variáveis aleatórias e independem do m_μ . Assim, pelo teorema do limite central $N^{-1} \sum_i$ é efetivamente uma média sobre o ruído gaussiano $\sum_{\mu \neq 1}^p \xi_i^1 \xi_i^\mu m_\mu$.

A média na desordem pode então ser reescrita como:

$$m = \left\langle \left\langle \frac{1}{\sqrt{2\pi}\sigma} \int dx \exp - \left(\frac{x^2}{2\sigma^2} \right) \tanh \left[\beta \left(m(1 - \Delta/2) + \xi^1 x - \xi^1 \Delta/2 \right) \right] \right\rangle \right\rangle_{\xi}, \quad (2.10)$$

onde $\langle \langle \dots \rangle \rangle_{\xi}$ representa a média sobre a variável aleatória ξ^1 . A variância está dada por:

$$\sigma^2 = \sum_{\mu} \sum_{\rho} m_{\mu} m_{\rho} \langle \xi_i^{\mu} \xi_i^{\rho} \rangle = \sum_{\mu} \sum_{\rho} m_{\mu} m_{\rho} \delta_{\mu\rho} = \sum_{\mu} (m_{\mu})^2. \quad (2.11)$$

Fazendo a transformação $x/\sigma = z$ e realizando a média sobre o ξ^1 , a equação para o m torna-se:

$$m = \frac{1}{2} \int Dz \left(\tanh(\beta L_z^+) + \tanh(\beta L_z^-) \right), \quad (2.12)$$

onde $L_z^{\pm} = (1 - \frac{\Delta}{2})m \pm \frac{\Delta}{2} + \sigma z$ e

$$Dz = \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}}.$$

Para determinar a variância devemos calcular a superposição do estado do sistema com os padrões não correlacionados. Usando um procedimento similar ao anterior, e separando no argumento da tanh os termos $\mu = 1$ e $\mu = \nu$, obtemos a seguinte expressão para o m_{ν}

$$m_{\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^{\nu} \xi_i^1 \tanh \left[\beta \left(m + \xi_i^1 \xi_i^{\nu} m_{\nu} + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^{\mu} m_{\mu} - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right]. \quad (2.13)$$

Note que o segundo termo no argumento da tanh (onde $m_{\nu} \simeq 1/\sqrt{N}$) é desprezível comparado aos demais, de modo que podemos expandir a tanh com relação a este termo

$$m_{\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^{\nu} \xi_i^1 \tanh \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^{\mu} m_{\mu} - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right] + \frac{\beta}{N} \sum_{i=1}^N \left\{ 1 - \tanh^2 \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^{\mu} m_{\mu} - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right] \right\} m_{\nu}. \quad (2.14)$$

Para tornar o tratamento auto-consistente, vamos introduzir os parâmetros:

$$q = \frac{1}{N} \sum_{i=1}^N \tanh^2 \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right] \quad (2.15)$$

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1}^p (m_\nu)^2 ,$$

onde q pode ser identificado como o parâmetro de Edwards-Anderson e r é a superposição quadrática média do estado do sistema com os $p - 1$ padrões restantes [46]. Note que o fator $\alpha^{-1} = N/p$ torna o r realmente uma média sobre as p superposições quadráticas e cancela a dependência esperada $1/\sqrt{N}$ para cada m_ν . Da definição acima, o parâmetro r está relacionado com a variância pela expressão: $\sigma^2 = \alpha r$. Identificando o parâmetro q na segunda linha da equação (2.14), podemos reescrevê-la como:

$$m_\nu = \frac{N^{-1} \sum_{i=1}^N \xi_i^\nu \xi_i^1 \tanh \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right]}{1 - \beta(1 - q)} . \quad (2.16)$$

Substituindo (2.16) na expressão do r obtem-se:

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1}^p \frac{1}{[1 - \beta(1 - q)]^2} \frac{1}{N^2} \sum_i \sum_j \xi_i^1 \xi_j^1 \xi_i^\nu \xi_j^\nu \times$$

$$\tanh \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right] \times$$

$$\tanh \left[\beta \left(m + \xi_j^1 \sum_{\mu \neq 1, \nu}^p \xi_j^\mu m_\mu - \xi_j^1 \frac{\Delta}{2} (1 + \langle S_j \rangle) \right) \right] . \quad (2.17)$$

Notemos que os argumentos das \tanh não dependem do índice ν . Portanto, é fácil verificar que os únicos termos que contribuirão para a média do r serão aqueles com $i = j$

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1}^p \frac{1}{[1 - \beta(1 - q)]^2} \frac{1}{N^2} \sum_i \tanh^2 \left[\beta \left(m + \xi_i^1 \sum_{\mu \neq 1, \nu}^p \xi_i^\mu m_\mu - \xi_i^1 \frac{\Delta}{2} (1 + \langle S_i \rangle) \right) \right] , \quad (2.18)$$

ou ainda,

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1} \frac{1}{N} \frac{q}{[1 - \beta(1 - q)]^2} = \frac{q}{[1 - \beta(1 - q)]^2}. \quad (2.19)$$

O mesmo procedimento realizado no cálculo do m (2.12) pode ser usado para calcular o parâmetro q , resultando:

$$q = \frac{1}{2} \int Dz [\tanh^2(\beta L_z^+) + \tanh^2(\beta L_z^-)] , \quad (2.20)$$

onde L_z^\pm já foi definido anteriormente. Deste modo, as equações (2.12), (2.19) e (2.20), onde $\sigma = \sqrt{\alpha r}$, podem ser resolvidas simultaneamente para m , q e r em função dos parâmetros T , α e Δ constantes.

Estas expressões assumem que o sistema evolui para um atrator ponto fixo macroscópico. É importante esclarecer que um ponto fixo macroscópico não implica necessariamente pontos fixos microscópicos. Em geral, os estados microscópicos estão mudando no tempo. Uma situação típica de um ponto fixo macroscópico é que os estados microscópicos estão evoluindo no tempo mas todos os estados visitados têm, estatisticamente, a mesma superposição com o padrão memorizado, levando então para um mesmo valor da superposição m .

Note que para o caso particular $\Delta = 0$ reobtemos as equações para o modelo de Hopfield [46], as quais também concordam com as obtidas por Amit et al [6] através da termodinâmica de equilíbrio (aproximação de réplicas simétricas).

2.3.1 O caso determinista $T = 0$

Começamos analisando o caso ($T = 0$) para o qual realizamos simulações numéricas. Neste limite, é claro da equação (2.20) que $q \rightarrow 1$, mas a quantidade $\beta(1 - q) \equiv c$ em (2.19) permanece finita. Para calcularmos as integrais das equações (2.12) e (2.20), usamos as seguintes identidades:

$$\lim_{T \rightarrow 0} \int \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}} \tanh \beta(az + b) = \operatorname{erf}\left(\frac{b}{\sqrt{2a}}\right) \quad (2.21)$$

onde a função erro foi definida em (1.7), e

$$\begin{aligned}
\int \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}} [1 - \tanh^2 \beta(az + b)] &\simeq \frac{\exp(-z^2/2)}{\sqrt{2\pi}} \Big|_{\tanh^2 \beta(az+b)=0} \times \\
&\int dz [1 - \tanh^2 \beta(az + b)] \\
&= \frac{\exp(-b^2/2a^2)}{\sqrt{2\pi}} \frac{1}{\alpha\beta} \int dz \frac{\partial}{\partial z} \tanh \beta(az + b) \\
&= \sqrt{\frac{2}{\pi}} \frac{1}{\alpha\beta} \exp(-b^2/2a^2). \tag{2.22}
\end{aligned}$$

As equações resultantes neste limite são:

$$m = \frac{1}{2} \operatorname{erf} \left[\frac{(1 - \frac{\Delta}{2})m + \frac{\Delta}{2}}{\sqrt{2\alpha r}} \right] + \frac{1}{2} \operatorname{erf} \left[\frac{(1 - \frac{\Delta}{2})m - \frac{\Delta}{2}}{\sqrt{2\alpha r}} \right] \tag{2.23}$$

$$c = \frac{1}{\sqrt{2\pi\alpha r}} \left[\exp \left(-\frac{((1 - \frac{\Delta}{2})m + \frac{\Delta}{2})^2}{2\alpha r} \right) + \exp \left(-\frac{((1 - \frac{\Delta}{2})m - \frac{\Delta}{2})^2}{2\alpha r} \right) \right] \tag{2.24}$$

$$r = \frac{1}{(1 - c)^2}. \tag{2.25}$$

Na figura 2.1 apresentamos m como função de α para diferentes valores de Δ . Para $\Delta < 1$ sempre existe um valor crítico α_c abaixo do qual o sistema reconhece o padrão armazenado com uma pequena fração de erro ϵ . Para cada $\alpha_c(\Delta)$ o sistema sofre uma transição descontínua da fase de reconhecimento (na qual a dinâmica é governada por atratores ponto fixo) para uma fase de não reconhecimento onde nossa aproximação analítica não é mais válida; as soluções das equações de auto-consistência não mais correspondem a atratores ponto fixo, os quais foram nossa suposição original. Observe que α_c diminui a medida que Δ cresce. Tomando o limite de $\alpha \rightarrow 0$ nas equações acima, verifica-se que $c \rightarrow 0$, $r \rightarrow 1$ e o comportamento assintótico de m (o argumento da função erf torna-se muito grande) pode ser obtido facilmente. A fração de erro na transição, definida como $\epsilon_c = \frac{1}{2}(1 - m)$, tende a 0 de acordo com a seguinte expressão:

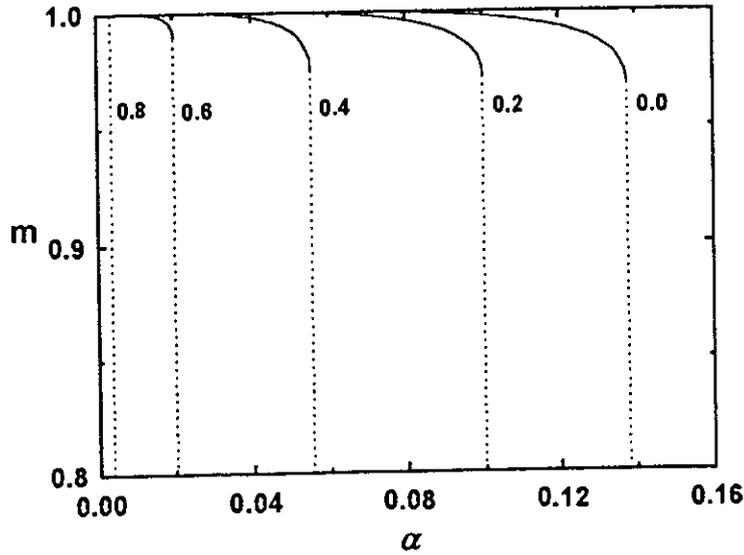


Figura 2.1: Gráfico m vs. α para $T = 0$ e diferentes valores de Δ . Em $\alpha_c(\Delta)$ o sistema sofre uma transição descontínua da fase de reconhecimento para uma fase de não reconhecimento.

$$\epsilon = \frac{1}{2} \sqrt{\frac{\alpha}{2\pi}} \left(\exp\left(-\frac{1}{2\alpha}\right) + \frac{\exp\left(-\frac{(1-\Delta)^2}{2\alpha}\right)}{1-\Delta} \right) \quad (2.26)$$

2.3.2 O caso $T \neq 0$

Também analisamos as soluções de ponto fixo na presença de ruído ($T \neq 0$). Na figura 2.2 apresentamos o diagrama de fase (T, α) para diferentes valores de Δ . Para $\Delta = 0$ recuperamos o diagrama de fase obtido em [6]. Ao longo das linhas $T_c(\alpha)$ o sistema passa por transições descontínuas da fase de reconhecimento (abaixo) para a fase de não reconhecimento (acima). Note que a fase de reconhecimento decresce com o crescimento de Δ ; ou seja, o principal efeito ao introduzir estes períodos refratários, é a desestabilização do reconhecimento por meio de atratores ponto fixo.

Na figura 2.3 apresentamos a linha crítica T versus Δ para $\alpha = 0$. Neste limite, a equação para m (2.12) é dada pela seguinte expressão:

$$m = \frac{1}{2} [\tanh \beta (m(1 - \Delta/2) - \Delta/2) + \tanh \beta (m(1 - \Delta/2) + \Delta/2)] \quad (2.27)$$

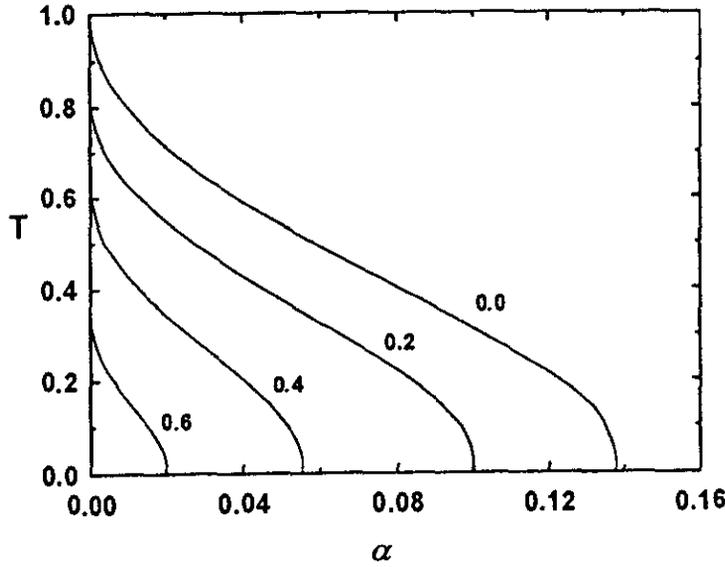


Figura 2.2: Diagrama de fase T vs. α para $\Delta = 0.0, 0.2, 0.4$ e 0.6 . Abaixo das linhas críticas o sistema reconhece com ponto fixo e a transição para a fase de não reconhecimento é de primeira ordem.

Expandindo a equação acima até segunda ordem para valores pequenos de m ($m \rightarrow 0$), obtemos a expressão analítica para a linha crítica

$$m^2 \sim \frac{T^2[(1 - \Delta/2)(1 - L^2) - T]}{(1 - \Delta/2)^3(1 - L^2)(1/3 - L^2)}, \quad (2.28)$$

onde $L = \tanh(\beta\Delta/2)$. O m torna-se nulo à temperatura crítica definida por

$$T_c = (1 - \Delta/2)(1 - \tanh^2(\Delta/2T_c)), \quad (2.29)$$

que é válida somente para $\tanh(\Delta/2T_c) < 1/\sqrt{3}$. Para valores de T e Δ para os quais esta condição não é verificada, o sistema apresenta uma transição descontínua. Assim, existe um ponto tricrítico dado por

$$T^* = \frac{2}{3 + 2 \operatorname{arctanh}(\sqrt{3})} \simeq 0.463$$

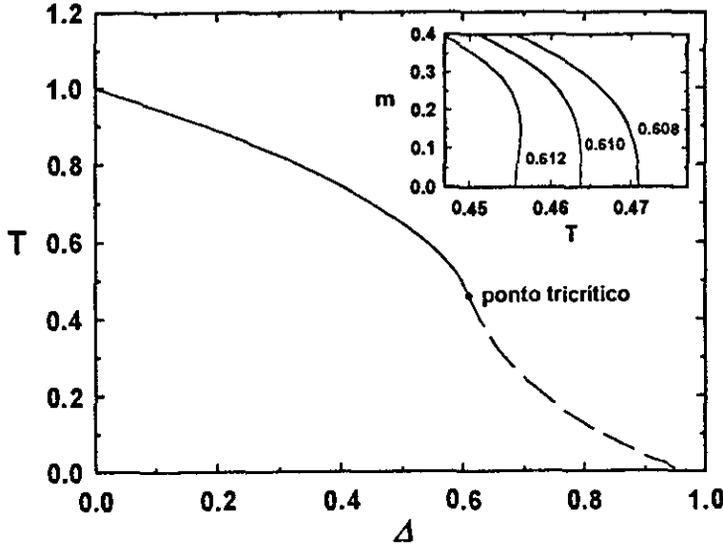


Figura 2.3: Linha crítica $T = f(\Delta)$ para $\alpha = 0$. $(T^*, \Delta^*) \simeq (0.463, 0.610)$ é um ponto tricrítico. Para $\Delta < \Delta^*$ a transição é de segunda ordem (linha cheia) enquanto que para $\Delta > \Delta^*$ a transição é descontínua (linha tracejada).

e

$$\Delta^* = 2T_* \operatorname{arctanh}(\sqrt{3}) \simeq 0.610 ,$$

que separa a linha crítica em duas partes. No detalhe mostramos o comportamento da superposição m próximo ao ponto tricrítico como função de T .

Para caso $\alpha \neq 0$ a transição entre as fases de reconhecimento e não reconhecimento é sempre de primeira ordem como acontece no caso $\Delta = 0$.

2.4 Simulações Numéricas

Nesta seção apresentaremos um estudo numérico da habilidade de reconhecimento e das propriedades dinâmicas do modelo para $T = 0$ e comparamos com os resultados analíticos obtidos na seção anterior. As simulações foram realizadas para redes com $N = 800, 1600$ e 3200 neurônios usando dinâmica paralela e sequencial. Para tornar mais eficiente nossas simulações usamos a técnica introduzida por Pena e Oliveira [17]. Apesar desta técnica permitir usar redes bem maiores [16], nossa maior limitação foi devida ao grande número

de simulações necessárias para construir o diagrama de fases e aos grandes transientes envolvidos.

Na simulação numérica, partindo de uma configuração inicial, deixamos o sistema evoluir durante um certo transiente (que em alguns casos, próximo da transição, chegaram até 5000 passos de Monte Carlo) e armazenamos sua última configuração, que chamaremos de configuração de referência). Para caracterizar o comportamento dinâmico, ou seja, se o sistema evolui para um atrator ponto fixo ou uma órbita periódica, monitoramos sua evolução num período de 100 passos de Monte Carlo e comparamos, a cada passo de tempo, sua configuração atual com a configuração de referência. Dependendo dos valores dos parâmetros e do tamanho da rede, o sistema pode retornar ou não para esta configuração. Se ele retorna os atratores podem ser ponto fixo ou periódicos, caso contrário, dizemos que ele está em uma órbita caótica. Apesar de não termos feito uma análise detalhada para determinar se estas órbitas são realmente caóticas ou se são órbitas de períodos grandes (maior que 100), em todas as simulações realizadas o tempo de convergência para os atratores ponto fixo ou atratores periódicos foi relativamente pequeno (menor que 100).

Para analisar a habilidade de reconhecimento calculamos para cada amostra a média temporal da superposição entre os padrões armazenados e o estado do sistema no atrator. Se o sistema atinge uma órbita cíclica de período t_c , medimos no atrator a seguinte quantidade:

$$m_\mu = \frac{1}{t_c} \sum_{t=t_0}^{t_0+t_c} m_\mu(t) . \quad (2.30)$$

Escolhendo o estado inicial como sendo a primeira memória, dizemos que a rede reconhece quando

$$m = m_1 \sim 1 \quad m_\mu \sim \mathcal{O}(1/\sqrt{N}), \text{ for } \mu > 1 . \quad (2.31)$$

Para realizar a média configuracional do m , repetimos este procedimento para 100 amostras diferentes onde usamos diferentes memórias, configurações iniciais e sequências de números aleatórios. Para caracterizar o comportamento dinâmico, calculamos a frequência na qual cada espécie de atrator aparece e a atividade média, definida como o número médio de neurônios ativos no atrator

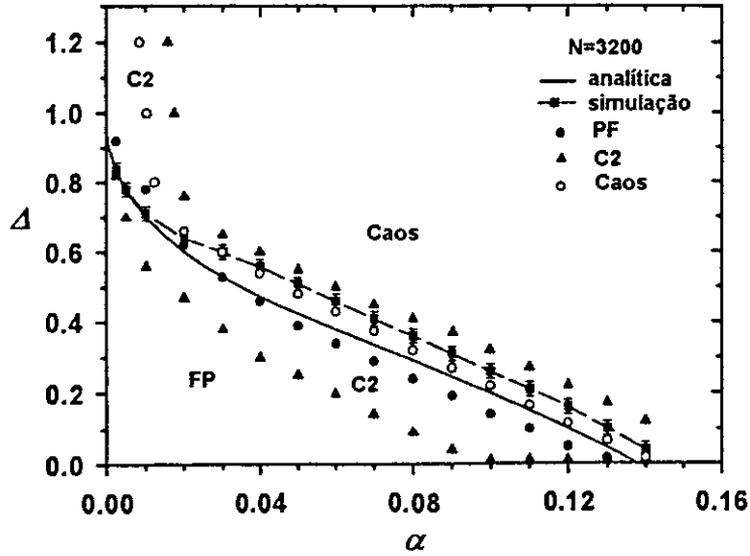


Figura 2.4: Diagrama de fase numérico (Δ, α) para $T = 0$ e $N = 3200$ neurônios, mostrando as regiões com atratores PF (abaixo dos círculos cheios), periódicos (entre as duas formações de triângulos) e caóticos (acima dos círculos vazios). A linha obtida da simulação (linha tracejada) e a analítica (linha cheia), separam a fase de reconhecimento (abaixo) da fase de não reconhecimento (acima).

$$a = \frac{1}{2N} \sum_{i=1}^N (1 + S_i) . \quad (2.32)$$

Na figura 2.4 apresentamos o diagrama de fase Δ vs. α para $N = 3200$ e dinâmica paralela. Para $\Delta = 0$ o sistema apresenta somente atratores ponto fixos (PF). Para α fixo e valores crescentes de Δ encontramos que:

1. para pequenos valores de Δ a dinâmica é governada somente por atratores ponto fixo . Os círculos cheios delimitam a região acima da qual os atratores ponto fixo deixam de existir;
2. a região entre as formações de triângulos, caracteriza a predominância de ciclos de período dois (C2). Eles surgem inicialmente na formação de triângulos inferior e desaparecem na superior;
3. os círculos vazios indicam os valores de Δ acima do qual o comportamento caótico

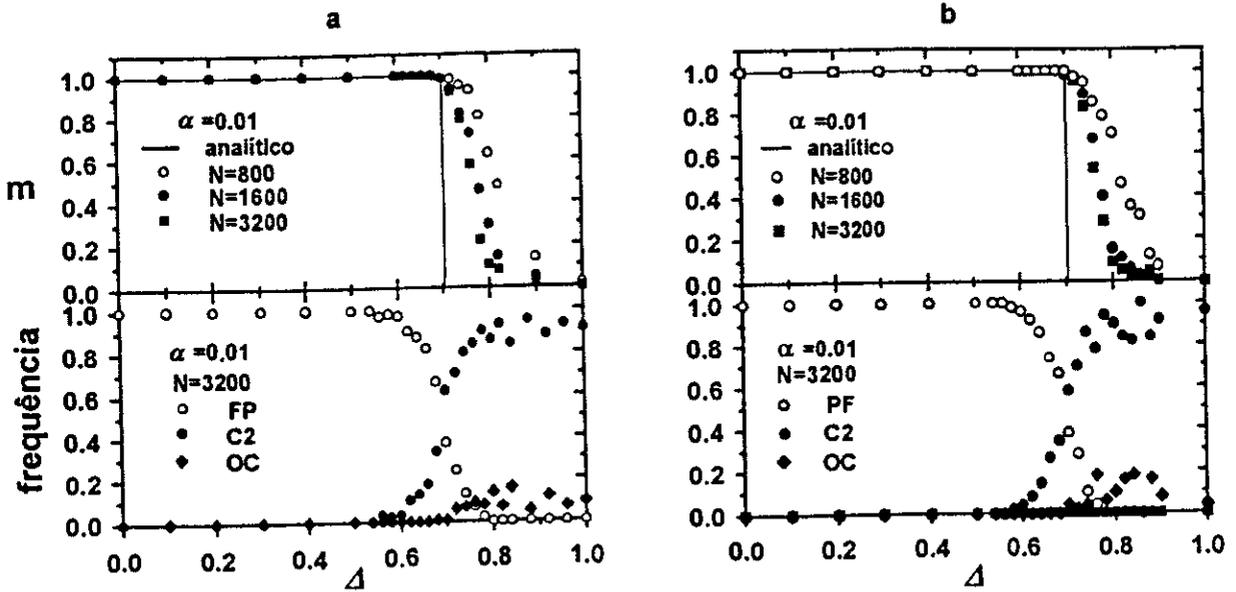


Figura 2.5: A superposição m e a frequência na qual cada tipo de atrator aparece como função de Δ para $\alpha = 0.01$: a) dinâmica paralela; b) dinâmica sequencial. A linha cheia corresponde a solução analítica.

emerge.

Observamos que existem muitas regiões de coexistência de atratores. Na região entre o C2 e Caos encontramos também ciclos de períodos maiores que dois que denominaremos de outros ciclos (OC), mas que não estão representados no diagrama de fase.

Independente do comportamento dinâmico, estudamos também a capacidade crítica de reconhecimento da rede. A linha tracejada, obtida numericamente, separa a fase de reconhecimento (abaixo) da fase de não reconhecimento (acima) e a linha cheia corresponde ao resultado analítico da seção anterior. A curva da simulação reproduz muito bem os resultados analíticos somente para valores pequeno de α .

Para entender a diferença qualitativa entre os resultados analíticos e numéricos para valores grandes de α , analisamos o comportamento do sistema ao longo de dois cortes para α fixo: $\alpha = 0.01$ e 0.04 . Na Figura 2.5 mostramos a superposição m e a frequência na qual cada espécie de atrator aparece como função do Δ para $\alpha = 0.01$ nos casos a) dinâmica paralela e b) dinâmica sequencial. A primeira coisa a notar é que a região de PF coincide com a fase de reconhecimento, e a região de C2 com a fase de não reconhecimento.

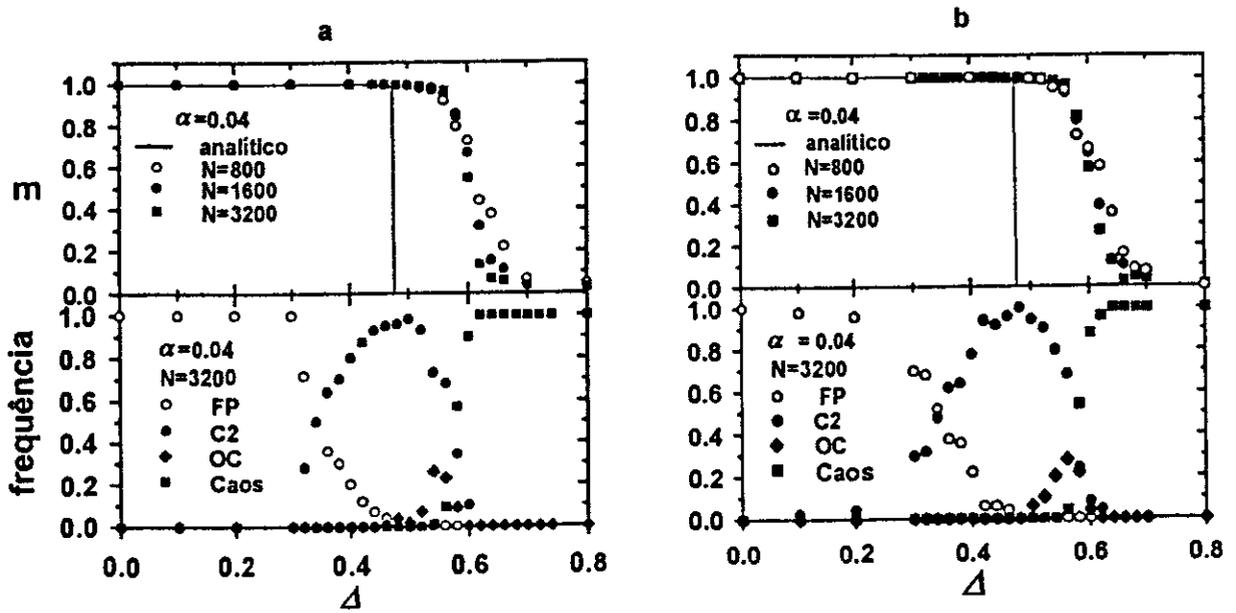


Figura 2.6: A superposição m e a frequência na qual cada tipo de atrator aparece como função de Δ para $\alpha = 0.04$: a) dinâmica paralela; b) dinâmica sequencial. A linhas cheia corresponde a solução analítica.

No caso em que o sistema reconhece a memória somente através de PF, a curva analítica prediz muito bem a transição. Por outro lado, na figura 2.6 apresentamos os mesmos gráficos para $\alpha = 0.04$. Note agora que a fase de reconhecimento apresenta dois diferentes comportamentos dinâmicos: para pequenos valores de Δ o sistema evolui para PF enquanto que para valores intermediários os C2 predominam. Ao contrário do caso $\alpha = 0.01$, agora a curva teórica não mais prediz a transição para a fase de não reconhecimento e sim, a transição de PF para C2. Em ambos os casos, analisamos o efeito de tamanho finito e usamos redes com tamanhos $N = 800, 1600$ e 3200 . Note que quando N cresce, os resultados da simulação para o m na transição tende a decair mais abruptamente caracterizando a transição de primeira ordem encontrada no cálculo analítico. Ambas as dinâmicas (paralela e sequencial) apresentam resultado muito parecidos, mostrando que os comportamentos oscilatórios (como esperado) são devidos ao termo refratário no campo local.

Na figura 2.7 apresentamos a atividade média como função do Δ para valores de $\alpha = 0.01$ e 0.04 . Podemos notar que na região de reconhecimento com pontos fixos

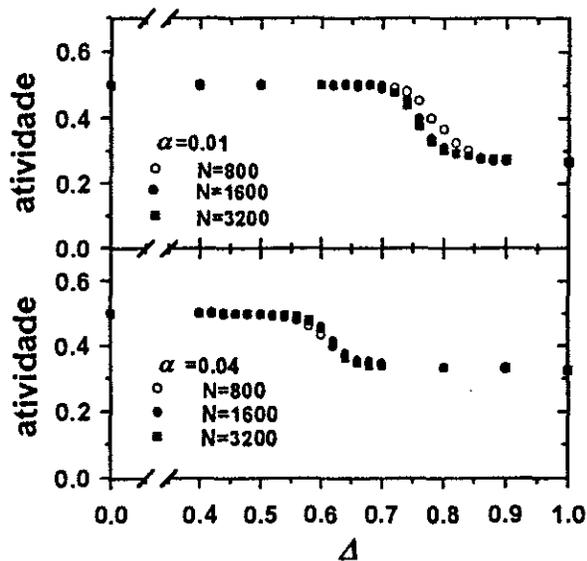


Figura 2.7: Atividade média vs. Δ para $\alpha = 0.01$ e $\alpha = 0.04$.

e órbitas periódicas, a atividade média permanece em torno do valor ~ 0.5 (variáveis aleatórias) e este valor decresce somente na transição para a fase de não reconhecimento. Isto mostra que o parâmetro Δ não somente danifica a habilidade do reconhecimento com ponto fixo, mas melhora esta habilidade permitindo o aparecimento e reconhecimento de atratores mais complicados (ciclos de ordem maior que dois), que certamente são mais importantes do ponto de vista biológico.

2.5 Conclusões

Neste capítulo estudamos analiticamente e através de simulação numérica um modelo de memória associativa onde incorporamos na dinâmica da rede uma nova espécie de limiar que simula o efeito dos períodos refratários. O principal resultado é que o parâmetro Δ que ativa este limiar produz o aparecimento de atratores caóticos e periódicos. Contudo, o sistema parece reconhecer somente nos regimes de ponto fixo e ciclos de ordem dois. Somente em uma pequena região o sistema reconhece com ciclos de ordem maior que dois e caos (ciclos maior que 100), mas este reconhecimento aparece justamente na fronteira para a fase de não reconhecimento. Um estudo mais detalhado nesta região seria interessante

para elucidar se a existência destes atratores são devido ao efeito do tamanho finito ou não. Nas nossas simulações, quando aumentamos o tamanho do sistema, estes atratores não desaparecem e portanto suspeitamos que eles devam existir no limite termodinâmico.

Na fase de reconhecimento (para Δ pequeno), o PPS é forte o suficiente para levar o sistema para atratores estáveis; PF e órbitas periódicas, onde a superposição com o padrão memorizado é próximo de 1. Crescendo o parâmetro Δ o sistema é drasticamente afetado, e nestas regiões a dinâmica é dominada por ciclos de períodos longos ou trajetórias caóticas.

A simulação numérica reproduz muito bem os resultados analíticos para valores de α pequenos, onde a transição ocorre de ponto fixo (PF) para ciclo de ordem dois (C2). Ambas as dinâmicas (paralela e sequencial) apresentam resultado muito parecidos, mostrando que o comportamento oscilatório deve-se a presença do termo refratário no campo local. Finalmente, observamos que somente na transição a atividade média decresce com o crescimento do Δ .

Gostaríamos de ressaltar que a definição de reconhecimento utilizada nesta tese, de comparar com um padrão fixo, pode não ser a definição mais adequada para se estudar o reconhecimento de memórias que sejam constituídas por ciclos ao invés de estados estacionários. Uma definição adequada de reconhecimento para padrões caracterizados por ciclos poderia nos permitir entender melhor esta nova fase sem estados estacionários (ponto fixo).

Capítulo 3

Rede de Neurônios com Diluição Assimétrica e Períodos Refratários

Neste capítulo, seguiremos analisando a influência dos períodos refratários, introduzidos no capítulo anterior, e estudaremos a versão ultra diluída do modelo de Hopfield [49].

Relembrando que no modelo de Hopfield, o fato dos acoplamentos serem simétricos e a rede ser totalmente conectada tornam o problema do reconhecimento de padrões matematicamente tratável. Podemos também, como foi visto, introduzir um parâmetro de ruído que modela os processos aleatórios biológicos, e tratá-lo como uma temperatura generalizada na dinâmica Monte Carlo estocástica. Neste caso, o comportamento de longo tempo do modelo pode ser facilmente obtido, através de uma análise termodinâmica em analogia com a teoria de campo médio para o modelo de vidro de spin tipo Ising [5, 6].

Contudo, as redes de neurônios biológicas são assimétricas e altamente diluídas. Isto significa que os dois elementos, acoplamentos simétricos e total conectividade, presentes no modelo de Hopfield, limitam seriamente sua utilidade do ponto de vista biológico. Para uma maior aproximação do modelo com as redes reais, diferentes modificações têm sido introduzidas na literatura, muitas delas preservando a prescrição Hebbiana para o acoplamento sináptico mas adicionando outros ingredientes como diluição e assimetria

[7, 23, 24, 50, 51, 52]. A principal consequência matemática de remover a simetria nas sinapses é que, agora, a condição de balanço detalhado não é mais satisfeita, e estamos restritos a um estudo dinâmico do sistema.

Um tratamento analítico da dinâmica do modelo de redes neuronais é, em geral, mais difícil que um tratamento termodinâmico, dado que não existe uma prescrição similar à distribuição de Boltzmann-Gibbs. Porém, Derrida et al. [7] mostraram que a dinâmica do modelo de Hopfield com diluição assimétrica aleatória, no regime conhecido como *ultra diluído* (no qual a conectividade média por neurônio C e o número de neurônios N satisfazem a condição $C \ll \ln N$) é simples de tratar (ver apêndice A).

Apresentamos neste capítulo um estudo da dinâmica do modelo de Hopfield ultra diluído com períodos refratários, que permite preservar a dependência do sistema ao longo de toda sua história, o qual, como veremos, é fundamental no nosso modelo. Deste modo, a ação dos períodos refratários pode modelar o efeito de fadiga verificados nos neurônios reais. Similarmente ao capítulo anterior, os períodos refratários são introduzidos como um limiar dependente do tempo que atua somente sobre os neurônios ativos no instante anterior. É importante destacar que neste estudo estamos modelando simultaneamente três ingredientes biológicos importantes.

3.1 O Modelo

O modelo é similar ao tratado no capítulo anterior. Consiste de N neurônios binários $S_i = \pm 1$ ($i = 1, \dots, N$) cuja evolução é governada por uma dinâmica estocástica paralela dada pela seguinte regra:

$$P_r(S_i(t+1)) = \frac{1}{2} (1 + S_i(t+1) \tanh \beta_0 h_i(t)) \quad , \quad (3.1)$$

onde $\beta_0 = 1/T_0$ mede o nível de ruído da rede e $h_i(t)$ representa o potencial pós-sináptico (PPS) sobre o i -ésimo neurônio no tempo t

$$h_i(t) = h_i^H(t) - \frac{\Delta_0}{2}(S_i(t) + 1). \quad (3.2)$$

O $h_i^H(t)$ corresponde ao PPS usual de Hopfield

$$h_i^H(t) = \sum_{j \neq i}^N J_{ij} S_j(t), \quad (3.3)$$

e a matriz sináptica assimétrica J_{ij} é definida pela regra Hebbiana

$$J_{ij} = C_{ij} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (3.4)$$

Os ξ_i^μ 's são variáveis aleatórias independentes que valem ± 1 com igual probabilidade. O ξ_i^μ representa o estado do i -ésimo neurônio no μ -ésimo padrão armazenado ($\mu = 1, \dots, p$). Os C_{ij} 's introduzem a diluição e a assimetria das sinapses e são variáveis aleatórias escolhidas de acordo com a seguinte distribuição de probabilidade:

$$\rho(C_{ij}) = \frac{C}{N} \delta(C_{ij} - 1) + \left(1 - \frac{C}{N}\right) \delta(C_{ij}). \quad (3.5)$$

O C é a conectividade média por neurônio. $C_{ij} = 1$ significa que, após a diluição o neurônio i recebe estímulos do neurônio j com eficiência sináptica J_{ij} , enquanto que $C_{ij} = 0$ significa que esta sinapse foi cortada. Note que, como os C_{ij} e C_{ji} são estatisticamente independentes, eles podem ser diferentes com probabilidade $2(C/N)(1 - C/N)$ e idênticos com probabilidade $(C/N)^2 + (1 - C/N)^2$, introduzindo assim assimetria nos acoplamentos.

Como no capítulo anterior, o segundo termo em (3.2) é incluído para modelar os períodos refratários. Similar ao caso não diluído, este modelo também não distingue entre o período refratário absoluto (PRA) e relativo (PRR). Consideramos uma espécie de média, de forma tal que Δ_0 grande significa que Δt é da ordem de 2 ms, entanto que Δ_0 pequeno significa Δt maior que 4 ms.

Devido a violação da condição de balanço detalhado no modelo, não é mais possível desenvolver um formalismo termodinâmico. Assim, nos restringiremos à considerar o comportamento dinâmico do sistema, procurando uma equação de recorrência para a superposição macroscópica entre o estado do sistema e as memórias:

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i(t) \rangle, \quad (3.6)$$

onde $\langle \dots \rangle$ representa as médias sobre as condições iniciais e sobre o ruído.

É importante notar que no nosso modelo, a violação da condição de balanço detalhado não é somente devida à assimetria nas sinapses, mas também devido a inclusão do limiar que depende do estado do mesmo neurônio em tempos diferentes. Na ausência deste termo e no limite $N \rightarrow \infty$, com $C \ll \ln N$, é possível obter uma equação de recorrência exata para m^μ que dependa somente do instante de tempo anterior. Neste limite ultra diluído, a desordem congelada (quenched) pode ser tratada como num modelo recozido (annealed), no qual as variáveis aleatórias (sinapses) flutuam na mesma escala de tempo em que flutuam as variáveis dinâmicas. Mas, quando os períodos refratários são incluídos na dinâmica (como um termo adicional que dependa do estado do mesmo neurônio em tempos anteriores), isto não é mais válido e somos forçados a levar em conta toda a história da rede.

Um tratamento analítico exato da dinâmica nesta linha é difícil de ser implementado, de modo que introduziremos um método aproximado para estudar a habilidade de reconhecimento da rede. A validade da nossa aproximação será testada na seção 3.3 onde comparamos os nossos resultados analíticos com os obtidos através de simulações numéricas.

3.2 A Dinâmica

Começamos assumindo que a configuração inicial $\{S_i(0)\}$ tem superposição macroscópica somente com a primeira memória ($m^1(0) = m$ e $m^\mu(0) \sim \mathcal{O}(N^{-1/2})$ para $\mu > 1$). Isto

é, as $p - 1$ superposições restantes serão sempre macroscopicamente nulas. Realizando a média térmica por meio de (3.1), a equação de recorrência para a superposição com a primeira memória $m^1(t + 1) \equiv m(t + 1)$ pode ser escrita como:

$$m(t + 1) = \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh \beta_0 \left[h_i^H(t) - \frac{\Delta_0}{2} (S_i(t) + 1) \right]. \quad (3.7)$$

Mesmo já tendo realizado a média térmica, na equação (3.7) ainda permanece a dependência explícita com $S_i(t)$ do termo refratário. Numa aproximação de campo médio, onde assumimos que $S_i(t)$ é dado por $\xi_i^1 m(t)$, os resultados para valores grandes de Δ_0 não expressam corretamente o comportamento da rede quando comparados com os resultados obtidos nas simulações numéricas. Afim de preservar a dependência com a história do sistema, realizamos novamente a média por meio de (3.1), resultando na seguinte expressão para $m(t + 1)$

$$\begin{aligned} m(t + 1) &= \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh \beta_0 (h_i^H(t) - \Delta_0) \times P_\tau(S_i(t) = +1) \\ &\quad + \frac{1}{N} \sum_{i=1}^N \xi_i^1 \tanh \beta_0 (h_i^H(t)) \times P_\tau(S_i(t) = -1) \\ &= \frac{1}{2N} \sum_{i=1}^N \xi_i^1 [\tanh \beta_0 (h_i^H(t) - \Delta_0) + \tanh \beta_0 (h_i^H(t))] + \\ &\quad \frac{1}{2N} \sum_{i=1}^N \xi_i^1 [\tanh \beta_0 (h_i^H(t) - \Delta_0) - \tanh \beta_0 (h_i^H(t))] \times \\ &\quad \tanh \beta_0 [h_i^H(t) - \frac{\Delta_0}{2} (S_i(t - 1) + 1)]. \end{aligned} \quad (3.8)$$

Note que $m(t + 1)$ depende agora do tempo $t - 1$. O mesmo procedimento pode ser repetido, até obtermos uma dependência com o estado inicial $S_i(0)$. Após tomar o limite termodinâmico $N \rightarrow \infty$, é fácil verificar que este mecanismo leva à seguinte expressão:

$$\begin{aligned} m(t + 1) &= \frac{1}{2} \langle \langle \xi_i^1 A_i^t \rangle \rangle + \frac{1}{2^2} \langle \langle \xi_i^1 B_i^t A_i^{t-1} \rangle \rangle + \frac{1}{2^3} \langle \langle \xi_i^1 B_i^t B_i^{t-1} A_i^{t-2} \rangle \rangle + \dots \\ &\quad + \frac{1}{2^{t+1}} \langle \langle \xi_i^1 B_i^t B_i^{t-1} \dots B_i^1 A_i^0 \rangle \rangle \end{aligned}$$

$$= \langle\langle \xi_i^1 \sum_{l=0}^t \frac{1}{2^{l+1}} A_i^{t-l} \prod_{q=1}^l B_i^{t-(q-1)} \rangle\rangle, \quad (3.9)$$

onde $\langle\langle \dots \rangle\rangle$ representa a média configuracional sobre as variáveis aleatórias $\{\xi_i^\mu\}$ e os A_i^t e B_i^t são dados por:

$$A_i^t = \tanh [\beta_0(h_i^H(t) - \Delta_0)] + \tanh [\beta_0 h_i^H(t)] \quad (3.10)$$

$$B_i^t = \tanh [\beta_0(h_i^H(t) - \Delta_0)] - \tanh [\beta_0 h_i^H(t)].$$

Um desenvolvimento analítico exato ao longo desta linha, é uma tarefa bastante difícil de ser implementada. Neste ponto, fazemos uma aproximação de modo a tornar o problema matematicamente tratável: consideramos que as correlações temporais podem ser desprezadas, isto é, em cada um dos termos da equação (3.9) os fatores para diferentes tempos podem ser tratados independentemente.

$$\langle\langle A_i^{t-l} \prod_{q=1}^l B_i^{t-(q-1)} \rangle\rangle = \langle\langle A_i^{t-l} \rangle\rangle \prod_{q=1}^l \langle\langle B_i^{t-(q-1)} \rangle\rangle. \quad (3.11)$$

Sob estas suposições, e lembrando que no limite ultra diluído as variáveis dinâmicas são todas independentes, $h_i(t)$ pode ser considerado como uma variável aleatória que, no limite $C \rightarrow \infty$, $p \rightarrow \infty$ (com $\alpha \equiv p/C$ constante) tem uma distribuição gaussiana cuja média e variância são m e α respectivamente (ver apêndice A). A equação aproximada para a recorrência do $m(t+1)$ após ter sido realizada a média nos ξ 's é:

$$m(t+1) = \sum_{l=0}^t \frac{1}{2^{l+2}} \left[A_-^{t-l} \prod_{q=1}^l B_-^{t-(q-1)} + (-1)^l A_+^{t-l} \prod_{q=1}^l B_+^{t-(q-1)} \right], \quad (3.12)$$

com

$$A_{\pm}^t = \int Dz \left[\tanh (\beta(m(t) + z\sqrt{\alpha} \pm \Delta)) + \tanh (\beta(m(t) + z\sqrt{\alpha})) \right]$$

(3.13)

$$B_{\pm}^t = \int Dz \left[\tanh(\beta(m(t) + z\sqrt{\alpha} \pm \Delta)) - \tanh(\beta(m(t) + z\sqrt{\alpha})) \right].$$

Aqui, $\beta \equiv 1/T = C/T_0$ e $\Delta = \Delta_0/C$ são definidos como os parâmetros reduzidos (temperatura e o termo refratário respectivamente) e Dz é dado por

$$Dz = \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}.$$

Para $\Delta = 0$, $B_{\pm}^t = 0$ para todo t e somente o termo $l = 0$ sobrevive na equação (3.12). Deste modo, recuperamos a expressão obtida por Derrida et al. em [7].

No limite $t \rightarrow \infty$, a equação (3.12) depende de um número infinito de termos. Uma primeira simplificação seria truncar a série até uma dada ordem. Realizamos tal procedimento e observamos que o $m(t+1)$ sempre converge para um atrator ponto fixo, independentemente da ordem de truncamento. Verificamos também que, para valores pequenos do parâmetro Δ , os resultados permanecem quase inalterados a partir do termo de ordem 4 da série ($l = 4$). Já para valores grandes de Δ a dependência com a história do sistema é importante, e portanto, é necessário mais termos para descrever satisfatoriamente o comportamento da rede. A seguir vamos assumir que o sistema sempre evolui para atratores ponto fixo (isto não implica em assumir que o sistema atinge um ponto fixo microscópico). A equação para o valor do m no ponto fixo é dada por:

$$m = \frac{1}{4} \sum_{l=0}^{\infty} \frac{1}{2^l} [A_- B_-^l + (-1)^l A_+ B_+^l], \quad (3.14)$$

onde A_{\pm} e B_{\pm} são dados por (3.13) com $m(t) = m$ para todo t . Na seção (3.3) mostraremos, numericamente, que esta suposição é válida para uma larga faixa de valores de Δ . É fácil verificar que no limite $l \rightarrow \infty$ estas séries convergem para a seguinte expressão:

$$\begin{aligned} m &= F(m) \\ &= \frac{1}{2} \sum_{\sigma=\pm} \frac{\int Dz [\tanh(\beta(m + z\sqrt{\alpha} - \sigma\Delta)) + \tanh(\beta(m + z\sqrt{\alpha}))]}{2 - \sigma \int Dz [\tanh(\beta(m + z\sqrt{\alpha} - \sigma\Delta)) - \tanh(\beta(m + z\sqrt{\alpha}))]} \end{aligned} \quad (3.15)$$

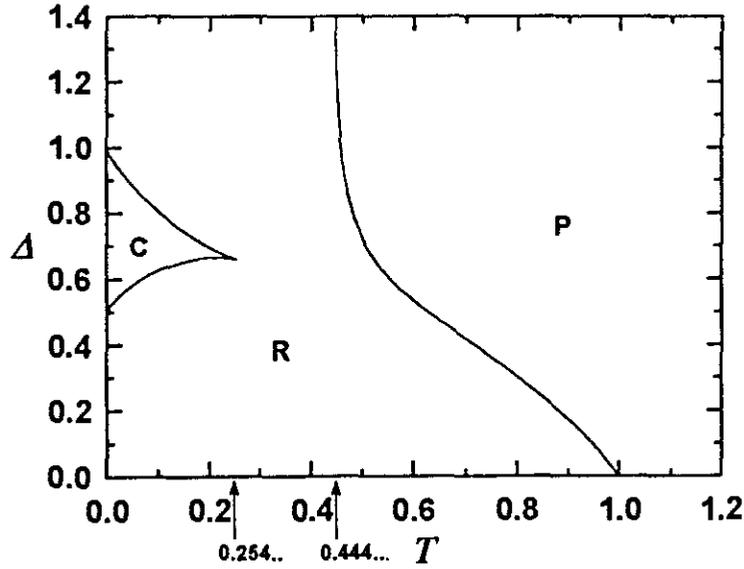


Figura 3.1: Diagrama de fase Δ vs. T para $\alpha = 0$.

3.2.1 Número finito de memórias

Vamos considerar o caso $\alpha = 0$. A equação (3.15) tem agora a seguinte forma

$$m = \frac{1}{2} \sum_{\sigma=\pm} \frac{\tanh(\beta(m - \sigma\Delta)) + \tanh(\beta m)}{2 - \sigma [\tanh(\beta(m - \sigma\Delta)) - \tanh(\beta m)]}. \quad (3.16)$$

A figura 3.1 mostra o diagrama de fase Δ vs. T . Para todos os valores de Δ o sistema sofre uma transição de fase de segunda ordem da fase de reconhecimento (R), caracterizada pelo atrator $m \neq 0$, para a fase paramagnética (P) caracterizada pela solução $m = 0$. Dado que a transição é contínua, podemos expandir a equação (3.16) para valores pequenos de m , obtendo a seguinte expressão para a linha crítica:

$$T_c = \frac{4 + 2 \tanh(\beta_c \Delta) - 2 \tanh^2(\beta_c \Delta)}{(2 + \tanh(\beta_c \Delta))^2}. \quad (3.17)$$

Quando $\Delta \rightarrow 0$ obtemos o resultado já conhecido do modelo de Hopfield ultra diluído $T_c = 1$. Por outro lado, para $\Delta \rightarrow \infty$ encontramos que $T^* = 4/9 = 0.444\dots$. Na fase de reconhecimento observamos que existem três diferentes comportamentos. Para valores pequenos de Δ a recuperação é sempre perfeita ($m \approx 1$). Para valores grandes

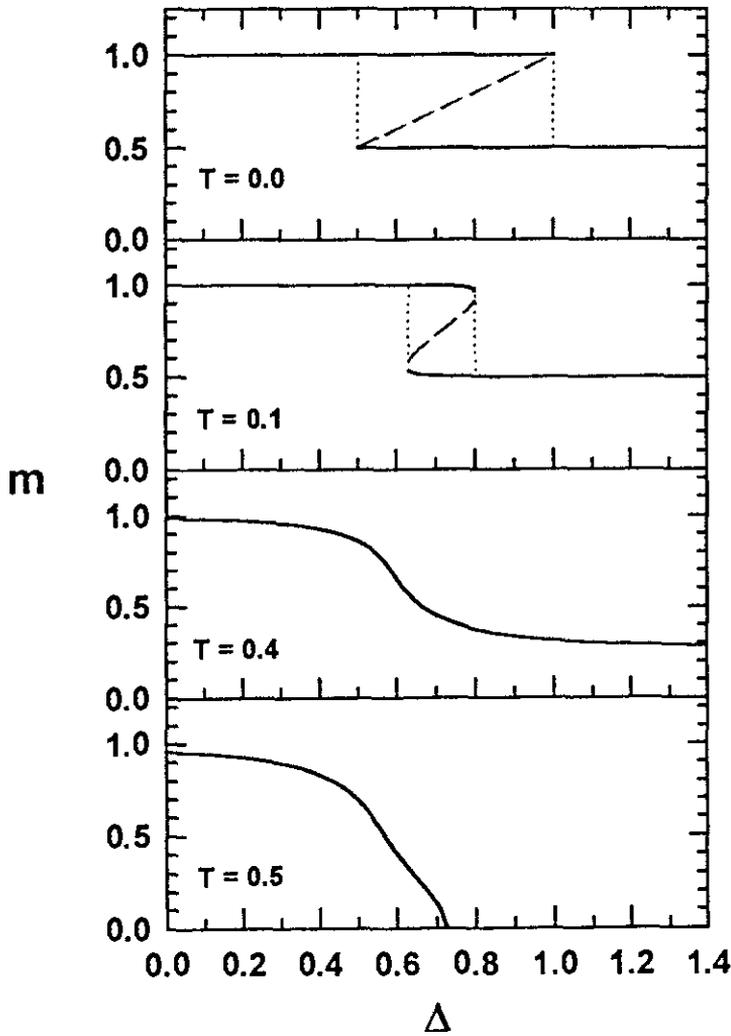


Figura 3.2: m como função do Δ para $\alpha = 0$ e diferentes valores de T (0.0, 0.1, 0.4 e 0.5).

de Δ a habilidade do reconhecimento é bastante pobre, mesmo para $T = 0$, com $m \leq 0.5$. Finalmente, na região intermediária ($0.5 < \Delta < 1$) o regime de reconhecimento depende da temperatura: para $T < 0.254$ encontramos uma pequena região (C) na qual coexistem os dois regimes anteriores, um próximo de 1 e o outro próximo de 0.5. O sistema escolhe um ou outro dependendo do o valor inicial da superposição m_0 . Para $T > 0.254$ e aumentando o valor de Δ , o sistema passa continuamente de um regime com alto reconhecimento para um outro com reconhecimento pobre.

Na figura 3.2 apresentamos o comportamento da superposição m como função de Δ

para diferentes valores de T . Para $T = 0$ e $T = 0.1$ observamos os três diferentes regimes descritos na figura 3.1. Para Δ pequeno o sistema tem somente uma solução estável $m \approx 1$, ou seja, para qualquer superposição macroscópica inicial m_0 o sistema sempre reconhece perfeitamente o padrão memorizado. Quando Δ cresce o sistema apresenta uma transição dinâmica de primeira ordem onde um novo atrator de reconhecimento surge com $m \approx 0.5$. A linha tracejada corresponde as soluções instáveis que separam as bacias atratoras das duas soluções estáveis coexistentes. Aumentando o valor do Δ , o sistema experimenta uma nova transição de fase dinâmica onde a solução com $m = 1$ (alto reconhecimento) desaparece descontinuamente. Nesta fase, o sistema reconhece pobremente o padrão memorizado $m = 0.5$ independente do valor inicial m_0 . Para $T = 0.4$ o sistema sempre reconhece, mas o m decresce para um valor assintótico quando Δ cresce. O mesmo comportamento é observado no intervalo $0.254 < T < 0.444 \dots$. Finalmente, para $T = 0.5$ o sistema apresenta uma transição dinâmica de segunda ordem da fase de reconhecimento para a fase paramagnética, e este comportamento é o mesmo para todos os valores de temperatura no intervalo $0.444 < T < 1$ (Figura 3.1).

3.2.2 Número infinito de memórias

Nesta seção consideramos o caso $\alpha \neq 0$. Vamos inicialmente analisar o sistema na ausência de ruído térmico $T = 0$, caso para o qual temos realizado simulações numéricas. No limite $T \rightarrow 0$ e usando a identidade (2.21), a equação (3.15) assume a seguinte forma:

$$m = \frac{1}{2} \sum_{\sigma=\pm} \frac{\operatorname{erf}\left(\frac{m-\sigma\Delta}{\sqrt{2\alpha}}\right) + \operatorname{erf}\left(\frac{m}{\sqrt{2\alpha}}\right)}{2 - \sigma \left[\operatorname{erf}\left(\frac{m-\sigma\Delta}{\sqrt{2\alpha}}\right) - \operatorname{erf}\left(\frac{m}{\sqrt{2\alpha}}\right) \right]}. \quad (3.18)$$

Na figura 3.3 apresentamos o diagrama de fase Δ vs. α . Note que neste caso, o parâmetro α desempenha um papel similar ao do ruído térmico (no limite $\alpha = 0$). As linhas de transições, que separam a fase de reconhecimento (R) da fase paramagnética (P), são de segunda ordem e correspondem a diferentes ordens de truncamentos da série ($l = 4, 6$ e 8) e a série completa ($l = \infty$). Note que, como ressaltamos anteriormente, para l igual ou

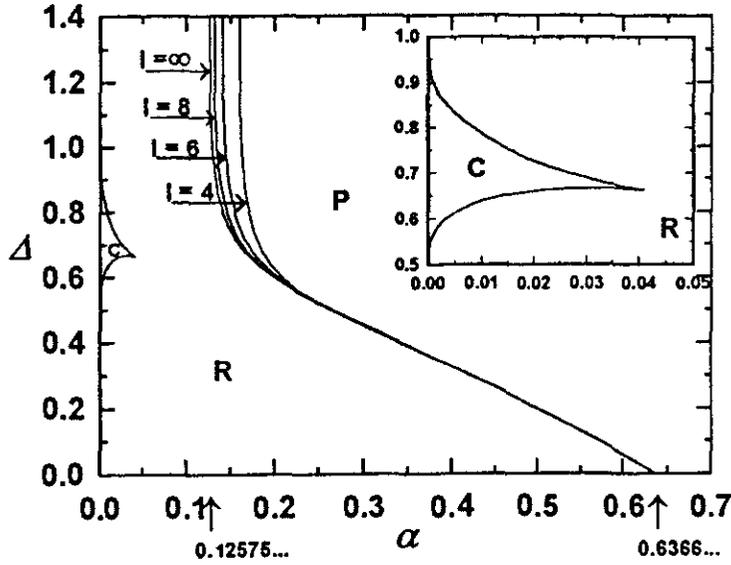


Figura 3.3: Digrama de fase Δ vs. α para $T = 0$. As linhas críticas separando a fase de recuperação da fase paramagnética correspondem a $l = 4, 6, 8$ e ∞ .

maior que 4 as linhas de transições se superpõem para valores pequenos de Δ , mas para Δ grande temos que somar sobre todos os termos. Expandindo a equação (3.18) para pequenos valores de m , obtemos a seguinte expressão para a linha crítica nos parâmetros Δ e α :

$$\alpha_c = \frac{8}{\pi} \left[\frac{1 + \operatorname{erf}\left(\frac{\Delta}{\sqrt{2\alpha_c}}\right) + e^{-\Delta^2/2\alpha_c}}{\left[2 + \operatorname{erf}\left(\frac{\Delta}{\sqrt{2\alpha_c}}\right)\right]^2} \right]^2. \quad (3.19)$$

Para $\Delta = 0$ recobramos a capacidade máxima de armazenamento $\alpha_c = 2/\pi$ obtida em [7]. No limite oposto $\Delta \rightarrow \infty$, a linha de transição tende assintoticamente para $\alpha^* = 32/81\pi \approx 0.12575\dots$

Na figura 3.4 apresentamos o comportamento da superposição m como função de α e Δ para uma configuração inicial próxima do padrão de reconhecimento ($m_0 \approx 1$). Para $\Delta = 0$ recuperamos a linha transição obtida em [7, 22]. Na região para valores de Δ grandes o sistema reconhece pobremente o padrão memorizado, onde $m = 0.5$ para $\alpha = 0$ e vai a zero continuamente para $\alpha^* = 0.12575\dots$

Estudamos também o caso $T \neq 0$. Da equação (3.15), a estabilidade da solução $m = 0$

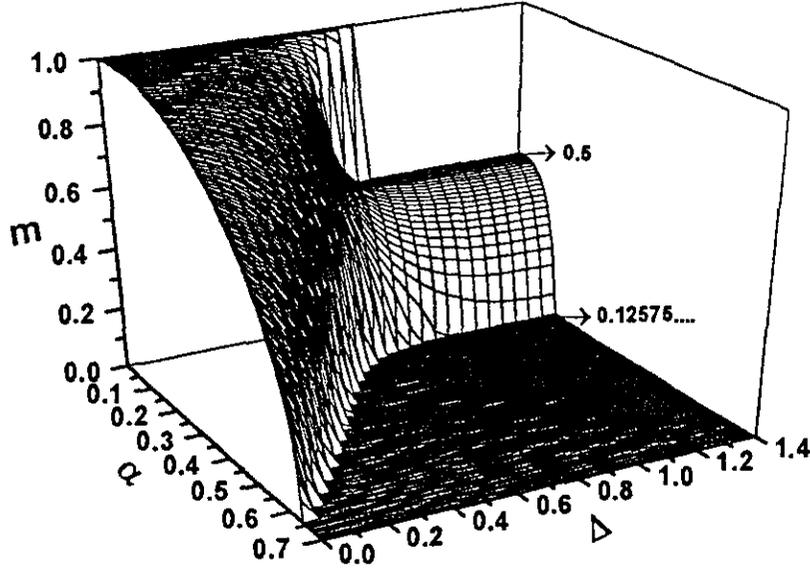


Figura 3.4: A superposição m como função de α e Δ a $T = 0$ com $m_0 = 1$.

é determinada pela condição:

$$\lambda = \left. \frac{\partial F(m)}{\partial m} \right|_{m=0} . \quad (3.20)$$

Para $\lambda > 1$ (< 1) a solução $m = 0$ é instável (estável). O caso $\lambda = 1$ corresponde a superfície crítica no espaço (T, α, Δ)

$$T_c = 2 \left[\frac{2 + \int Dz(f_+ - f_0) - \int Dz f_0^2(1 + \int Dz f_+) - \int Dz f_+^2(1 - \int Dz f_0)}{[2 + \int Dz(f_+ - f_0)]^2} \right] , \quad (3.21)$$

onde f_+ e f_0 são dados por

$$f_+ = \tanh \left(\frac{\sqrt{\alpha} z + \Delta}{T_c} \right) \quad f_0 = \tanh \left(\frac{\sqrt{\alpha} z}{T_c} \right) .$$

Na figura 3.5 mostramos o diagrama de fase no espaço T, α e Δ . Para $\Delta = 0$ e aumentando o valor do α , a linha crítica toca o eixo $T = 0$ em $\alpha_c = 2/\pi$ [7]. Ao longo da superfície crítica o sistema apresenta uma transição de fase de segunda ordem da fase de reconhecimento (abaixo) para a fase paramagnética (acima). Aumentando Δ , a superfície crítica tende a uma forma constante, com valores assintóticos $\alpha = 0$ ($T^* = 0.444\dots$) e $T = 0$ ($\alpha^* = 0.12575\dots$). Note que, abaixo desta superfície crítica, existe um volume

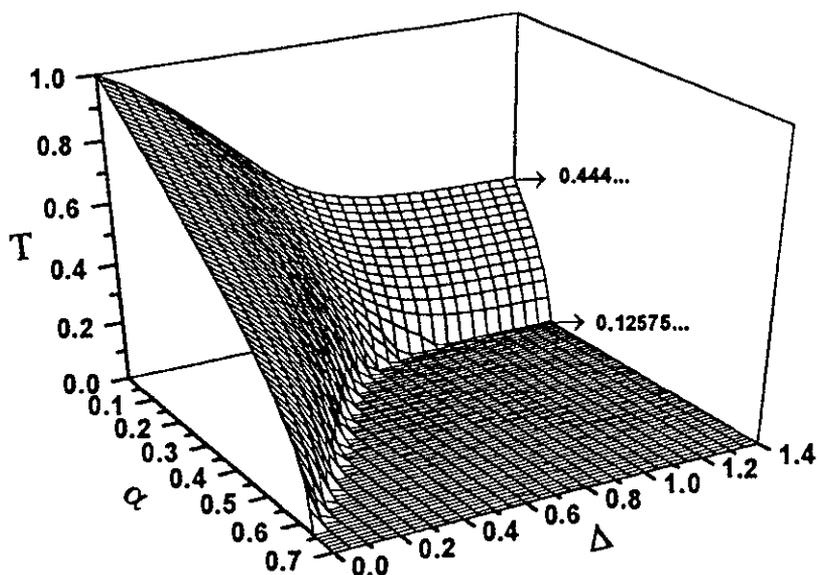


Figura 3.5: Superfície crítica separando a fase de reconhecimento (abaixo) da fase paramagnética (acima) no espaço dos parâmetros (T, α, Δ) .

onde as duas soluções de reconhecimento coexistem, mas não é mostrado neste diagrama.

3.3 Simulação Numérica

Para testar nossa aproximação analítica, realizamos simulações numéricas do modelo para o caso $T = 0$. Apesar da impossibilidade de implementar numericamente este modelo, devido as condições $C \ll \ln N$ e $C \gg 1$ (limite de ultra diluição), trabalhos recentes realizados para conectividade C pequena e com grandes redes têm mostrado boa concordância com os resultados analíticos [22, 31, 53]. Desta forma, a condição de ultra diluição, no que se refere às propriedades de reconhecimento, pode relaxar para uma outra menos rígida $C \ll N$. Para implementar a evolução dinâmica do sistema usamos uma atualização em paralelo. O campo local pode ser convenientemente escrito em termos da superposição entre o estado $\{S_i(t)\}$ do sistema e a primeira memória ξ_i^1 (ver equação

3.6) como segue:

$$S_i(t+1) = \text{Sinal} \left[\xi_i^1 m(t) + \frac{1}{C} \sum_{q=1}^C \sum_{\mu \neq 1}^p \xi_i^\mu \xi_q^\mu S_q(t) - \frac{\Delta}{2} (S_i(t) + 1) \right]. \quad (3.22)$$

O primeiro termo (sinal) tende a alinhar o sistema com o padrão memorizado, o segundo atua como um ruído e representa os padrões não correlacionados com a memória e o último termo simula o efeito dos períodos refratários (somente os estados com $S(t) = 1$ contribuem).

Para analisar a habilidade de reconhecimento partimos de uma configuração inicial correlacionada com a primeira memória. Após um transiente inicial, medimos a média temporal da superposição $m(t)$ entre o estado do sistema e o padrão de reconhecimento. Este procedimento foi repetido para 50 diferentes amostras, usando diferentes memórias, condições iniciais e sequências de números aleatórios, afim de calcularmos a média configuracional do m . Usamos redes com $N = 40000$ e 80000 e conectividade $C = 40$ e 80 . Estes tamanhos de rede estão além dos utilizados usualmente na literatura. Tamanhos maiores podem ser implementados (chegamos até $N = 180000$), mas o cálculo de médias configuracionais requer de grandes tempos de CPU.

Na figura 3.6 apresentamos o diagrama de fase Δ vs. α para $T = 0$. A linha cheia corresponde a linha crítica analítica no limite $C \rightarrow \infty$ e os círculos cheios são os resultados da nossa simulação numérica para $N = 80000$ e $C = 80$. Note que para pequenos e grandes valores de Δ ambos os resultados concordam muito bem. No primeiro caso (Δ pequeno), o efeito de realimentação devido aos estados $S = 1$ não é muito importante comparado ao termo de sinal. Nesta região, um processo Markoviano dado por uma aproximação de campo médio apresenta um resultado similar.

Quando Δ é maior que o termo de sinal, a evolução temporal leva o sistema para um regime onde:

- todos aqueles neurônios para o qual a primeira memória está em um estado inativo ($\xi_i^1 = -1$) alinham-se com a memória ($S_i = -1$);

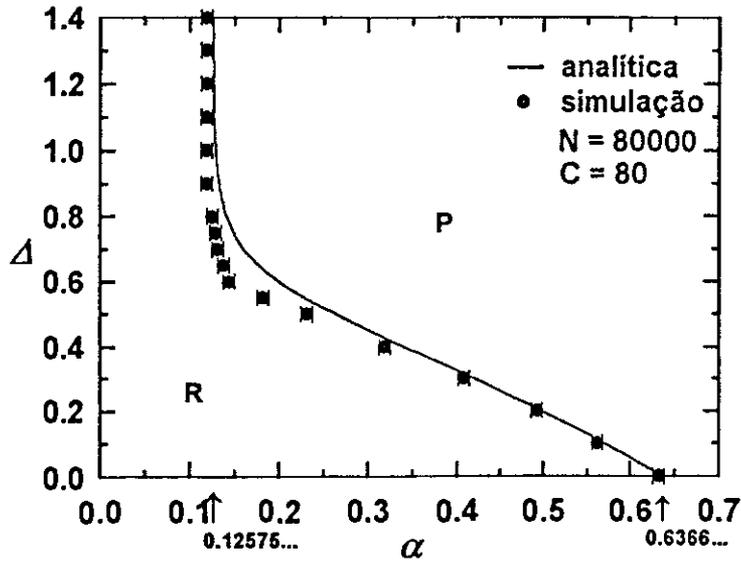


Figura 3.6: Linha crítica separando a fase de reconhecimento da fase paramagnética no plano Δ vs. α a $T = 0$. A linha cheia corresponde a solução analítica para $l \rightarrow \infty$ e os círculos cheios correspondem aos resultados da simulação numérica com $C = 80$ e $N = 80000$.

- todos aqueles neurônios para o qual a primeira memória está em um estado ativo ($\xi_i^1 = 1$) oscilam entre os estados ativo e inativo.

Estes comportamentos são desestabilizados pelo efeito das outras memórias quando α cresce. Para α pequeno o sistema é caracterizado por um regime periódico de ciclo dois em torno de $m(t \rightarrow \infty) \simeq 0.5$.

Nas figuras 3.7 e 3.8 apresentamos a superposição como função de α para $T = 0$ ao longo de dois cortes, $\Delta = 0.2$ e 0.7 respectivamente. Observe que para $\Delta = 0.2$ ambas as soluções analítica (linha cheia) e numérica (círculos) concordam bem. Para $\Delta = 0.7$ (figura 3.8), apesar de não esperarmos bons resultados com a nossa aproximação nesta região (devido a forte competição entre o termo do sinal produzido pelo padrão memorizado e o termo refratário), a solução analítica descreve qualitativamente bem os resultados numéricos. Em particular, note que nesta região a simulação também descreve a coexistência de fases. A linha tracejada corresponde a solução instável (analítica) que separa as duas bacias atratoras.

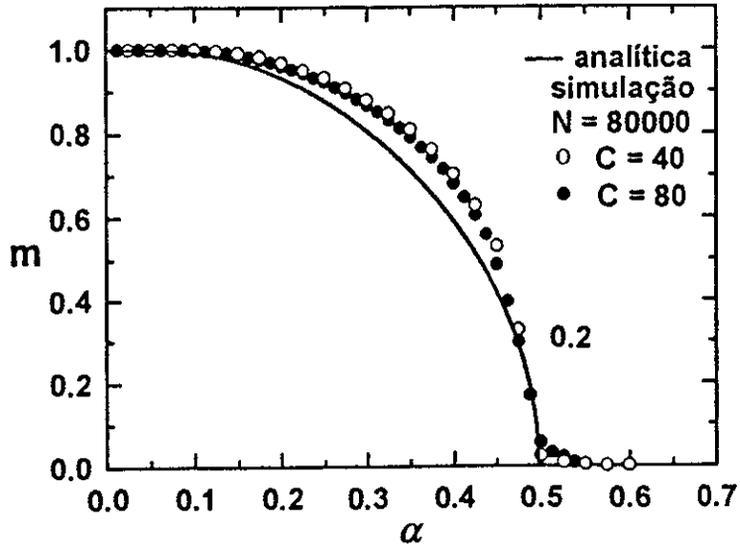


Figura 3.7: m vs. α para $\Delta = 0.2$, $m_0 = 1$ e $T = 0$. A linha cheia corresponde à solução analítica e os círculos aos resultados das simulações para $N = 80000$ com $C = 40$ (círculos vazios) e $C = 80$ (círculos cheios).

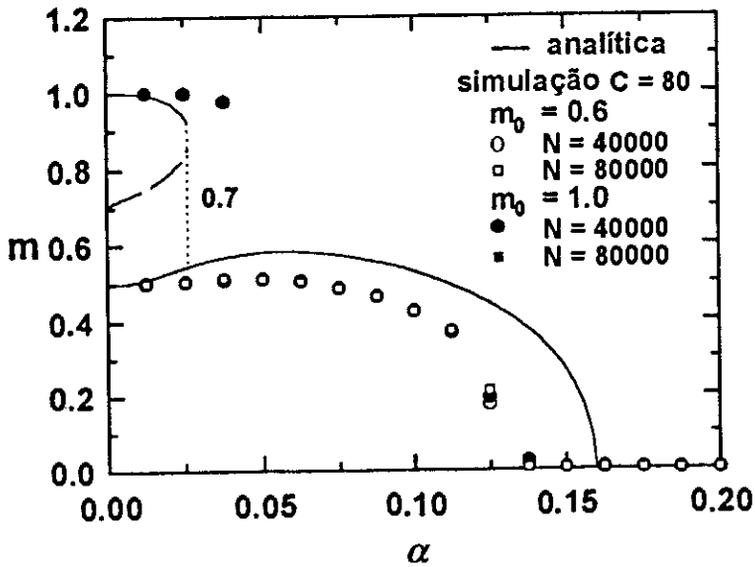


Figura 3.8: m vs. α para $\Delta = 0.7$, $C = 80$ e $T = 0$. As linhas cheias correspondem à solução analítica e os símbolos aos resultados das simulações: os círculos e os quadrados vazios correspondem a $N = 40000$ e $N = 80000$ respectivamente, com $m_0 = 0.6$. Os círculos e os quadrados cheios correspondem a $N = 40000$ e $N = 80000$ respectivamente, com $m_0 = 1$.

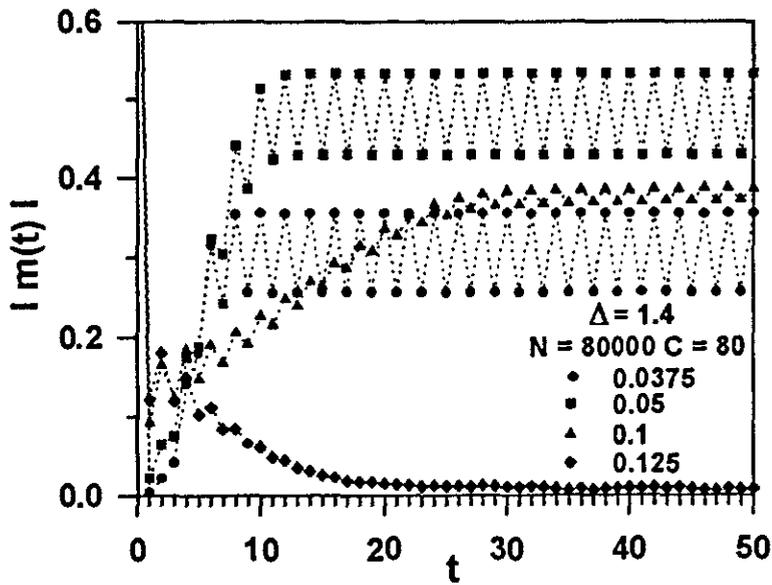


Figura 3.9: $|m(t)|$ vs. t para $C = 80$, $N = 80000$, $\Delta = 1.4$ e $\alpha = 0.0375$ (círculos), 0.05 (quadrados), 0.1 (triângulos) and 0.125 (lozângulos).

Na figura 3.9 representamos $|m(t)|$ versus o tempo para $\Delta = 1.4$, $N = 80000$, $C = 80$ e diferentes valores de α . Para valores grandes de Δ ($\Delta > 1$), no primeiro passo de tempo, para α pequeno, a dinâmica leva o sistema para um novo estado com $m \approx 0$ (positivo ou negativo), ou seja, o sistema pode evoluir para a memória ou anti-memória. Como em ambos os casos o sistema sempre reconhece, preferimos tomar o módulo da superposição limitando assim o movimento do sistema no espaço dos padrões. Para $\alpha = 0.0375$ a superposição converge para um ciclo de ordem dois com valor médio em torno de $m \simeq 0.3$, sinalizando a existência de estados espúrios (simétricos ou assimétricos, não previsto no desenvolvimento analítico). Quando α cresce, os estados espúrios desaparecem e a superposição atinge o atrator cujo valor médio é $m \simeq 0.5$. A amplitude de oscilação dos atratores ciclo dois diminuem com α , permanecendo sempre próximo do seu valor médio. Para α grande, o atrator converge para um regime com ciclos de ordem maior que dois ou órbitas caóticas.

3.4 Conclusões

Neste capítulo estudamos a dinâmica de uma rede de neurônios com diluição aleatória e assimetria nos acoplamentos sinápticos, onde incluímos um limiar para modelar os períodos refratários. Este limiar, que depende do tempo, produz um efeito de realimentação sobre a dinâmica do sistema. Desenvolvemos um método aproximado que permite estudar o comportamento de longo tempo do sistema. Desprezando as correlações temporais, é possível considerar toda a história do sistema e portanto analisar os efeitos dos períodos refratários na dinâmica e a habilidade de reconhecimento do modelo.

Obtivemos o diagrama de fase completo do modelo e comparamos com os resultados obtidos através da simulação numérica. Ao contrário do modelo de Hopfield totalmente conectado e com períodos refratários [45], o qual para valores grandes de Δ desestabiliza completamente os padrões armazenados, na versão ultra diluída verificamos que, para pequenos valores de α , o sistema pode reconhecer sempre os padrões, independentemente do valor de Δ . Para Δ grande, na fase de reconhecimento, os atratores são periódicos (ciclo dois) e o valor médio de m concorda muito bem com os valores obtidos no desenvolvimento analítico (atratores ponto fixo). Para valores intermediários de Δ e para α pequeno, uma inesperada fase de reconhecimento aparece e a rede pode reconhecer o padrão memorizado com alta e baixa qualidade, dependendo de se o estado inicial está próximo ou longe da memória. Nesta região, o termo de Hopfield (o sinal) e o termo refratário competem, dando lugar a estas duas diferentes soluções.

Nossos resultados analíticos concordam muito bem com os obtidos na simulação numérica para valores pequenos e grandes do parâmetro Δ . Já na região intermediária $0.5 < \Delta < 1.0$, onde não esperamos bons resultados com a nossa aproximação, pois as correlações temporais são importantes nesta região, este prediz qualitativamente bem os resultados.

Capítulo 4

Propagação de Danos em Uma Rede de Neurônios Diluída e Assimétrica

4.1 Propagação de Danos

Começamos este capítulo descrevendo a técnica de *propagação de dano* e fazendo uma pequena resenha sobre os principais resultados obtidos com ela. Embora tenha sido introduzida originalmente para estudar a sensibilidade às condições iniciais em autômatos celulares, esta técnica mostrou ser também de grande utilidade no estudo das propriedades dinâmicas de diferentes modelos magnéticos.

A técnica consiste basicamente em monitorar a evolução temporal da distância de Hamming ou dano entre duas réplicas inicialmente diferentes de um dado sistema, submetidas ambas à mesma dinâmica. Quando a dinâmica é probabilística, então utiliza-se também uma única sequência de números aleatórios para atualizar as duas réplicas. A distância de Hamming é definida como a fração de sítios nos quais as duas configurações diferem. A dependência do comportamento assintótico $t \rightarrow \infty$ do dano com o valor inicial da distância de Hamming e com qualquer outro parâmetro relevante do sistema permite caracterizar a dinâmica do modelo e construir um diagrama de fases dinâmico.

Operacionalmente, o dano $dh(t)$ entre duas configurações binárias quaisquer, como a

que trataremos neste capítulo, pode ser definido como:

$$dh(t) = \frac{1}{2N} \sum_{i=1}^N |\sigma_i(t) - \rho_i(t)| ,$$

onde $\sigma_i(t)$ e $\rho_i(t)$ representam as duas réplicas de um único sistema (em caso não binários esta definição pode ser adequadamente generalizada).

Esta técnica foi introduzida, como já fora dito, afim de estudar a sensibilidade às condições iniciais em sistemas com infinitos graus de liberdade e nos quais o tempo, o espaço e as variáveis dinâmicas são todos discretos. Neste caso, calcular quantidades tais como expoentes de Liapunov torna-se muito difícil. Se inicialmente as duas réplicas diferem num único sítio ($dh(t=0) \rightarrow 0$ quando $N \rightarrow \infty$), pode-se clasificar a sua dinâmica como:

- **caótica:** se $dh(\infty) \rightarrow \text{cte} \neq 0$ quando $dh(0) \rightarrow 0$;
- **congelada:** se $dh(\infty) \rightarrow 0$ quando $dh(0) \rightarrow 0$.

Assim, a técnica de propagação de danos foi aplicada a diferentes autômatos celulares, como por exemplo o modelo de Kauffman [20, 54], (Kauffman introduziu o conceito de autômatos celulares para modelar um sistema de genes [55, 56]), e o modelo de Domany–Kinzel [57, 58, 59, 60], entre outros.

Muitos modelos magnéticos, como por exemplo o modelo de Ising, compartilham com os autômatos celulares o fato de terem suas variáveis dinâmicas discretizadas e estas estarem associadas a uma rede discreta. Ainda mais: por não possuírem uma dinâmica própria, é comum designar uma dinâmica estocástica na qual o tempo de atualização também é discreto. Assim surgiu o interesse de aplicar a mesma técnica de propagação de danos usada em autômatos celulares nestes modelos magnéticos.

Entre os modelos magnéticos, o modelo de Ising ferromagnético [61, 62, 63, 64, 65] foi o primeiro a ser estudado. Ao contrário do que acontece com as quantidades de equilíbrio, observou-se que o comportamento da distância de Hamming dependia drasticamente da dinâmica estocástica utilizada. Em particular, quando utilizaram um processo

banho térmico, encontrou-se que o comportamento era sempre não caótico (ou seja, se $dh(0) \rightarrow 0$ então $dh(\infty) \rightarrow 0$). Mas, estudando o caso em que o dano inicial era macroscópico, observou-se que, enquanto para baixas temperaturas a distância de Hamming final depende do dano inicial, para altas temperaturas o dano é sempre nulo. Estudos muito cuidadosos mostraram que a temperatura de transição entre estes dois regimes coincide com a temperatura de transição termodinâmica entre as fases ferro e paramagnética. Este comportamento foi observado para os casos $d = 2$ e $d = 3$.

Quando a mesma técnica foi aplicada ao mesmo modelo mas utilizando uma dinâmica de *Glauber*, o resultado foi completamente diferente. Mesmo começando com um dano inicial muito pequeno ($dh(0) = 1/N$) foi possível distinguir duas fases, uma congelada de baixas temperaturas e uma caótica de altas temperaturas. Neste caso encontrou-se que a temperatura de transição dinâmica *não coincide* com a temperatura de transição termodinâmica.

Muito recentemente foi mostrado que esta técnica não somente permite caracterizar a dinâmica do modelo ou estudar possíveis correlações entre as transições dinâmicas e estáticas, mas também é muito útil para calcular de forma muito eficiente diferentes expoentes dinâmicos do sistema [66, 67].

Quando se estudou o efeito de colocar o sistema num campo magnético externo uniforme, observou-se que este campo não é o campo conjugado ao parâmetro de ordem (o dano). Pelo contrário, na presença de campo magnético a transição não é destruída. Tsallis e Martins introduziram um campo externo para o caso do autômato probabilístico de Domany–Kinzel. Ele é definido como a fração de vezes que as duas réplicas são atualizadas utilizando diferentes números aleatórios. Assim, no limite de campo máximo ($h = 1$), ambas as réplicas evoluem independentemente e o dano é sempre não nulo. No limite em que este campo vai para zero, eles mostraram que a susceptibilidade associada diverge na transição dinâmica e a transição desaparece, ou seja, eles mostraram que este campo é efetivamente o campo externo conjugado ao parâmetro de ordem. O mesmo campo foi aplicado aos modelos de Ising e Potts ferromagnéticos bidimensionais mostrando-se que é o campo conjugado ao dano.

Muitos outros modelos magnéticos foram estudados utilizando esta técnica, e em particular os modelos de vidro de spin ganharam particular interesse [68, 69, 70, 71, 72, 73]. Recentemente também se aplicou esta técnica ao modelo de Potts 3 estados [74] e q-estados [75].

Nos modelos de redes neuronais, alguns trabalhos foram desenvolvidos usando esta técnica de propagação de dano [7, 53, 76, 77]. Ela é particularmente interessante nestes casos já que permite obter valiosa informação sobre a estrutura do espaço de fases de modelos que não podem ser resolvidos analiticamente. Como mostraremos, o espaço de fase destes sistemas, assim como acontece nos vidros de spin, apresenta um comportamento dinâmico muito rico, principalmente em redes com conexões sinápticas assimétricas.

No mesmo trabalho no qual Derrida et al. [7] resolveram a dinâmica do modelo de Hopfield diluído e assimétrico, eles obtiveram também em forma analítica uma expressão para a correlação $q(t)$ entre duas configurações inicialmente diferentes

$$q(t) = \frac{1}{N} \sum_{i=1}^N \langle S_i(t) \bar{S}_i(t) \rangle ,$$

válida para todo tempo t . No limite C e $p \rightarrow \infty$ com $\alpha \equiv (p-1)/C$ ($C \ll \ln N$), a evolução do $q(t+1)$ em função do $q(t)$ e das superposições $m(t)$ e $\bar{m}(t)$, é dada por (ver equação 27 em [7]):

$$q(t+1) = \int_{-\infty}^{\infty} Dy \int_{-\infty}^{\infty} Dz \tanh \beta (m(t) - a_+ y - a_- z) \tanh \beta (m(t) - a_+ y + a_- z)$$

onde $T = \beta^{-1}$ é a temperatura generalizada $T = T_0/C$,

$$a_{\pm} = \sqrt{\alpha(1 \pm q(t))}$$

e

$$Du = \frac{du}{\sqrt{2\pi}} e^{-u^2/2} \quad u = y \text{ e } z .$$

No limite $t \rightarrow \infty$ $m(t)$ e $\bar{m}(t) \rightarrow m$ dado por (1.32), enquanto que $q(t)$ converge para

um ponto fixo da equação acima. No limite $T \rightarrow 0$ obtem-se a seguinte equação para a correlação:

$$q = -1 + \sqrt{\frac{2}{\pi}} \int_0^\infty dz \operatorname{erf}\left(\frac{za_+}{\sqrt{2a_-}}\right) \left[e^{-\frac{1}{2}(z+\frac{m}{a_+})^2} + e^{-\frac{1}{2}(z-\frac{m}{a_+})^2} \right] \quad (4.1)$$

onde a_\pm já foi definido anteriormente. Eles encontraram que na fase de reconhecimento, esta correlação é sempre menor que 1 para qualquer valor de α finito. Em particular, para valores pequenos de α eles obtiveram:

$$q(t \rightarrow \infty) \simeq 1 - \frac{8}{\pi^2} \exp(-1/\alpha).$$

Isto quer dizer que as duas configurações não se encontram no atrator, indicando a existência de uma nova classe de atratores de reconhecimento, muito mais complexa que os atratores ponto fixo usualmente encontrados no modelo de Hopfield. Eles sugeriram que o modelo deveria reconhecer através de uma *nuvem de pontos fixos* ou mesmo *atratores caóticos*.

Recentemente, Wong e Ho [77] desenvolveram novas ferramentas analíticas para estudar a natureza dos atratores em sistemas dinâmicos. Estas ferramentas são a distribuição da atividade da rede, a evolução temporal do dano, atividade de dano e a correlação temporal de dano. Eles estudaram os atratores de reconhecimento do modelo de Hopfield ultra diluído e mostraram que estes atratores de reconhecimento não são nuvens de atratores, mas consistem de um simples atrator caótico para cada padrão armazenado. Este resultado é muito interessante do ponto de vista biológico, dado que atratores ponto fixo são altamente improváveis em sistemas de neurônios reais.

Neste capítulo, estudaremos através de simulação numérica, a natureza dos atratores de reconhecimento do modelo de Hopfield diluído e assimétrico, no caso em que a conectividade por neurônio C e o número de neurônios N satisfazem a relação $C \ll N$, na ausência de ruído. Recentemente foi mostrado [22], mediante um estudo numérico, que a capacidade de reconhecimento do modelo diluído assimetricamente no caso $C \ll N$ é

bem descrita pelas equações obtidas no limite de ultra diluição $C \ll \ln N$. Isto é muito importante do ponto de vista prático, pois enquanto que o limite de ultra diluição é um modelo teórico impossível de ser implementado tanto em simulações como em dispositivos, o caso $C \ll N$ não somente é simples de simular, mas também é muito eficiente em termos de simulação numérica, se comparado com outros modelos neuronais. Nós estudaremos se a mesma coisa acontece com a natureza caótica dos atratores de reconhecimento, ou seja, se no caso $C \ll N$ o sistema reconhece somente por meio de atratores caóticos como foi mostrado em [7, 77]. Para isto usaremos a técnica de propagação de dano, medindo a distância de Hamming entre duas configurações, inicialmente muito próximas uma da outra (em termos da distância de Hamming) e ambas tendo uma projeção inicial finita com um único padrão armazenado. A análise do comportamento assintótico do dano caracteriza a natureza dinâmica dos atratores de reconhecimento. Mostraremos que o sistema apresenta, na fase de reconhecimento, uma transição entre dois regimes: um de reconhecimento por atratores pontos fixos e outro de reconhecimento por meio de atratores caóticas. Generalizamos também o conceito de campo externo h conjugado ao dano introduzido recentemente na literatura [59] no contexto de autômatos celulares probabilísticos (ver também [60, 65, 74]) de forma que seja aplicável a sistemas determinísticos como o modelo de Hopfield diluído na ausência de ruído ($T = 0$). Finalmente calculamos a *susceptibilidade de dano* associada ao parâmetro dh e mostramos que ela tende a divergir na de transição caótica–congelada.

4.2 O Modelo

O modelo a estudar é o mesmo introduzido no capítulo anterior mas sem o termo que modela os períodos refratários. Os neurônios são atualizados em paralelo segundo a regra determinística

$$S_i(t+1) = \text{Sinal} \left(\sum_{j \neq i}^N T_{ij} S_j(t) \right) . \quad (4.2)$$

T_{ij} é a matriz sináptica que conecta os neurônios pré e pós-sinápticos j e i e cujos elementos são dados por

$$T_{ij} = C_{ij} J_{ij} . \quad (4.3)$$

Os C_{ij} são variáveis aleatórias cuja distribuição de probabilidade é dada por (3.5), a qual introduz diluição e assimetria simultaneamente na rede. Para $C_{ij} = 1$, (correspondendo as sinapses que sobreviveram no processo de diluição) o T_{ij} é dado pela matriz sináptica usual de Hopfield

$$T_{ij} = J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu . \quad (4.4)$$

Os $\xi_i^\mu = \pm 1$ ($\mu = 1, 2, \dots, p$) indicam os valores das p memórias do i -ésimo neurônio, geradas aleatoriamente, que serão armazenadas na rede. Por outro lado, $C_{ij} = 0$ indica que a conexão sináptica foi cortada no processo de diluição.

4.3 Método e Resultados Numéricos

Para construir a rede, escolhemos aleatoriamente as p memórias e os C vizinhos de cada neurônio e construímos a matriz sináptica (de tamanho $N \times C$). Consideramos duas cópias idênticas da rede, e fazemos o sistema evoluir segundo a dinâmica 4.2. Partimos de duas configurações iniciais diferentes $\{S_i(0)\}$ e $\{\bar{S}_i(0)\}$, ambas tendo superposição finita somente com um dos padrões (o primeiro) e superposição macroscópica nula com todas as outras memórias. Após um transiente inicial t_0 , que depende sobre às condições iniciais, medimos a distância de Hamming

$$dh(t) = \frac{1}{2N} \sum_{i=1}^N |S_i(t) - \bar{S}_i(t)| \quad (4.5)$$

entre as duas réplicas. Seguindo, calculamos a média temporal sobre um certo número de passos de tempo t_{max}

$$\bar{dh} = \frac{1}{t_{max} - t_0} \sum_{t=t_0+1}^{t_{max}} dh(t) . \quad (4.6)$$

Repetimos este procedimento para 50 amostras diferentes, usando diferentes condições iniciais e números aleatórios, para calcularmos a média configuracional dh de \bar{dh} . A distância de Hamming dh está relacionada ao parâmetro q dado em (4.1) através da seguinte relação:

$$dh = \frac{1}{2}(1 - q). \quad (4.7)$$

Como estamos interessados em estudar as possíveis relações entre as transições dinâmicas e de reconhecimento, e em particular suas dependências com o tamanho do sistema N e com a conectividade C , calculamos a média configuracional da superposição m com o primeiro padrão ($\mu = 1$) definido como:

$$m = \frac{1}{t_{max} - t_0} \sum_{t=t_0+1}^{t_{max}} m^1(t), \quad (4.8)$$

onde

$$m^1(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^1 S_i(t) . \quad (4.9)$$

Na figura 4.1 mostramos o comportamento do dh e do m como função do α para $C = 40$ e para dois valores de N (40000 e 80000). Os resultados numéricos (no regime $C \ll N$) são representados por símbolos (cujos tamanhos são da ordem das barras de erro) e as linhas sólidas correspondem aos resultados analíticos no regime de extrema diluição [7]. Com relação ao comportamento do $m(\alpha)$, como mostrado em [22], confirmamos que ambos os resultados concordam muito bem. Note que na região para $\alpha < 0.5$ os parâmetros dh e m parecem ser insensíveis ao tamanho N do sistema, enquanto que próximo do α_c , esta dependência é bastante acentuada. Por outro lado, o comportamento do dh apresenta três fases diferentes no regime $C \ll N$.

- para $\alpha < \alpha_d$ o dano final é sempre zero, isto é, o sistema reconhece com atratores não caóticos;
- para valores intermediários ($\alpha_d < \alpha < \alpha_c$) o dano final é diferente de zero e menor que 0.5, indicando que os atratores são cíclicos ou caóticos mais eles ainda são

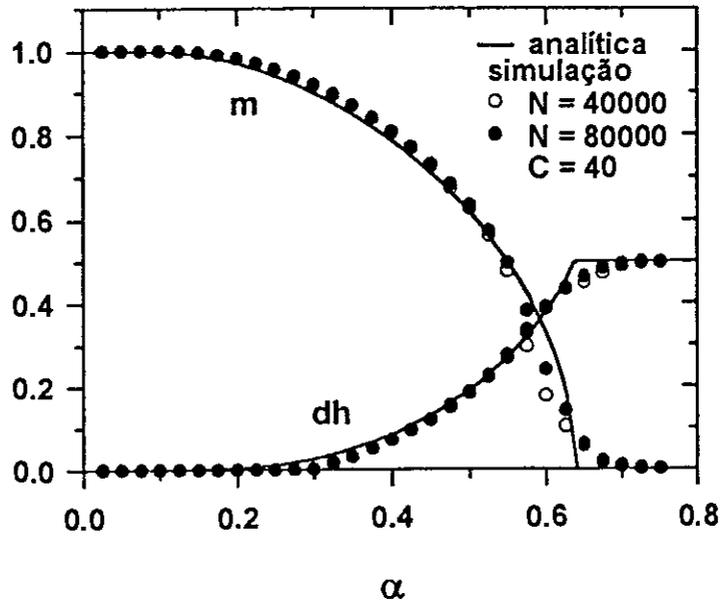


Figura 4.1: A distância de Hamming dh e a superposição m em função do α para $C = 40$ e $N = 40000$ (círculos vazios) e 80000 (círculos cheios). As linhas sólidas correspondem as soluções analíticas no regime de ultra diluição.

correlacionados um ao outro e com o padrão memorizado, isto é, o sistema reconhece através de atratores cíclicos ou caóticos;

- para $\alpha > \alpha_c$, na fase paramagnética, as réplicas são descorrelacionadas entre si ($q = 0$ e $dh = 0.5$) e com a memória.

Observamos que para α pequeno (atratores não caóticos) o limite de extrema diluição não prediz este comportamento, ele prediz $dh \neq 0$ para qualquer $\alpha > 0$. O α_d corresponde ao valor de α da transição dinâmica caótica-não caótica.

Na figura 4.2 apresentamos dh e m como função de α para $N = 80000$ e diferentes conectividades $C = 10, 20$ e 40 . Note que para valores intermediários de α (na região da transição caótica-não caótica) dh e m apresentam uma forte dependência com a conectividade, e aumentando C os resultados da simulação aproximam-se do resultado analítico. Próximo do α_c a dependência com C é bem menor.

Para estudar a natureza dos atratores nas diferentes regiões, após um dado transiente, armazenamos a última configuração microscópica do sistema (configuração de referência) e deixamos evoluir uma das réplicas até que o sistema volte a esta configuração, desde

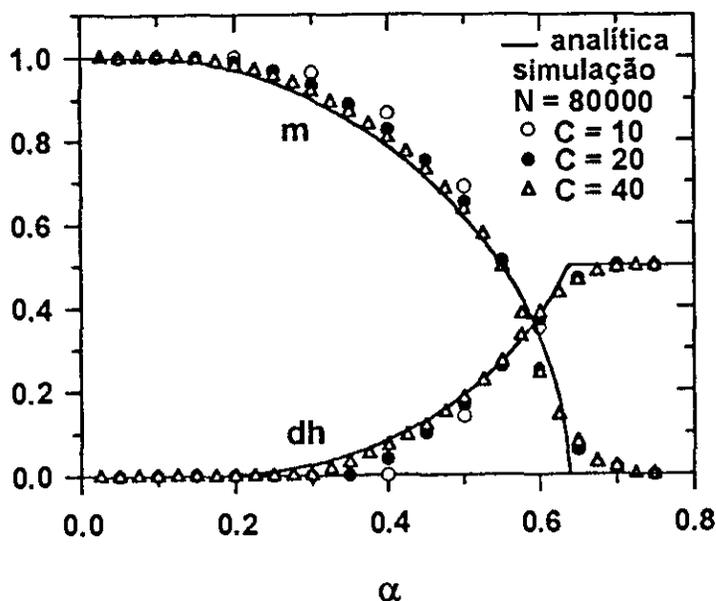


Figura 4.2: dh e m em função do α para $N = 80000$ e para $C = 10$ (círculos vazios), $C = 20$ (círculos cheios) e $C = 40$ (triângulos). As linhas sólidas correspondem as soluções analíticas no regime de ultra diluição.

que seja, no máximo, 100 passos de tempo (em unidade de Monte Carlo). Se nesse tempo o sistema não volta para a configuração de referência, dizemos que o atrator é caótico. Caso contrário, surgem outros regimes atratores como ponto fixo e ciclos de períodos finito (menor que 100). Neste caso, observamos nas simulações que o tempo de convergência é pequeno (menor que 60 passos de tempo) e que quase todos os ciclos finitos são de períodos dois ou de múltiplos de 2 menores que 20.

Na figura 4.3 apresentamos a frequência com a qual surgem os diferentes tipos de atratores, como função de α para $N = 80000$ e diferentes conectividades C ($C = 20$ no topo, $C = 40$ no meio e $C = 80$ na base). Classificamos somente três regimes: atratores ponto fixo (PF), ciclos de período finito menor ou igual a 100 (CF) e ciclos de período maior que 100 ou "atratores caóticos" (Caos). Podemos notar três comportamentos bem distintos: uma região (para α pequeno) dominada por atratores ponto fixo, uma outra intermediária onde coexistem os três regimes (PF, CF e Caos) e a terceira onde existe somente atratores caóticos. Notemos que a transição dinâmica observada nas figuras 4.1 e 4.2 corresponde à transição entre PF e Caos (os CF que aparecem somente nesta fronteira,

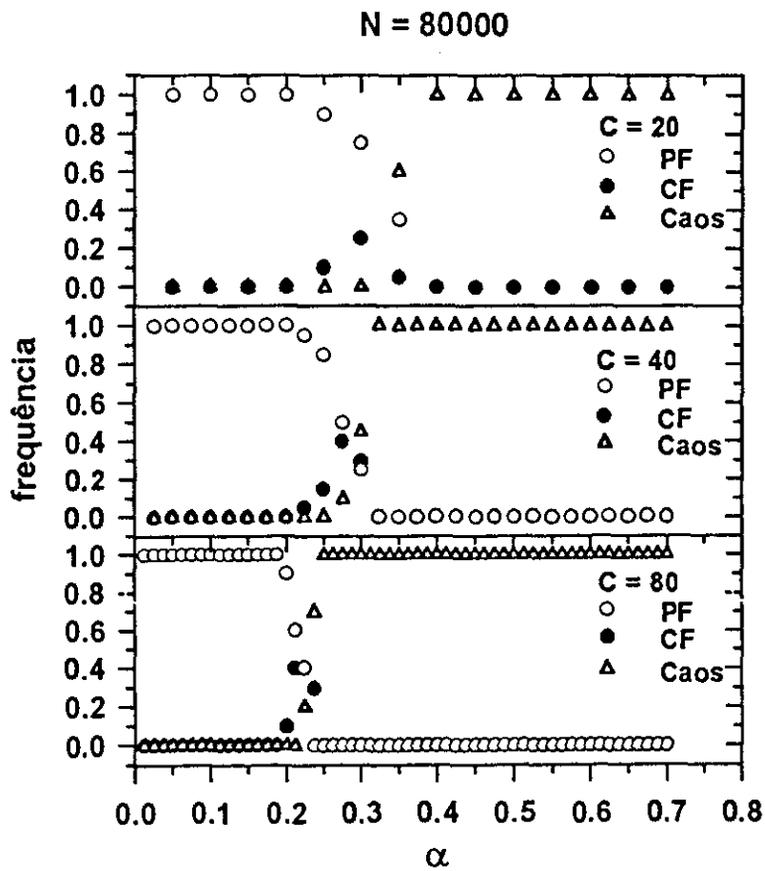


Figura 4.3: A frequência com a qual aparecem os diferentes tipos de atratores em função de α para $N = 80000$ e diferentes valores de C : 20 (no topo), 40 (no meio) e 80 (na base).

tendem a desaparecer com o aumento da conectividade). Um fato interessante é que o estreitamento da banda de CF com o crescimento de C , ocorre devido ao avanço da fase caótica sobre os CF. Este comportamento não era esperado. Pelo contrário, se supomos que ao aumentar C nos aproximamos do modelo de Hopfield, esperaríamos que a fase PF avance sobre a fase caótica. Para C finito e aumentando o N (não mostrado nas figuras) a fronteira da fase caótica permanece inalterada porém, ocorre uma diminuição na frequência dos ciclos finitos e um crescimento na frequência de ponto fixo, de modo que no limite $N \rightarrow \infty$, a transição dinâmica parece ser entre Caos e ponto fixo, sem que haja ciclos finitos.

A fim de caracterizar melhor a transição dinâmica caótica–não caótica, estudamos o efeito do campo externo h (conjugado ao parâmetro de ordem dh), introduzido recentemente na literatura por Tsallis e Martins [59]. Eles mostraram que o campo h quebra a transição dinâmica no autômato celular de Domany–Kinzel e que a susceptibilidade de dano, calculada numericamente, tende a divergir no limite $N \rightarrow \infty$. Este campo foi também aplicado para estudar transições dinâmicas em sistemas magnéticos; modelo de Ising ferromagnético em duas dimensões [65] e Potts [74]. O campo h é definido como *a frequência na qual diferentes números aleatórios são usados para atualizar as duas réplicas*. No nosso caso, dado que a dinâmica é determinística, escolhemos aleatoriamente alguns sítios (os mesmos para as duas réplicas) e trocamos estes sítios copiando outros escolhidos aleatoriamente dentro de cada réplica. Isto permite introduzir uma fonte de dano e preservar as características macroscópicas da rede como por exemplo a atividade média. Desta forma, a fração de sítios alterados define o campo externo h .

A susceptibilidade associada ao dano dh é definida como:

$$\chi_{dh} = \left. \frac{\partial dh}{\partial h} \right|_{h=0} \quad (4.10)$$

Dado que não existe por enquanto uma generalização do teorema de flutuação–dissipação que permita calcular a susceptibilidade a campo nulo, introduzimos um pequeno campo e estudamos o limite em que o campo vai para zero, ou seja, calculamos a derivada numérica

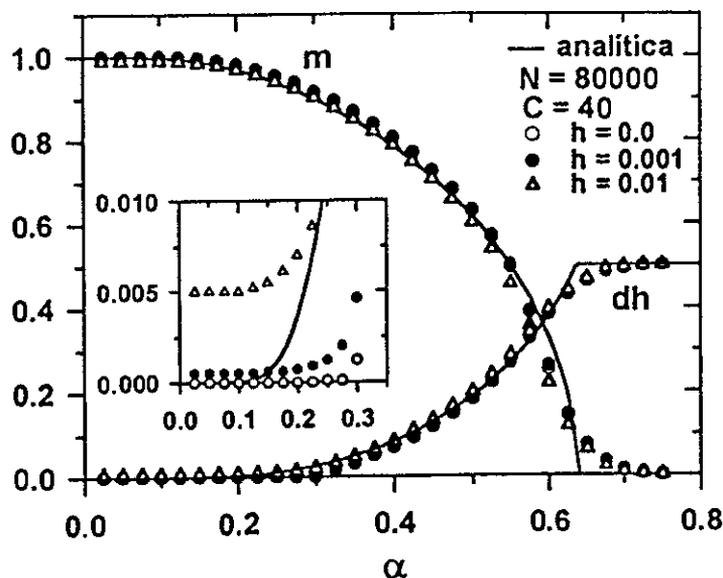


Figura 4.4: dh e m em função de α para $N = 80000$, $C = 40$ e $h = 0.0$ (círculos vazios), 0.001 (círculos cheios) e 0.01 (triângulos). As linhas sólidas correspondem ao regime ultra diluído. Note que para $h > 0$ o dh é sempre diferente de zero (detalhe).

do dano com respeito ao campo fazendo o campo tender para zero.

Na figura 4.4 apresentamos dh e m como função de α para $N = 80000$, $C = 40$ e diferentes valores do campo $h = 0.0$, 0.001 e 0.01 . Note que a transição da fase de reconhecimento é preservada e que h quebra a transição dinâmica do dh (ver detalhe). Um campo externo da forma $h = a \xi_i^1$ [6] destói a transição para a fase de não reconhecimento da mesma forma que o campo magnético destói a transição para a fase paramagnética em sistemas magnéticos. Realizamos simulações para alguns valores do parâmetro a e verificamos que este campo ($h = a \xi_i^1$) não afeta a transição dinâmica do dano. Um resultado similar foi observado por Le Caër [64] e Grassberger [78] no qual mostraram que o campo magnético não destrói a transição dinâmica apresentada pelo dano no modelo de Ising, usando uma dinâmica de Glauber.

A figura 4.5 mostra a susceptibilidade de dano χ_{dh} em função do α para $C = 40$, $h = 0.001$ e para $N = 20000$, 40000 e 80000 . Observe que o valor de α , no qual o χ_{dh} apresenta um pico (tende a divergir), coincide com o valor de α_d da transição dinâmica caótica–não caótica. Outro fato importante, já citado anteriormente, é que o α_d não muda

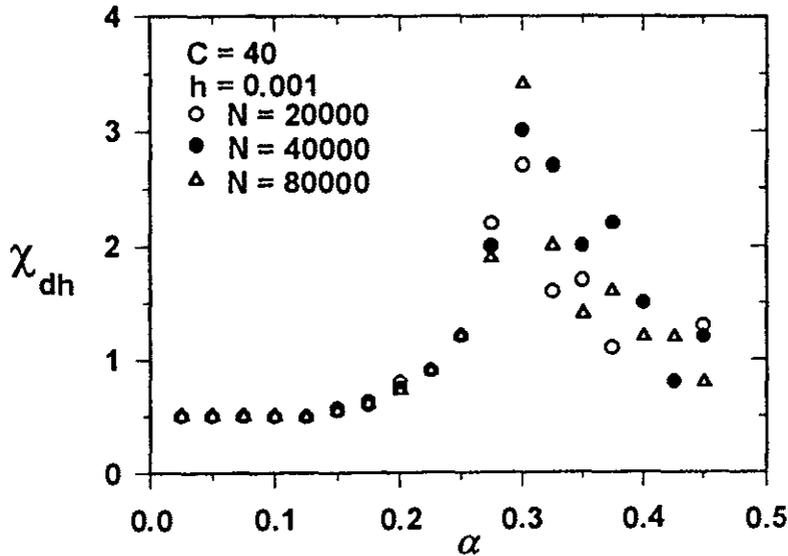


Figura 4.5: A susceptibilidade de dano χ_{dh} vs. α para $C = 40$, $h = 0.001$ e $N = 20000$ (círculos vazios), 40000 (círculos cheios) e 80000 (triângulos).

com N . O efeito em aumentar o tamanho da rede é similar ao efeito de diminuir o valor do campo h na simulação, isto é, quanto maior o N (ou quanto menor o campo h) mais acentuado e o pico na transição.

Finalmente, na figura 4.6 apresentamos o comportamento do χ_{dh} em função do α para $h = 0.001$, $N = 80000$ e conectividade $C = 20, 40$ e 80 . Como podemos observar, aumentando a conectividade a divergência na susceptibilidade desloca-se na direção do $\alpha = 0$. Efetivamente, a conectividade é quem determina a transição dinâmica e no limite do C muito grande com $C \ll N$, esperamos que $\alpha_d \rightarrow 0$. Ou seja, em toda fase de reconhecimento os atratores são caóticos, como no caso ultra diluído.

4.4 Conclusões

Neste capítulo, usamos o método de propagação de dano para estudar a natureza dos atratores de reconhecimento do modelo de Hopfield diluído e assimétrico. O comportamento dinâmico foi obtido através de simulações numéricas, sujeito a condição $C \ll N$ e para $T = 0$. Calculamos a distância de Hamming final entre duas réplicas (inicialmente muito

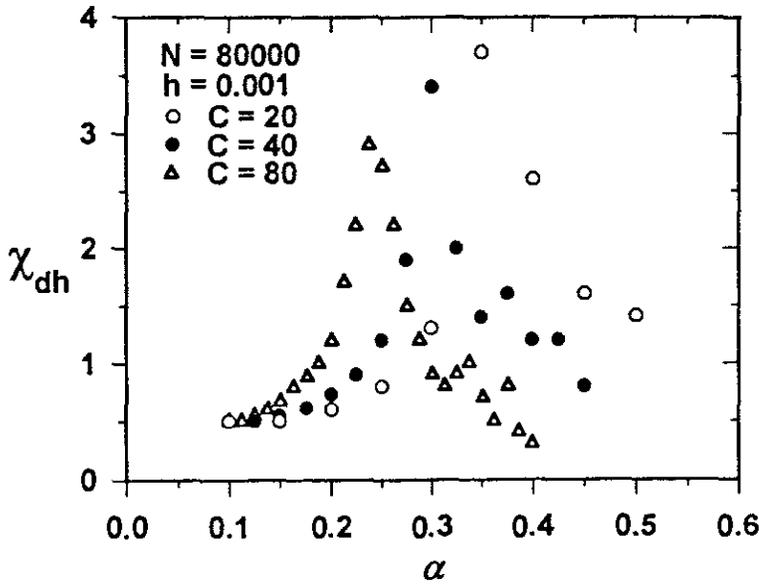


Figura 4.6: O χ_{dh} em função do α para $h = 0.001$, $N = 80000$ e conectividade $C = 20$ (círculos vazios), 40 (círculos cheios) e 80 (triângulos).

próximas e ambas correlacionadas entre si e com um único padrão) e a superposição do estado do sistema com este padrão, em função de α . Neste regime $C \ll N$, embora o sistema apresente as mesmas propriedades de reconhecimento que no regime ultra diluído $C \ll \ln N$, do ponto de vista dinâmico o sistema apresenta uma nova transição (dentro da fase de reconhecimento) no qual os atratores do sistema mudam de ponto fixo para atratores caóticos. Observamos que a transição dinâmica depende somente da conectividade C e isto sugere que a existência de atratores ponto fixo não é uma consequência do efeito do tamanho finito do sistema. Por outro lado, a existência de ciclos finitos parece depender do tamanho N , e que para $N \rightarrow \infty$, acreditamos que o sistema apresente uma transição dinâmica Caos-ponto fixo sem apresentar ciclos. O mesmo comportamento ocorre quando fixamos o N e aumentamos a conectividade. Neste caso, como mostrado na figura 4.3, o estreitamento da região de coexistências dos atratores PF, CF e Caos, ocorre devido ao avanço da região de atratores caóticos sobre as regiões de coexistência e ponto fixo. Este comportamento não era esperado já que aumentando a conectividade estamos caminhando na direção do modelo de Hopfield sem diluição, ou seja, a região de atratores ponto fixo deveria crescer. Um estudo mais detalhado sobre a dependência da

conectividade C e o tamanho do sistema N seria interessante para entender como se dá a transição do modelo diluído para o modelo não diluído.

Aplicamos também um campo externo h e mostramos que ele é o campo conjugado do parâmetro de ordem dinâmico dh . Verificamos que para qualquer valor arbitrário $h \neq 0$ ($dh \neq 0$), quebra a transição dinâmica associada ao dano (o sistema só apresenta atratores caóticos), e que este campo não afeta a transição de reconhecimento. Mostramos também que para um valor de C fixo a susceptibilidade apresenta um pico em α_d , que independe do tamanho do sistema. Quando fixamos N , observamos que α_d depende da conectividade, e no limite C muito grande com $C \ll N$ o $\alpha_d \rightarrow 0$ como no caso ultra diluído. Este campo é de validade geral, podendo ser utilizado em diferentes modelos determinísticos, e é muito apropriado para determinar com melhor precisão os valores críticos das transições dinâmicas.

Capítulo 5

Generalização em uma Rede de Neurônios com Diluição Assimétrica

Nos capítulos anteriores estudamos, fundamentalmente, propriedades das redes neuronais relacionadas ao reconhecimento de padrões por associação. No entanto, outras funções biológicas tem sido também modeladas nos últimos anos, como por exemplo o problema do aprendizado e generalização. A maior parte dos trabalhos relacionados com este tema, foram desenvolvidos numa arquitetura particular de redes neuronais, denominada de "Perceptron". Nessas redes, a informação flui em uma única direção e o objetivo é treiná-las através de exemplos de modo a torná-la capaz de abstrair uma regra (aprender). Um parâmetro importante no estudo destas redes é o *erro de generalização* (ϵ_g), que representa o desvio entre a resposta da rede (saída) e a resposta desejada (a função ϵ_g deve ser nula sempre que ambas as respostas sejam coincidentes). Um estudo detalhado deste modelo está fora do contexto desta tese. As referências [46, 79] apresentam um resumo do formalismo e seus principais resultados.

O problema da Generalização nas redes atratoras tipo Hopfield, originou-se de uma forma diferente que das redes tipo perceptron. Na busca de uma associação com as redes reais (no cérebro as memórias são armazenadas de uma forma ordenada e interligadas entre si), surge a necessidade de armazenar as memórias de acordo com as suas simila-

ridades, ou seja as memórias têm uma certa estrutura de correlação entre si. Uma das principais limitações do modelo de Hopfield com relação a capacidade de armazenamento e reconhecimento dos padrões, é que a rede só trabalha com relativo sucesso quando as memórias são descorrelacionadas. Neste sentido, vários trabalhos, usando uma estrutura de árvore hierárquica, tentaram superar esta limitação, e tal fato implicou em consideráveis modificações da regra de aprendizado de Hebb [80, 81, 82, 83, 84].

Nesta linha, mas com um novo ponto de vista, Fontanari [28] mostrou que, para o modelo de Hopfield totalmente conectado, quando os padrões armazenados são correlacionados, a rede é capaz de criar um novo atrator que tem uma mesma correlação com todas as memórias, e que pode ser interpretado como *resumindo* todas as propriedades comum a elas. Este atrator ou *conceito*, que do ponto de vista de reconhecimento é um atrator espúrio (no nosso caso representado pelas soluções simétricas), ganha assim um novo interesse ao permitir identificar configurações iniciais com categorias, ou seja, com grupos de memórias correlacionadas. Se a rede é exposta a um treinamento com exemplos de um determinado conceito, dependendo do número de exemplos correlacionados e do parâmetro de correlação, ela é capaz de extrair informações valiosas sobre este conceito, mesmo sem que ele apareça explicitamente na regra de aprendizagem.

Fontanari, utilizando a técnica de réplica, demonstrou que a transição para a fase de generalização é descontínua e que o erro de generalização decai exponencialmente com o número de exemplos. Em um estudo numérico com redes de até 512 neurônios, E. N. Miranda [29], tendo como configuração inicial um dos conceitos, confirmou as predições gerais do trabalho de Fontanari. Porém, ele encontrou uma transição contínua para a fase de generalização e que nesta fase, o erro de generalização, vai como uma potência do número de exemplos. M. C. Branchtein e J. J. Arenzon [85] estudaram o mesmo modelo para redes maiores, utilizando como condição inicial não somente o conceito (como no trabalho de Miranda), mas também exemplos. Eles mostraram que para $\alpha = 0$ a transição para a fase de generalização depende da condição inicial ser o conceito ou exemplos. Partindo de um exemplo, eles recuperaram a transição descontínua prevista pela aproximação de réplicas simétricas e, no caso da condição inicial ser um conceito, obtiveram

uma transição contínua. Na fase de generalização as curvas coincidem para ambos os casos, e o erro de generalização decai exponencialmente com o número de exemplos. Um comportamento similar para o caso em que a condição inicial é um conceito, foi obtido por D. Stariolo e F. Tamarit [30] para uma rede de neurônios analógica. Mais recentemente, P. R. Krebs e W. K. Theumann [86] estudaram este modelo analiticamente e numericamente e obtiveram todo o diagrama de fase.

Neste capítulo, apresentaremos uma extensão da dinâmica do modelo estudado nos capítulos anteriores (modelo de Hopfield ultra diluído) [7], para analisar a capacidade de generalização (categorização) da rede. O comportamento dinâmico é obtido exatamente, e o erro de generalização para o regime de longo tempo pode ser calculado. Finalmente, realizamos simulações numéricas e comparamos com os resultados analíticos.

A definição do erro de generalização, como veremos a seguir, não está relacionada com a definição dada no contexto das redes tipo perceptrons. Aqui, ele mede a habilidade do sistema para extrair uma regra a partir de um dado conjunto de exemplos comuns. Neste sentido, o uso do termo categorização é mais apropriado. Contudo, esta denominação tem sido vastamente empregada em todos os trabalhos científicos, de modo que manteremos esta terminologia e usaremos indistintamente as palavras generalização e categorização para expressar as propriedades comuns de uma determinada classe.

5.1 O Modelo

Seguindo o desenvolvimento dos dois últimos capítulos, vamos assumir que as contribuições da matriz sináptica que sobrevivem após a diluição T_{ij} , são dadas pela regra Hebbiana [28]:

$$T_{ij} = \sum_{\mu=1}^p \sum_{\nu=1}^s \xi_i^{\mu\nu} \xi_j^{\mu\nu} , \quad (5.1)$$

onde os $\xi_i^{\mu\nu} = \pm 1$ são variáveis aleatórias independentes correspondendo às sp configurações armazenadas $\{\xi_1^{\mu\nu}, \xi_2^{\mu\nu}, \dots, \xi_N^{\mu\nu}\}$ ($\nu = 1, 2, \dots, s$ e $\mu = 1, 2, \dots, p$). Estas memórias são agrupadas em p categorias (representada pela letra μ), onde cada uma contem s configurações (representada pela letra ν) que representa os diferentes elementos ou exem-

plos dessa categoria. Estas categorias são também representadas por p configurações $\{\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu\}$ ($\mu = 1, 2, \dots, p$) que representa as características comuns destes grupos e que serão chamados de conceitos. É importante enfatizar que estes conceitos não estão presentes na prescrição do armazenamento T_{ij} (5.1).

Vamos assumir que os exemplos $\{\xi_i^{\mu\nu}\}$ são variáveis aleatórias independentes cuja distribuição de probabilidade seja dada por:

$$P(\xi_i^{\mu\nu}) = \frac{1}{2}(1 + \xi_i^\mu b)\delta(\xi_i^{\mu\nu} - 1) + \frac{1}{2}(1 - \xi_i^\mu b)\delta(\xi_i^{\mu\nu} + 1), \quad (5.2)$$

onde os ξ_i^μ são variáveis aleatórias independentes que podem valer ± 1 com a mesma probabilidade e representa os p conceitos. O parâmetro b mede a correlação (similaridade) entre um conceito e seus exemplos

$$\langle \xi_i^{\mu\nu} \xi_j^{\gamma\delta} \rangle = b\delta_{\mu\gamma}\delta_{\nu\delta}, \quad (5.3)$$

e também entre dois exemplos:

$$\langle \xi_i^{\mu\nu} \xi_j^{\gamma\rho} \rangle = \delta_{ij} \delta_{\mu\gamma} [b^2 + \delta_{\nu\rho} (1 - b^2)]. \quad (5.4)$$

Desta forma, as memórias armazenadas são organizadas numa estrutura hierárquica. Na próxima seção vamos seguir o desenvolvimento introduzido por Derrida et al. [7], e resolver a dinâmica no contexto de generalização.

5.2 As Equações de Recorrência

Como estamos interessados na habilidade de generalização da rede vamos supor que a configuração inicial, corresponde a um exemplo da primeira categoria. Olharemos então para a solução com superposição macroscópica somente com o primeiro conceito e com mesma superposição $m^{\mu\nu}$ com todos seus exemplos armazenados, isto é $m^{1\nu}(0) = m_s$ para

qualquer ν e $m^{\mu\nu}(0) \sim \mathcal{O}(N^{-1/2})$ para $\mu > 1$, onde

$$m^{\mu\nu}(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^{\mu\nu} \langle S_i(t) \rangle . \quad (5.5)$$

Aqui, $\langle \dots \rangle$ denota a média térmica a temperatura T e sobre um conjunto de condições iniciais. A habilidade de categorização é medida pelo erro de generalização

$$\varepsilon = \frac{1 - m^1}{2} , \quad (5.6)$$

onde m^1 é a superposição entre o primeiro conceito e o estado da rede

$$m^1(t) = \frac{1}{N} \sum_i \xi_i^1 \langle S_i(t) \rangle . \quad (5.7)$$

Note que o ε é o fração de sítios na qual $\{S_i(t)\}$ e $\{\xi_i^1\}$ diferem.

Vamos olhar para as duas funções f e g que irão descrever o comportamento de longo tempo dos parâmetros de ordem

$$m_s(t+1) = f(m_s(t)) \quad (5.8)$$

$$m^1(t+1) = g(m_s(t)) \quad (5.9)$$

e, conseqüentemente, do ε .

Da definição (5.5) e (5.7) e após ter realizado a média térmica, obtemos as seguintes equações

$$m_s(t+1) = \left\langle \left\langle \frac{1}{s} \sum_{\nu=1}^s \xi_i^{1\nu} \tanh[\beta_o h_i(t)] \right\rangle \right\rangle \quad (5.10)$$

$$m^1(t+1) = \left\langle \left\langle \xi_i^1 \tanh[\beta_o h_i(t)] \right\rangle \right\rangle , \quad (5.11)$$

onde $\langle \langle \dots \rangle \rangle$ é a média sobre os padrões ξ_i^μ e $\xi_i^{\mu\nu}$ e o termo $1/s$ na equação (5.10) vem do

fato que estamos olhando as soluções simétricas $\{m_{11} = m_{12} = \dots = m_{1s} \equiv m_s\}$, onde

$$\sum_{\nu=1}^s m_{1\nu}(t+1) = s m_s(t+1) .$$

Se após a diluição o neurônio i está conectado a k outros neurônios, denotados por $\{j_1, j_2, \dots, j_k\}$, o campo local $h_i(t)$ pode ser escrito como:

$$\begin{aligned} h_i(t) &= \sum_{q=1}^k \sum_{\mu=1}^p \sum_{\nu=1}^s \xi_i^{\mu\nu} \xi_{j_q}^{\mu\nu} S_{j_q}(t) \\ &= \sum_{q=1}^k \sum_{\nu=1}^s \xi_i^{1\nu} \xi_{j_q}^{1\nu} S_{j_q}(t) + \sum_{\mu>1}^p \sum_{\nu=1}^s \xi_i^{\mu\nu} \sum_{q=1}^k \xi_{j_q}^{\mu\nu} S_{j_q}(t) , \end{aligned} \quad (5.12)$$

onde na última expressão separamos no somatório o termo $\mu = 1$. As equações de recorrências para o $m_s(t+1)$ e para $m^1(t+1)$ são dadas pelas seguintes expressões (ver descrição detalhada no apêndice A):

$$\begin{aligned} m_s(t+1) &= \frac{1}{s} \frac{1}{2^{s+1}} \sum_{k=0}^{\infty} \sum_{l=0}^s \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} (s-2l) F_{kmn}^+(s, b) \times \\ &\quad \tanh[\beta_o((s-2l)(k-2n) + ks(p-1) - 2m)] \end{aligned} \quad (5.13)$$

$$\begin{aligned} m^1(t+1) &= \frac{1}{2^{s+1}} \sum_{k=0}^{\infty} \sum_{l=0}^s \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} F_{kmn}^-(s, b) \times \\ &\quad \tanh[\beta_o((s-2l)(k-2n) + ks(p-1) - 2m)] \end{aligned} \quad (5.14)$$

onde

$$\begin{aligned} F_{kmn}^{\pm} &= \frac{C^k e^{-C}}{k! 2^{k(sp-s+1)}} \binom{s}{l} \binom{k}{n} \binom{ks(p-1)}{m} \times \\ &\quad [(1+m_s(t))^{k-n} (1-m_s(t))^n] [(1+b)^l (1-b)^{s-l} \pm (1-b)^l (1+b)^{s-l}] \end{aligned} \quad (5.15)$$

5.3 Limite de Conectividade Infinita

No limite de conectividade infinita ($C \rightarrow \infty$ onde $C \ll \ln N$) e $p \rightarrow \infty$, com $\alpha \equiv p/C$ constante e s finito, a equação 5.12, é uma soma de infinitas variáveis aleatórias independentes, cuja soma pode ser representada por uma distribuição Gaussiana. Neste limite, as equações para $m_s(t+1)$ e $m^1(t+1)$ são dadas por (ver apêndice A):

$$m_s(t+1) = \frac{1}{s} \sum_{l=0}^s G_l^+(s, b) (s-2l) \int Dy \tanh \left[\beta((s-2l)m_s(t) + \sqrt{\alpha' sy}) \right] \quad (5.16)$$

e

$$m^1(t+1) = \sum_{l=0}^s G_l^-(s, b) \int Dy \tanh \left[\beta((s-2l)m_s(t) + \sqrt{\alpha' sy}) \right] \quad (5.17)$$

onde $T \equiv \beta^{-1} = T_o/C$ é a temperatura reduzida, Dy é dado por

$$Dy = \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} , \quad (5.18)$$

α' é definido como:

$$\alpha' = \alpha \left[1 + (s-1)b^4 \right] , \quad (5.19)$$

e

$$G_l^\pm(s, b) = \frac{1}{2^{s+1}} \binom{s}{l} \left[(1+b)^l (1-b)^{s-l} \pm (1-b)^l (1+b)^{s-l} \right] . \quad (5.20)$$

Note que para $s = 1$ e $b = 1$ recuperamos a expressão obtida por Derrida et al [7] no problema de reconhecimento de padrões. Vamos resolver estas equações e obter as propriedades de generalização do sistema para diferentes valores de s , b , T e α .

5.3.1 Número finito de conceitos

Quando consideramos um número finito de conceito ($\alpha = 0$) as equações acima tornam-se:

$$m_s(t+1) = \frac{1}{s} \sum_{l=0}^s G_l^+(s, b) (s-2l) \tanh \left[\beta((s-2l)m_s(t)) \right] \quad (5.21)$$

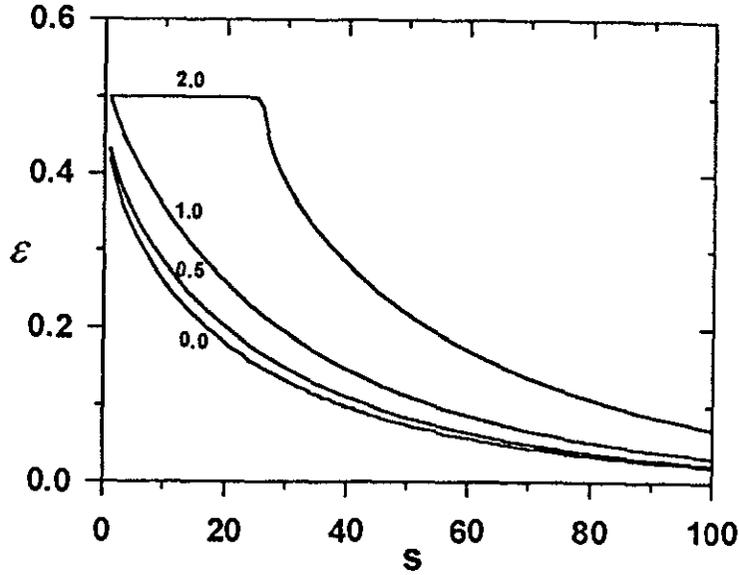


Figura 5.1: O erro de generalização ϵ em função do número de exemplos s para $\alpha = 0$, $b = 0.2$ e $T = 0.0, 0.5, 1.0$ e 2.0 .

$$m^1(t+1) = \sum_{i=0}^s G_i^-(s, b) \tanh[\beta((s-2i)m_s(t))] \quad (5.22)$$

Resolvendo iterativamente estas equações, podemos obter o erro de generalização ϵ dado por (5.6). A figura 5.1 mostra o ϵ vs. s para diferentes valores de T e $b = 0.2$. Para toda temperatura T existe sempre um valor finito s_c no qual o sistema passa continuamente da fase de não generalização (NG) para a fase de generalização (G) e ϵ tende a zero exponencialmente. Na figura 5.2 apresentamos as linhas críticas separando as duas fases para $b = 0.2$ e 0.4 . Observe que o ruído térmico afeta a capacidade de generalização da rede ou seja, quando o T cresce um número maior de exemplos tem que ser apresentado para a rede generalizar. Por outro lado, quando b cresce melhora a habilidade do sistema. Neste limite, recuperamos as equações obtidas para a rede de Hopfield sem diluição estudado em [28] para o caso $T = 0$ e em [86] para o caso $T \neq 0$. Contudo, os nossos resultados parecem contraditórios quando comparados com aqueles apresentados nos trabalhos acima [28, 86], onde eles encontraram que a transição para a fase de generalização é descontínua e que o ruído térmico melhora a habilidade do sistema. Na verdade, não existe esta contradição e a discrepância é devido ao fato que eles consideraram a configuração inicial como sendo

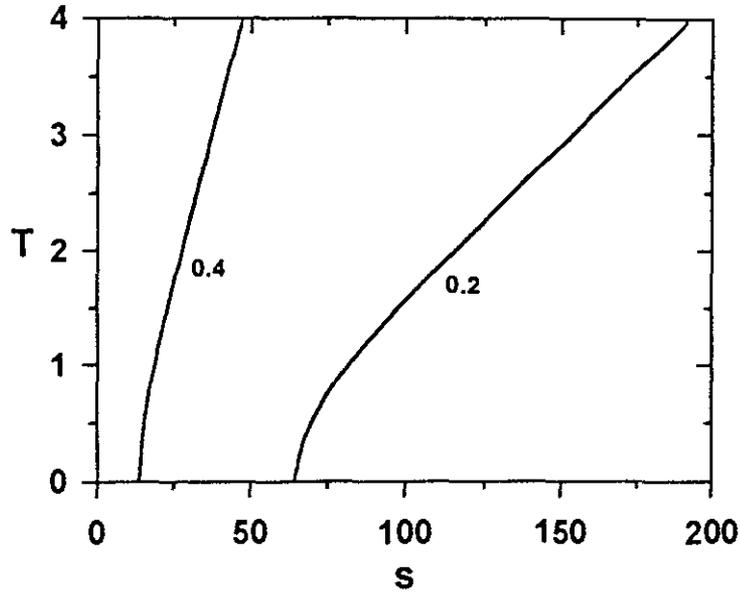


Figura 5.2: Diagrama de fase T vs. s para $\alpha = 0$ para dois valores do parâmetro de correlação: $b = 0.2$ e 0.4 . As linhas separando as fases de generalização e não generalização correspondem a ε menor que 5%.

um dos exemplos. Esta dependência da transição sobre a condição inicial foi observada numericamente por Branchtein et al. [85] e ambas as soluções apresentaram o mesmo comportamento para $s \gg 1$.

5.3.2 Número extensivo de conceito

Nesta seção vamos analisar o caso ($\alpha \neq 0$) com um número finito de exemplos. Vamos iniciar considerando o limite $T = 0$. A figura 5.3 mostra ε como função de s para $b = 2$ e diferentes valores de α/α_0 , onde $\alpha_0 = 2/\pi$ é o valor crítico acima do qual o sistema não reconhece os padrões descorrelacionados [7]. Para pequenos valores de α o erro de generalização é uma função monotonicamente decrescente de s . Aumentando α , este comportamento muda e ε apresenta um máximo para $s > 1$. Para $\alpha > \alpha_p$ o erro de generalização atinge um platô com $\varepsilon = 0.5$ para $s^- < s < s^+$ e em s^+ o sistema passa suavemente para a fase de generalização. Assumindo que para $s^+ \gg 1$, a distribuição binomial de z_s pode ser aproximada por uma distribuição Gaussiana, encontramos as seguintes equações de recorrência para m_s e m^1 (ver apêndice A)

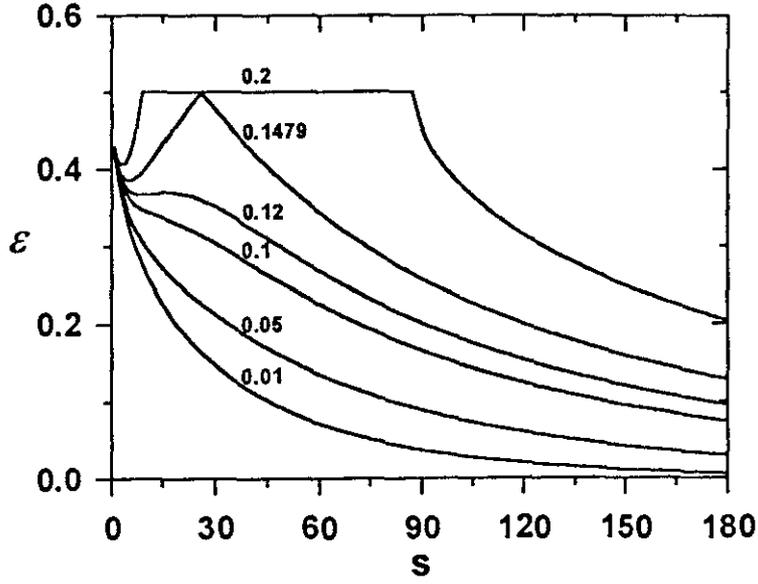


Figura 5.3: O erro de generalização ε em função do número de exemplos s para $T = 0$, $b = 0.2$ e $\alpha/\alpha_0 = 0.01, 0.05, 0.1, 0.12, 0.1479$ e 0.2 .

$$m_s(t+1) = b \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}} \frac{(1-b^2)\Delta}{bs} \exp\left(-\frac{\Delta^2}{2}\right) \quad (5.23)$$

$$m^1(t+1) = \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right), \quad (5.24)$$

onde

$$\Delta = \frac{b\sqrt{s}m_s(t)}{[\alpha' + m_s^2(t)(1-b^2)]^{1/2}}. \quad (5.25)$$

Desde que a transição para a fase de generalização é contínua, podemos expandir a equação para m_s (na transição $m_s \ll 1$), resultando na seguinte relação entre os parâmetros α , b e s :

$$\alpha = \frac{2[1 + b^2(s-1)]^2}{\pi s[1 + (s-1)b^4]}. \quad (5.26)$$

Note que a correlação b desempenha um papel fundamental no mecanismo de generalização do sistema. Observe também que para $s = 1$ recuperamos o valor crítico $\alpha_0 = 2/\pi$ obtido

em [7]. Da equação (5.26) é possível obter as seguintes expressões para s^- e s^+ :

$$s^\pm = -\frac{\alpha_0(1-b^2)}{b^2(\alpha_0-\alpha)} + \frac{\alpha(1-b^4)}{2b^4(\alpha_0-\alpha)} \left[1 \pm \left(1 - \frac{4\alpha_0 b^2}{\alpha(1+b^2)^2} \right)^{1/2} \right]. \quad (5.27)$$

O valor de α_p no qual surge o platô é

$$\alpha_p = \frac{8b^2}{\pi(1+b^2)^2}. \quad (5.28)$$

Para $\alpha = \alpha_p$ o erro de generalização atinge o valor 0.5 em um único ponto

$$s_p = \frac{1+b^2}{b^2}, \quad (5.29)$$

e para $\alpha < \alpha_p$ ele nunca atinge o platô ($\varepsilon < 0.5$). Na figura 5.3 com $b = 0.2$ obtemos $s_p = 26$ e $\alpha_p/\alpha_0 \approx 0.148$.

Para $s \gg 1$, $m^1 \rightarrow 1$ e $m_s \rightarrow b$ e ε decresce exponencialmente como

$$\varepsilon = \frac{1}{(2\pi)^{1/2}\Delta} \exp\left(-\frac{\Delta^2}{2}\right), \quad (5.30)$$

onde Δ é dado pela equação (5.25).

Quando os resultados acima são comparados com aqueles obtidos do modelo de Hopfield para $T = 0$ [28], observamos algumas novas propriedades. A transição para a fase de generalização é sempre contínua e para pequenos valores de s (poucos exemplos) o comportamento do erro de generalização ε em função de s é diferente nos dois limites. Dependendo do valor de b e α , ε cresce até atingir um máximo ou mesmo um regime de platô.

Na tabela 5.1 comparamos os valores de ε para os dois modelos, obtidos para $\alpha = 0.028$, $b = 0.4$ e diferentes valores de s . Para estes valores, a versão sem diluição generaliza para $s_c = 14$ com um erro de generalização $\varepsilon \sim 0.09$, enquanto que o Hopfield diluído atinge o mesmo valor para $s_c = 9$. Portanto, como acontece no contexto de reconhecimento de padrões a diluição também melhora o desempenho do sistema, ou seja, a rede generaliza

s	Hopfield	Hopfield diluído
10	0.500	0.1046
20	0.500	0.0321
30	0.032	0.0123
40	0.021	4×10^{-3}
50	0.017	9×10^{-4}

Tabela 5.1: O erro de generalização ε para $b = 0.4$, $\alpha = 0.028$ e diferentes valores de s , para ambos os modelos; Hopfield e Hopfield diluído.

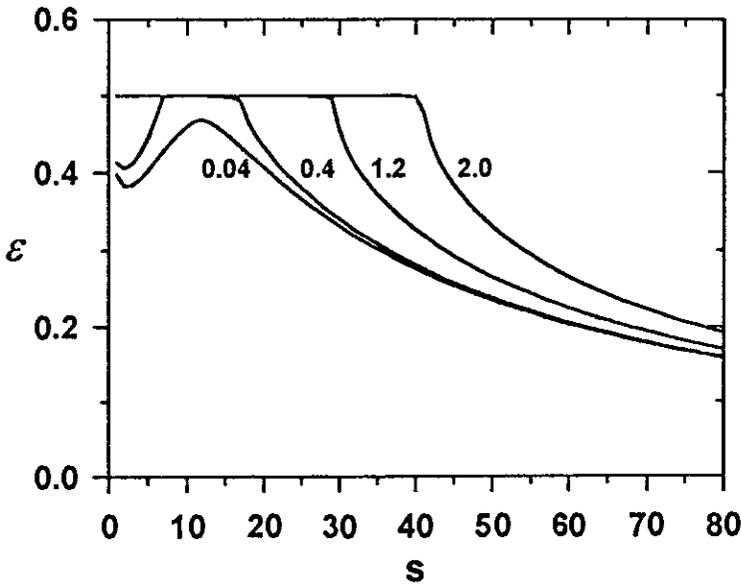


Figura 5.4: O erro de generalização ε em função do número de exemplos s para $b = 0.3$, $\alpha/\alpha_0 = 0.3$ e $T = 0.04, 0.4, 1.2$ e 2.0 .

com um número menor de exemplos. É importante ressaltar que, apesar das definições para o α serem diferentes ($\alpha = p/C$ no nosso caso e $\alpha = p/N$ em [28]), eles medem a mesma quantidade. Note também que os resultados são todos apresentados em diferentes unidades de α ($\alpha_0 = 2/\pi \sim 0.64$ para a versão diluída e ~ 0.138 para o caso não diluição).

Seguindo, apresentaremos os resultados para α e $T \neq 0$. A figura 5.4 mostra ε vs. s para $\alpha/\alpha_0 = 0.3$, $b = 0.3$ e diferentes valores de T . A temperatura desempenha um papel similar ao parâmetro α no caso $T = 0$. Quando T cresce, ε passa de um decréscimo mo-

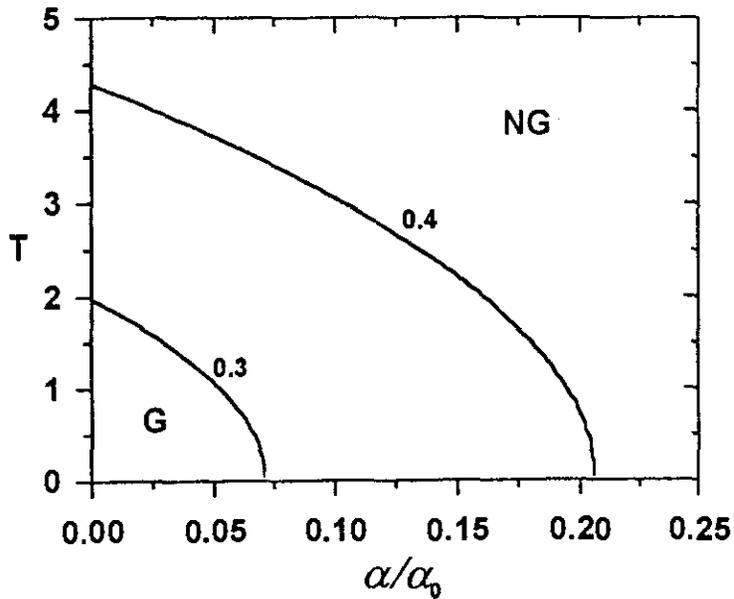


Figura 5.5: Linhas críticas T vs. α/α_0 separando a fase de generalização (G) da fase de não generalização (NG) para 50 exemplos, $b = 0.3$ e $b = 0.4$. As linhas correspondem a um erro $\varepsilon = 0.05$.

notóno para um comportamento tipo platô. Em todos os casos analisados neste trabalho, encontramos que o ruído térmico sempre reduz a habilidade de generalização do sistema ou seja, quando T cresce é necessário apresentar um número maior de exemplos para que a rede inicie a generalização. Finalmente, na figura 5.5 mostramos a linha crítica T versus α/α_0 abaixo da qual o sistema é capaz de generalizar com somente 50 exemplos, ε menor que 0.05 e para $b = 0.3$ e $b = 0.4$. Como esperado, quando b cresce a fase de generalização também cresce.

5.4 Simulações Numéricas

Nesta seção, estudamos através de simulações numéricas a habilidade de generalização do modelo para $T = 0$ e comparamos com os resultados analíticos. As simulações foram realizadas para $\alpha/\alpha_0 = 0.12$ e 0.20 , para redes com $N = 10000$ e 20000 . O número de conceitos foi sempre $p = 2$ e a conectividade foi $C = 26$ para $\alpha/\alpha_0 = 0.12$ e $C = 16$ para $\alpha/\alpha_0 = 0.20$.

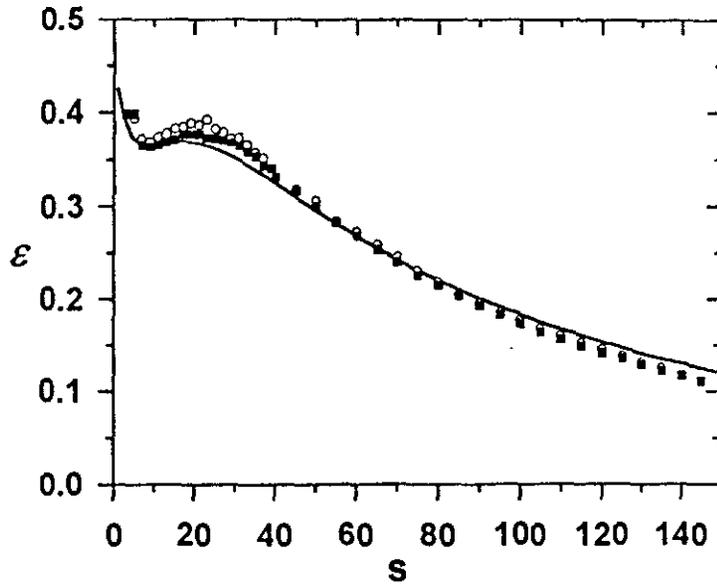


Figura 5.6: O erro de generalização ϵ vs. o número de exemplos s para $b = 0.2$ e $\alpha/\alpha_0 = 0.12$. A linha sólida corresponde as soluções analíticas. Para as simulações usamos $p = 2$, $C = 26$ e $N = 10000$ (círculos vazios) e 20000 (quadrados cheios).

O algoritmo usado foi o seguinte: inicialmente um conjunto de p conceitos não correlacionados foi criado. Para cada conceito, um exemplo ($s = 1$) é gerado e armazenado na rede. O estado inicial é escolhido como o primeiro conceito e a rede é atualizada em paralelo durante um número fixo de unidades de tempo, até que as flutuações da superposição m^{11} sejam desprezíveis (o sistema pode não alcançar um ponto fixo). Então, a média temporal de ϵ é calculada sobre 50 unidades de tempo. Seguindo, outro exemplo é criado ($s \rightarrow s + 1$) e o mesmos passos são repetidos. Para obter a média configuracional, este procedimento é repetido para 20 diferentes amostras, usando diferentes conceitos, exemplos e configurações iniciais.

Nas figuras 5.6 e 5.7 apresentamos os resultados para $\alpha/\alpha_0 = 0.12$ e 0.20 , respectivamente. As linhas sólidas são as soluções das equações (5.16) e (5.17) (válida para $C \rightarrow \infty$). Note que para valores pequenos de s a simulação reproduz muito bem os resultados analíticos, mesmo para pequenos sistemas. Quando s cresce, o comportamento apresenta uma pequena dependência sobre o tamanho do sistema. Para $s \gg 1$ observamos uma pequena discrepância com os dados teóricos, e este comportamento não parece ser

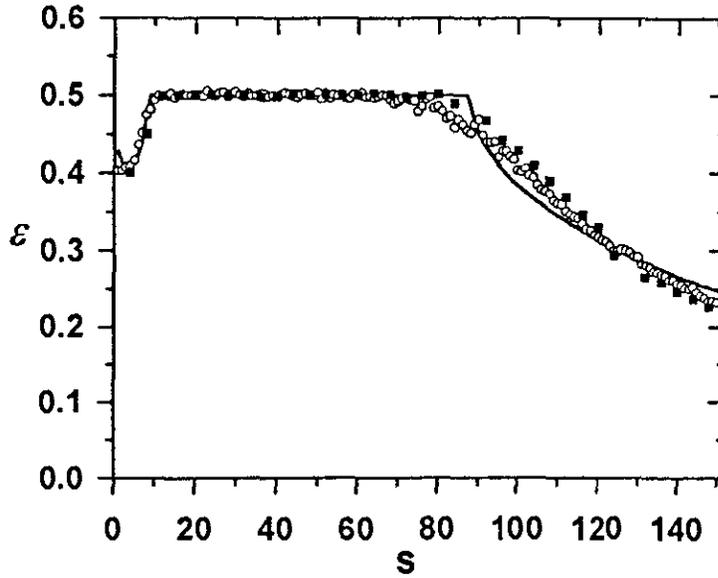


Figura 5.7: O erro de generalização ε vs. o número de exemplos s para $b = 0.2$ e $\alpha/\alpha_0 = 0.2$. A linha sólida corresponde as soluções analíticas. Para as simulações usamos $p = 2$, $C = 16$ e $N = 10000$ (círculos vazios) e 20000 (quadrados cheios).

devido ao efeito do tamanho finito. Acreditamos que esta diferença é devido ao fato que nesta região as simulações correspondem a $s \gg p$, enquanto que os resultados analíticos são válidos para $p \gg s$. De qualquer modo, note que o valor crítico de s^+ obtido na simulação para $\alpha/\alpha_0 = 0.20$ (figura 5.7) tende para o valor exato $s^+ = 88$ quando N cresce.

5.5 Conclusões

Neste capítulo estudamos a habilidade de generalização do modelo de Hopfield ultra diluído, onde a rede é treinada com exemplos de um dado conjunto de conceitos. Mostramos que após um certo número de exemplos ter sido treinado, o sistema é capaz de generalizar, isto é, de criar um atrator da dinâmica com superposição macroscópica com um dos conceitos. A dinâmica é resolvida exatamente e o comportamento de longo tempo para o erro de generalização, foi obtido através da evolução temporal da superposição m_1 entre o estado da rede no tempo t e o primeiro conceito ($\mu = 1$). Analisamos os casos $\alpha = 0$ e

$\alpha \neq 0$.

Resumindo nossos resultados, podemos dizer que, além de ser mais realista do ponto de vista biológico, o modelo diluído apresenta algumas características fundamentais como instrumento de generalização. Primeiro é que ele generaliza melhor que o modelo de Hopfield padrão (sem diluição), isto é, ele necessita de um número menor de exemplos para classificar corretamente em categorias. Este comportamento é observado tanto para dinâmica com ou sem ruído térmico para $\alpha = 0$ e $\alpha \neq 0$. Outros comportamentos interessantes neste modelo são que para $T = 0$ o sistema generaliza para alguns valores de α e que a inclusão do ruído diminui a habilidade do sistema como instrumento de generalização (para α , b e s fixos, $\varepsilon(T_1) \leq \varepsilon(T_2)$ se $T_1 \leq T_2$). Podemos enfatizar que este comportamento é consequência da escolha das soluções simétricas do modelo: se considerarmos as soluções assimétricas, para $\alpha = 0$ e para valores de T num certo intervalo, o sistema pode facilmente generalizar com poucos exemplos e para temperaturas altas. Também, para valores pequenos de s e dependendo sobre o valor de α a curva $\varepsilon \times s$ apresenta um comportamento diferente: após um breve decréscimo, ε cresce monotonamente até atingir um máximo ou o platô $\varepsilon = 0.5$ em $s = s^-$, e em $s = s^+$ o erro de generalização decresce exponencialmente.

Por fim, as simulações numéricas confirmam nossos resultados analíticos.

Conclusões

Neste capítulo faremos um resumo dos principais resultados obtidos nesta tese. No capítulo de introdução foi feita uma breve apologia ao desenvolvimento da física estatística, ao longo das últimas décadas, e em particular da teoria dos fenômenos críticos: seu papel de destaque no surgimento e desenvolvimento de novos modelos e teorias, e suas contribuições, que muitas vezes transcendem o plano acadêmico. Em particular, o estudo das funções do cérebro tornou-se um campo de interesse, consolidando uma nova linha de pesquisa conhecida como "Rede de Neurônios". No capítulo 1, primeiramente, descrevemos de maneira muito simplificada, a biologia de um neurônio, suas complexas conexões sinápticas e o processo de memorização e reconhecimento de padrões por associação. Num segunda parte apresentamos o modelo de Hopfield, e seu tratamento termodinâmico e da dinâmica. Apresentamos também os seus principais resultados analíticos e numéricos, quanto a sua capacidade de armazenamento e reconhecimento de padrões.

Nos capítulos 2 e 3 estudamos, analiticamente e através de simulações numéricas, como são afetadas as propriedades do modelo padrão de Hopfield não-diluído e diluído respectivamente, quando incorporamos na dinâmica da rede os períodos refratários. Este novo ingrediente biológico é adicionado ao PPS na forma de um limiar que depende do estado do neurônio nos tempos anteriores. No 2 capítulo usamos uma aproximação de campo médio (em analogia aos modelos magnéticos), para derivar a equação auto-consistente que descreve a superposição m do sistema em função dos parâmetros T , α e Δ . Este último é o parâmetro que ativa este limiar (períodos refratários). Para $\Delta = 0$ estas equações reproduzem os resultados obtidos mediante uma análise termodinâmica [6]. Para caso $\alpha \neq 0$ a transição entre as fases de reconhecimento e não reconhecimento é

sempre de primeira ordem como acontece no caso $\Delta = 0$, e para $\alpha = 0$ o diagrama de fase da figura 2.3 apresenta um ponto tricrítico cujos valores são $T^* \simeq 0.463$ e $\Delta^* \simeq 0.610$. Para $\Delta < \Delta^*$ a transição é de segunda ordem enquanto que para $\Delta > \Delta^*$ a transição é descontínua.

O principal resultado é que o parâmetro Δ , induz o surgimento de atratores caóticos e periódicos. Contudo, o sistema parece reconhecer principalmente nos regimes de ponto fixo e ciclos de ordem dois. Somente em uma pequena região o sistema reconhece com ciclos de ordem maior que dois e caos (ciclos maior que 100), mas este reconhecimento aparece na fronteira para a fase de não reconhecimento. Um estudo mais detalhado nesta região seria interessante para elucidar se a existência destes atratores é devido a efeitos do tamanho finito ou não. Nas nossas simulações, quando aumentamos o tamanho do sistema, estes atratores não desaparecem e portanto suspeitamos que eles possam existir no limite termodinâmico.

Na fase de reconhecimento (para Δ pequeno), o PPS de Hopfield é forte o suficiente para levar o sistema para atratores estáveis; PF e órbitas periódicas, onde a superposição com o padrão memorizado é próximo de 1. Aumentando o parâmetro Δ o sistema é drasticamente afetado, e nestas regiões a dinâmica é dominada por ciclos de períodos longos ou trajetórias caóticas.

A simulação numérica reproduz muito bem os resultados analíticos para valores de α pequenos, onde a transição ocorre de ponto fixo (PF) para ciclo de ordem dois (C2). Ambas as dinâmicas (paralela e sequencial) apresentam resultados muito parecidos, mostrando que o comportamento oscilatório deve-se a presença do termo refratário no campo local. Finalmente, observamos que somente na transição a atividade média decresce com o crescimento do Δ .

No capítulo 3 estudamos o modelo de Hopfield com diluição e assimetria nas sinapses. Neste caso, o limiar que modela os períodos refratários produz um efeito de realimentação sobre a dinâmica do sistema. Desenvolvemos um método aproximado que melhora amplamente os resultados de campo médio. Este método permitiu estudar o comportamento dinâmico de longo tempo do sistema e obter o diagrama de fase completo do modelo.

Ao contrário do modelo totalmente conectado (capítulo 2), que para valores grandes de Δ desestabiliza completamente os padrões armazenados, na versão ultra diluída verificamos que, para pequenos valores de α , o sistema pode reconhecer sempre os padrões, independentemente do valor de Δ . Para Δ grande, na fase de reconhecimento, o sistema reconhece pobremente as memórias com $m \simeq 0.5$, e apresenta uma forte dependência com a história passada dos neurônios. Nesta região, a superfície crítica em termos dos parâmetros T , α e Δ tende a ser constante (ver figura 3.5) com valores assintóticos $\alpha = 0$ ($T^* = 0.444\dots$) e $T = 0$ ($\alpha^* = 0.125\dots$). Os resultados numéricos mostram que os atratores são periódicos (ciclo dois) e o valor médio de m concorda muito bem com os valores obtidos no desenvolvimento analítico (atratores ponto fixo).

Para valores intermediários de Δ , uma inesperada fase de reconhecimento aparece e a rede pode reconhecer o padrão memorizado com alta e baixa qualidade, dependendo se o estado inicial está próximo ou longe da memória. Nesta região, o termo de Hopfield (o sinal) e o termo refratário competem, dando lugar a estas duas diferentes soluções. Apesar de não esperar bons resultados da nossa aproximação, pois as correlações temporais são importantes nesta região, os resultados numéricos predizem qualitativamente bem os resultados analíticos, mostrando a coexistência de fases.

Para Δ pequeno o efeito de realimentação não é muito importante comparado ao termo de sinal, e para $\Delta = 0$ (ver figura 3.5) recuperamos o valor $\alpha_c = 0.636\dots$ para $T = 0$ e o valor $T = 1$ para $\alpha = 0$ em [7]. Nesta região os resultados analíticos reproduzem muito bem os resultados numéricos (ver figura 3.6). A transição para a fase com $m = 0$ é sempre de segunda ordem, e dentro da fase de reconhecimento a transição para o regime no qual o sistema reconhece com alta ou baixa qualidade as memórias, é de primeira ordem.

Para valores pequenos de α e para Δ grande, apesar da baixa qualidade no reconhecimento, as bacias de atração são robustas e o sistema apresenta estados espúrios simétricos ou assimétricos de memórias, que desaparecem com α . É importante destacar que neste estudo estamos modelando simultaneamente três ingredientes biológicos verificados experimentalmente.

Na região para Δ grande, dado que as contribuições para a superposição deve-se aos

estados com $S = -1$ ($m \simeq 1/2$ para atividade inicial $1/2$), um estudo do modelo com baixa atividade deve melhorar o reconhecimento. Como a atividade dos neurônios pode ser implementada diretamente na distribuição dos ξ_i^μ , este modelo pode ser resolvido usando um desenvolvimento analítico [87].

No capítulo 4, usamos o método de propagação de dano para estudar a natureza dos atratores de reconhecimento do modelo de Hopfield diluído e assimétrico. O comportamento dinâmico foi obtido através de simulações numéricas sujeitas a condição menos restritiva $C \ll N$ (em comparação ao regime ultra diluído $C \ll \ln N$) e para $T = 0$. Calculamos a distância de Hamming final entre duas réplicas, inicialmente próximas entre si e ambas correlacionadas com um único padrão, e a superposição do estado do sistema com este padrão, em função de α . Neste regime $C \ll N$, embora o sistema apresente as mesmas propriedades de reconhecimento que no regime ultra diluído, do ponto de vista dinâmico o sistema apresenta uma nova transição (dentro da fase de reconhecimento) no qual o sistema passa de atratores ponto fixo para atratores caóticos. Observamos que a transição dinâmica depende somente da conectividade C e isto sugere que a existência de atratores ponto fixo não é uma consequência do efeito de tamanho do sistema. Por outro lado, a existência de ciclos finitos parece depender do tamanho N e que, para $N \rightarrow \infty$, acreditamos que o sistema apresente uma transição dinâmica caos-ponto fixo sem apresentar ciclos. O mesmo comportamento ocorre quando fixamos N e aumentamos a conectividade. Neste caso, como mostrado na figura 4.3, o estreitamento da janela onde coexistem atratores PF (ponto fixo), CF (ciclos de períodos finitos menores que 100) e Caos (atratores caóticos, ou ciclos de período maior que 100), ocorre devido ao prolongamento da janela dos atratores caóticos sobre as regiões de coexistência de atratores cíclicos e de ponto fixo. Este comportamento não era esperado já que aumentando a conectividade estamos indo na direção do modelo de Hopfield não diluído, ou seja, a região de atratores ponto fixo, ao contrário, deveria crescer. Um estudo mais detalhado sobre a dependência da conectividade C e o tamanho N do sistema seria interessante para entender como se dá a transição do modelo diluído para a outra fronteira, no caso, o modelo totalmente conectado (Hopfield).

Introduzimos um novo campo externo h , já estudado em autômatos celulares e modelos magnéticos, e mostramos que ele é o conjugado do parâmetro de ordem dinâmico dh (ver figura 4.4). Verificamos que para qualquer valor arbitrário $h \neq 0$, o dh é sempre diferente de zero, e que este campo não afeta a fase de reconhecimento, ou seja o m . Para um valor fixo de C , a susceptibilidade associada ao dano apresenta uma pico que tende a divergir em α_d (o valor de α da transição dinâmica caótica–não caótica), que independe do tamanho do sistema, e para N fixo e aumentando C , observamos que α_d depende da conectividade. No limite para C muito grande, com $C \ll N$ esperamos que $\alpha_d \rightarrow 0$ como no caso ultra diluído (os atratores de reconhecimento são caóticos). Este campo é de validade geral, podendo ser utilizado em diferentes modelos determinísticos, e é apropriado para determinar com certa precisão os valores críticos das transições dinâmicas.

No capítulo 5 estudamos a habilidade de generalização (categorização) do modelo de Hopfield ultra diluído, onde a rede é treinada com exemplos de um dado conjunto de conceitos. Após um certo número de exemplos ter sido apresentado, o sistema é capaz de generalizar, isto é, de criar um atrator da dinâmica com superposição macroscópica com um dos conceitos, sem que este apareça explicitamente na regra de aprendizado. A dinâmica é resolvida exatamente, e o comportamento de longo tempo para o erro de generalização ε é obtido em função da superposição m_1 (5.17) entre o estado da rede e o primeiro conceito ($\mu = 1$). Analisamos os casos $\alpha = 0$ e $\alpha \neq 0$.

O modelo diluído, além de ser mais realista do ponto de vista biológico, apresenta algumas características fundamentais como instrumento de generalização. Comparado com o modelo totalmente conectado, ele *generaliza melhor*, isto é, ele necessita de um número menor de exemplos para classificar corretamente uma dada categoria. Este comportamento é observado para todos os casos analisados. Observamos que a inclusão do ruído diminui a capacidade do sistema para generalizar, ou seja, para a capacidade de armazenamento (α), o número de exemplos (s) e o parâmetro de correlação (b) fixos, o erro de generalização $\varepsilon(T_1) \leq \varepsilon(T_2)$ se $T_1 \leq T_2$. Este comportamento é consequência da escolha das soluções simétricas do modelo; se considerarmos as soluções assimétricas, para $\alpha = 0$ e para valores de T num certo intervalo, o sistema pode facilmente generalizar

com poucos exemplos e para temperaturas altas. Também, para valores pequenos de s e dependendo sobre o valor de α e do T , as curvas $\varepsilon \times s$ (ver figuras 5.3 e 5.4) apresentam um comportamento diferente; após um breve decréscimo, ε cresce monotonamente até atingir um máximo ou o platô $\varepsilon = 0.5$ em $s = s^-$, e em $s = s^+$ o erro de generalização decresce exponencialmente.

Os resultados obtidos da simulação numérica confirmam os nossos resultados analíticos.

Apêndice A

Cálculo da Dinâmica do Modelo de Hopfield Diluído e Assimétrico

A.1 Reconhecimento de Padrões

Vamos encontrar a equação de recorrência para a superposição do sistema com o padrão $\mu = 1$, que permite estudar as propriedades do Modelo de Hopfield diluído e assimétrico, válidas para todo tempo t . Vamos partir da equação (1.23). O primeiro passo é calcular o valor de $m(t = 1)$ em função do $m(t = 0)$, para posteriormente generalizarmos para todo tempo t .

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

onde a média $\langle \dots \rangle$ é no ruído térmico a temperatura T_0 e sobre as condições iniciais, e $\langle \langle \dots \rangle \rangle$ é a média sobre as variáveis ξ . O termo $\xi_i^1 h_i(t = 0)$ é dado por:

$$\xi_i^1 h_i(t = 0) = \sum_{j_r \neq i}^k \xi_{j_r}^1 S_{j_r}(t = 0) + \xi_i^1 R_{ij}^\mu(t = 0). \quad (\text{A.2})$$

O R_{ij}^μ é o ruído devido as memórias não correlacionadas ($\mu \neq 1$), e é dado pela expressão:

$$R_{ij}^\mu(t=0) = \sum_{j_r \neq i}^k \sum_{\mu > 1}^p \xi_i^\mu \xi_{j_r}^\mu S_{j_r}(t=0). \quad (\text{A.3})$$

Os dois termos da equação (A.2) são somas de k e $k(p-1)$ termos respectivamente podendo valer ± 1 . Seja n o número de termos negativos na primeira soma e s o número de termos negativos na segunda soma. A expressão para o campo local pode tomar valores:

$$\xi_i^1 h_i(t=0) = k - 2n + k(p-1) - 2s = kp - 2(n+s),$$

onde $n = 0, 1, \dots, k$ e $s = 0, 1, \dots, k(p-1)$. Se no tempo $t=0$ a configuração inicial do sistema está correlacionada com a primeira memória, a probabilidade P_r de $S_{j_r} \xi_{j_r}^1$ ser ± 1 é:

$$P_r(S_{j_r} \xi_{j_r}^1 \pm 1) = \frac{(1 \pm m_0)}{2}, \quad (\text{A.4})$$

onde $m_0 = m(t=0)$. Dado que a configuração inicial está correlacionada com o primeiro padrão, a probabilidade de ter n termos negativos na primeira soma é dada por

$$P_r(n) = \sum_{n=0}^k \binom{k}{n} \frac{(1+m_0)^{k-n} (1-m_0)^n}{2^k}. \quad (\text{A.5})$$

Para a segunda soma o cálculo da probabilidade é mais simples, pois a configuração inicial não está correlacionada com as memórias restantes ($\mu \neq 1$). Assim, a probabilidade de ter s termos negativos na segunda soma é

$$P_r(s) = \sum_{s=0}^{k(p-1)} \binom{k(p-1)}{s} \frac{1}{2^{k(p-1)}}. \quad (\text{A.6})$$

Como em $t = 0$ as duas somas são independentes, a probabilidade total está dada pela expressão:

$$P_r(n, s) = \sum_{n=0}^k \sum_{s=0}^{k(p-1)} \frac{1}{2^{kp}} \binom{k(p-1)}{s} \binom{k}{n} (1 + m_0)^{k-n} (1 - m_0)^n. \quad (\text{A.7})$$

Usando (A.7) podemos reescrever a equação (A.1) na forma:

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

onde $P_r(k)$ é a probabilidade do neurônio i está conectado a k outros neurônios após a diluição

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

onde

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

Tomando o limite $N \rightarrow \infty$ o $P_r(k)$ torna-se uma distribuição de Poisson, cuja expressão está dada por:

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

A generalização da equação (A.8) para tempos quaisquer, só é possível desde que possamos garantir que as correlações entre os S_{j_r} na expressão do campo local (A.2) sejam nulas, ou seja:

- que não exista circuitos fechados entre os neurônios;
- e que todos os neurônios sejam independentes para todo tempo t com probabilidade 1 (cada um dos neurônios tenha sua própria árvore de ancestrais independente das árvores dos demais).

As condições acima serão satisfeitas se:

$$C \ll \ln N .$$

Desta forma, a expressão (A.8) pode ser generalizada para tempos arbitrários:

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

Devido a impossibilidade de realizar esta soma, vamos nos limitar ao caso particular $C \rightarrow \infty$.

A.1.1 Limite de conectividade infinita

Após tomar o limite termodinâmico, toma-se os limites C e $p \rightarrow \infty$ com $\alpha \equiv (p-1)/C$. Nestes limites a expressão do campo local (A.2) é uma soma de infinitas variáveis aleatórias independentes. Pelo teorema do limite central esta soma pode ser representada por uma distribuição gaussiana com média Cm e variância dada pela soma das variâncias de cada termo (o termo de ruído tem média nula).

$$\langle \xi_i^1 h_i(t) \rangle = \sum_{j_r \neq 1}^C \langle \xi_{j_r}^1 s_{j_r}(t) \rangle = Cm(t) ,$$

e a variância σ^2 é dada por

$$\sigma^2 = \langle (\xi_i^1 h_i(t))^2 \rangle - \langle \xi_i^1 h_i(t) \rangle^2 = C(p-1) \equiv C^2 \alpha .$$

Usando a representação Gaussiana, a equação para a superposição $m(t+1)$ é dada pela seguinte expressão:

$$m(t+1) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} dx e^{-\frac{(x-Cm(t))^2}{2\sigma^2}} \tanh \left(\frac{x}{T_0} \right) , \quad (\text{A.12})$$

onde $\sigma = \sqrt{\alpha}C$. Por último, fazendo a transformação

$$z = \frac{x - Cm(t)}{\sigma}$$

obtemos a equação (1.32) dada no capítulo 1.

$$m(t+1) = \int_{-\infty}^{\infty} Dz \tanh \beta (m(t) + z\sqrt{\alpha}) , \quad (\text{A.13})$$

onde $T = \beta^{-1}$ é a temperatura generalizada $T = T_0/C$ e

$$Dz = \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}} .$$

A.2 Generalização

A equação do campo local definido em (5.12) pode ser escrita como:

$$h_i(t) = Z_s \sum_{q=1}^k \xi_{j_q}^{1\nu} S_{j_q}(t) + \sum_{\mu>1}^p \sum_{\nu=1}^s \xi_i^{\mu\nu} \sum_{q=1}^k \xi_{j_q}^{\mu\nu} S_{j_q}(t) \quad (\text{A.14})$$

onde $Z_s = \sum_{\nu=1}^s \xi_i^{1\nu}$. Os dois termos acima, são somas de ks e $ks(p-1)$ termos, onde cada um pode valer ± 1 . Seja n e m o número de termos negativos de cada soma respectivamente. O campo local $h_i(t)$ pode ser expresso na forma:

$$h_i(t) = Z_s (k - 2n) + ks(p-1) - 2m . \quad (\text{A.15})$$

Como a configuração inicial está correlacionada com o ν -ésimo exemplo do primeiro conceito ($\mu = 1$), a probabilidade $P_r(\xi_{j_q} = \pm S_{j_q})$ é igual $(1 \pm m_s)/2$, onde $m_{j_q} = m_{j_1} = m_{j_2} = \dots = m_s$ (solução simétrica). As probabilidades de ter n termos negativos na primeira soma e m termos negativos na segunda são dadas respectivamente por:

$$P_r(n) = \sum_{n=0}^k \binom{k}{n} \frac{(1 + m_s)^{k-n} (1 - m_s)^n}{2^k} , \quad (\text{A.16})$$

$$P_r(m) = \sum_{m=0}^{ks(p-1)} \binom{ks(p-1)}{m} \frac{1}{2^{ks(p-1)}}. \quad (\text{A.17})$$

Inserindo a expressão do campo local (A.14) em (5.10) e (5.11), obtemos as seguintes equações de recorrências para as superposições do estado do sistema com os exemplos e o primeiro conceito respectivamente:

$$m_s(t+1) = \sum_{k=0}^{\infty} \frac{P_r(k)}{s2^{k(ps-s+1)}} \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} \binom{k}{n} \binom{ks(p-1)}{m} (1+m_s)^{k-n} (1-m_s)^n \times \left\langle \left\langle Z_s \tanh \left(\frac{Z_s(k-2n) + ks(p-1) - 2m}{T_0} \right) \right\rangle \right\rangle \quad (\text{A.18})$$

e

$$m^1(t+1) = \sum_{k=0}^{\infty} \frac{P_r(k)}{2^{k(ps-s+1)}} \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} \binom{k}{n} \binom{ks(p-1)}{m} (1+m_s)^{k-n} (1-m_s)^n \times \left\langle \left\langle \xi_i^1 \tanh \left(\frac{Z_s(k-2n) + ks(p-1) - 2m}{T_0} \right) \right\rangle \right\rangle. \quad (\text{A.19})$$

Aqui, já tomamos o limite termodinâmico e realizamos a média sobre todas as possíveis conectividades k do neurônio i (o $P_r(k)$ está dado pela equação (A.10)), e estamos supondo que as correlações entre os termos S_{j_r} do campo local (A.14) são nulas para todo tempo t , ou seja, que a condição ($C \ll \ln N$) é satisfeita. A média $\langle \langle \dots \rangle \rangle$ é sobre os $\xi_i^{1\nu}$ e ξ_i^1 .

Seja $P_r(Z_s = s - 2l)$ ($l = 0, 1, \dots, s$) a probabilidade de Z_s ter l termos negativos. Usando a distribuição de probabilidade dos $\xi_i^{\mu,\nu}$ definida em (5.2) obtemos para $P_r(Z_s)$

$$P_r(Z_s) = \sum_{l=0}^s \binom{s}{l} \frac{(1 + b\xi^1)^{s-l} (1 - b\xi^1)^l}{2^s}. \quad (\text{A.20})$$

Substituindo a expressão acima em (A.18) e (A.19) e por último realizando a média sobre

os ξ_i^1 , obtemos as equações (5.13) e (5.14) do capítulo 5.

$$m_s(t+1) = \frac{1}{s 2^{s+1}} \sum_{k=0}^{\infty} \sum_{l=0}^s \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} (s-2l) F_{kmn}^+(s, b) \times \tanh[\beta_o((s-2l)(k-2n) + ks(p-1) - 2m)] \quad (\text{A.21})$$

$$m^1(t+1) = \frac{1}{2^{s+1}} \sum_{k=0}^{\infty} \sum_{l=0}^s \sum_{n=0}^k \sum_{m=0}^{ks(p-1)} F_{kmn}^-(s, b) \times \tanh[\beta_o((s-2l)(k-2n) + ks(p-1) - 2m)] \quad (\text{A.22})$$

onde

$$F_{kmn}^{\pm} = \frac{C^k e^{-C}}{k! 2^{k(sp-s+1)}} \binom{s}{l} \binom{k}{n} \binom{ks(p-1)}{m} \times [(1+m_s(t))^{k-n} (1-m_s(t))^n] [(1+b)^l (1-b)^{s-l} \pm (1-b)^l (1+b)^{s-l}]. \quad (\text{A.23})$$

Fazendo $s = 1$ e $b = 1$, nas equações acima, obtem-se a equação para o $m(t+1)$ no contexto de reconhecimento de padrões (A.11) [7].

A.2.1 Limite de conectividade infinita

Este limite segue o mesmo desenvolvimento da seção anterior. Mantendo o s finito e para C e $p \rightarrow \infty$ com $\alpha \equiv (p-1)/C$, o segundo termo da expressão do campo local (A.14) é uma soma de infinitas variáveis aleatórias independentes. Pelo teorema do limite central esta soma pode ser representada por uma distribuição Gaussiana com média nula e variância dada por

$$\begin{aligned} \sigma^2 &= \left\langle \sum_{\mu, \mu' > 1}^p \sum_{\nu, \nu'}^s \xi_i^{\mu\nu} \xi_i^{\mu'\nu'} \sum_{q, q'}^k \xi_{j_q}^{\mu\nu} \xi_{j_{q'}}^{\mu'\nu'} S_{j_q}(t) S_{j_{q'}}(t) \right\rangle \\ &= C^2 \alpha s + C^2 \alpha s (s-1) b^4, \end{aligned} \quad (\text{A.24})$$

onde o último termo ($C^2 \alpha s(s-1)b^4$) é devido as correlações entre exemplos de um mesmo conceito. A distribuição de probabilidade para o campo local é portanto a distribuição Gaussiana deslocada em $Z_s C m(t)$:

$$P_r(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-Z_s C m(t))^2}{2\sigma^2}}, \quad (\text{A.25})$$

onde usamos

$$\lim_{C \rightarrow \infty} Z_s \sum_{q=1}^C \xi_{j_q}^{1\nu} S_{j_q}(t) = Z_s C m(t).$$

Usando a distribuição binomial para o Z_s e por último realizando a média sobre as variáveis $\{\xi\}$, obtem-se as equações (5.16) e (5.17) da capítulo 5.

$$m_s(t+1) = \frac{1}{s} \sum_{l=0}^s G_l^+(s, b) (s-2l) \int Dy \tanh[\beta((s-2l)m_s(t) + \sqrt{\alpha' s y})] \quad (\text{A.26})$$

e

$$m^1(t+1) = \sum_{l=0}^s G_l^-(s, b) \int Dy \tanh[\beta((s-2l)m_s(t) + \sqrt{\alpha' s y})] \quad (\text{A.27})$$

onde $T \equiv \beta^{-1} = T_0/C$ é a temperatura reduzida, Dy é dado por

$$Dy = \frac{dy \exp(-y^2/2)}{\sqrt{2\pi}},$$

α' é definido como:

$$\alpha' = \alpha [1 + (s-1)b^4], \quad (\text{A.28})$$

e

$$G_l^\pm(s, b) = \frac{1}{2^{s+1}} \binom{s}{l} [(1+b)^l (1-b)^{s-l} \pm (1-b)^l (1+b)^{s-l}]. \quad (\text{A.29})$$

A.2.2 Número extensivo de conceitos

Nesta seção vamos obter as equações (5.23) e (5.24) do capítulo 5. Primeiro vamos olhar o caso em que um grande número de exemplos são apresentados a rede. Para $s \gg 1$, a distribuição binomial do Z_s pode ser aproximada por uma distribuição Gaussiana cuja

média é dada por

$$\langle Z_s \rangle = \sum_{\nu=1}^s \langle \xi_i^{1\nu} \rangle = \sum_{\nu=1}^s \xi_i^1 b = s \xi_i^1 b , \quad (\text{A.30})$$

onde usamos a distribuição do $\xi_i^{1\nu}$ dado em (5.2). A variância é definida como:

$$\sigma_s^2 = \langle (Z_s)^2 \rangle - \langle Z_s \rangle^2 = s - sb . \quad (\text{A.31})$$

Desta forma, as superposições $m_s(t+1)$ e $m^1(t+1)$ são dadas pelas equações:

$$m_s(t+1) = \frac{1}{s\sqrt{2\pi}\sigma_s} \left\langle \left\langle \int dx e^{-\frac{(x-sb\xi_i^1)^2}{2\sigma_s^2}} x \int Dz \tanh \beta(x m_s + \sqrt{\alpha'} sz) \right\rangle \right\rangle \quad (\text{A.32})$$

e

$$m^1(t+1) = \frac{1}{\sqrt{2\pi}\sigma_s} \left\langle \left\langle \xi_i^1 \int dx e^{-\frac{(x-sb\xi_i^1)^2}{2\sigma_s^2}} \int Dz \tanh \beta(x m_s + \sqrt{\alpha'} sz) \right\rangle \right\rangle . \quad (\text{A.33})$$

Fazendo a transformação

$$y = \frac{x - sb\xi_i^1}{\sigma_s}$$

e por último realizando a média sobre o ξ_i^1 encontramos:

$$m_s(t+1) = \frac{1}{2} \int \int Dy Dz \left\{ (\sqrt{L}y + b) \tanh \left[\frac{\Theta^+(y, z)}{T} \right] + (\sqrt{L}y - b) \tanh \left[\frac{\Theta^-(y, z)}{T} \right] \right\} \quad (\text{A.34})$$

e

$$m^1(t+1) = \frac{1}{2} \int \int Dy Dz \left\{ \tanh \left[\frac{\Theta^+(y, z)}{T} \right] - \tanh \left[\frac{\Theta^-(y, z)}{T} \right] \right\} , \quad (\text{A.35})$$

onde

$$\Theta^\pm(y, z) = sm_s(\sqrt{L}y \pm b) + \sqrt{\alpha'} sz ,$$

α' é dada em (A.28) e

$$L = \frac{(1 - b^2)}{s} .$$

No limite $T \rightarrow 0$ é necessário usar as seguintes identidades para calcular as integrais das equações acima:

$$\lim_{T \rightarrow 0} \int_{-\infty}^{\infty} \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}} \tanh\left(\frac{az+b}{T}\right) = \operatorname{erf}\left(\frac{b}{\sqrt{2a}}\right)$$

onde a função erro foi definida em (1.7),

$$\int_{-\infty}^{\infty} \frac{dz \exp(-z^2/2)}{\sqrt{2\pi}} \operatorname{erf}(az+b) = \operatorname{erf}\left(\frac{b}{\sqrt{1+2a^2}}\right)$$

e

$$\int_{-\infty}^{\infty} dz e^{-\left(\frac{c^2 z^2}{2}\right)} z \operatorname{erf}(az+b) = \frac{a e^{-\left(\frac{b^2 c^2}{a^2+c^2}\right)}}{c^2 \sqrt{a^2+c^2}}.$$

As equações resultantes deste limite são

$$m_s(t+1) = b \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}} \frac{(1-b^2)\Delta}{bs} \exp\left(-\frac{\Delta^2}{2}\right) \quad (\text{A.36})$$

$$m^1(t+1) = \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right), \quad (\text{A.37})$$

onde

$$\Delta = \frac{b\sqrt{s}m_s(t)}{[\alpha' + m_s^2(t)(1-b^2)]^{1/2}}. \quad (\text{A.38})$$

Para determinar a equação crítica em função dos parâmetros α , b e s , vamos expandir a função erro da equação (A.36), que para valores pequenos do argumento

$$\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \simeq \left(\frac{2}{\pi}\right)^{1/2} x \quad (x \text{ pequeno}).$$

Na transição para a fase de generalização $m_s \ll 1$. Em primeira ordem de m_s obtemos:

$$m_s = \left(\frac{2}{\pi}\right)^{1/2} \frac{b\sqrt{s}m_s}{\sqrt{\alpha'}} + \left(\frac{2}{\pi}\right)^{1/2} \frac{(1-b^2) b\sqrt{s}m_s}{\sqrt{bs\alpha'}}. \quad (\text{A.39})$$

É fácil verificar que esta equação produz o seguinte resultado

$$\alpha = \frac{2 [1 + b^2(s - 1)]^2}{\pi s [1 + (s - 1)b^4]} , \quad (\text{A.40})$$

que corresponde a equação (5.26) do capítulo 5.

Referências

- [1] W. A. Little, *Math. Biosci.* **19**, 101 (1974).
- [2] J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79**, 91 (1982).
- [3] D. Sherrington and S. Kirkpatrick, *Phys. Rev. Lett.* **35**, 1792 (1975).
- [4] S. F. Edward and P. W. Anderson, *J. Phys. F* **5**, 965 (1975).
- [5] D. J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. A.* **32**, 1007 (1985).
- [6] D. J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. Lett.* **55**, 1530 (1985),
Ann. Phys. **173**,30 (1987).
- [7] B. Derrida, E. Gardner and A. Zippelius, *Europhys. Lett.* **4**, 167 (1987).
- [8] M. Abelles, *Studies in Brain Function* (Spring - Verlag, New York 1982).
- [9] B. Alberts, D. Bray, J. Lewis, M. Ralf, K. Roberts and J. Watson, *Molecular Biology of the Cell* (Garland Publishing Inc., New York, 1983).
- [10] J. Darnel, H. Lodish and B. Baltimore, *Molecular Cell Biology* (Scientif American Books, New Yorks,1990).
- [11] D. J. Amit *Modeling Brain Function* (Cambridge University Press, Cambridge,1989).
- [12] P. Peretto, *An Introduction to the modeling of Neural Network* (Cambridge University Press, Cambridge 1994).

- [13] D. O. Hebb, *The Organization of Behaviour: A Neuropsychological Theory* (Wiley, New York 1949).
- [14] A. Crisanti, D. J. Amit and H. Gutfreund, *Europhys Lett.* **2**, 337 (1986).
- [15] G. Parisi, *J. Phys. A* **13**, 1101 (1980).
- [16] G. A. Kohring, *J. Stat. Phys.* **59**, 1077 (1990).
- [17] T. J. P. Pena and P. M. C. Oliveira, *J. Phys. A: Math. Gen.* **22**, L719 (1989).
- [18] P. Peretto, *Biol. Cybern.* **50**, 51 (1984).
- [19] J. F. Fontanari and R. Höberle, *J. Phys. France* **49**, 13 (1988).
- [20] B. Derrida and G. Weisbuch, *J. Physique* **47**, 1297 (1986).
- [21] H. J. Hilhorst and M. Nijmijer, *J. Physique* **48**, 185 (1987).
- [22] J. J. Arenzon and N. Lemke, *J. Phys. A: Math. Gen.* **27**, 5161 (1994).
- [23] H. Gutfreund and M. Mezard, *Phys. Rev. Lett.* **61**, 235 (1988).
- [24] H. Smpolinsky and I. Kanter, *Phys. Rev. Lett.* **57**, 2861 (1986).
- [25] J. Buhmann K. Schulten, *Europhys. Lett.* **4**, 1205 (1987).
- [26] D. Horn and M. Usher, *Phys. Rev. A* **40**, 1036 (1989); D. Horn, *Physica A* **200**, 594 (1993).
- [27] F. Zertuche, R. López-Peña and H. Waelbroech, *J. Phys. A: Math. Gen.* **27**, 5879 (1994).
- [28] J. F. Fontanari, *J. Physique I* **51**, 2412 (1990).
- [29] E. N. Miranda, *J. Phys. I France* **1**, 999 (1991).
- [30] D. A. Stariolo and T. A. Tamarit, *Phys. Rev. A* **43**, 5249 (1992).

- [31] C. R. da Silva, F. A. Tamarit, N. Lemke, J. J. Arenzon and E. M. F. Curado, *J. Phys. A: Math. Gen.* **28**, 1593 (1995).
- [32] K. Huang, *Statistical Mechanics* (New York, Wiley 1987).
- [33] R. M. C. de Almeida and J. R. Igleasis, *Phys. Lett. A* **146**, 239 (1990).
- [34] J.J. Arenzon, R. M. C. de Almeida and J. R. Igleasis, *J. Stat. Phys.* **69**, 385 (1992).
- [35] J. J. Arenzon and R. M. C. de Almeida, *Phys. Rev. E* **48**, 4060 (1993).
- [36] J. Buhmann K. Schulten, *Biol. Cybern.* **54**, 319 (1986); **56**,313 (1987).
- [37] D. J. Amit, M. R. Evans and M. Abeles, *Network* **1**, 381 (1990).
- [38] J. Buhmann, *Phys. Rev. A* **40**, 4145 (1989).
- [39] M.Y. Choi, *Phys. Rev. Lett.* **61**, 2809 (1988).
- [40] K. Aihara, T. Takabe and M. Toyoda, *Phys. Lett. A* **144**, 333 (1990).
- [41] W. Gerstner and J. L. van Hemmen, *Network* **3**, 139 (1992).
- [42] F. A. Tamarit, D. A. Stariolo, S. A. Cannas and P. Serra, *Phys. Rev. E* **53**, 5146 (1996).
- [43] I. Opris, *Phys. Rev. E* **51**, 2619 (1995).
- [44] J. Moreira and D. Auto, *Europhys. Lett.* **21** 693 (1993).
- [45] C. R. da Silva, F. A. Tamarit and E. M. F. Curado, *Int. J. Mod. Phys. C* **7**, 43 (1996).
- [46] J. Hertz, A. Krogh and R. G. Palmer *Introduction to the Theory of Neural Computation* (Reading, MA: Addison-Wesley 1991).
- [47] T. Geszti, *Physical Models of Neural Network*, (Singapore: World Scientific, 1990).

- [48] P. Peretto, *J. Phys. France* **49**, 711 (1988).
- [49] C. R. da Silva, F. A. Tamarit and E. M. F. Curado, *to be published in Phys. Rev. E* **55** (1997)
- [50] T. L. H. Watking and D. Sherrington, *Europhys. Lett.* **14**, 791 (1991).
- [51] T. Fukai and M. Shiino, *Phys. Rev. Lett.* **64**, 1465 (1990).
- [52] A. C. C. Coolen and D. Sherrington, *J. Phys. A: Math. and Gen.* **25**, 5493 (1992).
- [53] C. R. da Silva, L. da Silva and F. A. Tamarit, *submetido*
- [54] B. Derrida and Y. Pomeau, *Europhys. Lett.* **1**, 45 (1986).
- [55] S.A. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
- [56] S.A. Kauffman, *Physica D* **10**, 145 (1984).
- [57] E. Domany and W. Kinzel, *Phys. Rev. Lett.* **53**, 447 (1984).
- [58] M. L. Martins, H. F. V. de Resende, C. Tsallis and A. C. N. de Magalhães, *Phys. Rev. Lett.* **66**, 2045 (1991).
- [59] C. Tsallis and M. L. Martins, *Europhys. Lett.* **27**, 415 (1994).
- [60] T. Tomé, *Physica A* **212**, 99 (1994).
- [61] H. Stanley, J. Kertés and H. Herrmann, *Phys. Rev. Lett.* **59**, 2326 (1987).
- [62] U. M. S. Costa, *J. Phys. A: Math. Gen.* **20**, L-583 (1987).
- [63] A. Neumann and B. Derrida, *J. Phys. France* **49**, 1647 (1988).
- [64] G. Le Caer, *J. Phys. A: Math. Gen.* **22**, L-647 (1989).
- [65] F. A. Tamarit and L. da Silva, *J. Phys. A: Math. Gen.* **27**, L-809 (1994).
- [66] F. Wang, N. Hatano and M. Suzuki, *J. Phys. A: Math. Gen.* **28**, 4543 (1995).

- [67] P. Grassberger, *J. Phys. A: Math. Gen.* **28**, L-67 (1995).
- [68] R. N. Bhatt and A. P. Young, *Phys. Rev. Lett.* **54**, 924 (1985).
- [69] A. T. Ogielski and I. Morgenstein, *Phys. Rev. Lett.* **54**, 928 (1985).
- [70] B. Derrida and G. Weisbuch, *Europhys. Lett.* **4**, 657 (1987).
- [71] M. N. Barber and B. Derrida, *J. Stat. Phys.* **51**, 877 (1988).
- [72] H. R. da Cruz, U. M. S. da Costa and E. M. F. Curado, *J. Phys. A: Math. Gen.* **22**, L-651 (1989).
- [73] F. A. Tamarit and E. M. F. Curado *J. Phys A: Math. Gen.* **27**, 671 (1994), *Lett.* **27**, 415 (1994).
- [74] L. da Silva, F. A. Tamarit and A. C. N. de Magalhães, *submetido*
- [75] M. F. A. Bibiano, F. B. moreira and A. M. Mariz, *submetido*
- [76] F. A. Tamarit, *Tese de Doutorado* (Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro 1992).
- [77] K. Y. M. Wong and C. Ho, *J. Phys. A: Math. Gen.* **27**, 5167 (1994).
- [78] P. Grassberger, *Physica A* **214**, 547 and **217**, 227 (erratum) (1995).
- [79] D. A. Stariolo, *Tese de Doutorado* (Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro 1993).
- [80] N. Parga and J. Virasoro, *J. Phys.* **47**, 1857 (1986).
- [81] M. Feigelman and L. Ioffe, *Int. J. Mod. Phys. B* **1**, 51 (1987).
- [82] H. Gutfreund, *Phys. Rev. A* **37**, 570 (1988).
- [83] M. A. Virasoro, *Europhys. Lett.* **7**, 293 (1988).

- [84] F. A. Tamarit and E. M. F. Curado, *J. Stat. Phys.* **62**, 473 (1991).
- [85] M. C. Branchtein and J. J. Arenzon, *J. Phys. I France* **2** 2019 (1992).
- [86] P. R. Krebs and W. K. Theumann, *J. Phys. A* **26** 3983 (1993).
- [87] C. R. da Silva, em preparação.

“PROPRIEDADES DINÂMICAS DE REDE DE NEURÔNIOS COM PERÍODOS REFRAATÁRIOS E DILUIÇÃO ASSIMÉTRICA”

CRISOGONO RODRIGUES DA SILVA

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

Evaldo Mendonça Fleury Curado - Presidente

Francisco Antonio Tamarit

Rita Maria Cunha de Almeida

Tadeu Josino Pena

Constantino Tsallis

Nami Fux Svaiter

Rio de Janeiro, 10 de abril de 1997