

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

Redes de Neurônios com Multi-Interações[†]

Jeferson J. Arenzon

Tese realizada sob a orientação da Dra. Rita M. C. de Almeida, apresentada ao Instituto de Física da UFRGS em preenchimento parcial dos requisitos para a obtenção do grau de Doutor em Ciências.

Porto Alegre
1995

[†]Trabalho parcialmente financiado por CNPq, FINEP e FAPERGS.

Aos meus avós,
Rebeca, Rosa e Jacó

AGRADECIMENTOS

- À Rita M.C. de Almeida pela orientação, colaboração e amizade.
- À Claudinha pelo amor, apoio e paciência.
- Aos meus pais, Sara Lea e Bernardo, por me trazerem até aqui.
- Ao Ney Lemke pelo companheirismo, colaboração e pela paciência com que me explicou um monte de coisas.
- Ao Roberto Iglesias pelo apoio nestes anos todos.
- Ao Américo T. Bernardes, Miguel C. Branchtein, Marco Aurélio P. Idiart, José Carlos M. Mombach, Paulo Murilo C. de Oliveira, Thadeu J.P. Penna, Jorge A. Quillfeldt e Francisco A. Tamarit pelas discussões e/ou colaborações.
- Ao pessoal da M202 pela companhia nestes anos.
- À turma da biblioteca, especialmente a Ana, Sílvia e Zuleika.

RESUMO

Conexões de ordem mais alta do que dois entre neurônios são estudadas, principalmente para duas regras de aprendizado, a saber, o modelo de Hopfield generalizado e o modelo Truncado. Ambos modelos podem ser considerados como correções ao modelo de Hopfield original, que contém somente interações entre pares de neurônios. Para o modelo Truncado, as regras de aprendizado contêm misturas de memórias, diferentemente de todas prévias generalizações do modelo de Hopfield. As propriedades de equilíbrio são analisadas usando teoria de campo médio (réplicas) e comparadas com simulações numéricas. A dinâmica, no limite de extrema diluição, é também investigada.

Nas versões completamente conexas, enquanto o comportamento de um dos modelos qualitativamente se assemelha ao de Hopfield, o outro apresenta um novo e rico comportamento: dependendo da temperatura T e do peso ε das conexões de quarta ordem, o sistema apresenta duas regiões distintas de recuperação, separadas por uma lacuna, assim como várias transições de fase. Além disso, aparentemente os estados de vidro de spin desaparecem acima de um certo valor do parâmetro de armazenamento α . Um algoritmo ótimo de aprendizado é introduzido, o qual aumenta a capacidade de armazenamento sem aumentar os pesos dos termos de ordem mais alta.

A dinâmica pode apresentar pontos fixos, órbitas periódicas e caóticas, dependendo novamente dos pesos ε , do nível de ruído T e da carga da rede α . Como no caso anterior, há um valor ótimo no valor dos pesos que aumenta a capacidade do sistema.

ABSTRACT

High order connections between neurons are studied, mainly for two different learning rules, the ones from the generalized Hopfield and Truncated models. Both terms may act as corrections to the original, two neuron interaction Hopfield model. In the truncated model, the learning rules contain mixing of memories, differently from all generalizations of the Hopfield model. Equilibrium properties are analyzed using the replica mean-field theory and compared with numerical simulations. The dynamics, in the extreme dilution limit, is also investigated.

In the fully connected versions, while the behavior of one model qualitatively resembles the original Hopfield one, the other presents a new and very rich behavior: depending on temperature and the strength ε of the fourth order connections, the system presents two distinct retrieval regions separated by a gap, as well as several phase transitions. Also, spin glass states seems to disappear above a certain value of the load parameter α . An optimal learning algorithm for fourth order connections is given and the storage capacity is improved without increasing the weight of the higher order term.

The dynamics may display fixed points, periodic and chaotic orbits depending on the weight of the high order connections ε , the noise level T and the network load. As in the related fully connected case, there is an optimal value of the weight ε that improves the storage capacity of the system.

Índice

1	Redes de Neurônios	8
1.1	Introdução	8
1.2	Redes de Neurônios Biológicas	10
1.3	O Ponto de Vista Físico	13
1.4	O Modelo de Hopfield	16
1.4.1	Simulação Numérica	21
1.5	Conclusão	26
2	Modelos com Multi-interações	28
2.1	Multi-interações Em Redes Biológicas	28
2.2	O Modelo de Hopfield Generalizado	29
2.2.1	O Caso $\varepsilon_2 \neq 0$	31
2.2.2	O Caso $\varepsilon_\ell = \delta_{\ell k}, k > 2$	37
2.2.3	O Caso $k \rightarrow \infty$	40
2.2.4	Evolução Dinâmica	41
2.3	A Pseudo-Inversa Generalizada	44
2.4	Conclusões	47
3	O Modelo Completo – Casos OPS e PAS	49
3.1	Formulação do Modelo Geral	49
3.2	O Caso OPS	51
3.2.1	Simulação Numérica	53
3.3	O Caso PAS	54
3.3.1	O Limite $\alpha = 0$	55
3.3.2	O Limite $T = 0$	56
3.3.3	O Caso $T \neq 0$	57
3.3.4	Simulações Numéricas	58

Índice	6
3.4 Padrões com atividade	60
3.5 Conclusão	62
4 O Modelo Truncado – a Estática	64
4.1 O Modelo	64
4.2 Teoria de Campo Médio	66
4.2.1 O Limite $T = 0$	67
4.2.2 O Caso $T \neq 0$	73
4.3 Simulações Numéricas	76
4.4 Conclusões	82
5 O Modelo Truncado - A Dinâmica	84
5.1 O Modelo Truncado Diluído	84
5.2 O Limite $T = 0$	86
5.3 O Caso $T \neq 0$	92
5.4 Recuperação de duas memórias	93
5.5 Conclusão	97
6 Conclusões	98
A O Método Das Réplicas	105
A.1 O Método	105
A.2 A Média Sobre As Memórias Microscópicas	106
A.3 O Modelo Truncado	108
A.4 O Modelo de Hopfield Generalizado	112
A.5 O Modelo Completo	114
B Expansões para o Modelo Truncado	116
B.1 Transições Contínuas em α_c^\pm	116
B.2 O limite $\varepsilon \rightarrow \infty$	117
B.3 Temperatura de vidro de spin	117
C O Modelo de Kauffman	119
D Cálculo da Dinâmica	123

E	O Modelo de Energia Randômicas (REM)	129
E.1	REM	129
E.2	O limite $k \rightarrow \infty$ do modelo GH	129
	Referências Bibliográficas	134

Capítulo 1

Redes de Neurônios

1.1 Introdução

O sistema nervoso e suas partes integrantes foram e são intensamente estudados, fornecendo uma enorme quantidade de dados experimentais sobre sua anatomia e comportamento. À medida em que se coloca o problema da modelagem teórica de tais sistemas, é preciso distinguir quais destas informações são importantes e quais são essenciais, sendo essenciais aquelas sem as quais o comportamento do sistema é trivial e importantes, as responsáveis pela aproximação dos modelos com o sistema real, mas cuja ausência não compromete o comportamento complexo apresentado. A própria distinção entre o que é importante e o que é essencial depende do funcionamento dos modelos e do grau desejado de aproximação com a realidade. A idéia é tentar capturar, em modelos simples, propriedades do cérebro tais como aprendizagem, armazenamento e recuperação de padrões, manipulação simbólica, pensamento criativo, etc. Hoje em dia, porém, a grande maioria dos modelos teóricos se atêm às tarefas de aprendizado, armazenamento e recuperação de informação.

Redes de neurônios têm chamado a atenção da comunidade científica há alguns anos. O assunto, claramente interdisciplinar, tem sido alvo da abordagem por cientistas em várias frentes: biólogos, psicólogos, especialistas em inteligência artificial e, mais recentemente, físicos. Embora o assunto tenha sido matéria de estudo no final dos anos sessenta, somente na década passada houve um renascer do campo com o trabalho seminal de J.J. Hopfield [1], onde é feita a analogia entre um sistema formado por neurônios simplificados e a sua contraparte magnética, os vidros de spin. O preço pago pelo ganho em tratabilidade do sistema, oriundo das simplificações introduzidas, é o afastamento da realidade. A partir do modelo original, à medida que este era entendido mais profundamente, o caminho de volta à realidade foi sendo novamente percorrido. Propriedades dos sistemas reais, que antes não

eram tomadas em consideração, foram aos poucos ganhando espaço. O uso da mecânica estatística mostrou-se muito poderoso na análise de tais sistemas, permitindo estudar as propriedades de equilíbrio e dinâmicas destas redes, as quais são formadas por um número muito grande de elementos simples, os neurônios.

O entusiasmo dos físicos não passou despercebido, como bem descreve Amit [2]: “Biólogos começaram a se perguntar se as propriedades antecipadas pelos físicos podiam de fato ser observadas e se forneciam guias teóricos úteis para a investigação empírica da atividade cerebral; cientistas de computação não desprezaram estas idéias como candidatas para o processamento paralelo; psicólogos e neurologistas têm esperado novas metáforas úteis para a interpretação de disfunções comportamentais;... e tecnólogos têm acrescentado, é claro, estes tipos de modelos à lista dos futuros produtos a serem industrializados e vendidos”.

A aplicação tecnológica das redes de neurônios é um campo de pesquisa que vem recebendo altos investimentos, tanto de pessoal quanto financeiros por parte das indústrias, principalmente aquelas com interesse em controle de qualidade automatizado, reconhecimento de padrões ou fala, etc. As implementações são feitas tanto via programas quanto via mecanismos especializados (redes construídas com amplificadores e resistores ou mecanismos óticos). Existem vários exemplos [3,4], como detectores de explosivos para aeroportos, predição de séries temporais, predição da estrutura secundária de proteínas, diagnósticos médicos, verificação de assinaturas, controle de qualidade de motores por meio do som produzido, etc.

É comum afirmar que o cérebro é uma espécie de computador (na acepção mais geral da palavra) e para entender o seu funcionamento é necessário compreender os princípios computacionais que o sistema nervoso utiliza, ou seja, devemos entender como as células interagem para representar, transformar e armazenar informações. Por outro lado, as diferenças que existem entre um cérebro e um computador são enormes, como por exemplo, a memória em um computador é localizada e a recuperação é feita por endereçamento, ao passo que no cérebro acredita-se que seja distribuída e recuperada por associação. Como o sistema nervoso, os modelos teóricos necessitam ser, entre outros aspectos, robustos (a qualidade da recuperação não deve ser severamente afetada quando uma porção não muito grande da rede é danificada) e associativos (o reconhecimento das memórias é feito pelo seu conteúdo). Embora não haja ainda qualquer modelo que explique exatamente como vemos e aprendemos, há uma crescente convicção de que as atuais linhas de pesquisa estão caminhando em direção às respostas.

O objetivo do trabalho desenvolvido nesta tese é discutir novos modelos para sistemas de neurônios acoplados através de multi-sinapses os quais se comportam de maneira quali-

tativamente distinta dos modelos anteriormente conhecidos. Os resultados obtidos para os vários modelos são comparados entre si, realçando-se a importância deste tipo de conexão para o comportamento das redes.

A tese está assim dividida: nas próximas seções apresentamos brevemente alguns fatos básicos relacionados às redes biológicas e os pontos fundamentais da modelagem de tais sistemas. Também neste capítulo é apresentado o mais estudado modelo de redes de neurônios recorrentes, o modelo de Hopfield, e a sua solução analítica, realizada por Amit *et al* [5–7], juntamente com alguns resultados de simulação numérica. As generalizações deste modelo envolvendo interações entre mais de dois neurônios são estudadas no capítulo 2. Resultados para a estática e a dinâmica são apresentados e comparados com simulações numéricas.

No capítulo 3 os resultados para um modelo recente para multi-interações [8], obtidos via métodos analíticos [9–11] e numéricos [12, 13], são brevemente apresentados. Duas situações foram extensamente estudadas: quando somente as memórias são armazenadas e quando tanto elas quanto as suas inversas são aprendidas. Estes dois casos se comportam distintamente e a inclusão das anti-memórias é fundamental para que o sistema se comporte adequadamente como memória associativa. No capítulo 4 uma versão truncada da energia do modelo acima é estudada utilizando-se o método das réplicas e comparando estes resultados com simulações numéricas. Em seguida, no capítulo 5, a evolução dinâmica deste mesmo modelo, no caso de extrema diluição, é detalhada.

Finalmente, no capítulo 6, resumimos os resultados obtidos, comparamos com os apresentados para os outros vários modelos, indicamos as possíveis extensões do trabalho e concluímos.

Essencialmente, o trabalho original apresentado na tese está compreendido nos capítulos 4 e 5, embora vários outros resultados estejam dispersos nos capítulos restantes.

1.2 Redes de Neurônios Biológicas

O cérebro humano contém cerca de 10^{10} células nervosas ou neurônios [2, 4]. Um neurônio típico é mostrado na figura 1.1, onde suas partes principais são identificadas: o corpo celular (núcleo), o axônio e os dendritos. As conexões entre diferentes neurônios são feitas por meio de sinapses, que são contatos entre os terminais do axônio de uma célula (pré-sináptica) e os dendritos (ou corpo celular) de outra (pós-sináptica), e cujo número médio, para um único neurônio, é da ordem de 10^4 (embora em algumas células possa chegar a 150.000). O sinal chega à célula através dos seus dendritos (processos aferentes), é processado no corpo celular e, se for o caso, é enviado através do axônio em direção

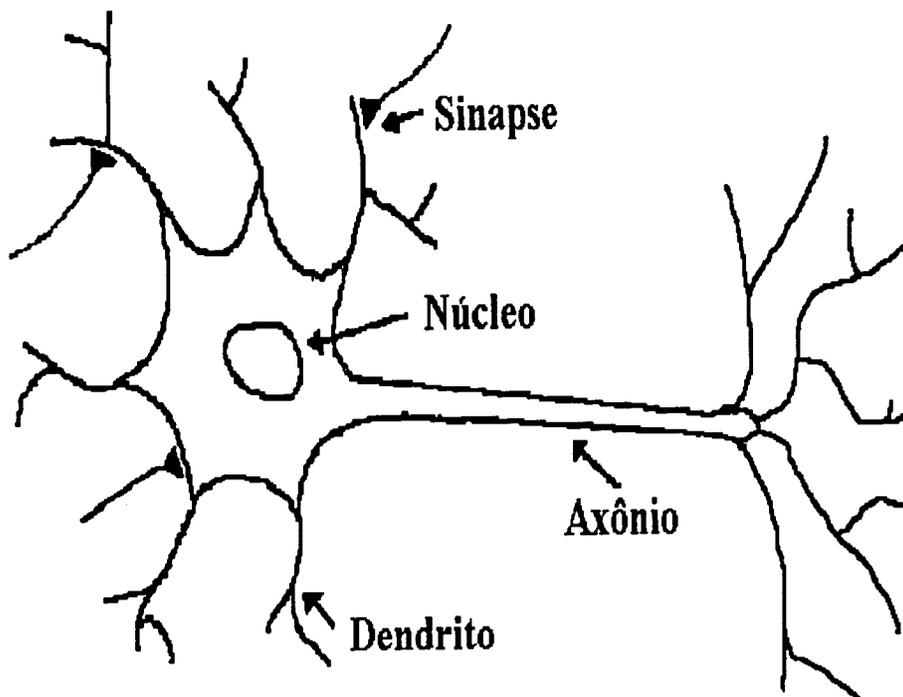


Figura 1.1: Célula piramidal típica encontrada no cérebro. A informação passa através dos dendritos até o corpo celular a partir do qual, se o potencial acumulado ultrapassar um determinado limiar, um pulso é enviado ao longo do axônio [2,4].

aos outros neurônios (processos eferentes). O fato do corpo celular enviar ou não um sinal é determinado pela soma, ponderada pelas intensidades sinápticas, das atividades dos neurônios pré-sinápticos. Em geral, admite-se que este processo de integração dos sinais que ocorre no corpo celular é linear. As sinapses podem ser fracas ou fortes, excitatórias ou inibitórias (neste último caso, a atividade da célula pré-sináptica inibe o funcionamento da pós-sináptica) e o seu funcionamento é basicamente o seguinte: a chegada de um sinal na membrana pré-sináptica causa a liberação de neurotransmissores que se difundem através da sinapse até atingir as moléculas receptoras na membrana pós-sináptica, as quais retransmitem o sinal. Existem algumas evidências [4] de que todas as sinapses de um dado axônio são *ou* excitatórias *ou* inibitórias (lei de Dale). Os dendritos, por sua vez, formam uma estrutura altamente ramificada, como os galhos de uma árvore, sendo que o axônio usualmente também termina em um padrão semelhante.

Os sinais nervosos são transmitidos eletricamente no interior dos axônios a partir do corpo celular até as sinapses com os neurônios vizinhos. Quando em repouso, há um excesso de cargas negativas no interior do neurônio, cujo potencial em relação ao meio exterior é $-70mV$ (a membrana celular é impermeável aos íons de Na^+ que estão no líquido intercelular). Os sinais que chegam ao neurônio causam uma pequena diminuição neste potencial (despolarização) e, quando este atinge $-60mV$, a membrana torna-se permeável aos íons

positivos, os quais penetram no interior da célula e atenuam o potencial. Esta descarga pode ser tão violenta que o neurônio adquire um potencial positivo em relação ao meio. Logo após, a membrana começa o processo de recuperação e regenera o potencial original (o que leva alguns milissegundos). Durante este tempo, o qual é chamado de período refratário, o neurônio é incapaz de ser novamente excitado. Este processo ocorre inicialmente no corpo celular e em seguida o sinal viaja ao longo do axônio (é unidirecional pois as regiões que vão ficando para trás estão se recuperando e não podem ser novamente excitadas) até atingir uma das sinapses onde o processo de lançamento de neurotransmissores é inicializado.

Uma sinapse pode alterar o sinal que está sendo transmitido através dela e, mais importante ainda, as propriedades das sinapses podem também ser modificadas pelos sinais que as atravessam [14]. Esta plasticidade é o ponto básico do aprendizado: as sinapses são criadas ou têm suas intensidades alteradas quando novos padrões são aprendidos. De acordo com o próprio Hebb [14]:

“Quando um axônio da célula *A* está próximo o suficiente para excitar *B* e repetida ou persistentemente tomar parte no seu disparo, algum processo de crescimento ou 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 aumentada¹”.

Este processo de condicionamento sináptico já havia sido enunciado no século passado por William James [15]:

“Quando dois processos neurais ocorrem em rápida sucessão, um tende a evocar o outro²”.

De qualquer maneira, os mecanismos biológicos responsáveis pela alteração necessária nas conexões são bastante complicados e ainda incompreendidos [16].

O córtex cerebral, ou neocórtex, consiste de camadas de células com aproximadamente 3 mm de espessura, as quais são dobradas várias vezes para permitir sua acomodação no crânio [2,4]. Também há evidências de que os neurônios se agrupam em colunas com cerca de 0.5 mm de diâmetro, perpendiculares à superfície cortical.

¹When an axon of cell *A* is near enough to excite cell *B* and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that *A*'s efficiency, as one of the cells firing *B*, is increased.

²When two neural processes occur in close succession, one tends to evoke the other.

Resta falar da atividade do sistema nervoso, a qual é uma medida da fração dos neurônios que estão ativos (emitindo sinal) em um instante de tempo. A média espacial do nível de atividade do córtex é relativamente baixa: nas regiões onde são detectadas funções associativas, a quantidade de neurônios ativos é de 4 – 5% [17].

Diferentemente da arquitetura dos atuais computadores, no cérebro as memórias estão distribuídas pelas sinapses que compõem o sistema [2, 4, 18]. A recuperação de informação é associativa, ou auto-endereçada, e as memórias são recuperadas pela semelhança de seu conteúdo e não pelo endereço: a dinâmica da rede, ao ser colocada em um estado que se assemelha a um determinado padrão aprendido, se encarrega de corrigir os erros e convergir para aquela memória.

Muitos outros aspectos do funcionamento do cérebro já foram estudados e a quantidade de dados gerada até hoje é enorme, ficando além dos objetivos desta tese analisar em detalhe a fisiologia e o comportamento de neurônios reais. Sendo assim, vamos passar a discutir o modelamento de tais sistemas.

1.3 O Ponto de Vista Físico

Redes de neurônios formais (simplificados) servem de paradigma para estudar o comportamento complexo de sistemas nervosos reais. Embora alguns tipos de cérebros sejam dotados de funções mentais de alto nível, a maioria dos modelos físicos existentes se ocupa em reproduzir as capacidades de aprendizado e recuperação de informação como um passo inicial em direção à descrição de processos mais elevados. O comportamento de tais modelos, longe de ser considerado trivial, apresenta várias características que são comuns aos seus análogos biológicos.

Em 1943, McCulloch e Pitts [19] mostraram que qualquer operação do cálculo proposicional pode ser efetuada por neurônios simplificados combinados em pequenas seqüências temporais. A atividade do neurônio de saída representa o valor de verdade da operação lógica efetuada pelos neurônios. Isto pode ser feito escolhendo-se adequadamente os valores das conexões sinápticas e dos limiares de cada componente. Nos casos mais simples como a operação lógica AND, bastam dois neurônios de entrada e um de saída (perceptron), mas no caso do XOR (“ou” exclusivo) uma camada a mais precisa ser introduzida. Estes neurônios formais são representados pela variável S_i , cujo valor pode ser +1 (ativo) e -1 (inativo) ou, equivalentemente, 1 e 0. Uma rede de neurônios formais é um conjunto formado por N destes elementos simples, altamente conectados, cujo estado será então associado a um vetor N -dimensional $\mathbf{S} = (S_1, \dots, S_N)$.

As propriedades das conexões entre dois neurônios i e j podem ser modeladas por um número real, J_{ij} . Uma conexão sináptica é excitatória quando $J_{ij} > 0$ e inibitória se $J_{ij} < 0$. Hebb [14] propôs que o sistema nervoso pode aprender e fazer associações através de modificações seletivas das conexões sinápticas. Isto se reflete, nos modelos físicos, na forma funcional de J_{ij} , que depende das memórias armazenadas. É importante salientar que na literatura, o termo ‘redes de neurônios’ as vezes é usado com sentido um pouco mais amplo do que o definido aqui, para englobar sistemas cujas conexões são escolhidas ao acaso, sem que haja um conjunto de memórias a serem aprendidas, mas que no restante obedecem os princípios aqui expostos.

Uma rede de neurônios é um sistema dinâmico, cuja evolução é governada por regras que dependem do modo com que um neurônio responde aos sinais que chegam até ele [2,18,20]. O potencial pós-sináptico, ou campo local, no i -ésimo neurônio é a soma, ponderada pelas intensidades das conexões sinápticas, da atividade do resto da rede e, quando consideramos apenas interações entre dois neurônios, pode ser escrito como

$$h_i = \sum_{j \neq i} J_{ij} S_j(t) \quad . \quad (1.1)$$

Além disso, cada célula possui o seu próprio limiar θ_i . Se o campo local ultrapassar o limiar, a probabilidade de que este neurônio envie um sinal é

$$Pr(S_i = +1) = f(h_i - \theta_i) \quad , \quad (1.2)$$

onde f em geral é uma função sigmóide (e.g., \tanh). A largura da região de f que vai de 0 a 1 (ou de -1 a $+1$) é uma medida da estocasticidade do processo ou, em outras palavras, do ruído na rede, o qual, daqui por diante, será denotado pela temperatura $T = \beta^{-1}$. As origens físicas deste ruído podem ser as mais variadas possíveis [21]: variações randômicas no número e no tamanho dos pacotes de neurotransmissores químicos lançados nas sinapses, “vazamento” destes neurotransmissores (que pode acontecer mesmo se o neurônio estiver inativo) entre outras flutuações estocásticas que podem ocorrer ao longo do processo.

A figura 1.2 apresenta esquematicamente a forma de um neurônio. As linhas J_{ij} , $j \neq i$, são as sinapses que o conectam com os outros N neurônios (quadrados pequenos) da rede, h_i é o potencial pós-sináptico atuando em i e θ_i é o limiar. A função $f(h_i - \theta_i)$ determina o estado final S_i do neurônio.

No limite de ruído nulo ($T = 0$), a dinâmica da rede obedece à seguinte regra determinística:

$$S_i(t+1) = \text{sgn} \left[\sum_{j \neq i} J_{ij} S_j(t) - \theta_i \right] \quad , \quad (1.3)$$

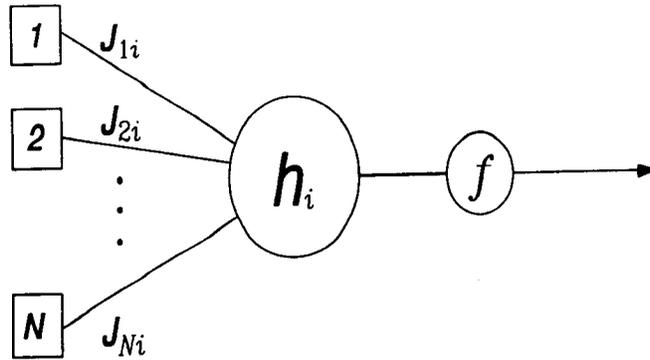


Figura 1.2: Estrutura lógica de um neurônio [2]. Os sinais provenientes dos neurônios $j = 1, \dots, N$ são totalizados no núcleo da célula e o resultado, o campo local, sofre a ação de uma função geralmente não linear, $f(h_i - \theta_i)$, sendo em seguida transmitido para os neurônios seguintes.

onde a função sinal vale ± 1 se o argumento for positivo ou negativo, respectivamente.

A dinâmica ditada pela eq.(1.3) pode ser implementada em série (assíncrona) ou em paralelo (síncrona). No esquema serial, em cada unidade de tempo, um neurônio é escolhido aleatoriamente e alinhado com o seu campo local, enquanto que na dinâmica em paralelo, em cada instante de tempo todos os neurônios são atualizados simultaneamente.

A eq.(1.3) pode levar a estados estáveis, $S_i(t+1) = S_i(t), \forall i$ (pontos fixos da dinâmica) onde $S_i h_i > 0$, isto é, todos os neurônios estão alinhados com os campos que atuam sobre eles. Nas redes de neurônios com atratores ('attractor neural network' ou simplesmente ANN), a matriz J das conexões sinápticas é construída de tal modo que os elementos de um conjunto previamente escolhido de padrões (estados possíveis da rede), designados por $\xi^\mu = (\xi_1^\mu, \dots, \xi_N^\mu)$, $\mu = 1, \dots, P$, sejam os pontos fixos (ou estejam muito próximos destes) da equação (1.3). O limite da capacidade de armazenamento da rede, $\alpha = P/N$, é uma medida do número de memórias que podem ser armazenadas e recuperadas com relativo sucesso.

Estamos interessados nas propriedades do sistema que são independentes da particular escolha dos padrões a serem armazenados. Para tanto, faremos uma média sobre todos os possíveis conjuntos de memórias [2, 5, 20], a qual pode estar sujeita a alguma espécie de vínculo (atividade, correlação, etc). Consideramos as componentes ξ_i^μ dos P padrões como variáveis aleatórias independentes, cuja distribuição de probabilidade é dada por

$$P(\xi_i^\mu) = a\delta(\xi_i^\mu - 1) + (1 - a)\delta(\xi_i^\mu + 1) \quad , \quad (1.4)$$

onde a é, quando $N \rightarrow \infty$, a fração de neurônios ativos na rede, o que é chamado de

atividade. Boa parte do que será visto nas seções e capítulos seguintes considera redes com atividade 0.5, ou seja, com o mesmo número de neurônios ativos e inativos. Esta restrição não é muito realista, uma vez que, como vimos anteriormente, a atividade do córtex é muito menor do que isso ou, no caso de sistemas montados com a finalidade de reconhecer padrões, boa parte das memórias armazenadas possui um pano de fundo comum. Entretanto, este é o caso mais simples e, como veremos logo a seguir, vários dos modelos apresentados somente se comportam adequadamente quando $a = 0.5$.

A memória de uma ANN é associativa, ou auto-endereçada: partindo de um estado \mathbf{S} próximo a uma das memórias ξ^μ , o processo dinâmico (1.3) converge para o padrão armazenado. Isto acontece desde que o sistema se encontre numa região ao redor da memória ξ^μ , conhecida como bacia de atração. A medida de quão próximo um estado está da μ -ésima memória é dada pela superposição ('overlap') m_μ , definida como

$$m_\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i \quad , \quad (1.5)$$

a qual vale 1 se $\mathbf{S} = \xi^\mu$.

1.4 O Modelo de Hopfield

O modelo de Hopfield contém suposições bastante simples e, apesar disso, apresenta características de armazenamento e recuperação de memória que o tornam de especial interesse. Além das propriedades da seção anterior, o modelo considera que as interações ocorrem entre pares de neurônios e são simétricas, isto é

$$J_{ij} = J_{ji} \quad , \quad (1.6)$$

o que significa que a influência do neurônio j no neurônio i é igual à influência do i no j . Esta suposição, do ponto de vista biológico, não é realística uma vez que os canais de comunicação nas redes de neurônios reais são unidirecionais. A grande vantagem é que tal simetria garante a existência de uma função energia, cujos mínimos são os estados estáveis da dinâmica dada por (1.3) (em $T = 0$) e cuja existência permite usar várias técnicas da mecânica estatística. A forma particular de J_{ij} a ser considerada aqui, a qual leva o nome de regra de aprendizagem de Hebb (apesar de ser apenas uma das possíveis implementações do princípio enunciado acima), é

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \quad . \quad (1.7)$$

Normalmente assume-se que $J_{ii} = 0$, uma vez que os neurônios não costumam possuir sinapses consigo mesmo. Esta é uma regra local (o valor de J_{ij} depende somente de ξ_i^μ e ξ_j^μ), aditiva e, além disso, as memórias são aprendidas igualmente (não há distinções entre memórias novas e antigas).

O modelo de Hopfield é descrito pela seguinte função energia³

$$E = -\frac{1}{2} \sum_{\substack{i,j \\ j \neq i}} J_{ij} S_i S_j \quad , \quad (1.8)$$

onde J_{ij} é dado por (1.7). Substituindo (1.7) em (1.8), a energia pode ser reescrita em termos das superposições m_μ :

$$E = -\frac{N}{2} \sum_{\mu=1}^P m_\mu^2 + \frac{P}{2} \quad , \quad (1.9)$$

onde a constante aditiva $P/2$ vem do termo de auto-interação e pode ser desprezada.

Quando o número de memórias armazenadas é finito ($\alpha = 0$) o sistema apresenta o seguinte comportamento [5]: a altas temperaturas ($T > 1$), somente o estado paramagnético ($m = 0$) é estável, sendo que à medida que a temperatura diminui abaixo de 1, estados que se superpõem igualmente com um número ímpar (crescente) de memórias (estados simétricos) vão sucessivamente se tornando estáveis. O estado que representa uma superposição com somente um dos padrões é estável para todo valor de temperatura menor do que $T_c = 1$.

Para o caso em que P é uma quantidade extensiva ($\alpha \neq 0$), os resultados são obtidos por meio do método das réplicas (veja o apêndice A para os detalhes), o qual é utilizado quando temos que realizar uma média sobre as variáveis randômicas fixas ('quenched') de um sistema (no caso, as memórias) [23,24]. Para o modelo de Hopfield, a energia livre e as equações de ponto de sela, as quais descrevem os estados de equilíbrio do sistema são [6,7] (ver apêndice A):

$$\begin{aligned} f &= \frac{1}{2} \sum_{\mu} m_{\mu}^2 + \frac{\alpha}{2\beta} \ln[1 - \beta(1 - q)] - \frac{1}{2} \frac{\alpha q}{1 - \beta(1 - q)} \\ &+ \frac{1}{2} \alpha \beta r (1 - q) - \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \beta \left[m \cdot \xi + \sqrt{\alpha r} z \right] \right\rangle \right\rangle \end{aligned} \quad (1.10)$$

³Formalmente, a regra (1.7) e a função energia (1.8) foram introduzidas alguns anos antes de Hopfield por Pastur e Figotin [22], mas no contexto de vidros de spin, sendo que uma versão equivalente ao caso $\alpha = 0$ foi resolvida.

e

$$m = \langle\langle \xi \tanh \beta (m \cdot \xi + \sqrt{\alpha r} z) \rangle\rangle \quad (1.11a)$$

$$q = \langle\langle \tanh^2 \beta (m \cdot \xi + \sqrt{\alpha r} z) \rangle\rangle \quad (1.11b)$$

$$r = \frac{q}{[1 - \beta(1 - q)]^2} \quad (1.11c)$$

onde o símbolo $\langle\langle \dots \rangle\rangle$ significa duas médias: uma sobre o número finito de memórias que podem se condensar e outra sobre a variável gaussiana z , relacionada ao infinito número de memórias com superposição microscópica. O parâmetro de ordem q é uma medida do grau de congelamento da rede (apêndice A).

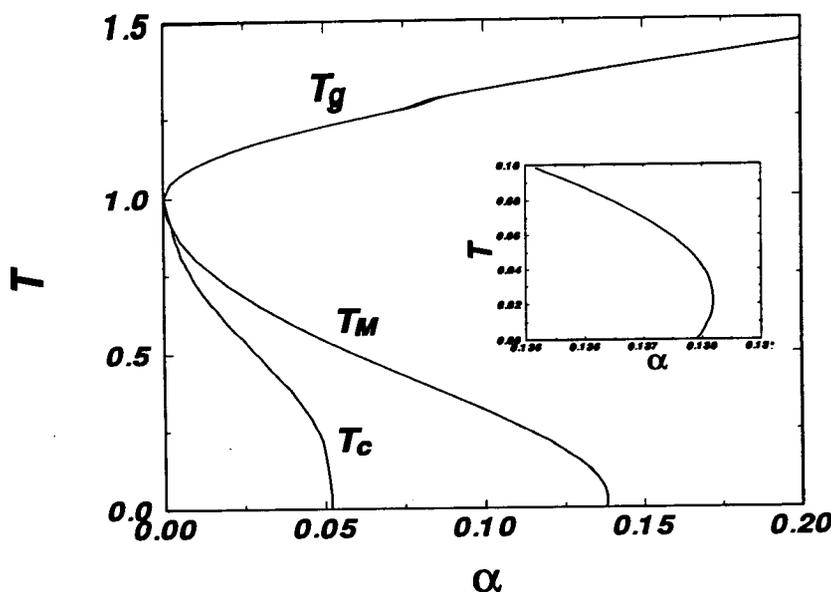


Figura 1.3: Diagrama de fases para o modelo de Hopfield [6, 7]. T_g é a temperatura de transição para o estado de vidro de spin, T_M é a temperatura abaixo da qual surgem os estados que se superpõem com apenas uma memória e T_c é a transição onde estes estados se tornam mínimos globais da energia livre.

Os principais resultados obtidos das equações acima quando $m_\mu = \delta_{1\mu} m$ são apresentados a seguir e deles resulta o diagrama de fases mostrado na figura 1.3. O principal resultado é que, permitindo uma pequena fração de erros, a capacidade da rede em $T = 0$, dada pelo valor máximo do parâmetro α , vale $\alpha_c \simeq 0.1379^4$. Abaixo de α_c o sistema

⁴Do ponto de vista matemático (veja a referência [25] para uma exposição sobre resultados rigorosos em redes de neurônios), alguns teoremas foram recentemente demonstrados definindo limites inferiores para o

efetivamente funciona como uma memória associativa, sendo que os estados estáveis não são mais os padrões originais (que são desestabilizados), mas estados muito semelhantes àqueles. Desde que α permaneça abaixo do limite acima, a superposição entre estes estados, os quais são chamados de estados de recuperação, e as memórias originais é maior do que $m \simeq 0.967$, que é o valor da superposição em α_c . À medida que α tende a 0, a superposição tende a crescer, sendo que para $T \rightarrow 0$ o crescimento é exponencial, $m \simeq 1 - e^{-1/2\alpha}$. Estes resultados são obtidos no *Ansatz* de simetria de réplicas (apêndice A), ou seja, quando os parâmetros que definem cada uma das réplicas do sistema são iguais entre si. Esta solução, porém, é instável a baixas temperaturas. Quando a simetria das réplicas é quebrada, o valor de α_c aumenta, embora ainda exista uma certa divergência sobre o valor exato que α_c assume: Crisanti *et al* [26] obtiveram $\alpha_c \simeq 0.144$ ao passo que Steffan *et al* [27] obtiveram $\alpha_c \simeq 0.1382$, ambos com um passo de quebra de simetria⁵. O valor obtido por meio de simulações numéricas [28] é $\alpha_c \simeq 0.143 \pm 0.001$.

Os estados de recuperação tornam-se estados metaestáveis para $T < T_M(\alpha)$, onde $T_M(\alpha)$ decresce desde 1 em $\alpha = 0$ até 0 em $\alpha = \alpha_c$. Abaixo de $T_c(\alpha)$, onde $T_c(\alpha) < T_M(\alpha)$, as memórias tornam-se mínimos globais da energia livre, ao passo que para $T_c < T < T_M$, os estados de vidro de spin ('spin-glass', SG) são os mínimos globais (e existem para todos os valores de α). Portanto, a recuperação das memórias é estável frente a uma quantidade não muito grande de ruído térmico. Quando α é finito, há dois tipos de estados espúrios: para α suficientemente pequeno temos os estados simétricos, que possuem superposições macroscópicas com diversos padrões. Além destes estados, temos ainda os estados de vidro de spin, os quais possuem uma superposição muito pequena com as memórias, $\mathcal{O}(1/\sqrt{N})$, e aparecem para $T < T_g = 1 + \sqrt{\alpha}$. Tanto na fase paramagnética quanto na de vidro de spin, as superposições são nulas com todas as memórias ($m_\mu = 0, \forall \mu$), embora na fase de vidro de spin, $q \neq 0$.

Quando a fração de neurônios ativos nas memórias difere de 0.5 ($a \neq 0.5$), introduzimos correlações⁶ entre os padrões armazenados. Toda componente ξ_i^μ em um padrão é ser escolhida independentemente com probabilidade $P(\xi_i^\mu)$ dada pela equação (1.4). Definimos

valor de α_c , ao passo que a obtenção rigorosa de um limite superior ainda é um problema em aberto. Em particular, pode-se demonstrar que $\alpha_c > 0.071$.

⁵As equações obtidas em ambos os casos são as mesmas, a diferença reside no método numérico de resolvê-las. Crisanti *et al* [26] minimizaram f no espaço dos parâmetros de ordem, ao passo que Steffan e Kühn [27] resolveram as equações de ponto de sela, obtendo os seguintes valores: 0.137905 (RS), 0.138186 (1RSB) e 0.138187 (2RSB).

⁶Aqui a palavra correlação não é utilizada no sentido correto pois $\langle\langle \xi_i^\mu \xi_j^\nu \rangle\rangle = \delta_{\mu\nu} \delta_{ij} + (1 - \delta_{\mu\nu} \delta_{ij}) \langle\langle \xi_i^\mu \rangle\rangle \langle\langle \xi_j^\nu \rangle\rangle$, mas este uso já se consolidou na literatura.

$b = 2a - 1$ que é o valor médio de ξ_i^μ e assume valores entre -1 e $+1$. Com esta distribuição de padrões armazenados, as memórias são correlacionadas do seguinte modo

$$\langle\langle \xi^\mu \xi^\nu \rangle\rangle = \delta_{\mu\nu} + b^2(1 - \delta_{\mu\nu}) \quad . \quad (1.12)$$

Quando considerando a regra de Hebb usual [17], eq.(1.7), os padrões originais são desestabilizados quando $P = 1 + b^{-2}$, ou seja, para o nível de atividade cortical ($\sim 5\%$), podemos armazenar perfeitamente somente 2 memórias! Na tentativa de contornar o problema, Amit *et al* [17] propuseram a seguinte modificação na regra de aprendizagem:

$$J_{ij} = \frac{1}{N} \sum_{\mu} (\xi_i^\mu - b)(\xi_j^\mu - b) \quad . \quad (1.13)$$

Agora a regra deixa de ser local, pois introduzimos uma característica que depende de toda a rede (a atividade). O modelo de Hopfield com tal regra possui o dobro de estados espúrios do que o caso não correlacionado ($b = 0$), pois tanto as soluções simétricas pares quanto as ímpares são estáveis em $T = 0$. Além disso, quando $|b|$ torna-se maior do que $\sqrt{2} - 1$, os estados simétricos tornam-se mínimos globais da energia. Quando $T \neq 0$, temos uma temperatura crítica $T_c(b) = 1 - b^2$, abaixo da qual as memórias se tornam instáveis para $b^2 > 1/3$. Quando a temperatura diminui, primeiro aparecem os estados espúrios estáveis. Somente a temperaturas mais baixas os padrões armazenados tornam-se estáveis, mas neste ponto os estados espúrios (pares) já possuem energias menores. Apesar disso, com a modificação, novamente temos armazenamento de um número de memórias proporcional a N .

Os resultados acima representam as propriedades estáticas do sistema, ou seja, seus estados de equilíbrio. Por outro lado, existe um limite onde resultados analíticos podem ser facilmente obtidos para a dinâmica do modelo de Hopfield: o de extrema diluição [29]. Neste caso, não mais temos cada neurônio conectado a todos os outros $N - 1$, mas somente a um grupo reduzido destes ($C \ll \ln N$). No apêndice D mostramos em detalhe como obter a evolução da superposição $m(t)$ no tempo, a qual é dada por

$$m(t+1) = \int_{-\infty}^{+\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2} \tanh [\beta (m(t) - y\sqrt{2\alpha})] \quad , \quad (1.14)$$

onde $\alpha = P/C$. Note que esta definição de α é compatível com a que vínhamos usando uma vez que no caso onde os neurônios se ligam todos com todos, $C = N$. Os pontos fixos exibidos pela equação apresentam dois regimes: abaixo de $\alpha_c(T)$ existe uma solução estável com $m \neq 0$ ao passo que acima, somente existe $m = 0$. Em $T = 0$, $\alpha_c(T = 0) = 2/\pi$. Na figura 1.4 apresentamos o diagrama T versus α do sistema.

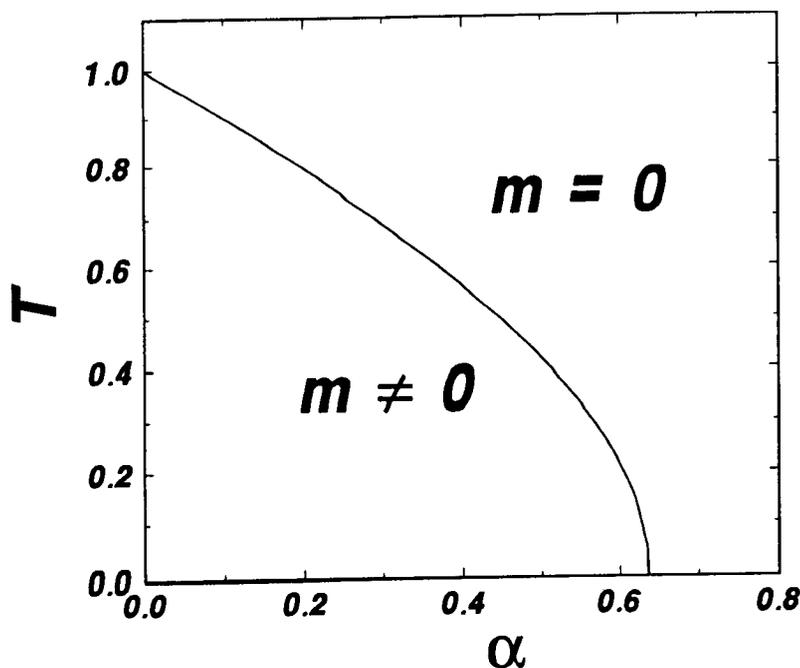


Figura 1.4: Diagrama de fases para a versão diluída do modelo de Hopfield. Existem duas fases: $m \neq 0$ para baixas temperaturas e pequenos valores de α e $m = 0$ nos outros casos. Embora $m = 0$ seja sempre solução de (1.14), ela somente é estável para $T > T_c(\alpha)$.

1.4.1 Simulação Numérica

Os tratamentos analíticos podem se tornar, em alguns casos, muito difíceis de serem levados a cabo e, de qualquer modo, seria interessante que os resultados assim obtidos pudessem ser testados de alguma forma. São nessas situações que o recurso à simulação numérica se faz necessário. Além disso, a simulação contribui para um maior conhecimento do modelo que está sendo estudado, servindo como guia para a intuição. Entre as quantidades passíveis de serem medidas, escolhemos umas poucas que melhor representam a habilidade do modelo de operar como uma memória associativa: a capacidade máxima de armazenamento α_c , os tamanhos das bacias de atração associadas às memórias, uma medida do tamanho relativo das bacias de atração dos estados espúrios e os tempos médios de convergência, isto é, o número de passos necessários a partir de um determinado estado inicial para atingir a memória escolhida. Associadas a estes tempos temos as dispersões ao redor do valor médio, as quais fornecem informações relevantes sobre a estrutura do espaço de fase que cerca os padrões armazenados.

Vamos iniciar, então, com a capacidade máxima de armazenamento, que é medida pelo parâmetro α , como vimos anteriormente. A obtenção numérica de α_c para o modelo de Hopfield foi feita primeiramente por Amit *et al* [5] e, posteriormente, com maior precisão,

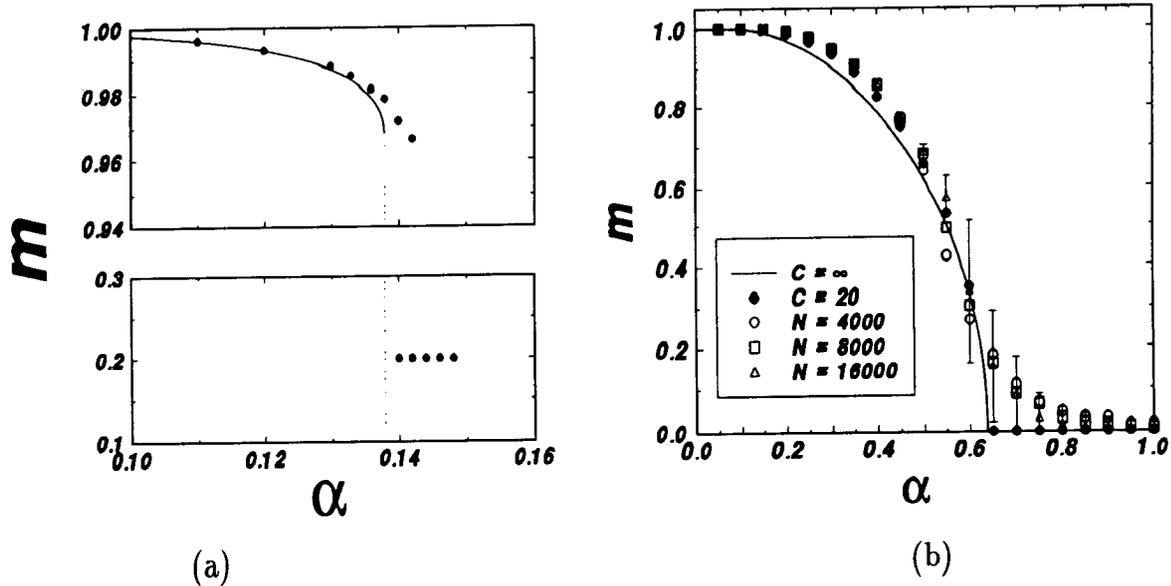


Figura 1.5: A superposição final m em função de α para o caso completamente conexo [28] (a), $C = N$, e para o caso diluído [30] (b), com $C = 20$.

por Kohring [28]. O procedimento é o seguinte: após armazenar P padrões em uma rede composta de N neurônios, escolhe-se um estado inicial igual a um dos padrões, e testa-se, sucessivamente, a estabilidade de cada um dos neurônios. Se a inversão do estado de algum neurônio diminui a energia, tal inversão é efetuada. Este processo é repetido até atingir-se uma configuração estável. Na figura 1.5a mostramos a superposição final contra α extrapolada para grandes valores de N [28], sendo que o valor obtido numericamente para α_c é 0.143 ± 0.001 . Note que existe uma região, logo abaixo de α_c , onde coexistem duas soluções, uma com $m \simeq 1$ e outra com $m \simeq 0.2$. Isto pode ser interpretado como se algumas das memórias já estivessem desestabilizadas nesta região, enquanto que as outras ainda permanecessem estáveis. Qualquer que seja a estimativa que tomemos para o valor teórico de α_c [26,27], a concordância é muito boa. A mesma medida pode ser feita no caso diluído sem que sejam necessárias redes exponencialmente grandes, como mostrado por Arenzon e Lemke [30]. Na figura 1.5b mostramos m contra α no caso em que $C = 20$. Junto com os resultados de simulação são também mostradas as soluções (pontos fixos em $T = 0$) da equação (1.14) ($C = \infty$) e da equação equivalente para C fixo (eq.(D.25) com $C = 20$).

Os tamanhos das bacias de atração que rodeiam cada uma das memórias também podem ser estimados [31]. O processo é o mesmo descrito acima, apenas que o estado inicial possui uma superposição m_0 com uma das memórias, escolhida ao acaso. Ao final do processo,

quando a dinâmica atingiu um estado estável, a superposição final m é registrada, a qual é mostrada na figura 1.6, para vários valores de α e N . Para $\alpha < \alpha_c$ as curvas tendem a aumentar a sua inclinação à medida que N aumenta, ao contrário do que acontece para $\alpha > \alpha_c$.

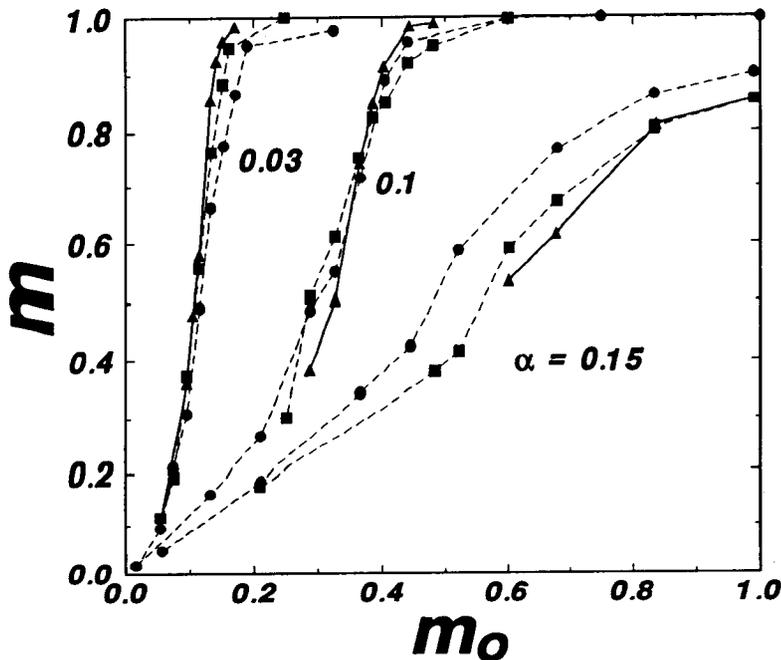


Figura 1.6: A superposição final m dos estados recuperados a partir de um estado inicial com superposição m_0 para vários valores de α . As médias foram feitas sobre mais de 1000 estados iniciais e os símbolos se referem a $N = 512$ (●), 1024(■) e 2048(▲) [31].

Vamos agora analisar os resultados encontrados para o tempo de convergência, definido como o tempo necessário para, a partir de um certo estado inicial, atingir a memória escolhida [32,33]. Na simulação descrita acima, pode-se medir o número de atualizações dos neurônios até que o estado estável final seja atingido. Tal número, dividido pelo tamanho da rede N , fornece o tempo de convergência $\langle T \rangle$, mostrado na figura 1.7 para $\alpha = 0.1$. No histograma da figura 1.7a pode-se notar que quanto maior for N , mais alta e menos espalhada será a distribuição de $\langle T \rangle$ e que, além disso, $\langle T \rangle$ é uma função crescente de N . Além do crescimento de $\langle T \rangle$ com N , na figura 1.7b podemos ver que para estados iniciais próximos da memória, aparentemente há uma diminuição no número de estados espúrios, implicando na diminuição da inclinação da curva para valores grandes de N . Na figura 1.8 a variância no tempo de convergência, $\sigma^2 = \langle T^2 \rangle - \langle T \rangle^2$, versus N é mostrada. Estas duas quantidades, $\langle T \rangle$ e σ^2 , estão relacionadas com a homogeneidade e suavidade do espaço de fase ao redor das memórias armazenadas. Por exemplo, dispersões altas são causadas pela existência de distintos caminhos levando até o fundo da bacia de atração, cada um deles

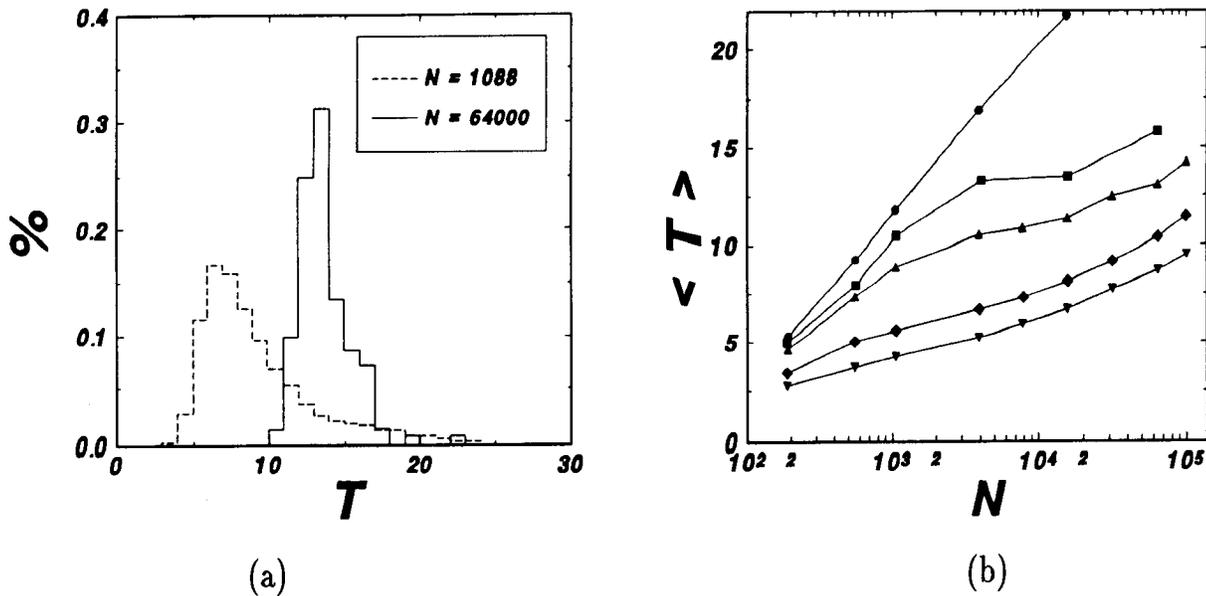


Figura 1.7: Tempos de convergência no modelo de Hopfield para $\alpha = 0.1$ [32]: a) histograma para $m_0 = 0.4$ e b) em função de N para vários valores de m_0 : (\bullet) $m_0 = 0.35$, (\blacksquare) $m_0 = 0.375$, (\blacktriangle) $m_0 = 0.4$, (\blacklozenge) $m_0 = 0.5$ e (\blacktriangledown) $m_0 = 0.6$.

com um tempo de convergência diferente. Isto é devido à existência de estados espúrios que retardam a evolução do sistema. O tempo de convergência foi estudado também quando α está próximo de α_c [34], e encontra-se que $\langle T \rangle$ aumenta exponencialmente ao passo que a distribuição se alarga a medida que $\alpha \rightarrow \alpha_c$. Os resultados podem ser razoavelmente descritos por $\langle T \rangle = \exp[-A(N)(\alpha_c - \alpha)^\beta]$, onde β é da ordem de 1 (o valor exato depende de m_0 e de N) e $A(N)$ é uma constante que depende do tamanho do sistema. De qualquer maneira, os tamanhos utilizados neste caso ($N = 16.000$) são pequenos e redes maiores são necessárias para resultados mais definitivos.

Uma outra estimativa do papel representado pelos estados espúrios pode ser obtida medindo-se \mathcal{M} , quantidade introduzida por Arenzon e de Almeida [35], a qual é proporcional à ocupação fracional do espaço de fase pelas bacias de atração dos estados espúrios⁷. Diferentemente das simulações anteriores, o estado inicial que antes possuía alguma semelhança com uma das memórias, agora é escolhido aleatoriamente. Após o sistema atingir um estado estável, registramos o valor da maior superposição com alguma das memórias,

⁷Poderíamos ter medido a fração de vezes em que um estado inicial aleatório converge para uma das memórias, mas além de termos que definir um critério de recuperação, o fato da transição ser contínua traria problemas próximo ao ponto de transição.

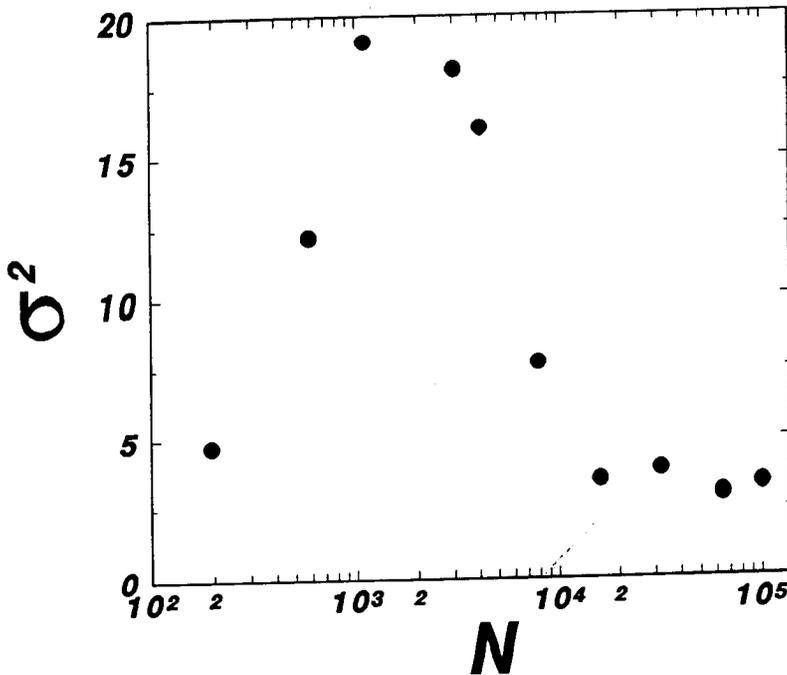


Figura 1.8: Variância da distribuição dos tempos de convergência versus N para $\alpha = 0.1$ e $m_0 = 0.4$ [32].

ou seja, $m_{ris} = \max(\{m_\mu\})^8$. A idéia é que se o sistema parte de posições aleatórias, o estado final ou é uma das memórias ou um estado espúrio (se existir algum). Portanto, se compararmos o valor obtido com aquele onde o estado inicial é uma das memórias, $m_{m_0=1}$, e após efetuarmos uma média sobre um número considerável de estados iniciais e conjuntos de memórias, obtemos \mathcal{M} :

$$\mathcal{M} \equiv \langle\langle m_{m_0=1} - m_{ris} \rangle\rangle \quad (1.15)$$

O símbolo $\langle\langle \dots \rangle\rangle$ significa que a média é tomada sobre vários conjuntos de memórias e estados iniciais. Se $\alpha > \alpha_c$, embora não exista um estado de recuperação, o estado final apresenta uma dependência na condição inicial [28], e \mathcal{M} não fornece nenhuma informação relevante. Como podemos ver na figura 1.9, abaixo de α_c , \mathcal{M} aumenta com α e N , sendo que a dependência com α é linear (o expoente depende de N). Ou seja, os estados espúrios e suas bacias ocupam uma porção considerável do espaço de fases e seu papel é extremamente importante. Para efeito de comparação, mostramos \mathcal{M} (fig. 1.9) para o caso diluído [30], onde o efeito dos estados espúrios aparentemente é devido ao tamanho finito do sistema, devendo desaparecer quando o limite termodinâmico é atingido.

⁸*ris* = random initial state, estado inicial aleatório.

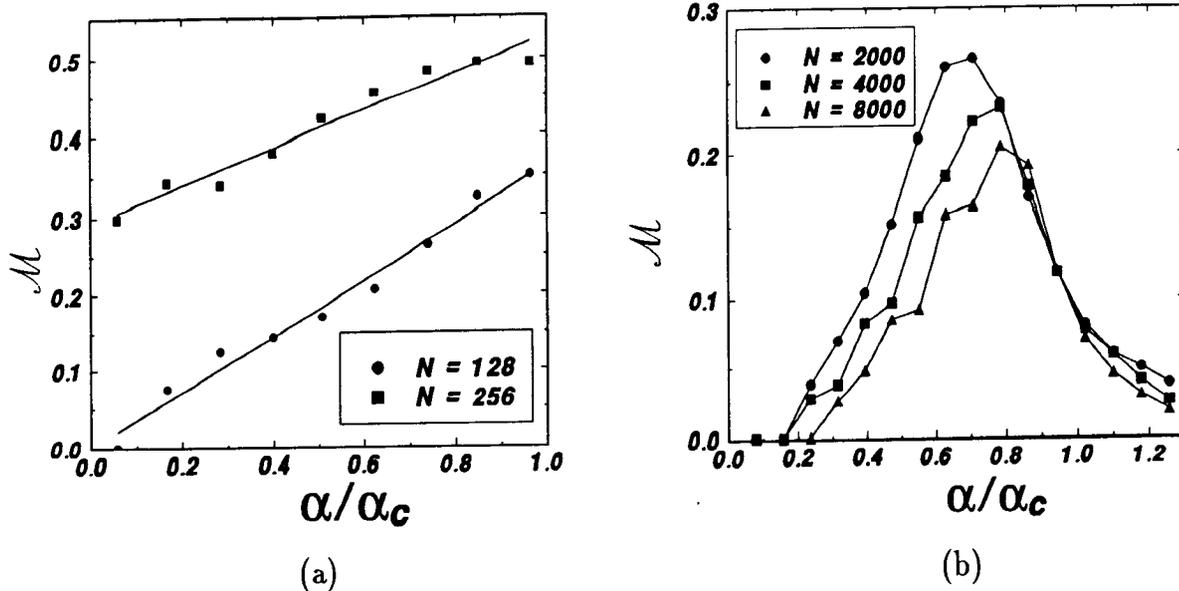


Figura 1.9: Resultados para M para o modelo de Hopfield não diluído [35] (a) e diluído [30] (b).

1.5 Conclusão

Apresentamos, neste capítulo, uma breve revisão dos principais conceitos e modelos de redes de neurônios. Existem muitas outras situações nas quais o modelo de Hopfield foi estudado, como por exemplo, quando as memórias ou os neurônios estão estruturados hierarquicamente [36–38], quando a rede é capaz de esquecer, isto é, memórias antigas vão sendo esquecidas à medida que novas são aprendidas [39–42], quando introduzimos assimetria ($J_{ij} \neq J_{ji}$) e/ou diluição (número de coordenação menor do que N) nas sinapses [43–46], quando queremos armazenar seqüências temporais de padrões (como um poema, por exemplo) [47], quando o sistema é capaz de categorização (classificação) [48–53], quando o limiar é uma variável dinâmica do sistema [54], etc. Além disso, existem outras regras de aprendizado e arquiteturas possíveis, como por exemplo, a regra da pseudo-inversa [55, 56] que é capaz de armazenar até N memórias linearmente independentes e o modelo Trion [57] que considera as colunas corticais como as unidades de processamento. Estes modelos e arquiteturas, entre muitos outros não citados, embora interessantes e relevantes, não guardam relação direta com o que discutiremos a seguir.

O resultado mais importante obtido através das técnicas de mecânica estatística é que existe um valor crítico da capacidade de armazenamento, acima do qual os padrões são desestabilizados, isto é, deixam de ser mínimos da energia e a rede pára de funcionar como memória associativa. Muitas foram as tentativas de, se não eliminar o problema, pelo menos diminuí-lo. Uma das possibilidades, a qual nos ateremos no restante desta tese, é a

introdução de sinapses conectando mais de dois neurônios, ou seja, multi-sinapses.

Capítulo 2

Modelos com Multi-interações

Neste capítulo revisamos os modelos para redes de neurônios que envolvem multi-interações entre os neurônios, bem como alguns novos resultados para tais sistemas. Essencialmente, estas sinapses são introduzidas na tentativa de aumentar a capacidade de armazenamento do sistema e para ser um simulacro de sinapses similares existentes em sistemas nervosos reais.

2.1 Multi-interações Em Redes Biológicas

Apesar de que sinapses conectando pares de neurônios tenham sido as mais extensamente estudadas, existem conexões de ordem mais alta em sistemas reais cujos efeitos são importantes de serem compreendidos. Entre várias possibilidades, podemos citar as seguintes:

- interações axônio-...-axônio-dendrito [58] e
- interações entre sinapses.

Conexões axônio-...-axônio-dendrito podem ser pensadas como sinapses efetivas de ordem mais alta do que dois e são relativamente comuns em sistemas nervosos reais (principalmente as de terceira ordem). O seu papel em sistemas nervosos de invertebrados foi enfatizado por Kandel, Carew e outros [59–61], e é provável que seus efeitos no comportamento de organismos superiores sejam também importantes. É possível também que tais conexões exerçam um efeito modulatório nas sinapses binárias nas quais se ligam uma vez que podem controlar o fluxo de neurotransmissores nas sinapses. Um efeito semelhante pode ocorrer a partir das interações entre sinapses, acoplamento este que também pode ser considerado como uma multi-interação, embora neste caso somente teremos conexões de ordem par. Na figura 2.1 mostramos uma visão esquemática destas multi-sinapses.

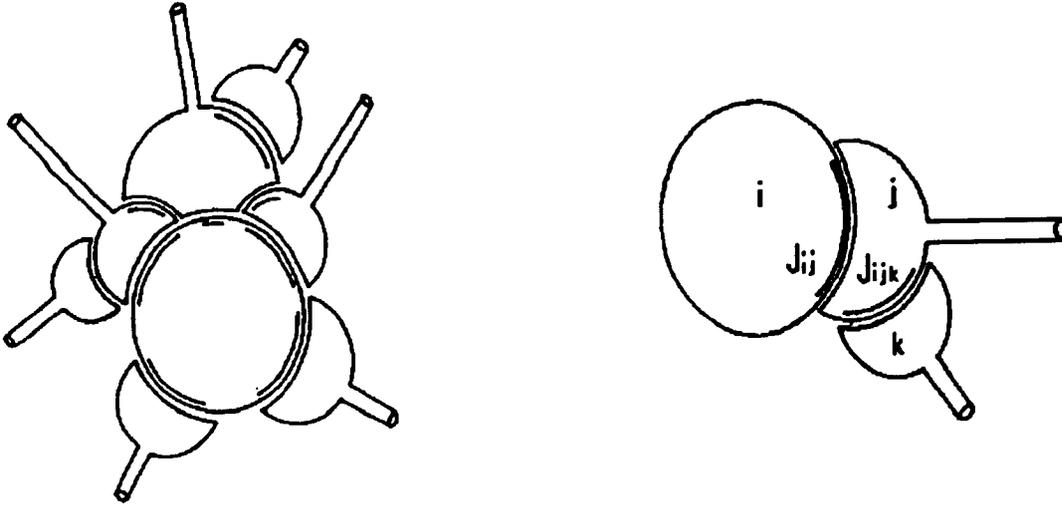


Figura 2.1: Esquema mostrando possíveis contatos multi-sinápticos [58].

De qualquer modo, sinapses de segunda ordem são altamente dominantes sendo interessante considerar o efeito dos termos de ordem mais alta apenas como uma perturbação.

Alternativamente, podemos considerar os termos de ordens mais altas no campo local como representando as possíveis não-linearidades existentes no processo de acumulação do potencial pós-sináptico.

2.2 O Modelo de Hopfield Generalizado

A maneira mais direta de se introduzir multi-interações é generalizar a energia do modelo de Hopfield, eq.(1.9), por um polinômio de grau $M > 2$ [62–67], a saber

$$E = -\frac{N}{2} \sum_{\ell=2}^M \varepsilon_{\ell} \sum_{\mu=1}^P m_{\mu}^{\ell} \quad , \quad (2.1)$$

onde a superposição m_{μ} é dada em (1.5) e ε_{ℓ} são os pesos relativos de cada ordem ℓ de interação. A equação acima pode ser reescrita explicitamente em função das memórias:

$$E = -\frac{1}{2} \sum_{\ell} \frac{\varepsilon_{\ell}}{N^{\ell-1}} \sum_{\mu} \sum_{i_1, \dots, i_{\ell}} \xi_{i_1}^{\mu} \dots \xi_{i_{\ell}}^{\mu} S_{i_1} \dots S_{i_{\ell}} \quad . \quad (2.2)$$

Vamos considerar nulos os termos de auto-interação (nos quais temos dois ou mais índices iguais), o que será conveniente pois simplificará os cálculos das seções posteriores:

$$E = -\frac{1}{2} \sum_{\ell} \sum_{i_1 < \dots < i_{\ell}} J_{i_1 \dots i_{\ell}} S_{i_1} \dots S_{i_{\ell}} \quad , \quad (2.3)$$

onde

$$J_{i_1 \dots i_\ell} = \frac{\ell! \varepsilon_\ell}{N^{\ell-1}} \sum_{\mu} \xi_{i_1}^{\mu} \dots \xi_{i_\ell}^{\mu} \quad , \quad (2.4)$$

e a regra de aprendizado, para qualquer $\ell \geq 2$, é:

$$J_{i_1 \dots i_\ell} \longrightarrow J_{i_1 \dots i_\ell} + \frac{\ell! \varepsilon_\ell}{N^{\ell-1}} \xi_{i_1}^{P+1} \xi_{i_2}^{P+1} \dots \xi_{i_\ell}^{P+1} \quad . \quad (2.5)$$

O campo local atuando sobre o i -ésimo neurônio neste caso é

$$\begin{aligned} h_i &= \sum_{\ell} \frac{\ell! \varepsilon_\ell}{N^{\ell-1}} \sum_{\mu} \sum_{i_2 < \dots < i_\ell} \xi_i^{\mu} \xi_{i_2}^{\mu} \dots \xi_{i_\ell}^{\mu} \xi_{i_2}^1 \dots \xi_{i_\ell}^1 \\ &= \sum_{\ell} \varepsilon_\ell (\ell \xi_i^1 + R_\ell) \end{aligned} \quad (2.6)$$

onde consideramos que $\mathbf{S} = \xi^1$. O termo de ruído R_ℓ vale

$$R_\ell = \frac{\ell!}{N^{\ell-1}} \sum_{\mu > 1} \sum_{i_2 < \dots < i_\ell} \xi_i^{\mu} \xi_{i_2}^{\mu} \dots \xi_{i_\ell}^{\mu} \xi_{i_2}^1 \dots \xi_{i_\ell}^1 \quad . \quad (2.7)$$

A menor ordem de interação, ℓ_{min} , é a contribuição mais significativa para o ruído causado pela presença de um número extensivo de memórias, e o número máximo de padrões que podem ser armazenados é $\mathcal{O}(N^{\ell_{min}-1})$. Em particular, a presença de termos de segunda ordem implica que $P \sim \mathcal{O}(N)$. Isto pode ser facilmente entendido a partir da dispersão do termo de ruído, o qual é um caminho aleatório:

$$\sigma^2(R_\ell) \simeq \frac{P \ell!}{N^{\ell-1}} \quad . \quad (2.8)$$

Portanto,

$$\frac{\sigma^2(R_\ell)}{\sigma^2(R_{\ell_{min}})} \sim \mathcal{O}(N^{\ell_{min}-\ell}) \quad (2.9)$$

e ficamos somente com a contribuição do termo de mais baixa ordem quando $N \rightarrow \infty$. Com isso, o campo local se reduz a

$$h_i = \sum_{\ell} \ell \varepsilon_\ell \xi_i^1 + \varepsilon_{\ell_{min}} R_{\ell_{min}} \quad (2.10)$$

onde $\ell_{min} = \{min(\ell) | \varepsilon_\ell \neq 0\}$. Logo, como a contribuição dos termos de ordens altas vai somente para o sinal, há um aumento na quantidade de informação que pode ser armazenada no sistema. Como o termo de sinal é $\mathcal{O}(1)$, o sistema pode armazenar memórias até que o termo de ruído seja da mesma ordem, ou seja, de acordo com (2.8), $\sigma(R_{\ell_{min}}) \sim \mathcal{O}(1)$ ou $P \sim \mathcal{O}(N^{\ell_{min}-1})$. A capacidade de armazenamento é então definida em função de ℓ_{min} :

$$\alpha = \frac{P}{N^{\ell_{min}-1}} \quad (2.11)$$

a qual se reduz ao caso usual P/N quando $\ell_{min} = 2$. Note que não levamos em conta a simetria existente nas conexões frente a permutações de índices de sítios e medimos a capacidade como a quantidade total de informação a ser armazenada (PN) dividida pelo número total de sinapses ($N^{\ell_{min}}$) de ordem mais baixa. Se considerarmos a simetria existente, a capacidade seria dada por $\alpha' = \ell_{min}! \alpha$ [62].

Vamos tratar três casos de interesse, a saber, (1) o caso acima quando vários termos de ordens diferentes estão presentes (inclusive $k = 2$), (2) quando temos somente um termo de interação ($\varepsilon_\ell = \delta_{\ell k}$, $k > 2$) e (3) quando o grau do polinômio é muito grande, isto é, $M \rightarrow \infty$ no caso (2) acima.

2.2.1 O Caso $\varepsilon_2 \neq 0$

A teoria de campo médio é análoga à feita para o modelo de Hopfield original por Amit *et al* [2] (ver apêndice A). É importante lembrar que o ruído é governado pelo termo de segunda ordem (estamos explicitamente considerando que $\varepsilon_2 \neq 0$) ao passo que os termos de ordens mais altas contribuem somente para o sinal. A energia livre é dada por

$$f = -\frac{1}{2} \sum_{\ell} \sum_{\mu} \varepsilon_{\ell} m_{\mu}^{\ell} + \sum_{\mu} t_{\mu} m_{\mu} + \frac{\alpha}{2\beta} \ln[1 - \beta(1 - q)] - \frac{1}{2} \frac{\alpha q}{1 - \beta(1 - q)} + \frac{1}{2} \alpha \beta r (1 - q) - \frac{1}{\beta} \langle\langle \ln 2 \cosh \beta [\mathbf{t} \cdot \boldsymbol{\xi} + \sqrt{\alpha r} z] \rangle\rangle, \quad (2.12)$$

onde o símbolo $\langle\langle \dots \rangle\rangle$ significa duas médias: uma sobre o número finito de memórias que podem se condensar e outra sobre a variável gaussiana z , relacionada ao infinito número de memórias com superposição microscópica. As equações de ponto de sela são as mesmas do modelo original exceto por t_{μ} e são dadas por

$$m = \langle\langle \boldsymbol{\xi} \tanh \beta (\mathbf{t} \cdot \boldsymbol{\xi} + \sqrt{\alpha r} z) \rangle\rangle \quad (2.13a)$$

$$t_{\mu} = \frac{1}{2} \sum_{\ell} \varepsilon_{\ell} \ell m_{\mu}^{\ell-1} \quad (2.13b)$$

$$q = \langle\langle \tanh^2 \beta (\mathbf{t} \cdot \boldsymbol{\xi} + \sqrt{\alpha r} z) \rangle\rangle \quad (2.13c)$$

$$r = \frac{q}{[1 - \beta(1 - q)]^2} \quad (2.13d)$$

Vamos considerar que a rede apresenta superposição macroscópica m com somente uma das memórias ($m_{\mu} = m \delta_{1\mu}$) e analisar o seu comportamento tanto em $T = 0$ quanto para $T \neq 0$.

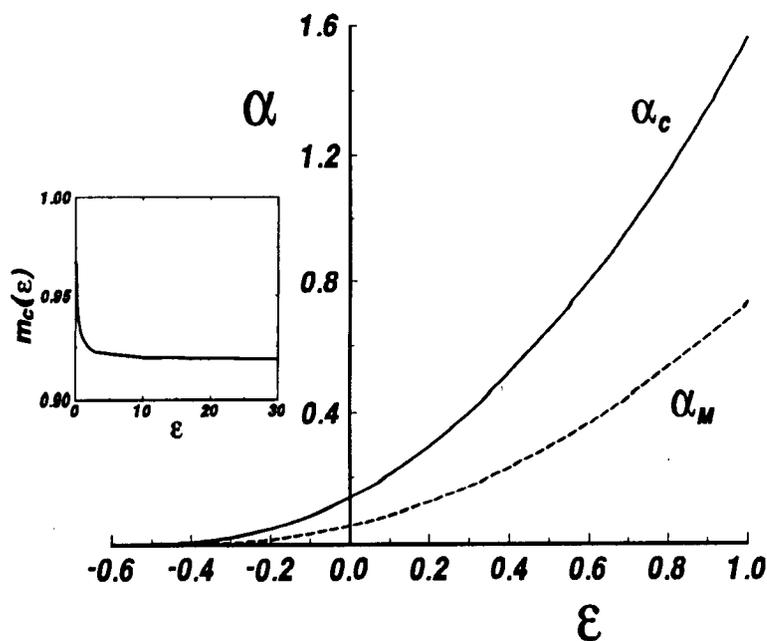


Figura 2.2: Valores críticos de α versus ϵ em $T = 0$, juntamente com a linha abaixo da qual as memórias são mínimos globais da energia livre, α_M (linha tracejada). A superposição na criticalidade, $m_c(\epsilon)$, é mostrada no detalhe e seu valor assintótico, $m_c(\infty)$, é 0.918. Quando $\epsilon \rightarrow \epsilon^{corte} = -0.5$ (por cima), $\alpha_c \rightarrow 0$ e $m_c \rightarrow 1$. Há outro valor de corte abaixo do qual as memórias nunca se tornam mínimos globais da energia livre: $\epsilon_M^{corte} \simeq -0.364$.

As equações acima no limite $T = 0$ ficam

$$m = \operatorname{erf}\left(\frac{t}{\sqrt{2\alpha r}}\right) \quad (2.14a)$$

$$t = \frac{1}{2} \sum_{\ell} \epsilon_{\ell} m^{\ell-1} \quad (2.14b)$$

$$C = \sqrt{\frac{2}{\alpha \pi r}} \exp\left(-\frac{t^2}{2\alpha r}\right) \quad (2.14c)$$

$$r = (1 - C)^{-2}, \quad (2.14d)$$

e a capacidade de armazenamento, α_c , é uma função monotonicamente crescente de M e ϵ_{ℓ} . A fig. 2.2 mostra α_c versus ϵ no caso $\epsilon_{\ell} = \delta_{2\ell} + \epsilon \delta_{4\ell}$, isto é, quando consideramos somente termos de segunda e quarta ordem. A linha α_M onde os estados de recuperação tornam-se mínimos globais de f e o valor de m na criticalidade, $m_c(\epsilon)$, são também mostrados. Quando $\epsilon \rightarrow \infty$, o valor assintótico de $m_c(\epsilon)$ tende para $m_c(\infty) = 0.918$ e o valor crítico de α (assim como α_M) cresce como $\alpha_c \sim \epsilon^2$ o que pode ser entendido como segue. Quanto maior for ϵ , maior é o termo de sinal (que agora passa a ser $\mathcal{O}(\epsilon)$). Mas a dispersão ainda é a mesma, ou seja, $\sigma^2 = 4\alpha$, eq.(2.8), portanto a capacidade de armazenamento é obtida

quando a dispersão do termo de ruído atinge o termo de sinal, isto é, $\sigma \sim \mathcal{O}(\varepsilon)$ ou $\alpha \sim \varepsilon^2$.

Para valores negativos de ε existe um corte inferior (neste valor de ε , t , dado pela eq.(2.14), é nulo): abaixo de ε^{corte} não há recuperação. Por exemplo, para o caso $\varepsilon_\ell = \delta_{2\ell} + \varepsilon\delta_{k\ell}$, $\varepsilon^{corte} = -2/k$. À medida que $\varepsilon \rightarrow \varepsilon^{corte}$ (por cima), $\alpha_c \rightarrow 0$ e $m_c \rightarrow 1$. No caso geral temos

$$\sum_{\ell} \ell \varepsilon_{\ell}^{corte} = 0 \quad , \quad (2.15)$$

definindo um hiper-plano no espaço dos ε_{ℓ} não nulos. Há também um corte inferior em α_M , abaixo do qual as memórias nunca são mínimos globais da energia livre¹. Por exemplo, para o mesmo caso acima ($k = 4$), $\varepsilon_M^{corte} = 2/\pi - 1 \simeq -0.364$.

Quando introduzimos ruído térmico no sistema ($T \neq 0$), observamos que existem algumas características gerais que não dependem dos valores particulares de ε_{ℓ} e M . Por exemplo, a linha abaixo da qual a solução de vidro de spin (SG) existe é

$$T_g = 1 + \sqrt{\alpha} \quad , \quad \forall \varepsilon_{\ell}, M \quad , \quad (2.16)$$

e $\varepsilon_2 \neq 0$. O diagrama de fases para $\varepsilon_{\ell} = \delta_{2\ell} + \delta_{4\ell}$ é apresentado na fig. 2.3. Há três linhas importantes: T_g , dada por (2.16), assinalando o aparecimento dos estados SG; T_M , onde os estados de recuperação primeiro aparecem ($m \neq 0$) e finalmente T_c , onde estes estados tornam-se mínimos globais da energia livre. Em $\alpha = 0$, T_g e T_M não se encontram, implicando que há uma transição entre a fase de recuperação e a paramagnética para valores pequenos de α . Este diagrama de fases é similar ao encontrado por Bollé *et al* [68] para redes de neurônios com spins de Potts de Q estados, devido a uma similaridade formal entre neurônios com multiestados e neurônios binários ligados por multi-interações [69].

A fase SG é reentrante (T_c e T_M não são unívocas) e portanto o máximo valor possível de α não ocorre em $T = 0$ ($\alpha_c \simeq 1.556$) mas em um valor não nulo de T : $\alpha_c^{max} \simeq 1.566$ para $T \simeq 0.126$. O grau de reentrância depende de ε : no limite $\varepsilon_{\ell} = \delta_{2\ell}$ (Hopfield) ele é muito pequeno [70] (ver figura 1.3). Como consequência da reentrância, uma pequena quantidade de ruído térmico aumenta a capacidade de armazenamento do sistema (na aproximação de simetria de réplicas). Isto também pode ser observado no comportamento da superposição m contra T próximo da região da reentrância: m em princípio aumenta para depois diminuir, indicando uma pequena melhora com o ruído. Estes efeitos podem ser um artefato da simetria de réplicas (supostamente mais forte próxima à região da reentrância): acredita-se que α_c^{max} é um limite inferior do valor real em $T = 0$ de α_c obtido

¹A energia livre das memórias próxima do ponto de corte é obtida lembrando que neste ponto, $m \rightarrow 1$ e $\alpha_c \rightarrow 0$. Com isso, $f = -(1 + \varepsilon)/2$. A energia livre dos estados SG vale $f_{SG} = -1/\pi$ e, igualando as duas, obtemos $\varepsilon_M^{corte} = 2/\pi - 1$.

quando a simetria é quebrada (a reentrância então desapareceria). Em outras palavras, $\alpha_c^{RSB} \geq \alpha_c^{max}$, o que é apoiado pelos resultados de simulação numérica neste caso [71].

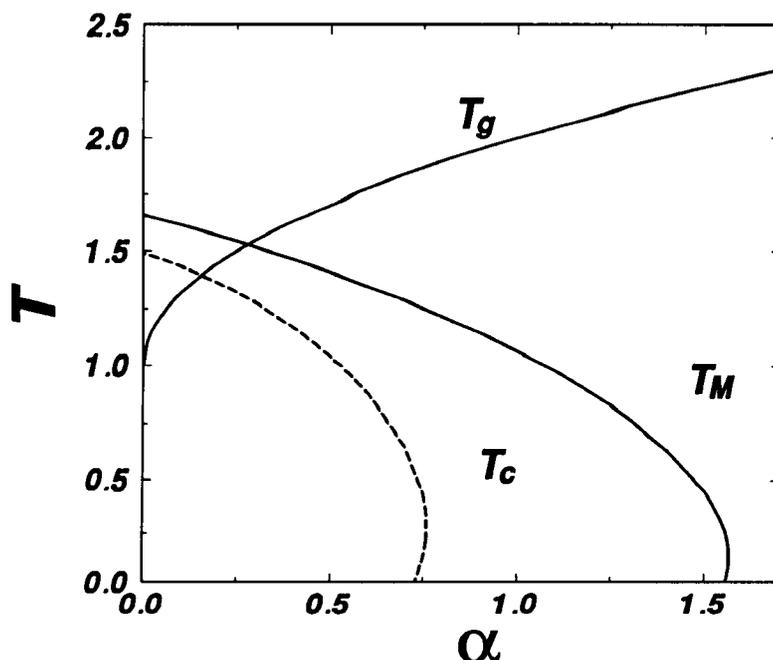


Figura 2.3: Diagrama de fase para $\varepsilon = 1$. Note a fase SG fortemente reentrante tanto para T_c quanto para T_M . A linha T_g é a mesma para todos os valores de ε e M . Note que neste caso há uma transição entre a fase de recuperação e a paramagnética para pequenos valores de α , o que não ocorre no modelo de Hopfield original (as três linhas se encontram em $\alpha = 0$).

Simulações numéricas foram feitas para o caso $\varepsilon_\ell = \delta_{2\ell} + \delta_{4\ell}$ com redes cujos tamanhos variavam de 128 a 1024 neurônios [71]. Os passos da simulação são os seguintes: um estado inicial com superposição m_0 com uma das memórias é criado e sequencialmente testamos cada neurônio, invertendo o seu estado toda vez que isso diminuir a energia. Este procedimento é repetido até atingirmos um ponto fixo da dinâmica, quando medimos várias quantidades de interesse, como por exemplo, a superposição final m e o número de passos necessários para atingir o equilíbrio (tempo de convergência).

Na figura 2.4, onde a superposição inicial é $m_0 = 1$, pode ser vista a superposição m contra α descontinuamente sofrer uma transição em α_c : acima deste valor não temos mais recuperação. Há uma pequena discrepância entre o valor encontrado na simulação e o obtido analiticamente para a transição, o que provavelmente é causado pela instabilidade da solução com simetria de réplicas em $T = 0$. Para $m_0 < 1$ obtemos informações sobre o tamanho das bacias de atração (fig. 2.5), as quais são tão menores quanto maior for α , como no modelo de Hopfield original [31]. Observe que as curvas para diferentes valores de

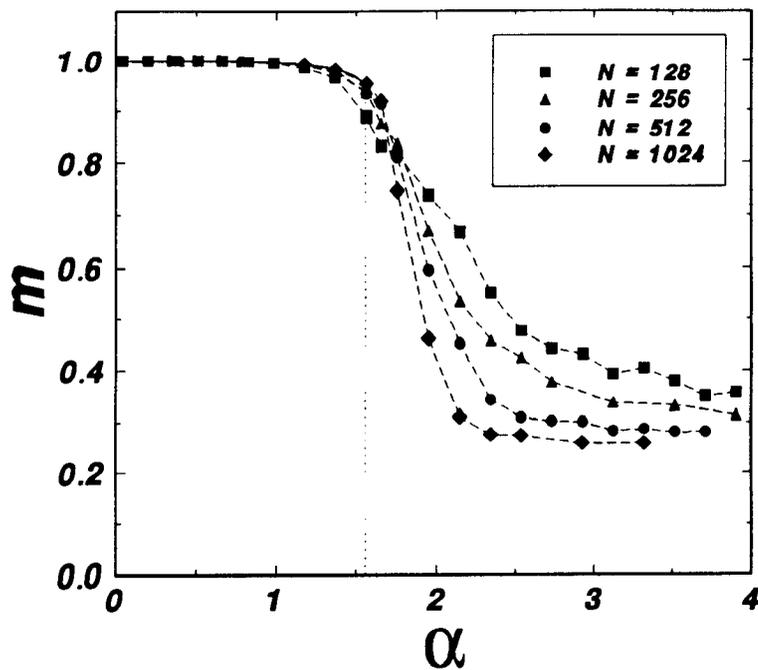


Figura 2.4: Superposição m contra α mostrando a transição de primeira ordem. A linha contínua é obtida a partir das equações (2.14)–(2.14).

N , mas com o mesmo α , se superpõem, permitindo uma análise de *finite-size scaling* [31] para obter $m_c(\alpha)$, a superposição inicial mínima crítica a partir da qual as memórias podem ser recuperadas.

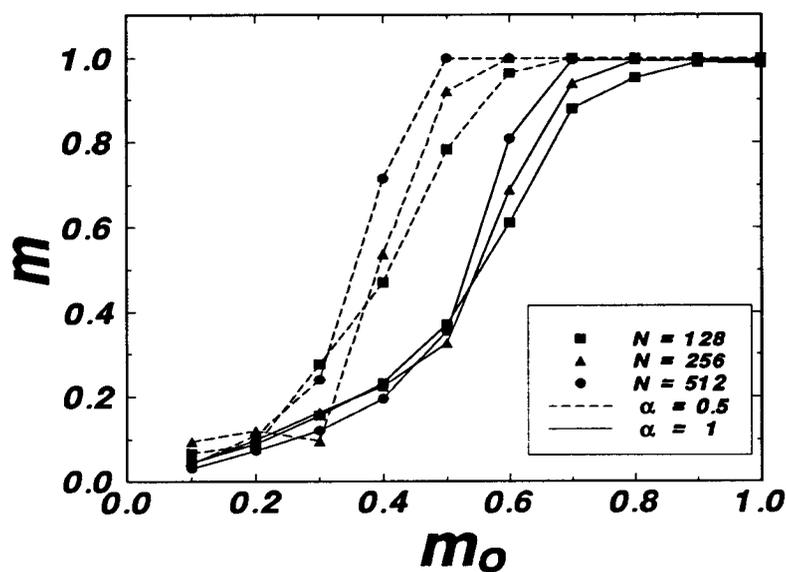


Figura 2.5: Superposição final contra a superposição inicial m_0 .

O tempo médio de convergência $\langle T \rangle$, ou seja, o número médio de passos (atualizações de toda a rede) necessários para atingir um estado estável, e a correspondente dispersão supostamente estão relacionados com as irregularidades do espaço de fase ao redor das memórias. Longos tempos de convergência e grandes dispersões estão relacionados ao número de estados espúrios: quando há muitos estados metaestáveis ao redor das memórias originais, existem muitos caminhos com diferentes comprimentos levando desde o estado inicial até o fundo da bacia mais próxima. E a diversidade de possíveis caminhos implica maiores dispersões. Na fig.2.6a mostramos $\langle T \rangle$ em função de m_0 : se o estado inicial está fora da bacia de atração, $\langle T \rangle$ aumenta com N , ao passo que dentro da bacia, um ou dois passos são suficientes para atingir o estado estável. Também medimos $\langle T \rangle$ em função do parâmetro α , fig. 2.6b, para dois valores de m_0 . Para $m_0 = 0.3$, $\langle T \rangle$ é independente de α a menos que α seja pequeno o suficiente. Para maiores valores de α , esta superposição inicial é muito pequena e o estado não se encontra dentro da bacia de atração da memória. Por outro lado, para $m_0 = 0.8$, $\langle T \rangle$ é pequeno até atingirmos α_c , quando ambas as curvas (para $m_0 = 0.3$ e 0.8) colapsam, indicando que as bacias de atração (e a estabilidade) das memórias estão destruídas.

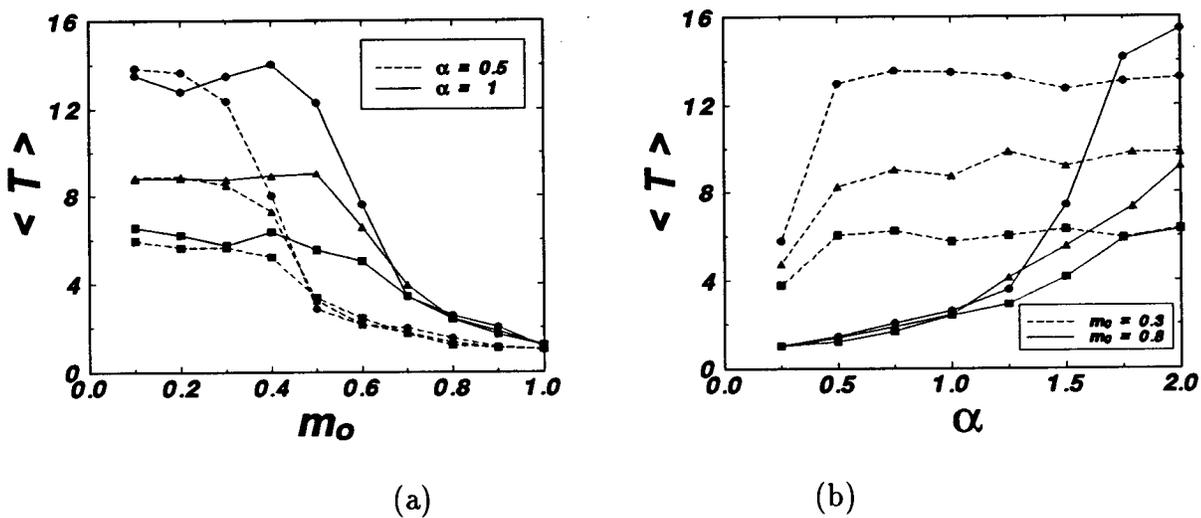


Figura 2.6: Tempo médio de convergência contra (a) m_0 e (b) α para os seguintes tamanhos de rede: 128 (■), 256 (▲) e 512 (●).

Para $\epsilon = 1$ também medimos \mathcal{M} , definido no capítulo anterior, que é proporcional à ocupação fracional do espaço de fase pelas bacias de atração dos estados espúrios (ver

figura 2.7). Para valores pequenos de α , \mathcal{M} está próximo de zero, crescendo com α até atingir um patamar onde aparentemente as variações são pequenas. De qualquer modo, as redes utilizadas, com $N = 128$ e 256 neurônios, são muito pequenas para tirar conclusões definitivas sobre o comportamento de \mathcal{M} , embora fique claro que $\mathcal{M} \neq 0$ para todos valores de $\alpha < \alpha_c$ (exceto $\alpha = 0$).

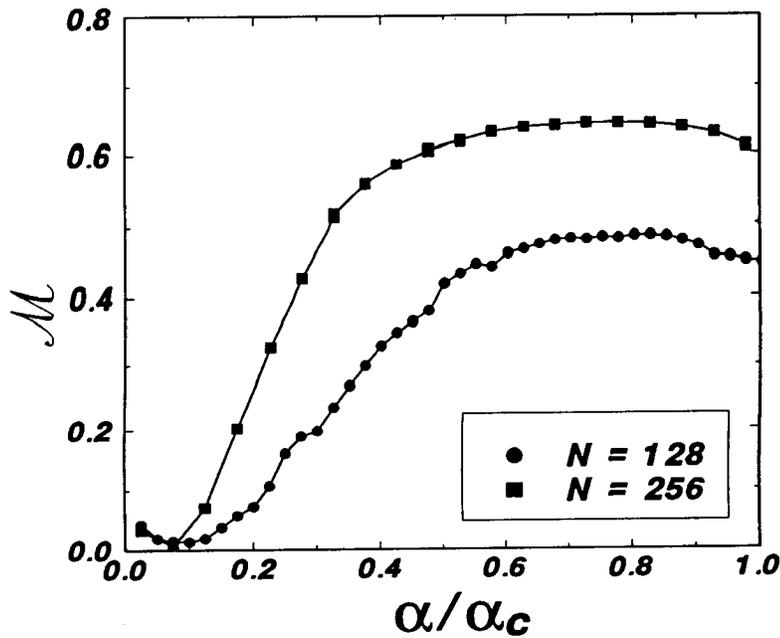


Figura 2.7: Resultados para \mathcal{M} com $\varepsilon_\ell = \delta_{2\ell} + \delta_{4\ell}$.

2.2.2 O Caso $\varepsilon_\ell = \delta_{\ell k}, k > 2$

A função energia (2.1), com $\varepsilon = \delta_{\ell k}$, isto é,

$$\begin{aligned}
 E &= -\frac{N}{2} \sum_{\mu} m_{\mu}^k \\
 &= -\frac{1}{2k!} \sum_{i_1, \dots, i_k} J_{i_1 \dots i_k} S_{i_1} \dots S_{i_k} \quad , \quad (2.17)
 \end{aligned}$$

e $J_{i_1 \dots i_k}$ dado por (2.4), foi estudada por vários autores [62–65]. Na verdade, para simplificar os cálculos (veja o apêndice A), tratou-se de uma versão ligeiramente diferente, onde as auto-interações foram eliminadas:

$$E' = -\frac{1}{2} \sum_{i_1 < \dots < i_k} J_{i_1 \dots i_k} S_{i_1} \dots S_{i_k} \quad . \quad (2.18)$$

A energia livre deste caso, conforme o apêndice A é:

$$f = -\frac{1}{2}m^k + \frac{\alpha\beta k!}{4}(q^k - 1) + tm + \frac{1}{2}\alpha\beta r(1 - q) - \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \beta(t\xi + \sqrt{\alpha r}z) \right\rangle \right\rangle \quad (2.19)$$

e as equações de ponto de sela são dadas por:

$$m = \left\langle \left\langle \tanh \beta(t + \sqrt{\alpha r}z) \right\rangle \right\rangle \quad (2.20a)$$

$$t = \frac{1}{2}km^{k-1} \quad (2.20b)$$

$$r = \frac{1}{2}k!kq^{k-1} \quad (2.20c)$$

$$q = \left\langle \left\langle \tanh^2 \beta(t + \sqrt{\alpha r}z) \right\rangle \right\rangle . \quad (2.20d)$$

Essencialmente a diferença para o caso anterior, quando tínhamos o termo de segunda ordem, está na média sobre as memórias microscópicas (ver apêndice A): aqui somente os cumulantes de segunda ordem contribuem ao passo que no caso anterior todos são significativos e a série pode ser somada. Isso explica porque as equações acima não valem quando $k = 2$.

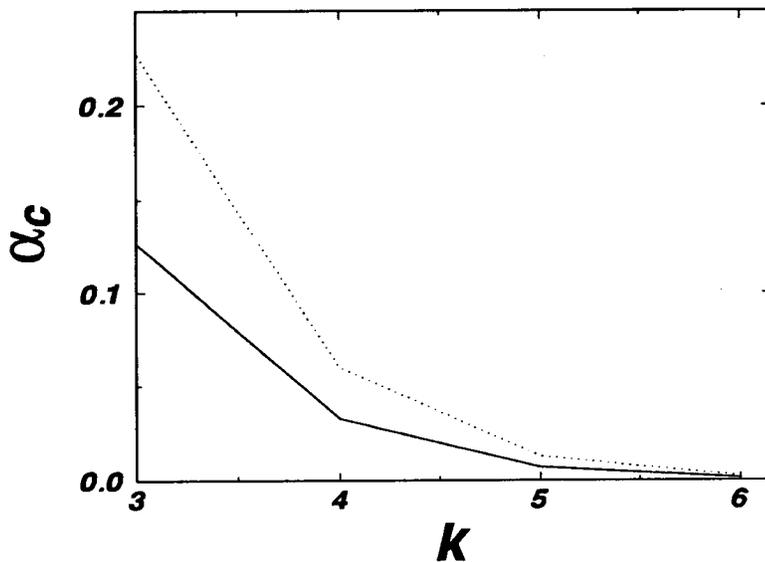


Figura 2.8: Valores críticos de α para vários k . A linha contínua representa o comportamento assintótico de α_c (válido para grandes valores de k) ao passo que a linha pontilhada mostra os valores de α_c calculados a partir da eq.(2.21).

Em $T = 0$ ficamos com somente uma equação a ser resolvida numericamente:

$$m = \operatorname{erf} \left[\frac{m^{k-1}}{\sqrt{2\alpha(k-1)!}} \right] \quad (2.21)$$

Na figura 2.8 mostramos os valores críticos de α para várias ordens de interação k , os quais diminuem à medida que k cresce. Apesar disso, se levássemos em conta na definição de α a simetria existente nas conexões, este α'_c cresceria com k . O valor de m na criticalidade está mais próximo de 1 quanto maior for k . Usando isso podemos expandir a equação acima para grandes valores de k , obtendo que

$$m > 1 - \frac{1}{k\sqrt{\pi \ln k}} \quad (2.22)$$

se

$$\alpha \leq \alpha_c = \frac{1}{2(k-1)! \ln k} \quad (2.23)$$

Esta curva também é mostrada na fig. 2.8 e concorda com os valores obtidos numericamente quando k é grande. Na figura 2.9 mostramos o comportamento de m contra α para $k = 3$ junto com resultados de simulação numérica [72].

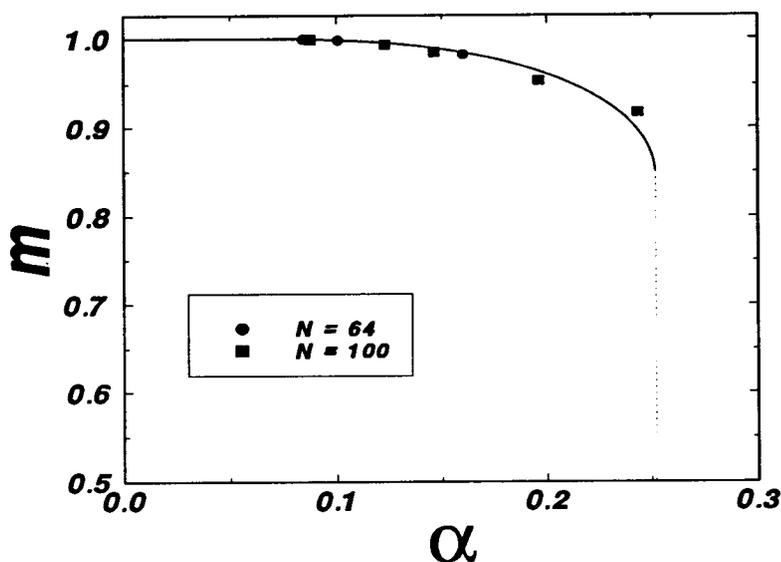


Figura 2.9: Superposição m contra α . A linha contínua é solução da eq.(2.21) e os pontos são resultados de simulação numérica [72].

2.2.3 O Caso $k \rightarrow \infty$

Vamos estudar aqui o caso $\varepsilon_\ell = \delta_{\ell k}$ no limite em que $k \rightarrow \infty$. Neste limite recaímos numa situação análoga ao modelo de energias randômicas (REM) de Derrida [73–76], e o diagrama de fases do sistema pode ser obtido sem o auxílio de réplicas. Primeiramente vamos mostrar que os níveis de energia do sistema são completamente descorrelacionados neste limite. Para isto precisamos calcular a probabilidade de uma configuração \mathcal{C}_1 possuir energia ε_c por sítio e superposição $\{m_\mu\}$ com um conjunto de padrões Γ , a qual é dada por [62] (ver apêndice E)

$$\begin{aligned} P(\varepsilon, \{m_\mu\}) &= \langle\langle \delta(\varepsilon - \varepsilon_c(\{m_\mu\})) \rangle\rangle \\ &= \left(\frac{N}{2\pi\alpha'}\right)^{1/2} \exp\left[-\frac{N}{2\alpha'} \left(\varepsilon + \sum_{\mu \in \Gamma} m_\mu\right)^2\right], \end{aligned} \quad (2.24)$$

onde $\langle\langle \dots \rangle\rangle$ significa uma média sobre as memórias que não pertencem a Γ . Analogamente, a probabilidade conjunta de duas configurações, \mathcal{C}_1 e \mathcal{C}_2 , possuírem superposições $\{m_\mu^1\}$, $\mu \in \Gamma_1$, e $\{m_\mu^2\}$, $\mu \in \Gamma_2$, com energias ε_1 e ε_2 , respectivamente, é

$$\begin{aligned} P_{12}(\varepsilon_1, \varepsilon_2) &= \frac{N}{2\pi\alpha'\sqrt{1-q^{2k}}} \exp\left\{-\frac{N}{2\alpha'(1-q^{2k})} \left[\left(\varepsilon_1 + \sum_{\mu \in \Gamma_1} (m_\mu^1)^k\right)^2 \right. \right. \\ &\quad \left. \left. + \left(\varepsilon_2 + \sum_{\mu \in \Gamma_2} (m_\mu^2)^k\right)^2 - 2q^k \left(\varepsilon_1 + \sum_{\mu \in \Gamma_1} (m_\mu^1)^k\right) \left(\varepsilon_2 + \sum_{\mu \in \Gamma_2} (m_\mu^2)^k\right) \right] \right\} \end{aligned} \quad (2.25)$$

onde

$$q = \frac{1}{N} \sum_i S_i^1 S_i^2 \quad (2.26)$$

é a superposição entre as duas configurações. Quando $k \rightarrow \infty$, a probabilidade acima descorrelaciona uma vez que $q^k \rightarrow 0$ se $q \neq 1$, ou seja, $P_{12} = P_1 P_2$. Portanto, os níveis de energia são totalmente descorrelacionados neste limite e o terreno de energia do sistema é extremamente rugoso, da mesma forma que o modelo de Kauffman quando todos os sítios estão ligados entre si [77].

No limite em que estamos interessados, somente precisamos considerar duas possibilidades para a superposição: ou $m = 1$ com, digamos, a primeira memória, ou $m = 0$. Qualquer outra solução para a superposição (incluindo as misturas) é degenerada com a solução $m = 0$. A função partição é obtida de maneira similar ao modelo de energias randômicas de Derrida [73–76] (veja o apêndice E para mais detalhes) e o diagrama de fases é mostrado na figura 2.10.

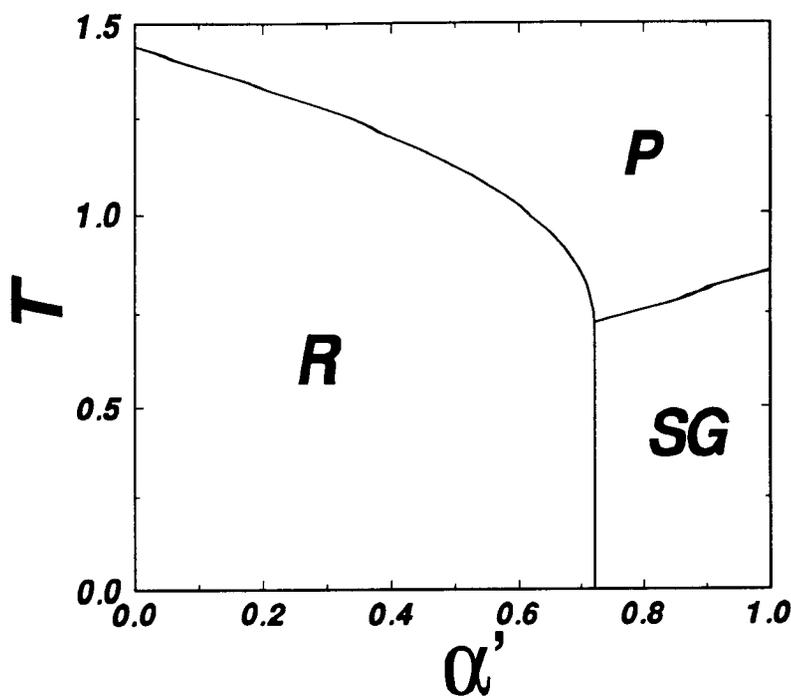


Figura 2.10: Diagrama de fases para o modelo de Hopfield generalizado com $\varepsilon_\ell = \delta_{k\ell}$ e $k \rightarrow \infty$.

A fronteira entre a fase de recuperação e a paramagnética é dada por

$$T_p = \frac{1}{2 \ln 2} \left[\frac{1}{4(\ln 2)^2} - \frac{\alpha'}{2 \ln 2} \right]^{1/2} \quad (2.27)$$

enquanto que a borda entre as fases paramagnética e vidro de spin é

$$T_g = \sqrt{\frac{\alpha'}{2 \ln 2}} \quad , \quad (2.28)$$

independentemente da temperatura. Finalmente, a fronteira entre a fase de recuperação e a de vidro de spin é

$$\alpha' = \frac{1}{2 \ln 2} \quad . \quad (2.29)$$

Note que não há nenhuma restrição para o valor máximo de α' , em concordância com o resultado da seção anterior que mostrava o crescimento de α' com k .

2.2.4 Evolução Dinâmica

Nesta seção discutiremos a dinâmica do modelo de Hopfield generalizado no limite de alta diluição [66,67], isto é, quando uma boa parte das conexões do sistema é severa e assimetricamente cortada (com probabilidade $1 - C/N^{k-1}$), onde a conectividade C satisfaz $C \ll \ln N$. Neste caso, a dinâmica da rede pode ser avaliada exatamente [29].

A evolução da variável que descreve o estado do sítio i no tempo t é governada pelo campo local atuando sobre este neurônio:

$$h_i = \sum_{i_2, \dots, i_k} J_{ii_2 \dots i_k} S_{i_2} \dots S_{i_k} \quad . \quad (2.30)$$

Os acoplamentos são:

$$J_{i_1 \dots i_k} = C_{i_1 \dots i_k} \sum_{\mu} \xi_{i_1}^{\mu} \dots \xi_{i_k}^{\mu} \quad (2.31)$$

e $C_{i_1 \dots i_k}$ é uma variável aleatória distribuída de acordo com

$$P(C_{i_1 \dots i_k}) = \frac{C}{N^{k-1}} \delta(C_{i_1 \dots i_k} - 1) + \left(1 - \frac{C}{N^{k-1}}\right) \delta(C_{i_1 \dots i_k}) \quad . \quad (2.32)$$

Para obtermos a equação que descreve a evolução da superposição com uma das memórias no tempo, $m_{t+1} = f(m_t)$, seguimos o procedimento mostrado no apêndice D. Após o limite $C \rightarrow \infty$, a equação que governa a evolução em paralelo do sistema é:

$$m_{t+1} = \int_{-\infty}^{+\infty} \mathcal{D}y \tanh \beta (m_t^{k-1} - \sqrt{2\alpha}y) \quad , \quad (2.33)$$

onde $\mathcal{D}y$ vale:

$$\mathcal{D}y = \frac{dy}{\sqrt{\pi}} e^{-y^2} \quad . \quad (2.34)$$

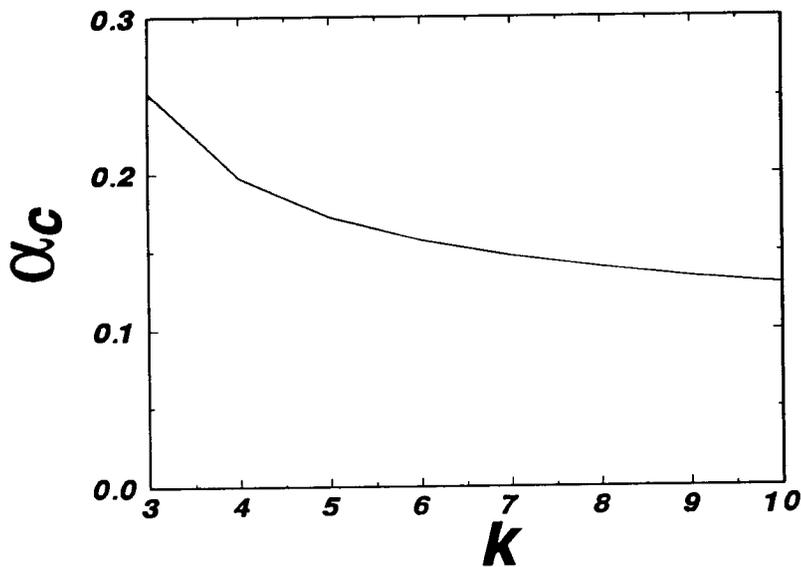


Figura 2.11: Valores críticos de α para vários valores discretos de k .

No limite $T = 0$, o mapa $f(m_t)$, eq.(2.33), fica

$$f(m_t) = \operatorname{erf} \left(\frac{m_t^{k-1}}{\sqrt{2\alpha}} \right) . \quad (2.35)$$

Os valores críticos de α , onde desaparece a solução com $m \neq 0$, são mostrados na figura 2.11 para vários valores de k . Embora α_c diminua com k , o valor de m na criticalidade aumenta, ou seja, em todo o intervalo onde existe recuperação, esta é bastante boa.

Wang e Ross [67] estudaram o modelo acima no caso em que o campo local contém vários termos de interação. Diferentemente dos outros casos apresentados aqui, o ruído externo foi introduzido através de uma variável gaussiana aleatória (com dispersão σ_0) que é adicionada ao campo local. Com isso, obtiveram a seguinte equação para a evolução do sistema:

$$m_{t+1} = \operatorname{erf} \left(\frac{1}{\sqrt{2}\sigma} \sum_{\ell=1}^M \varepsilon_{\ell} m_t^{\ell-1} \right) , \quad (2.36)$$

onde

$$\sigma^2 = \alpha \sum_{\ell=1}^M \varepsilon_{\ell}^2 + \left(\frac{\sigma_0}{C} \right)^2 . \quad (2.37)$$

Note que o ruído total tem duas contribuições: uma externa e outra devido ao ruído causado por todas as outras memórias armazenadas no sistema. Como é somente o ruído total que influencia a dinâmica, sem perda de generalidade podemos considerar $\sigma_0 = 0$ e escrever

$$\sigma^2 = \alpha \sum_{\ell=1}^M \varepsilon_{\ell}^2 . \quad (2.38)$$

O sistema apresenta uma transição contínua para $m = 0$ em

$$\alpha_c = \frac{2}{\pi} \frac{1}{\sum_{\ell} \varepsilon_{\ell}^2} . \quad (2.39)$$

Note que em termos de σ , a transição não depende dos coeficientes ε_{ℓ} : $\sigma_c^2 = 2/\pi$. Abaixo de α_c o comportamento do sistema é bastante complexo, podendo apresentar, além dos pontos fixos, soluções periódicas e caóticas. Na figura 2.12 mostramos o diagrama de bifurcações para o caso $\varepsilon_{\ell} = \delta_{2\ell} - 2\delta_{4\ell}$ ($\varepsilon_4 = -2$). A solução de recuperação ($m \sim 1$) é estável até $\alpha_c \sim 0.031$. Para $\varepsilon_4 > 0$ não temos soluções oscilatórias nem caóticas, somente pontos fixos.

Portanto, o sistema apresenta recuperação para alguns intervalos de α , podendo também apresentar ciclos periódicos e comportamento caótico.

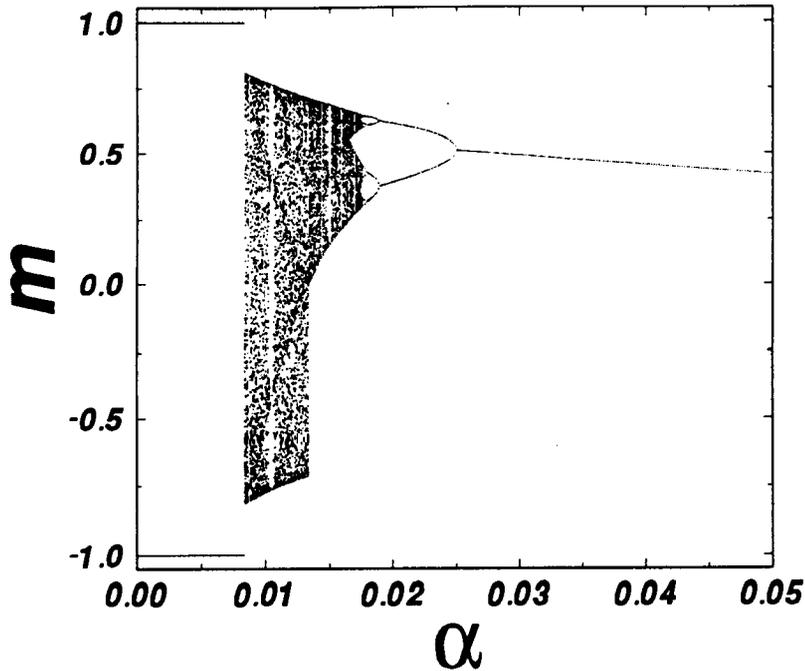


Figura 2.12: Diagrama de bifurcações para $\varepsilon_\ell = \delta_{2\ell} - 2\delta_{4\ell}$. O valor inicial de m_0 é 0.1.

2.3 A Pseudo-Inversa Generalizada

Da mesma maneira que o modelo de Hopfield, a regra da pseudo-inversa [55,56] também foi generalizada para englobar multi-interações [78]. Entretanto, somente o caso de interações ternárias e a performance do modelo quanto ao armazenamento perfeito das memórias foram estudados, sem que resultados relativos à capacidade, diagrama de fases e comportamento assintótico fossem obtidos [79].

Vamos generalizar aqui a versão apresentada na referência 78 com o seguinte campo local atuando sobre S_i :

$$h_i^k = \sum_{i_2 < \dots < i_k} J_{ii_2 \dots i_k} S_{i_2} \dots S_{i_k} \quad , \quad (2.40)$$

e definimos o vetor γ como

$$\gamma^T \equiv (S_1 S_2 \dots S_{k-1}, S_1 \dots S_{k-2} S_k, \dots, S_{N-k+2} S_{N-k+3} \dots S_N) \quad , \quad (2.41)$$

onde T significa que tomamos a transposta da matriz. As componentes de γ são todas as C_N^{k-1} combinações de $S_{i_2} \dots S_{i_k}$ sujeitas ao vínculo $i_2 < \dots < i_k$. Por exemplo, para $k = 3$:

$$\gamma = (S_1 S_2, \dots, S_1 S_N, S_2 S_3, \dots, S_2 S_N, \dots, S_{N-1} S_N) \quad . \quad (2.42)$$

Analogamente vamos definir uma matriz J com N linhas e C_N^{k-1} colunas. Para o caso $k = 3$

teríamos:

$$J = \begin{pmatrix} J_{112} & J_{113} & \dots & J_{11N} & J_{123} & \dots & J_{12N} & \dots & J_{1,N-1,N} \\ J_{212} & J_{213} & \dots & J_{21N} & J_{223} & \dots & J_{22N} & \dots & J_{2,N-1,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ J_{N12} & J_{N13} & \dots & J_{N1N} & J_{N23} & \dots & J_{N2N} & \dots & J_{N,N-1,N} \end{pmatrix} . \quad (2.43)$$

Então, para obtermos o campo local atuando em i , eq.(2.40), basta multiplicar a i -ésima linha de J por γ . A condição, em forma matricial, para que um determinado estado \mathbf{S} seja estável frente à dinâmica é de que o campo local subtraído do limiar, $J\gamma - \theta$, possua o mesmo sinal que o estado \mathbf{S} , ou seja

$$J\gamma - \theta = A\mathbf{S} \quad , \quad (2.44)$$

onde θ_i é o limiar do i -ésimo neurônio e A uma matriz diagonal com $A_{ii} \geq 0, \forall i$. Para

que todas as memórias ensinadas à rede sejam estáveis, esta relação deve ser satisfeita para cada uma delas. Vamos considerar o caso mais simples, isto é, $\theta_i = 0$ e $A = I$. Neste caso, a condição acima reduz-se a

$$J\gamma = \mathbf{S} \quad . \quad (2.45)$$

Agora definimos:

$$\gamma_\mu \equiv \gamma(\mathbf{S} = \xi^\mu) \quad (2.46a)$$

$$\Gamma \equiv (\gamma_1 \gamma_2 \dots \gamma_P) \quad (2.46b)$$

$$\Sigma \equiv (\xi^1 \xi^2 \dots \xi^P) \quad (2.46c)$$

onde Γ é uma matriz com P colunas e C_N^{k-1} linhas, ao passo que Σ possui também P colunas mas somente N linhas. Assim podemos exprimir a condição de que todas as memórias sejam estáveis por

$$J\Gamma = \Sigma \quad . \quad (2.47)$$

A solução geral para a equação acima é

$$J = \Sigma\Gamma^I + B(I - \Gamma\Gamma^I) \quad (2.48)$$

onde B é uma matriz arbitrária e Γ^I é a pseudo-inversa de Γ ,

$$\Gamma^I = (\Gamma^T \Gamma)^{-1} \Gamma^T \quad . \quad (2.49)$$

A equação acima é válida se as colunas de Γ forem linearmente independentes. Considerando $B = 0$ (*tabula rasa*), obtemos a solução formal do problema:

$$J = \Sigma(\Gamma^T \Gamma)^{-1} \Gamma^T \quad . \quad (2.50)$$

Definindo

$$C_{\mu\nu} \equiv \frac{1}{C_N^{k-1}} (\Gamma^T \Gamma)_{\mu\nu} = \frac{1}{C_N^{k-1}} \gamma_\mu \cdot \gamma_\nu \quad , \quad (2.51)$$

podemos obter uma solução explícita, escrevendo os produtos matriciais acima:

$$J_{i_1 i_2 \dots i_k} = \frac{1}{C_N^{k-1}} \sum_{\mu, \nu} \xi_{i_1}^\mu (C^{-1})_{\mu\nu} \xi_{i_2}^\nu \dots \xi_{i_k}^\nu \quad . \quad (2.52)$$

Note-se a não localidade da regra, uma vez que a conexão entre k neurônios depende de uma propriedade global da rede (a matriz C).

Os estados estáveis da dinâmica são aqueles onde os neurônios estão alinhados com os seus campos locais, eq. (2.40). Para o i -ésimo neurônio de um destes estados estáveis (memórias) o campo local é:

$$\begin{aligned} h_i^k &= \sum_{i_2 < \dots < i_k} J_{i i_2 \dots i_k} \xi_{i_2}^\mu \dots \xi_{i_k}^\mu = \frac{1}{C_N^{k-1}} \sum_{i_2 < \dots < i_k} \sum_{\lambda, \nu} \xi_i^\lambda (C^{-1})_{\lambda\nu} \xi_{i_2}^\nu \dots \xi_{i_k}^\nu \xi_{i_2}^\mu \dots \xi_{i_k}^\mu \\ &= \sum_{\lambda, \nu} \xi_i^\lambda C_{\mu\nu} (C^{-1})_{\lambda\nu} = \sum_{\lambda} \xi_i^\lambda \delta_{\mu\lambda} = \xi_i^\mu \quad . \end{aligned} \quad (2.53)$$

Esta equação implica que

$$\text{sgn} \left(\sum_{i_2 < \dots < i_k} J_{i i_2 \dots i_k} \xi_{i_2}^\mu \dots \xi_{i_k}^\mu \right) = \xi_i^\mu \quad , \quad (2.54)$$

isto é, todos os padrões são estáveis.

Como a estabilidade das memórias é garantida até que o seu número, P , atinja o número de vetores γ linearmente independentes, a capacidade de memória é proporcional a N^{k-1} , onde k é a ordem das multi-sinapses. Isto porque a dimensão dos vetores γ é $\mathcal{O}(N^{k-1})$.

Algumas simulações numéricas foram efetuadas com a pseudo-inversa generalizada consistindo de sinapses de segunda ordem J_{ij} e de ligações de ordem 3 entre os neurônios. Os resultados anteriores podem ser facilmente estendidos para incluir mais de uma ordem de interação. As conexões de terceira ordem foram diluídas para estudar como o modelo responde em diferentes situações variando-se o nível de diluição (ou seja, a dimensão do vetor γ). As memórias são escolhidas randomicamente com probabilidade 0.5 para cada um dos valores ± 1 e o número de neurônios é $N = 30$. O número de ligações de terceira ordem é fixo e tais ligações são escolhidas ao acaso. O sistema é inicialmente colocado em um estado com superposição m_0 com um dos padrões e, após o sistema atingir uma configuração estável, medimos a superposição final m com o mesmo padrão. Os resultados são mostrados na figura 2.13. Na figura 2.13b notamos que mais erros existem no estado

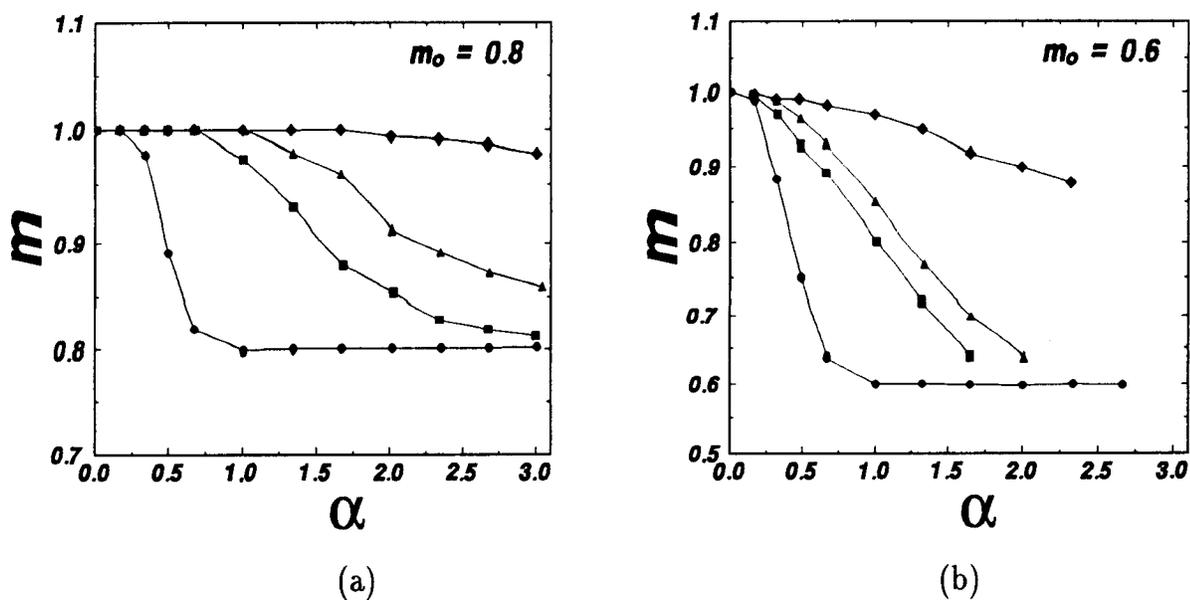


Figura 2.13: Efeito da diluição das multi-sinapses de terceira ordem. Para cada ponto foi efetuada uma média sobre $\min(p, 10)$ padrões e 30 relaxações para cada um dos padrões. As dimensões de γ utilizadas foram: $m = N = 30$ (●) (somente sinapses binárias); $m = 120$ (■); $m = 180$ (▲) e $m = 465$ (◆) (todas as interações de terceira ordem).

final pois partimos de um estado inicial mais distante da memória do que no caso da figura 2.13a. É claro das duas figuras que as bacias de atração são sensivelmente aumentadas com a introdução dos termos de ordem mais alta. Note que acima de um valor crítico de α , a superposição final é igual à inicial, indicando que todos os estados são estáveis [56].

A simulação mostra claramente que o comportamento da rede melhora ao introduzirmos conexões de ordem mais alta, como era esperado, mas ainda seria preciso estudar outros tamanhos de rede para obter resultados quantitativamente confiáveis.

2.4 Conclusões

Neste capítulo revisamos os modelos que contêm sinapses envolvendo mais de dois neurônios simultaneamente. Os diagramas de fases foram apresentados e as diferenças para o modelo de Hopfield original, com e sem ruído térmico, foram mostradas.

Quando multi-interações são introduzidas o sistema passa a armazenar um número de memórias cuja ordem vai depender do grau de interação entre os neurônios. Qualitativamente os resultados em muito se assemelham aos encontrados para o modelo de Hopfield original.

Apesar de intensamente estudados, várias perguntas ainda necessitam de respostas em relação a estes modelos. Por exemplo, para o modelo de Hopfield generalizado, no regime diluído, o diagrama de fases completo não foi obtido, embora não se espere nenhuma surpresa. Já para a pseudo-inversa generalizada quase nada foi feito: ainda é preciso obter o diagrama de fases utilizando-se o método das réplicas [79], estudar o limite quando a ordem de interação é muito grande para verificar se o comportamento assintótico é o mesmo do modelo de Hopfield generalizado, entre outras coisas. Em um caso mais geral, os pesos podem ser considerados diferentes para cada padrão [64], $\varepsilon_\ell = \varepsilon_\ell(\mu)$, generalizando o modelo estudado por Viana [80].

Capítulo 3

O Modelo Completo – Casos OPS e PAS

No capítulo anterior vários modelos considerando termos de multi-interações entre neurônios foram considerados, sendo que somente o estudado por Wang e Ross [67] e o modelo de Hopfield Generalizado apresentavam simultaneamente diferentes ordens de interações. Neste capítulo revisamos um outro modelo também constituído por várias ordens de interações, mas cujas regras de aprendizagem são diferentes das previamente apresentadas pois consideram misturas de memórias.

3.1 Formulação do Modelo Geral

A idéia geral é construir uma função energia da qual as memórias sejam garantidamente (e de preferência os únicos) mínimos globais em $T = 0$. Uma possível maneira de fazer isso é através do seguinte produto de P termos:

$$E = \frac{N}{2} \prod_{\mu=1}^P (1 - m_{\mu}^2) \quad , \quad (3.1)$$

onde as superposições m_{μ} são dadas pela eq.(1.5). O fator $1/2$ foi introduzido por conveniência para obtermos a mesma escala de temperatura do modelo de Hopfield. Aqui, tal como no modelo de Hopfield original, existe simetria frente à reversão dos estados dos neurônios, ou seja, o valor de energia que um certo estado S assume é o mesmo para o estado $-S$. Originalmente [8], o modelo foi definido como o produto do quadrado das distâncias euclidianas entre o estado da rede S e as P memórias armazenadas ξ^{μ} :

$$E = \frac{N}{2} \prod_{\mu=1}^P \left[\frac{1}{2N} \sum_{i=1}^N (\xi_i^{\mu} - S_i)^2 \right] = \frac{N}{2} \prod_{\mu=1}^P (1 - m_{\mu}) \quad (3.2)$$

e, neste caso, apesar de cada um dos P padrões ξ^{μ} ser um mínimo da energia, a rede não armazena os seus antípodas $-\xi^{\mu}$. A eq.(3.1) pode ser explicitamente obtida de (3.2)

se armazenarmos ξ^μ e $-\xi^\mu$. Em ambos os casos, eqs.(3.1) e (3.2), a energia E é sempre positiva, $E[S] \geq 0$, sendo a igualdade satisfeita quando $S = \xi^\mu$, para qualquer $\mu \leq P$. Portanto, os padrões armazenados sempre são mínimos de E , independentemente do seu número, o que sugere que a capacidade de armazenamento da rede (medida por $\alpha = P/N$) deve ser muito alta, como veremos a seguir.

Para mostrar o caráter de multi-interação apresentado pelo modelo nestas situações, expandimos a equação (3.1):

$$E = \frac{N}{2} \left(1 + \sum_{i_1, i_2} J_{i_1 i_2} S_{i_1} S_{i_2} + \sum_{i_1, i_2, i_3, i_4} J_{i_1 i_2 i_3 i_4} S_{i_1} S_{i_2} S_{i_3} S_{i_4} + \dots \right. \\ \left. \dots + \sum_{i_1, \dots, i_{2P}} J_{i_1 \dots i_{2P}} S_{i_1} S_{i_2} \dots S_{i_{2P}} \right), \quad (3.3)$$

onde as conexões sinápticas entre k neurônios são definidas como

$$J_{i_1 \dots i_K} = \frac{(-1)^{K/2}}{N^K} \sum_{\mu_1 < \dots < \mu_{K/2}} \xi_{i_1}^{\mu_1} \xi_{i_2}^{\mu_1} \dots \xi_{i_{K-1}}^{\mu_{K/2}} \xi_{i_K}^{\mu_{K/2}} \quad (3.4)$$

e $2P$ é a maior ordem de interação possível. Existe uma expansão análoga para a equação (3.2), com a diferença fundamental de que os termos de interação de ordem ímpar não são mais nulos (neste caso a maior ordem de interação possível é P) e cujas conexões são dadas por

$$J_{i_1 \dots i_K} = \frac{(-1)^K}{N^K} \sum_{\mu_1 < \dots < \mu_K} \xi_{i_1}^{\mu_1} \xi_{i_2}^{\mu_2} \dots \xi_{i_K}^{\mu_K}. \quad (3.5)$$

Note que embora o primeiro termo não trivial em (3.3) seja a função energia do modelo de Hopfield, os termos de ordens mais altas são diferentes dos modelos apresentados no capítulo anterior porque contêm misturas de memórias. Além disso, devido à presença do termo quadrático, o número de padrões que podem ser armazenados é $\mathcal{O}(N)$.

Estas são, portanto, as duas situações extremas possíveis: quando somente as memórias são aprendidas (caso OPS¹) e quando tanto a memória quanto a anti-memória são armazenadas (caso PAS²). Obviamente poderíamos ter situações intermediárias, com algumas anti-memórias sendo aprendidas e outras não, mas teríamos os mesmos problemas que ocorrem em (3.2), como veremos em breve. Ambos os casos foram estudados tanto por simulações numéricas [9, 12, 13] quanto por cálculos analíticos [9, 10, 68], cujos resultados apresentamos nas seções seguintes.

¹Only Patterns Stored, somente padrões armazenados.

²P pattern and Anti-pattern Stored, padrões e anti-padrões armazenados.

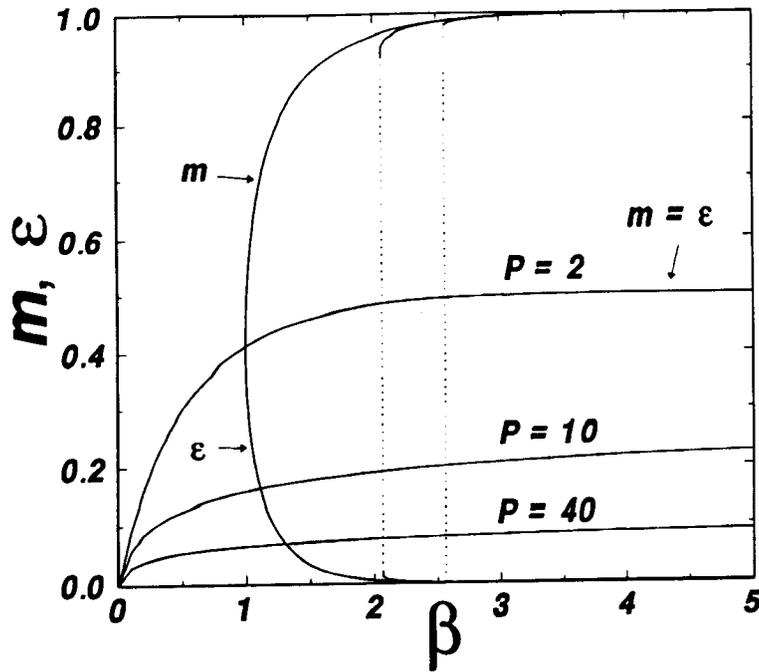


Figura 3.1: As soluções m e ϵ da eq.(3.7) contra β para vários valores de P .

3.2 O Caso OPS

O sistema OPS funciona como uma memória associativa no caso em que o número de padrões armazenados é finito, ou seja, $\alpha = 0$, como mostram os resultados dos cálculos de campo médio [9,10] e simulações numéricas [12,13]. As bacias de atração ao redor de cada mínimo são pequenas, embora estejam dentro de uma grande bacia comum cujo mínimo é o estado simétrico que corresponde à superposição idêntica com todas memórias [9,10]. A origem deste mínimo local de energia pode ser entendida analisando-se o primeiro termo da expansão de (3.2), o qual representa um campo J_i que tende a alinhar o neurônio S_i em sua direção e possui a seguinte forma:

$$h_1 = - \sum_i J_i S_i = - \sum_i \left(\sum_{\mu=1}^P \xi_i^\mu \right) S_i \quad .$$

Portanto, o estado S dado por

$$S_i = \text{sgn} \left(\sum_{\mu=1}^P \xi_i^\mu \right) \quad , \quad (3.6)$$

minimiza este termo do Hamiltoniano. Este estado, além de possuir superposições iguais com todos os padrões, representa o ponto equidistante a todas as memórias.

Supondo que o sistema apresente superposição m com um dos padrões e ϵ com os $P - 1$

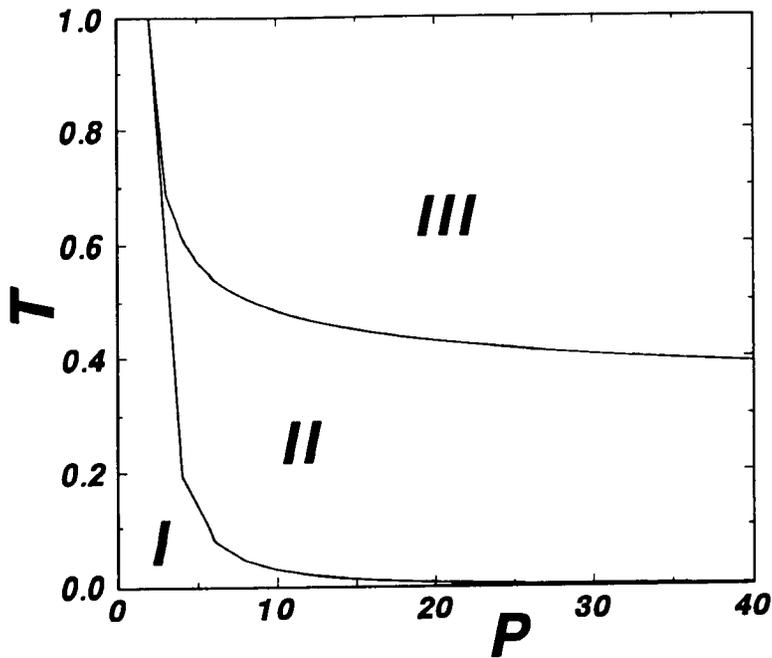


Figura 3.2: Diagrama de fases para a rede OPS. Na região III apenas a solução simétrica ($m = \epsilon$) é estável; na região II tanto esta solução quanto as memórias próprias são estáveis e na região I, as memórias próprias são estáveis e a solução simétrica é instável somente quando P é par.

restantes³, $\mathbf{m} = (m, \epsilon, \dots, \epsilon)$, as equações de campo médio em $\alpha = 0$ [9, 10],

$$m_\mu = \left\langle \left\langle \xi^\mu \tanh \left[\beta \sum_{\nu=1}^P \xi^\nu \prod_{\lambda \neq \nu} (1 - m_\lambda) \right] \right\rangle \right\rangle, \quad \mu \leq P, \quad (3.7)$$

possuem pelo menos duas soluções para cada valor de P : uma solução simétrica com $m = \epsilon$ (que representa o estado central), que existe em qualquer temperatura, e outra para as memórias próprias ($\epsilon \lesssim m$) que existe somente abaixo de uma certa temperatura $T(P)$ (ver figura 3.1). A situação pode ser visualizada no diagrama T versus P (fig. 3.2): na região III, onde a temperatura é alta, somente a solução simétrica é estável, na região II tanto a solução simétrica quanto as memórias próprias são estáveis e na região I, onde a temperatura é muito baixa, os estados simétricos (para P par) se tornam instáveis.

Temos então o seguinte cenário: um estado estável em todas as temperaturas que corresponde à mistura de todas as memórias e cuja bacia de atração é bastante grande, e estados correspondentes às memórias próprias, estáveis somente para $T < T(P)$, com bacias de atração menores.

³Solução esta que não permite estados de Mattis, isto é, superposição não nula com apenas uma das memórias, $\epsilon = 0$, exceto no limite em que $T \rightarrow 0$.

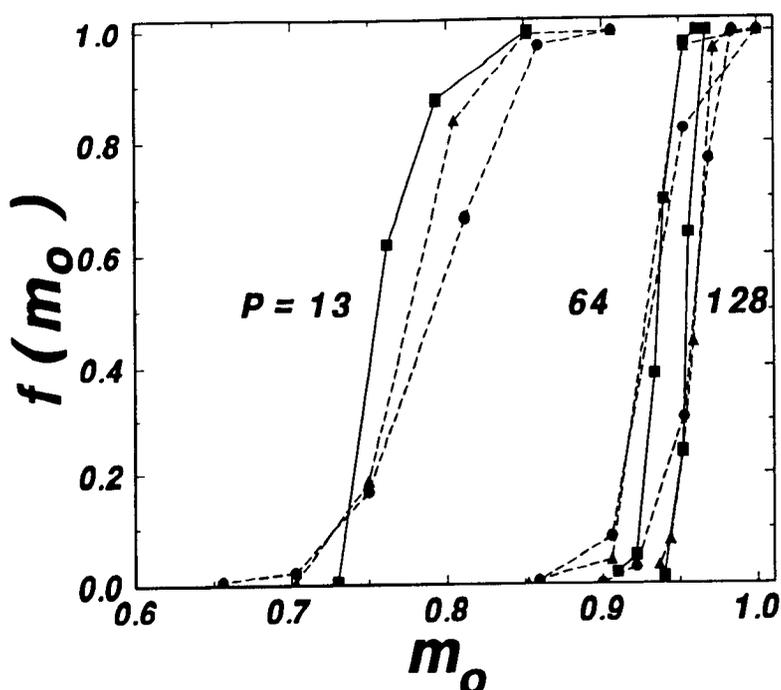


Figura 3.3: A fração $f(m_0)$ contra m_0 para $N = 128(\bullet)$, $256(\blacktriangle)$ e $512(\blacksquare)$ e $P = 13, 64$ e 128 para o caso OPS.

3.2.1 Simulação Numérica

O modelo OPS foi também extensamente estudado através de simulações numéricas a temperatura zero, com as bacias de atração das memórias sendo medidas seguindo o procedimento de Forrest [31]: medimos a fração $f(m_0)$ de vezes em que uma memória é recuperada a partir de um estado cuja superposição inicial com esta memória é m_0 . Se o estado inicial do sistema estiver dentro de sua bacia de atração, o sistema converge para a memória.

Na figura 3.3 apresentamos $f(m_0)$ para redes OPS com tamanhos de 128, 256 e 512 neurônios. Para cada ponto foram feitas 1000 relaxações usando 5 diferentes conjuntos de memórias, exceto para a rede de 512 elementos onde foram feitas 250 relaxações. Aqui, ao contrário do caso PAS e do modelo de Hopfield, as curvas com diferentes valores de N se superpõem para o mesmo valor de P (e não α), indicando que as bacias associadas às memórias tornam-se cada vez menores à medida que P aumenta, independentemente do tamanho da rede. Quando N aumenta, a inclinação das curvas aumenta e $f(m_0)$ se aproxima de uma função degrau, $\Theta(m_0 - m_c)$, a qual vale 1 se o argumento for positivo e 0 se for negativo.

A existência de um mínimo local localizado no centro de massa das memórias pode ser verificada na figura 3.4 onde, além da fração de memórias recuperadas f_M , apresentamos

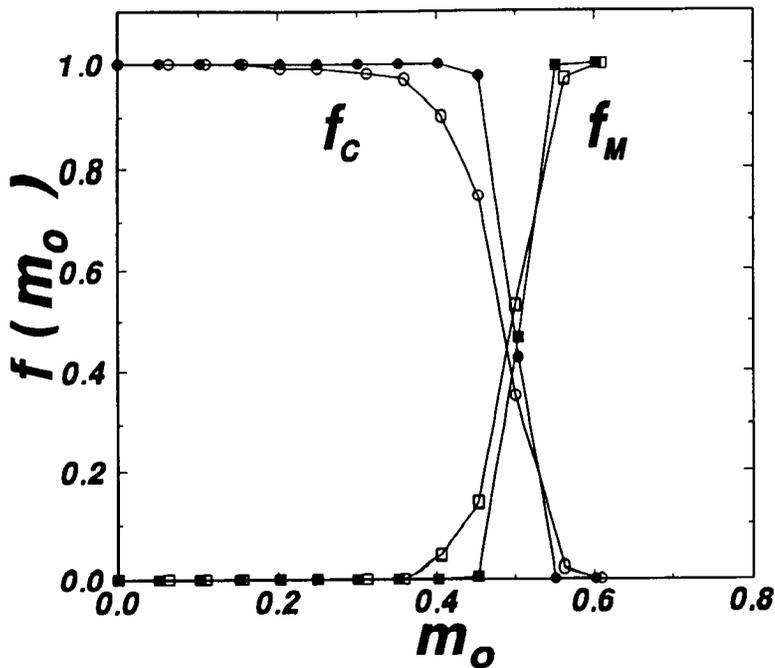


Figura 3.4: A fração de vezes com que o estado central é recuperado, f_C (o e ●), contra m_0 para $N = 128$ e 512 (símbolos cheios) e $P = 5$. Os símbolos □ e ■ representam a fração de vezes em que uma das memórias é recuperada.

a fração f_C de vezes com que este estado central, dado por (3.6), é recuperado.

Resumindo, o modelo OPS é incapaz de distinguir as memórias aprendidas quando o seu número não é muito pequeno ($\alpha \neq 0$) ou quando a informação inicial for bastante incompleta (bacias pequenas), sendo que em ambos os casos o sistema converge para um estado misto.

3.3 O Caso PAS

Alguns dos problemas apresentados pelo modelo OPS podem ser eliminados quando também as antimemórias são ensinadas à rede. Como veremos a seguir, o estado central que antes praticamente dominava a dinâmica no caso OPS agora não mais existe (se a atividade dos padrões for 0.5) e a capacidade da rede aumenta extremamente (o número de memórias que podem ser armazenadas por uma rede PAS cresce com o tamanho do sistema).

Seguindo o procedimento de Bollé *et al* [68], para que possamos utilizar o método das réplicas, supomos primeiramente que a rede apresenta somente superposição macroscópica com um número limitado de memórias, ℓ . Assim, a energia pode ser reescrita da seguinte

forma:

$$E = \frac{N}{2} \exp \left(- \sum_{\rho=\ell+1}^P m_\rho^2 \right) \prod_{\mu=1}^{\ell} (1 - m_\mu^2) . \quad (3.8)$$

Isto pode ser facilmente verificado expandindo-se a exponencial em sua série e comparando-a com a expansão de $\prod_{\rho}(1 - m_\rho^2)$, lembrando que $m_\rho \sim \mathcal{O}(N^{-1/2})$ para $\rho > \ell$.

A energia livre do sistema é (ver apêndice A):

$$\begin{aligned} f = & \frac{1}{2} S \tilde{S} + \frac{\alpha}{2\beta} \ln \left[1 - \frac{\beta \tilde{S}}{\alpha} (1 - q) \right] - \frac{1}{2} \frac{\tilde{S} q}{1 - \frac{\beta \tilde{S}}{\alpha} (1 - q)} + \frac{1}{2} e^{-\alpha S} \prod_{\mu} (1 - m_\mu^2) \\ & + \frac{1}{2} \sum_{\mu} m_\mu \tilde{m}_\mu - \frac{1}{4} \tilde{q} (1 - q) - \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \left(\frac{\beta}{2} \sum_{\mu} \tilde{m}_\mu \xi^\mu + \sqrt{\frac{\beta \tilde{q}}{2}} z \right) \right\rangle \right\rangle \end{aligned} \quad (3.9)$$

onde $\langle \langle \dots \rangle \rangle$ é a média sobre as memórias ξ^μ e sobre a variável gaussiana z . As equações de ponto de sela, considerando que a solução seja na forma $m_\mu = m \delta_{1\mu}$, são:

$$m = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh \left[\frac{\tilde{\beta}}{2} \left(\frac{m}{1 - m^2} + \sqrt{\alpha r} z \right) \right] \quad (3.10a)$$

$$q = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2 \left[\frac{\tilde{\beta}}{2} \left(\frac{m}{1 - m^2} + \sqrt{\alpha r} z \right) \right] \quad (3.10b)$$

$$r = \frac{q}{\left[1 - \frac{\tilde{\beta}}{2} (1 - q) \right]^2} \quad (3.10c)$$

$$S = \frac{1 - \frac{\tilde{\beta}}{2} (1 - q)^2}{\left[1 - \frac{\tilde{\beta}}{2} (1 - q) \right]^2} \quad (3.10d)$$

$$\tilde{\beta} = 2\beta e^{-\alpha S} (1 - m^2) . \quad (3.10e)$$

Vamos agora analisar as soluções destas equações para várias situações.

3.3.1 O Limite $\alpha = 0$

Quando $\alpha \rightarrow 0$ as equações ficam bastante simplificadas pois a equação para m se desacopla das outras variáveis:

$$m = \tanh(\beta m) \quad (3.11)$$

que é exatamente a equação satisfeita pelo modelo de Hopfield neste mesmo limite. A temperatura de transição, em ambos os casos, é $T = 1$: acima desta temperatura o sistema está em uma fase paramagnética e não é capaz de reconhecimento. Outras soluções podem ainda ser investigadas, como por exemplo, soluções simétricas correspondendo a estados

com idêntica superposição com um número finito de memórias, $m_\mu = m_n \sum_{\nu=1}^n \delta_{\mu\nu}$, mas pode-se mostrar [9, 10] que tais estados são instáveis, o que significa que a dinâmica da rede não se detém neles. No modelo de Hopfield, tais estados são estáveis (até uma certa temperatura crítica dependente de n) quando n é ímpar. Em resumo, para $\alpha = 0$, somente temos duas possibilidades: ou o sistema reconhece as memórias aprendidas (para baixas temperaturas) ou simplesmente não reconhece nada (altas temperaturas).

3.3.2 O Limite $T = 0$

Em $T = 0$ as equações (3.10) assumem a seguinte forma:

$$m = \operatorname{erf} \left[\frac{m}{(1 - m^2)\sqrt{2\alpha r}} \right] \quad (3.12)$$

$$r = (1 - C)^{-2} \quad (3.13)$$

$$C = \sqrt{\frac{2}{\pi\alpha r}} \exp \left[-\frac{m^2}{(1 - m^2)2\alpha r} \right] \quad (3.14)$$

onde $C \equiv \tilde{\beta}(1 - q)/2$. É interessante notar que $m = 1$ é sempre solução neste caso, independentemente do valor de α , ou seja, a capacidade é ilimitada. Além desta solução também temos sempre a solução paramagnética, $m = 0$. Podemos obter uma indicação de que a solução de recuperação é estável frente à quebra de simetria de réplicas observando o comportamento da entropia (S_0) neste caso:

$$S_0 \equiv \beta^2 \left. \frac{\partial f}{\partial \beta} \right|_{T=0} = -\frac{\alpha}{2} \ln(1 - C) - \frac{\alpha}{2} \frac{C}{1 - C} \quad (3.15)$$

Se $\alpha \neq 0$ e $C > 0$ a entropia é negativa, o que indica que a aproximação de simetria de réplicas falha neste limite e precisa ser quebrada ('replica symmetry breaking' ou RSB). Por outro lado, quando $m = 1$ temos $C = 0$ e, portanto, $S_0 = 0$. Logo, as soluções de recuperação são estáveis frente à quebra de simetria de réplicas.

Como no caso $\alpha = 0$, também aqui soluções simétricas podem ser consideradas. Diferentemente do caso anterior ($\alpha = 0$) e do modelo de Hopfield, tanto para n par quanto ímpar teremos misturas estáveis com um valor crítico de α não nulo $\alpha_c(n)$ [68]. Por exemplo, para $n = 2$ e 3 temos $\alpha_c(n) = 0.013$ e 0.021 , respectivamente.

Concluindo, o modelo PAS apresenta uma grande capacidade de armazenamento com recuperação perfeita ($m = 1$) para $T = 0$ devido à forma peculiar de sua função energia. À medida que a temperatura cresce esta capacidade é severamente prejudicada pois a temperatura crítica decresce muito rapidamente à medida que α cresce.

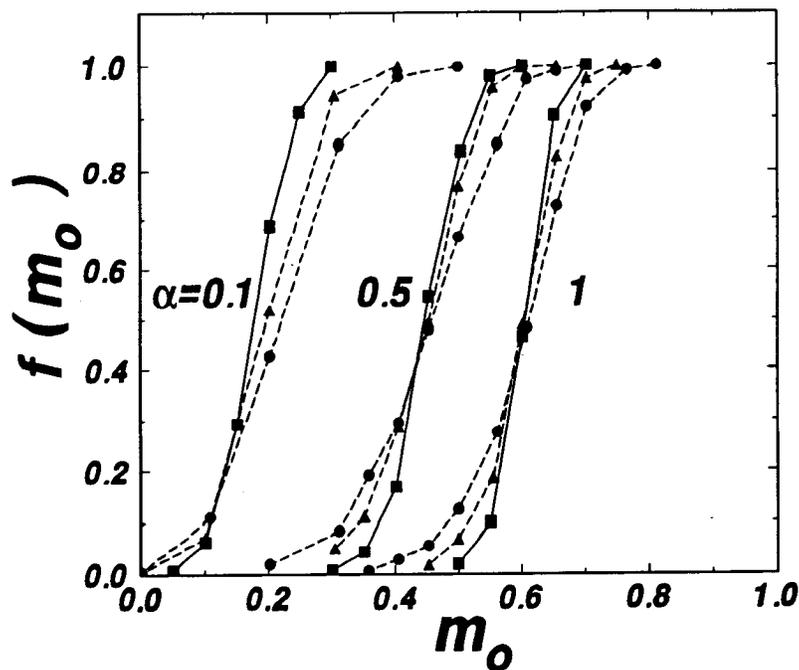


Figura 3.6: Fração $f(m_0)$ de padrões recuperados com uma superposição inicial m_0 para $N = 128(\bullet)$, $256(\blacktriangle)$ e $512(\blacksquare)$ e $\alpha = 0.1, 0.5$ e 1 no caso PAS.

3.3.4 Simulações Numéricas

Com relação aos tamanhos das bacias de atração, a figura 3.6 mostra $f(m_0)$ para redes PAS com 128, 256 e 512 neurônios e $\alpha = 0.1, 0.5$ e 1 . Como no caso anterior, são feitas 1000 relaxações, exceto na rede de 512 onde são feitas 250. Como as memórias são sempre mínimos da energia, a superposição final, quando a memória é recuperada, é sempre 1, ou seja, não há erros no estado recuperado, resultado este previsto na seção anterior. Como no modelo de Hopfield, as curvas correspondendo ao mesmo α se superpõem e, à medida que N aumenta, a inclinação da curva também aumenta. A diferença é que no modelo de Hopfield, se $\alpha > \alpha_c \simeq 0.14$, a inclinação diminui com N [31] e aqui, até $\alpha \simeq 1$ onde testamos, este efeito não existe. O valor de $m_c(\alpha)$, que é obtido através da extrapolação para $N \rightarrow \infty$ das curvas da figura 3.6 (é o ponto de giro das curvas pois neste limite obtemos uma função degrau), nos dá uma estimativa do tamanho das bacias de atração. Por exemplo, $m_c(0.1) \simeq 0.12$, $m_c(0.5) \simeq 0.43$ e $m_c(1) \simeq 0.61$. Em conexão com o que foi visto na seção anterior, não há indícios da existência de um valor limite para α , ou seja, também os resultados de simulação indicam que a capacidade de armazenamento é ilimitada.

Outro aspecto dinâmico analisado foi o tempo de convergência médio $\langle T \rangle$, isto é, o número de passos (cuja unidade é o tamanho da rede N) dispendidos para o sistema

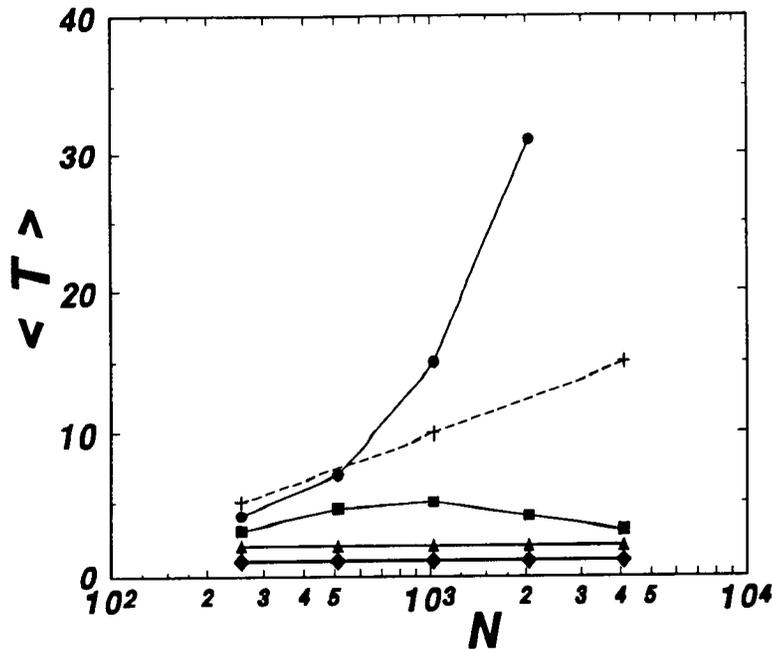


Figura 3.7: Tempo de convergência $\langle T \rangle$ em função de N com $\alpha = 0.1$ para diferentes superposições iniciais: 0.1(●), 0.2(■), 0.3(▲) e 0.8(◆). Os dados representados por + são para o modelo de Hopfield [32] com $m_0 = 0.35$.

atingir um estado estacionário. Também estimamos a variância no tempo de convergência, $\sigma^2 = \langle T^2 \rangle - \langle T \rangle^2$, a qual está relacionada com a isotropia das bacias de atração ao redor de cada memória. Se o espaço é isotrópico e suave, $\langle T \rangle$ e σ^2 devem ser ambos pequenos, ao passo que se tivermos estados meta-estáveis (espúrios) ao redor das memórias, vão existir vários caminhos diferentes que levam desde o estado inicial até a memória, no fundo da bacia. Além do aumento no tempo de convergência, esta diversidade de caminhos provoca uma maior dispersão [13, 32, 33]. Vale notar que isto permanece verdadeiro somente na situação onde os mínimos da função energia estão próximos das memórias armazenadas, isto é, quando a superposição final for aproximadamente 1 (veja os resultados de simulação para o modelo truncado onde isto não vale).

Foram simuladas redes com até 4096 neurônios, nas quais se mediu o tempo de convergência $\langle T \rangle$ e a variância σ^2 para diferentes superposições iniciais m_0 e capacidades α , tanto para a fase de recuperação ($m_0 > m_c$) quanto fora dela ($m_0 < m_c$). O número considerado de conjuntos de memórias e relaxações por conjunto depende do tamanho da rede e do valor de α . Por exemplo, para redes com $N = 4096$ foram utilizados 50 conjuntos e 10 relaxações para cada um, enquanto que para as menores redes foram usados 8 conjuntos e 50 relaxações. Na figura 3.7 podemos ver que na fase de recuperação, poucos

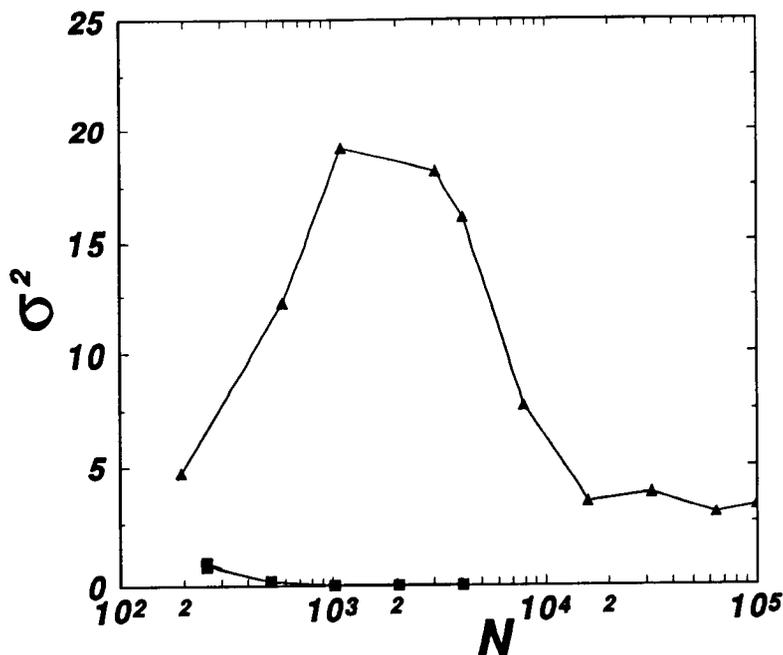


Figura 3.8: Variância da distribuição de $\langle T \rangle$ em função de N (\blacksquare). O símbolo \blacktriangle é para o modelo de Hopfield. A superposição inicial é $m_0 = 0.4$ e $\alpha = 0.1$.

passos são necessários para atingir o estado estável, não dependendo do tamanho da rede. Esta característica é importante para aplicações práticas, uma vez que permite um rápido reconhecimento após alguns poucos passos. A figura 3.8 mostra a variância σ^2 contra o tamanho da rede N para estados na fase de recuperação, ou seja, dentro da bacia de atração. O fato das dispersões se anularem mostra a isotropia e suavidade do espaço de fase ao redor das memórias, uma vez que diferentes caminhos levando ao centro da bacia não implicam em tempos de convergência diferentes.

Resumindo, as bacias de atração são bastante grandes, a capacidade de armazenamento é ilimitada e se a recuperação é bem sucedida, o tempo de convergência é muito pequeno e depende somente das condições iniciais e do parâmetro α , independentemente do tamanho da rede. A dispersão destes tempos é também pequena, indicando que o espaço de fase é homogêneo e suave.

3.4 Padrões com atividade

Como no modelo de Hopfield, aqui também temos problemas quando ξ_i^h pode assumir os valores ± 1 com diferentes probabilidades, a saber a e $1 - a$, respectivamente. Resultados de simulação numérica [9, 12] mostram que no caso PAS, curvas com mesmo P (e não mais com o mesmo α) se superpõem, como no caso OPS, e também surge um estado

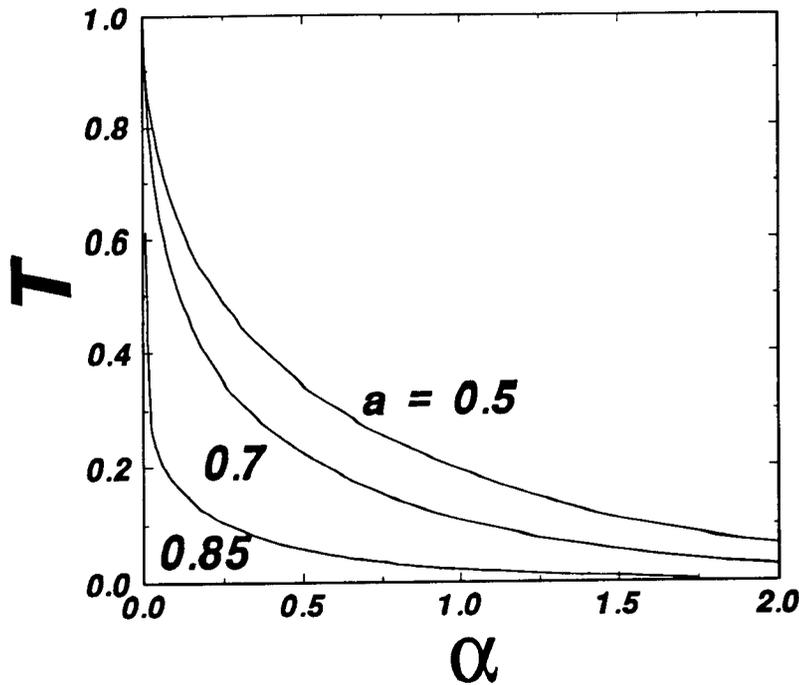


Figura 3.9: Curva de transição para a fase de recuperação para vários valores da atividade.

central ferromagnético (pois a maioria dos neurônios vai estar no mesmo estado). As bacias de atração tornam-se menores, como esperado, e $\langle T \rangle$ é menor (para a fase de não-recuperação) do que no caso decorrelacionado. Isto é devido ao fato das memórias estarem mais próximas umas das outras, o que faz com que a dinâmica leve rapidamente a uma delas ao invés de cruzar o espaço de fases.

Podemos tentar contornar alguns destes problemas redefinindo o modelo [68]. Como agora $\langle \xi_i^\mu \rangle = 2a - 1 = b$, descontamos de ξ_i^μ o seu valor médio. As superposições ficam:

$$m_\mu = \frac{1}{N} \sum_i \frac{(\xi_i^\mu - b)S_i}{1 - b^2} . \quad (3.17)$$

O fator $1 - b^2$ garante que a energia, expressa agora como

$$E = N(1 - b^2) \prod_\mu (1 - m_\mu^2) , \quad (3.18)$$

seja ainda zero nos seus mínimos.

O modelo modificado pode ser resolvido analogamente ao que foi feito anteriormente [68]. Novamente temos soluções com $m = 1$ em $T = 0$ para qualquer valor de b e α , isto é, capacidade de armazenamento infinita. A temperatura T_g agora depende de b e é dada por

$$T_g = (1 + \sqrt{\alpha}) \exp \left[-\frac{1 - b^2}{\sqrt{\alpha}(1 + \sqrt{\alpha})} \right] , \quad (3.19)$$

e no limite $\alpha \rightarrow 0$, $T_g \sim \sqrt{\alpha}$ e não linearmente como no caso $b = 0$. Na figura 3.9 comparamos as linhas de recuperação para vários valores de b .

Alternativamente, podemos tentar eliminar a bacia de atração do estado central modificando a energia de outra maneira, ou seja, aumentando a energia desse estado. Isto é feito em analogia ao processo de eliminação de estados espúrios no modelo de Hopfield (sonhos) [81, 82], com a diferença que aqui podemos escrever uma equação explícita para a energia modificada. No modelo de Hopfield a idéia era, a partir de estados iniciais aleatórios, obter os estados de mínimo (denotados por σ^μ) e descontá-los das conexões sinápticas:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu - \frac{\lambda}{N} \sum_{\rho=1}^D \sigma_i^\rho \sigma_j^\rho \quad , \quad (3.20)$$

onde λ pesa a importância do desconto ($\lambda < 0.02$) e D é o número total de sonhos. Como não se faz distinção se o estado final é uma memória ou não, existe um número ótimo de sonhos (D^{opt}) que satisfaz

$$D^{opt}(\alpha, N, \lambda) = \lambda^{-1}(c_0 N + c_1)\alpha \quad (3.21)$$

onde $c_0 \simeq 0.56$ e $c_1 \simeq 20$. Por exemplo, para $N = 400$, $\alpha = 0.5$ e $\lambda = 0.01$, teríamos $D^{opt} = 12.000$. Note que não há dependência na atividade dos padrões.

No nosso caso como somente queremos eliminar o estado central, reescrevemos a energia como [83]:

$$E = \frac{N}{2}(1 + C^2)^D \prod_{\mu=1}^P (1 - m_\mu^2) \quad , \quad (3.22)$$

onde

$$C = \frac{1}{N} \sum_i C_i S_i \quad (3.23)$$

e C_i é dado por (3.6). O número ótimo de sonhos agora satisfaz

$$D^{opt} = a^{-1.43}(c_1 \alpha + c_2 N + c_3 \alpha N + c_4) \quad (3.24)$$

onde $c_1 \simeq 4.86$, $c_2 \simeq 0.00481$, $c_3 \simeq 0.0175$ e $c_4 \simeq -0.77$. Uma das diferenças para o modelo de Hopfield é que aqui aparece a dependência na atividade das memórias, o que é esperado, pois quando $a = 0.5$ o estado central não existe.

3.5 Conclusão

Neste capítulo revisamos o modelo completo e apresentamos os principais resultados obtidos analiticamente e por simulações numéricas. Foram consideradas duas situações de

interesse: quando as memórias e as anti-memórias são aprendidas (caso PAS) e quando somente são consideradas as memórias (caso OPS).

Os sistemas OPS somente funcionam quando o número de padrões armazenados P é finito, resultado que é encontrado analiticamente e comprovado nas simulações. Estes resultados estão de acordo com o capítulo anterior, onde vimos que o número de memórias que podem ser armazenadas depende da menor ordem de interação, que aqui é $l_{min} = 1$, ou seja, $\alpha = P/N^0$. Este comportamento indica a importância do armazenamento das anti-memórias e sua origem está na forma característica do espaço de fase no caso OPS: uma grande bacia com pequenos furos, cada um deles correspondendo a uma memória. Quando as anti-memórias são também guardadas, todas as componentes J_i do campo que tende a alinhar o estado S_i com o estado central são nulas.

Por outro lado, na situação PAS o sistema funciona muito bem como memória associativa, podendo armazenar uma quantidade muito grande de padrões, com tempos de recuperação muito pequenos, embora a região de recuperação fique cada vez menor no diagrama de fase. Além disso, o espaço ao redor de cada memória, isto é, sua bacia de atração, é suave e isotrópico.

Um dos problemas do modelo é que a quantidade de sinapses envolvida é muito grande e aumenta muito para cada memória acrescentada (lembre que o termo de mais alta ordem é $2P$, ou seja, cada padrão introduz uma nova ordem de interação na energia). Com isso, apesar de não existir um valor crítico para α , a quantidade de informação por sinapse é pequena. No capítulo seguinte vamos mostrar como podemos contornar este problema truncando a expansão em uma ordem conveniente.

Capítulo 4

O Modelo Truncado – a Estática

Neste capítulo, onde inicia o núcleo da tese, introduzimos o modelo truncado e apresentamos os resultados obtidos por métodos analíticos [35] (usando o truque das réplicas na aproximação em que estas são simétricas) e de simulação para os estados de equilíbrio do sistema [71]. O modelo truncado se comporta diferentemente do caso completo e de outros modelos com multi-interações em vários aspectos notáveis, como veremos a seguir.

4.1 O Modelo

As multi-interações que aparecem na eq.(3.1) também podem ser escritas em função das superposições:

$$E = \frac{N}{2} \left[1 - \sum_{\mu_1} m_{\mu_1}^2 + \sum_{\mu_1 < \mu_2} m_{\mu_1}^2 m_{\mu_2}^2 + \dots + (-1)^P \sum_{\mu_1 < \dots < \mu_P} m_{\mu_1}^2 \dots m_{\mu_P}^2 \right] , \quad (4.1)$$

e, negligenciando o termo de ordem zero, definimos uma nova função energia com os próximos M termos ($M < P$). Esta função energia, introduzindo os pesos $\varepsilon_{2\ell}$, fica

$$E = \frac{N}{2} \sum_{\ell=1}^M (-1)^\ell \varepsilon_{2\ell} \sum_{\mu_1 < \dots < \mu_\ell} m_{\mu_1}^2 m_{\mu_2}^2 \dots m_{\mu_\ell}^2 . \quad (4.2)$$

A eq.(4.2) define um modelo que pode ser tomado como o modelo de Hopfield ($\ell = 1$) mais termos de correção de ordem par mais alta e, ao contrário de outros modelos, aqui não podemos ter $\varepsilon_2 = 0$ porque os termos remanescentes envolvem misturas de várias memórias e nenhum padrão pode ser recuperado somente com eles. No que segue vamos considerar somente a primeira correção ao modelo de Hopfield ($M = 2$). A eq.(4.2) pode então ser reescrita como ($\varepsilon_{2\ell} = \delta_{1\ell} + \varepsilon \delta_{2\ell}$)

$$E = -\frac{1}{2} \sum_{i,j} J_{ij} S_i S_j + \frac{\varepsilon}{2} \sum_{i,j,k,l} J_{ijkl} S_i S_j S_k S_l . \quad (4.3)$$

A regra de aprendizado para as conexões de segunda ordem J_{ij} segue a prescrição de Hebb [2, 18], eq.(2.5). As sinapses de quarta ordem, J_{ijkl} , por outro lado, podem ser implementadas através da seguinte regra de aprendizado

$$J_{ijkl} = \frac{1}{N^3} \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu \quad (4.4)$$

e, quando um novo padrão é aprendido, as sinapses são modificadas de acordo com

$$J_{ijkl} \longrightarrow J_{ijkl} + \frac{\epsilon}{N^2} J_{ij} \xi_k^{P+1} \xi_l^{P+1} \quad , \quad (4.5)$$

onde ξ^{P+1} é o $(P+1)$ -ésimo padrão a ser ensinado a rede. A eq.(4.5) pode ser interpretada como as multi-sinapses descritas na introdução: o último termo representando a ação de dois axônios sobre uma sinapse de segunda ordem, embora a analogia não deva ser levada muito adiante. Note que para $P = 1$ este modelo é equivalente ao modelo de Hopfield uma vez que as sinapses de quarta ordem não existem. Além disso, embora as sinapses binárias J_{ij} sejam simétricas ($J_{ij} = J_{ji}$), as de quarta ordem não apresentam simetria completa sobre todas as possíveis trocas de índices (somente $i \leftrightarrow j$ e $k \leftrightarrow l$, devido à presença de μ e ν em (4.4)). Por exemplo, $J_{ijkl} \neq J_{ikjl}$ como podemos ver de (4.4). Estes acoplamentos podem ser simetrizados (sem que isso altere a função energia) se forem reescritos como [84]

$$J'_{ijkl} = \frac{1}{3} (J_{ijkl} + J_{ljkil} + J_{kjiil}) \quad (4.6)$$

e agora as novas conexões J'_{ijkl} possuem simetria frente a trocas de quaisquer índices. Com isso, a energia (4.3) é uma função de Lyapunov para a dinâmica

$$S_i(t+1) = \text{sgn}[h_i(t)] \quad , \quad (4.7)$$

onde

$$h_i(t) = \sum_j J_{ij} S_j(t) - \epsilon \sum_{j,k,l} J'_{ijkl} S_j(t) S_k(t) S_l(t) \quad . \quad (4.8)$$

Ou seja, a dinâmica acima alinha o estado do neurônio S_i com o campo local h_i atuando sobre ele, o que implica diminuir a energia do sistema, levando mais cedo ou mais tarde a um mínimo de energia.

Devemos notar a importância dos auto-acoplamentos aqui: por exemplo, as sinapses J_{iikl} e J_{ijkk} dão origem a contribuições para a energia da mesma ordem que as sinapses J_{ij} , o que não acontece no modelo de Hopfield generalizado. Isto pode ser facilmente verificado reescrevendo os acoplamentos como

$$J_{ijkl} = \frac{1}{2N} J_{ij} J_{kl} - \frac{1}{2} J_{ijkl}^H \quad , \quad (4.9)$$

onde J_{ijkl}^H é dado por (2.4). Então, para os auto-acoplamentos mencionados acima, temos

$$J_{iikl} = \frac{\alpha}{2N} J_{kl} - \frac{1}{2N^2} J_{kl} \simeq \frac{\alpha}{2N} J_{kl} \quad , \quad (4.10)$$

onde o fator N^{-1} é compensado por uma soma sobre os sítios (i) em (4.3). Note que a contribuição de J_{iikl}^H é desprezível.

No restante do capítulo mostraremos que este modelo tem um comportamento completamente diferente dos anteriores.

4.2 Teoria de Campo Médio

Seguindo o mesmo procedimento do capítulo anterior podemos obter as equações para os estados de equilíbrio do sistema no limite termodinâmico [2]. No modelo de Hopfield generalizado, quando a energia é composta de mais de uma ordem de interação, o ruído proveniente das muitas memórias armazenadas é governado pelo termo de mais baixa ordem enquanto que os termos de ordem mais alta somente contribuem para o sinal. Por outro lado, para o modelo truncado há uma contribuição dos termos de ordem mais alta (proveniente dos termos de auto-acoplamento) para o ruído total. Até a quarta ordem, a função energia pode ser escrita como

$$E = -\frac{N}{2} \sum_{\mu} m_{\mu}^2 - \frac{N\varepsilon}{4} \sum_{\mu} m_{\mu}^4 + \frac{N\varepsilon}{4} \left(\sum_{\mu} m_{\mu}^2 \right)^2 \quad , \quad (4.11)$$

e a energia livre por neurônio pode ser obtida através do método das réplicas. Note que é o último termo na equação acima que introduz a diferença entre o modelo truncado e as generalizações do modelo de Hopfield. Assumindo que as réplicas são simétricas e tomando o limite $n \rightarrow 0$ obtemos (ver apêndice A)

$$\begin{aligned} f = & -\frac{1}{2}(1 - \varepsilon y) \sum_{\mu} m_{\mu}^2 - \frac{\varepsilon}{4} \sum_{\mu} m_{\mu}^4 - \frac{\varepsilon}{4} y^2 + \sum_{\mu} t_{\mu} m_{\mu} \\ & + \frac{\alpha}{2\beta} \ln[1 - \beta(1 - \varepsilon y)(1 - q)] - \frac{1}{2} \frac{\alpha q(1 - \varepsilon y)}{1 - \beta(1 - \varepsilon y)(1 - q)} \\ & + \frac{1}{2} \alpha \beta r(1 - q) - \frac{1}{\beta} \langle \langle \ln 2 \cosh \beta (t \cdot \xi + \sqrt{\alpha r} z) \rangle \rangle \quad , \end{aligned} \quad (4.12)$$

onde y é introduzido para linearizar o último termo na eq.(4.11). O símbolo $\langle \langle \rangle \rangle$ compreende duas médias: uma sobre o número finito de padrões que podem se condensar e outra sobre a variável gaussiana z , associada ao infinito número de memórias que não se

condensam. As equações de ponto de sela são

$$m = \langle \langle \xi \tanh \beta (\xi \cdot t + \sqrt{\alpha r z}) \rangle \rangle \quad (4.13a)$$

$$t_\mu = (1 - \varepsilon y) m_\mu + \varepsilon m_\mu^3 \quad (4.13b)$$

$$q = \langle \langle \tanh^2 \beta (\xi \cdot t + \sqrt{\alpha r z}) \rangle \rangle \quad (4.13c)$$

$$r = q \left[\frac{1 - \varepsilon y}{1 - \beta(1 - \varepsilon y)(1 - q)} \right]^2 \quad (4.13d)$$

$$y = \sum_\mu m_\mu^2 + \alpha \frac{1 - \beta(1 - \varepsilon y)(1 - q)^2}{[1 - \beta(1 - \varepsilon y)(1 - q)]^2} \quad (4.13e)$$

Este conjunto de equações não lineares acopladas será resolvido numericamente nas próximas subseções no caso em que a rede apresenta uma superposição macroscópica m com apenas uma das memórias ($m_\mu = m \delta_{1\mu}$).

4.2.1 O Limite $T = 0$

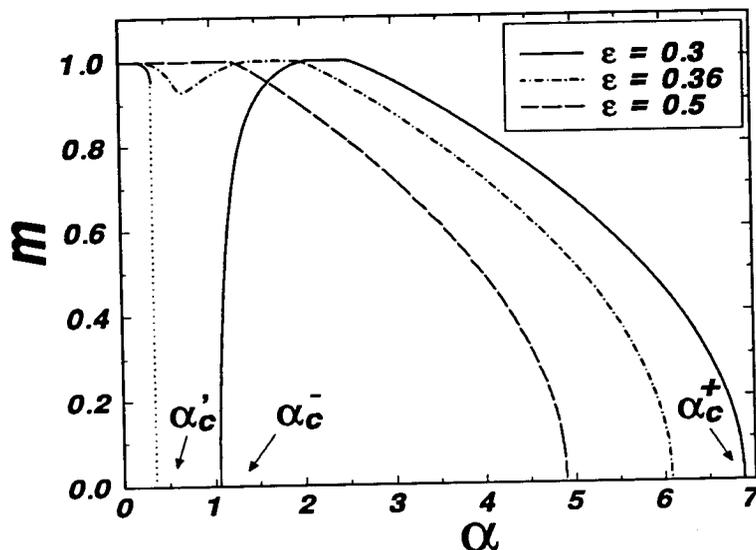


Figura 4.1: Superposição contra α para $\varepsilon = 0.3, 0.36$ e 0.5 . Quando $\varepsilon \rightarrow 0^+$ então $\alpha_c' \rightarrow 0.138$ e $\alpha_c^\pm \rightarrow \infty$, recuperando o modelo de Hopfield original.

O limite $T = 0$ ($\beta \rightarrow \infty$) das equações de ponto de sela (4.13a)–(4.13e) é

$$m = \operatorname{erf}\left(\frac{t}{\sqrt{2\alpha r}}\right) \quad (4.14a)$$

$$t = (1 - \varepsilon y)m + \varepsilon m^3 \quad (4.14b)$$

$$r = \left[\frac{1 - \varepsilon y}{1 - C(1 - \varepsilon y)}\right]^2 \quad (4.14c)$$

$$y = m^2 + \frac{\alpha r}{(1 - \varepsilon y)^2} \quad (4.14d)$$

$$C = \sqrt{\frac{2}{\alpha \pi r}} \exp\left(-\frac{t^2}{2\alpha r}\right), \quad (4.14e)$$

onde $q \rightarrow 1$ e $C \equiv \beta(1 - q)$. Estas equações são resolvidas numericamente para vários valores de α e ε e os resultados são comparados com as simulações cujos detalhes serão apresentados posteriormente.

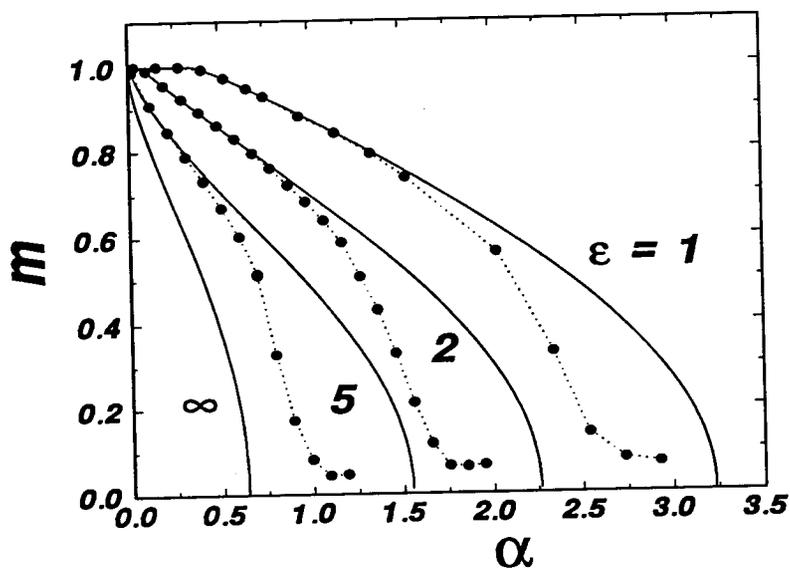


Figura 4.2: Superposição contra α para $\varepsilon = 1, 2, 5$ e ∞ . A qualidade de recuperação diminui para valores crescentes de ε e para $\varepsilon \rightarrow \infty$, $\alpha_c \rightarrow 2/\pi$. Os círculos são resultados de simulação numérica usando redes com $N = 512$. A diferença para $\alpha \simeq \alpha_c$ é devida a instabilidade da solução com simetria de réplicas.

O resultado mais interessante é que a dependência em ε , ao contrário do modelo de Hopfield generalizado, não é apenas quantitativa mas qualitativa. Nos modelos estudados no capítulo 2, observa-se que m é monotonicamente decrescente com α . Aqui, ao contrário,

foram encontrados essencialmente dois regimes distintos, dependendo se ε está abaixo ou acima de um certo valor crítico, $\varepsilon_c \simeq 0.3587$, e as figuras 4.1 e 4.2 mostram curvas típicas de recuperação para vários valores de ε . Para ε grande (~ 0.5) a superposição m decresce com α , indo monotonicamente de 1 até zero em $\alpha_c^+(\varepsilon)$, assinalando uma transição contínua. Ligeiramente acima de ε_c (~ 0.36), a superposição apresenta um mínimo local antes de finalmente ir a zero em $\alpha_c^+(\varepsilon)$. Se $\varepsilon < \varepsilon_c$, o mínimo origina uma lacuna separando duas regiões de recuperação.

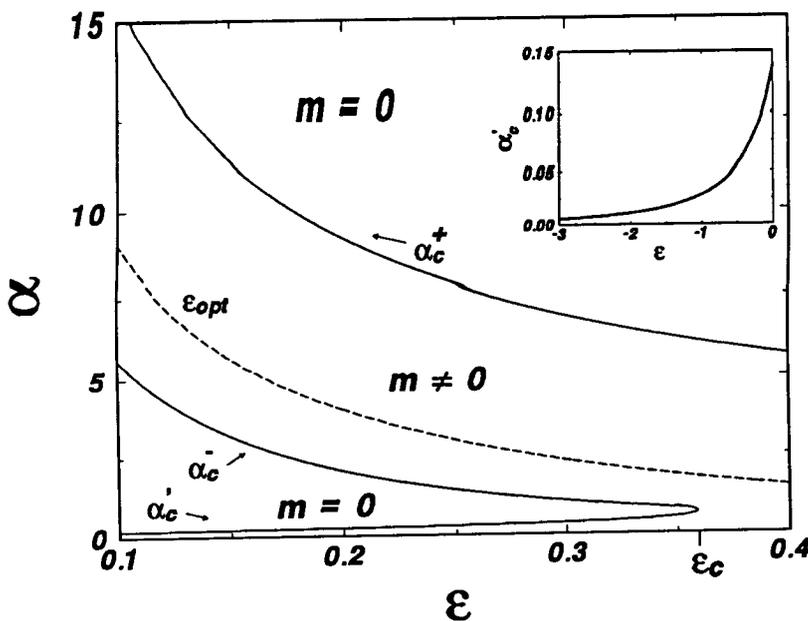


Figura 4.3: Diagrama de fases em $T = 0$. Para $\varepsilon > \varepsilon_c \simeq 0.3587$ a lacuna não existe, enquanto que para $\varepsilon \rightarrow 0$, $\Delta \rightarrow \infty$ como ε^{-1} . As linhas $\alpha_c^\pm(\varepsilon)$ são de segunda ordem enquanto que $\alpha'_c(\varepsilon)$ é de primeira ordem. A linha tracejada (ε_{opt}) é o valor de $\alpha \neq 0$ onde ocorre o pico $m = 1$. No detalhe, a linha de transição de primeira ordem α'_c para $\varepsilon < 0$.

Os valores críticos $\alpha_c^\pm(\varepsilon)$, $\varepsilon > 0$, onde m tende continuamente a zero ($m \sim |\alpha - \alpha_c^\pm(\varepsilon)|^{1/2}$ quando $\alpha \rightarrow \alpha_c^\pm(\varepsilon)$), são obtidos expandindo-se as equações (4.14a)–(4.14e) para pequenos valores de m (ver apêndice B):

$$\alpha_c^\pm(\varepsilon) = \left(\frac{1}{\sqrt{\varepsilon}} \pm \sqrt{\frac{2}{\pi}} \right)^2. \quad (4.15)$$

Note que o valor crítico de α é uma função decrescente de ε . Estas linhas são mostradas no diagrama de fases α versus ε (figura 4.3). Além disso, para $\varepsilon < \varepsilon_c$ também ocorre uma transição de primeira ordem em $\alpha'_c(\varepsilon)$.

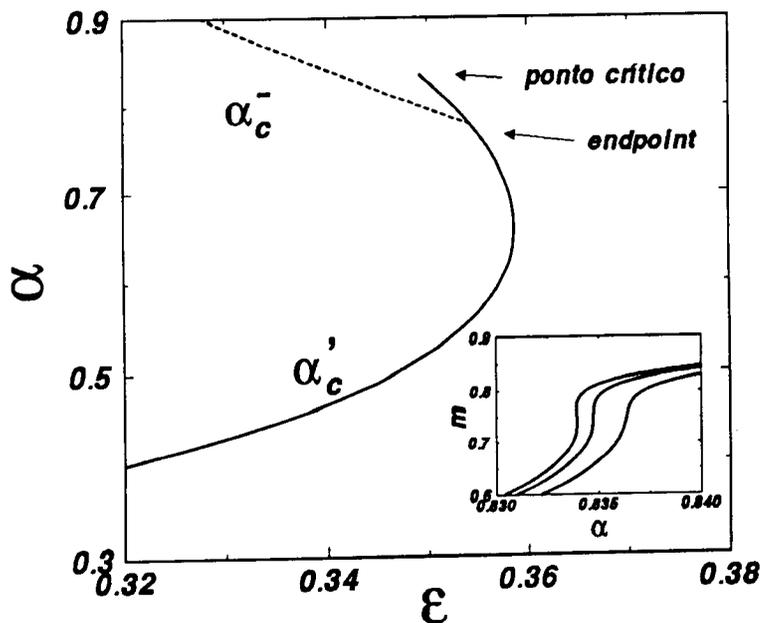


Figura 4.4: Região próxima do ponto crítico onde surge a lacuna. A linha tracejada é de segunda ordem enquanto que a sólida é de primeira ordem. A linha α'_c termina em um ponto crítico em $\varepsilon \simeq 0.3492$ e as linhas se cortam em um endpoint em $\varepsilon \simeq 0.3543$. Detalhe: superposição contra α perto do ponto crítico mostrando a segunda transição de primeira ordem para $\varepsilon = 0.3493, 0.3492$ e 0.3491 (da esquerda para a direita).

A lacuna é delimitada pelas linhas $\alpha_c^-(\varepsilon)$ e $\alpha'_c(\varepsilon)$, as quais se encontram no endpoint localizado em $(\varepsilon, \alpha) \simeq (0.3543, 0.7784)$ (fig. 4.4): a borda esquerda é sempre dada por $\alpha'_c(\varepsilon)$ enquanto que a direita é definida por $\alpha_c^-(\varepsilon)$ (segunda ordem) para $\varepsilon < 0.3543$ e por $\alpha'_c(\varepsilon)$ (primeira ordem) para $0.3543 < \varepsilon < 0.3587$. Como fica claro na figura 4.4, $\alpha'_c(\varepsilon)$ não é uma função pois existe um intervalo onde não é unívoca. A linha $\alpha'_c(\varepsilon)$ termina (na aproximação de simetria de réplicas) em um ponto crítico em $(\varepsilon, \alpha) \simeq (0.3492, 0.833)$. Quando ε tende a zero, a largura da lacuna entre a primeira e a segunda região de recuperação, $\Delta \equiv \alpha_c^- - \alpha'_c$, vai a infinito e para valores negativos de ε somente a transição de primeira ordem associada a $\alpha'_c(\varepsilon)$ está presente (fig. 4.3, detalhe). Quando a margem direita da lacuna é dada por α_c^- , a largura da segunda região de recuperação é

$$\Delta' \equiv \alpha_c^+(\varepsilon) - \alpha_c^-(\varepsilon) = 4\sqrt{\frac{2}{\varepsilon\pi}}. \quad (4.16)$$

À medida que $\varepsilon \rightarrow 0$, Δ vai a infinito como ε^{-1} e Δ' como $\varepsilon^{-1/2}$. Portanto, quando o modelo truncado recupera o modelo de Hopfield ($\varepsilon \rightarrow 0$), a localização da segunda região no eixo α vai a infinito e $\alpha'_c \rightarrow 0.138$.

Por outro lado, se $\varepsilon \rightarrow \infty$, a solução para m pode ser obtida como uma função de α

(ver apêndice B),

$$m = \operatorname{erf} \left(\sqrt{\frac{1}{2} \ln \frac{2}{\alpha\pi}} \right) \quad (4.17)$$

e

$$\alpha_c^+(\infty) = \frac{2}{\pi} \quad (4.18)$$

Neste caso limite, o sistema se comporta adequadamente como memória associativa ($m \simeq 1$) somente se o número de memórias armazenadas é finito ($\alpha = 0$), uma vez que m decresce rapidamente à medida em que α cresce.

Há um intervalo de ε , $0.3492 < \varepsilon < 0.3543$ onde o sistema sofre até quatro transições de fase (duas de primeira e duas de segunda ordem) uma vez que, para um valor fixo de ε neste intervalo, à medida que α aumenta, a linha α'_c é cruzada duas vezes.

O pico $m = 1$ na segunda região de recuperação ocorre porque o ruído no termo de Hopfield no campo local, eq.(4.8), é completamente compensado pelo ruído do termo de quarta ordem quando $\varepsilon = y^{-1}$ ($r = 0$). Para valores negativos isto nunca ocorre porque $1 + |\varepsilon|y$ é sempre positivo, o que explica porque não ocorre uma segunda região de recuperação quando $\varepsilon < 0$. A localização do pico, ou seja, do máximo valor de m na segunda região de recuperação, ocorre nos valores de α que permitem $m = 1$ em (4.14a)–(4.14e). Além da solução $\alpha = 0$ temos ainda

$$\alpha = \frac{1 - \varepsilon}{\varepsilon} \quad , \quad (4.19)$$

pois $r = C = 0$. O ruído de segunda ordem não contribui no pico e devemos levar em conta o próximo termo (quarta ordem).

Isto nos permite introduzir uma regra de aprendizado ótima escolhendo o peso ε como o valor que satisfaz (4.19):

$$\varepsilon_{opt} = \frac{1}{1 + \alpha} \quad , \quad (4.20)$$

o qual pode ser reintroduzido novamente na função energia. Mas para isso teríamos que permitir que a regra de aprendizagem se modificasse em função do número de memórias já aprendidas pela rede, o que, biologicamente, é razoável. De acordo com a eq.(4.5), quanto maior for α , menor será ε_{opt} e menores serão as modificações nas sinapses de quarta ordem: quanto mais sabemos mais fácil é aprender algo novo. Embora o termo de ruído (de segunda ordem) se anule, isso não significa que agora a capacidade seja governada pelo ruído de quarta ordem implicando que $P \sim \mathcal{O}(N^3)$ porque neste caso também o termo de sinal iria a zero (pois é proporcional a ε). Aqui aparece uma das diferenças fundamentais entre este modelo e as generalizações do modelo de Hopfield: lá, quanto maior fosse o peso do termo de ordem mais alta, mais informação poderia ser armazenada, ao passo que aqui,

com ε_{opt} , este peso torna-se cada vez menor à medida que mais e mais informação pode ser acrescentada ao sistema.

Os resultados acima podem ser entendidos qualitativamente também através de uma análise do campo local atuando sobre o i -ésimo neurônio quando o sistema está recuperando o primeiro padrão:

$$h_i \simeq (1 - \alpha\varepsilon)\xi_i^1 + (1 - \alpha\varepsilon - \varepsilon)\frac{1}{N} \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu \xi_j^1, \quad (4.21)$$

onde consideramos as contribuições das sinapses entre dois e quatro neurônios na eq.(4.3). O aumento de ε tem dois efeitos: ele age tanto sobre o sinal quanto sobre o ruído. Dependendo do valor de ε , um aumento em α pode tanto suprimir o sinal quanto o termo de ruído. Quando ε é suficientemente pequeno, a soma na eq.(4.21) pode suplantar o termo de sinal e teremos um mecanismo tal como no modelo de Hopfield para acabar com a capacidade de recuperação do sistema. Neste caso, o comportamento geral é qualitativamente similar a este modelo. Isto também corresponde à primeira região de recuperação para $\varepsilon < \varepsilon_c$. Por outro lado, quando $\varepsilon > \varepsilon_c$, junto com uma diminuição no termo de ruído, podemos observar uma diminuição no termo de sinal à medida que α aumenta na segunda região de recuperação. Agora o mecanismo para a perda das propriedades associativas é a aniquilação do termo de sinal. Isto explica o comportamento qualitativamente diverso da rede nesta região. Também, o segundo termo (ruído) tem média zero e variância $\sigma^2 = \alpha(1 - \alpha\varepsilon - \varepsilon)^2$. Um resultado importante é que a variância é zero se $\alpha = 0$ ou $\alpha = (1 - \varepsilon)/\varepsilon$: estes são os pontos, como já vimos, onde a recuperação é perfeita ($m = 1$). Termos de ruído de ordem mais alta, que não escrevemos, têm dispersões no mínimo $\mathcal{O}(N^{-1})$ vezes menores. Para valores crescentes de ε , o coeficiente do termo de sinal, $1 - \alpha\varepsilon$, troca de sinal para valores cada vez menores de α , explicando porque o valor crítico de α diminui.

Deve ser enfatizado aqui que a solução com simetria de réplicas é instável neste limite ($T = 0$), uma vez que a entropia

$$S_0 \equiv \beta^2 \left. \frac{\partial f}{\partial \beta} \right|_{T=0} = -\frac{\alpha}{2} \ln[1 - C(1 - \varepsilon y)] - \frac{\alpha}{2} \frac{C(1 - \varepsilon y)}{1 - C(1 - \varepsilon y)} \quad (4.22)$$

é sempre negativa¹ (exceto em alguns pontos particulares como veremos a seguir). Em módulo, o valor numérico da entropia (4.22) é maior do que no modelo de Hopfield, pos-

¹Isto pode ser mostrado a partir do fato que $C(1 - \varepsilon y) < 1$ pois, como $r \geq 0$:

$$\frac{C(1 - \varepsilon y)}{1 - C(1 - \varepsilon y)} = \sqrt{\frac{2}{\alpha\pi}} \exp\left(-\frac{t^2}{2\alpha r}\right) \leq \sqrt{\frac{2}{\alpha\pi}}$$

sivelmente indicando que os efeitos de quebra de simetria (RSB) aqui são maiores. Isto não vale para todos os valores de α : em $\alpha = 0$ e em $\alpha = (1 - \epsilon_{opt})/\epsilon_{opt}$ (onde $y = \epsilon^{-1}$) a entropia se anula ($S_0 = 0$), ou seja, para estes valores a quebra de simetria de réplica não trará nenhuma modificação, o que explica o fato das simulações numéricas concordarem tão bem com os resultados analíticos nestes pontos.

Um último comentário sobre a energia livre do modelo truncado em $T = 0$: os estados de recuperação são mínimos globais para todos os valores de α abaixo de α_c^+ se $\epsilon > \epsilon_c$. Para $\epsilon < \epsilon_c$ há um intervalo de α na primeira região de recuperação onde eles são mínimos locais (metaestáveis): $\alpha_M < \alpha < \alpha'_c$ (na segunda região, entretanto, são sempre mínimos globais). Para $\epsilon \rightarrow 0$, $\alpha_M \rightarrow 0.05$, como esperado [2].

4.2.2 O Caso $T \neq 0$

As equações de ponto de sela também foram estudadas quando o ruído térmico não é nulo, onde alguns resultados não usuais também ocorrem. A figura 4.5 ilustra o comportamento da superposição para dois valores da temperatura: note a existência de duas regiões de recuperação já aparentes em $T = 0$ e desconexas em $T = 0.15$. Também pode-se notar que a introdução de ruído térmico no sistema diminui o valor de α no qual m tem um máximo na segunda região de recuperação, além de fazer com que todas as transições sejam de primeira ordem.

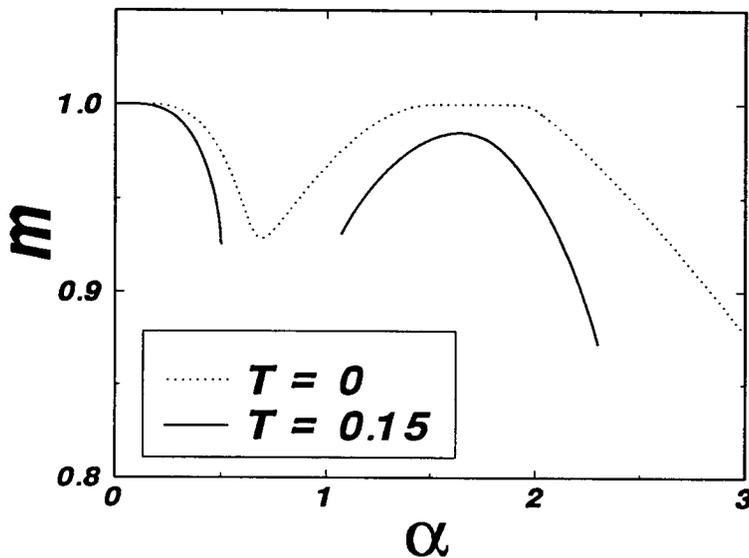


Figura 4.5: Superposição m para $\epsilon = 0.36$ e vários valores de T .

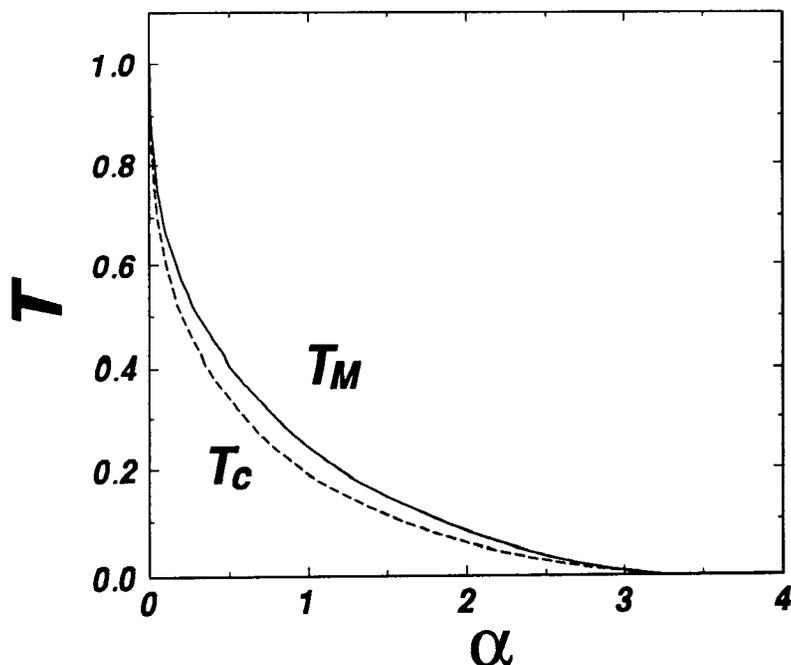


Figura 4.6: Diagrama de fases para $\varepsilon = 1$. Abaixo da linha T_M temos soluções com $m \neq 0$ (estados de recuperação) e estes estados tornam-se mínimos globais da energia livre abaixo de T_c (linha tracejada). Note que há somente uma região de recuperação e nenhuma fase reentrante.

O diagrama de fases para o modelo truncado para $\varepsilon = 1$ é mostrado na figura 4.6. As linhas T_M e T_c assinalam, respectivamente, quando os estados de recuperação primeiro aparecem e quando se tornam mínimos globais da energia livre. Na figura 4.7, a linha T_M é mostrada para dois valores de ε : 0.36 e 0.3. Para $\varepsilon > \varepsilon_c$, os estados de recuperação são sempre mínimos globais da energia livre para baixas temperaturas, embora para $\varepsilon < \varepsilon_c$ possam se tornar mínimos locais na primeira região de recuperação. Para $\alpha = 0$ e todos os valores de ε , $T_M = T_g = T_c = 1$. Entretanto, para $\alpha \neq 0$, as características qualitativas do diagrama de fases dependem de ε . Quando $\varepsilon = 1$ (fig. 4.6), T_M diminui monotonicamente e não há nenhuma fase reentrante. À medida que ε diminui, a linha T_M desenvolve um mínimo, implicando que para algum intervalo de temperatura há duas regiões de recuperação (e.g., para $\varepsilon = 0.36$, $0.086 < T < 0.202$), como mostrado na fig. 4.7. Para valores ainda menores de ε , o mínimo se transforma em uma lacuna (fig. 4.7) e a primeira região de recuperação é reentrante. Em $T = 0$ há três linhas de transição de fases no plano $\varepsilon \times \alpha$: uma de primeira ordem, α'_c e duas de segunda ordem α_c^\pm (fig. 4.3), mas para $T \neq 0$, as três linhas de transição de fase são de primeira ordem.

A temperatura na qual a solução de vidro de spin SG ($m = 0, q \neq 0$) continuamente

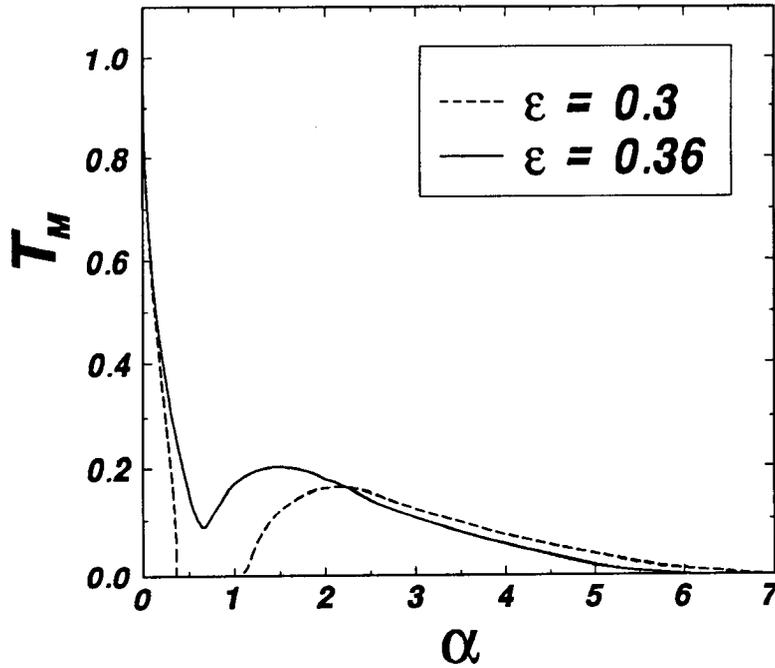


Figura 4.7: A linha T_M para $\varepsilon = 0.3$ e 0.36 . O último caso já mostra a estrutura de duas regiões de recuperação, mas ainda conexas, embora exista um intervalo de temperatura no qual temos duas regiões de recuperação: $0.086 < T < 0.202$ para $\varepsilon = 0.36$. Para $\varepsilon = 0.3$ as regiões se separam e a primeira é reentrante.

desaparece ($q \rightarrow 0$) é (ver apêndice B)

$$T_g(\alpha, \varepsilon) = \frac{1 + \sqrt{\alpha}}{\sqrt{\alpha}} \left[(1 - \alpha\varepsilon)(1 + \sqrt{\alpha}) - 1 \right] \quad . \quad (4.23)$$

Estas linhas são mostradas na figura 4.8 para vários valores de ε . Da equação acima podemos ver que $T_g(0, \varepsilon) = 1$ para todos valores de ε e para $\varepsilon = 0$ recuperamos a linha $T_g(\alpha, 0) = 1 + \sqrt{\alpha}$ do modelo de Hopfield. A linha T_g , para $\varepsilon \neq 0$, vai a zero em (fig. 4.8)

$$\alpha_g = \frac{1}{4} \left(1 - \sqrt{1 + \frac{4}{\varepsilon}} \right)^2 \quad . \quad (4.24)$$

O mesmo acontece no modelo da pseudo-inversa [56], embora naquele caso a transição seja descontínua e $\alpha_g \simeq 0.363$. Embora haja uma solução com $q = 0$ para a fase SG ao longo da linha $\alpha = \varepsilon^{-1}$, a linha crítica α_g , onde a fase SG desaparece, é tal que $\alpha_g \leq \varepsilon^{-1}$ (fig. 4.8). Uma consequência disto é que para grandes valores de α (acima de α_g) o efeito dos estados espúrios na dinâmica é severamente reduzido. É também interessante notar que $\alpha_{opt} \geq \alpha_g$ para $\varepsilon < 1/2$, isto é, quando temos o pico da segunda região de recuperação já não temos mais os estados de vidro de spin (pelo menos na aproximação de simetria de réplicas).

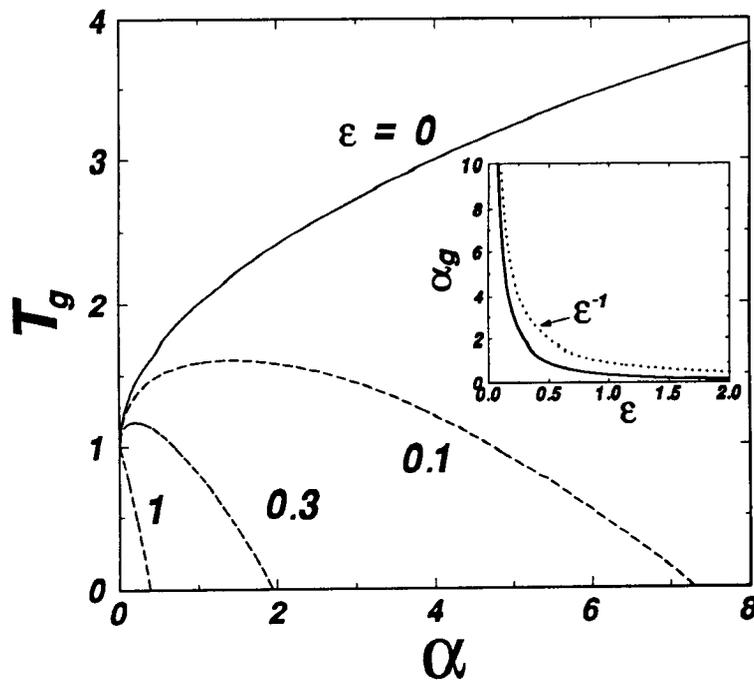


Figura 4.8: Temperatura de transição para as soluções SG, T_g , para vários valores de ϵ . Para $\epsilon \rightarrow 0$, recuperamos a linha $T_g = 1 + \sqrt{\alpha}$ do modelo de Hopfield. No detalhe mostramos os pontos α_g onde $T_g = 0$, assim como a linha ϵ^{-1} .

4.3 Simulações Numéricas

Os resultados das seções anteriores foram extensivamente confrontados com simulações numéricas que passamos agora a descrever. Essencialmente foram medidas:

- curvas de recuperação;
- tempos de convergência e
- efeito dos estados espúrios.

Estas quantidades foram obtidas da mesma maneira descrita nos capítulos anteriores para os outros modelos, utilizando-se redes de $N = 128, 256, 512$ e 1024 neurônios. Os passos são os seguintes: um estado inicial com superposição m_0 com uma das memórias é gerado e os neurônios são testados seqüencialmente para detectar uma possível diminuição de energia frente a uma inversão. Isto é repetido até que o sistema atinge um estado estável, quando então as várias quantidades de interesse são medidas.

O máximo valor de α para o qual a rede estabiliza em (ou muito próximo de) uma das memórias ($m_f \simeq 1$) nos fornece informações sobre a capacidade de armazenamento da rede. Partindo de superposições iniciais $m_0 = 1$ com uma dada memória, medimos a

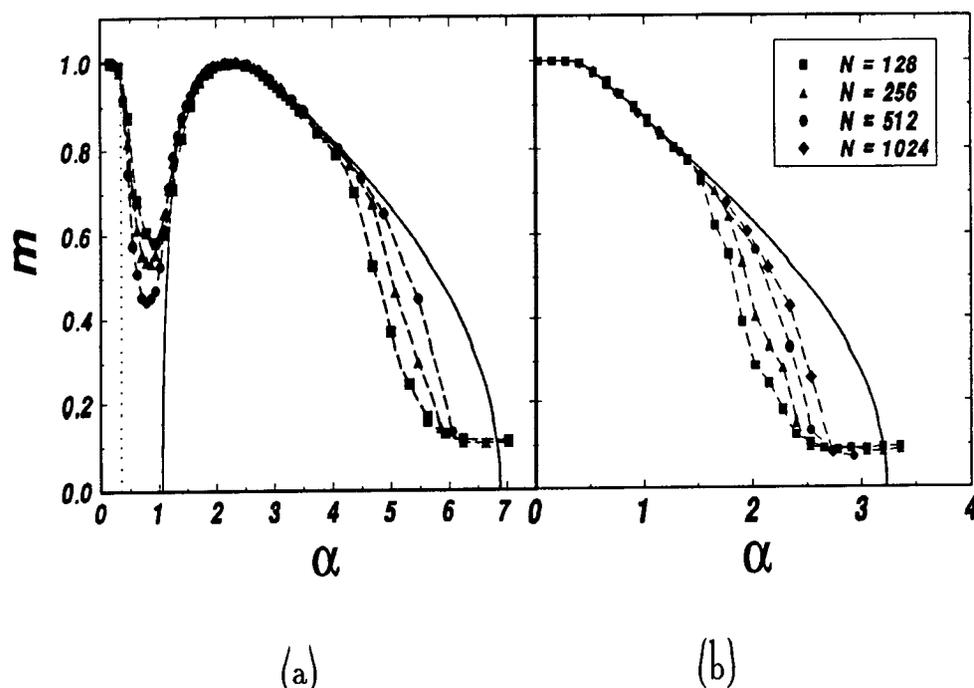


Figura 4.9: Superposição contra α para $\varepsilon = 0.3$ (a) e 1 (b). A linha cheia é a solução em $T = 0$ das eqs.(4.14a)-(4.14e) enquanto que os pontos são obtidos a partir de simulação numérica. Note a recuperação perfeita em $\alpha = 0$ e $\alpha = (1 - \varepsilon)/\varepsilon \simeq 2.33$ na figura (a).

superposição final com esta memória para vários valores de α . Os resultados são mostrados na figura 4.9 para dois valores de ε . Em ambos os casos a concordância é muito boa, confirmando inclusive a existência da lacuna. Contrariamente aos resultados analíticos das seções precedentes que previam $m = 0$ para $\alpha > \alpha_c^+(\varepsilon)$, aqui temos uma superposição residual (também chamada de magnetização remanente) da ordem de 0.1 (mas que diminui à medida que ε cresce, como mostra a figura 4.2). No modelo de Hopfield tal superposição vale aproximadamente 0.2. Comparando com os resultados analíticos observamos que há uma certa discrepância entre o valor previsto de α_c^+ e o obtido na simulação. Esta diferença é causada pela instabilidade da aproximação de simetria de réplicas em $T = 0$. Provavelmente tal diferença diminuiria em cada passo na quebra da simetria.

Investigamos o comportamento do sistema quando as superposições iniciais são diferentes de 1, $m_0 < 1$, de modo a obter informações sobre os tamanhos das bacias de atração. Na figura 4.10 mostramos os resultados para $\varepsilon = 1$: aparentemente as bacias são independentes de α uma vez que para diversos valores de α as curvas se superpõem. Neste caso, o valor crítico da superposição inicial, m_c , vale aproximadamente 0.1 para todo $\alpha < \alpha_c$.

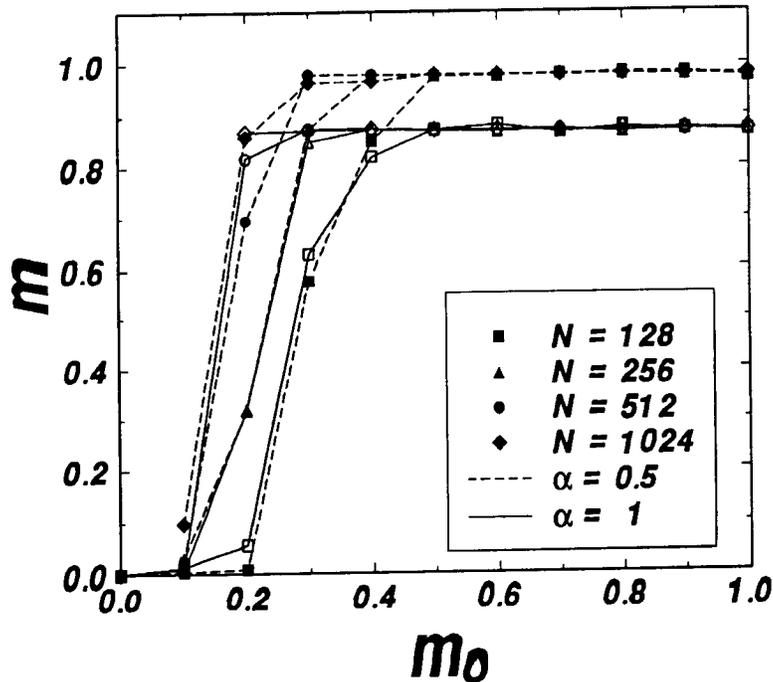


Figura 4.10: Superposição final contra a superposição inicial para $\varepsilon = 1$ e dois valores de α .

No entanto, apesar das bacias serem grandes, a qualidade de recuperação diminui com α . Para valores de $\varepsilon < \varepsilon_c$, onde existe a lacuna, o comportamento depende da região de recuperação em que o sistema se encontra (veja a figura 4.11). Na primeira região de recuperação o comportamento lembra o encontrado para o modelo de Hopfield [31]: o tamanho das bacias de atração depende do valor de α e diminui com este. Já na segunda região de recuperação o comportamento é análogo ao encontrado para grandes valores de α , as bacias sendo razoavelmente independentes de α . Na lacuna, como não há recuperação, as curvas tendem a se aproximar do eixo horizontal à medida que o tamanho da rede aumenta.

Como vimos nos capítulos precedentes, o tempo médio de convergência, $\langle T \rangle$, isto é, o número médio de passos (atualizações de toda a rede) necessários para atingir um estado estável, e a dispersão correspondente (σ) supostamente estão relacionados com as irregularidades do espaço de fase ao redor das memórias. Em alguns modelos, longos tempos de convergência e grandes dispersões na sua média indicam a presença de estados espúrios: quando temos muitos estados metaestáveis ao redor das memórias existem vários caminhos com diferentes comprimentos levando do estado inicial até a memória. Isto, além de aumentar o tempo médio de convergência, também implica em maiores dispersões. Por outro lado, esta interpretação permanece válida somente quando os padrões estão ou no fundo da bacia de atração ou muito próximos do fundo, isto é, a superposição final média é próxima de 1.

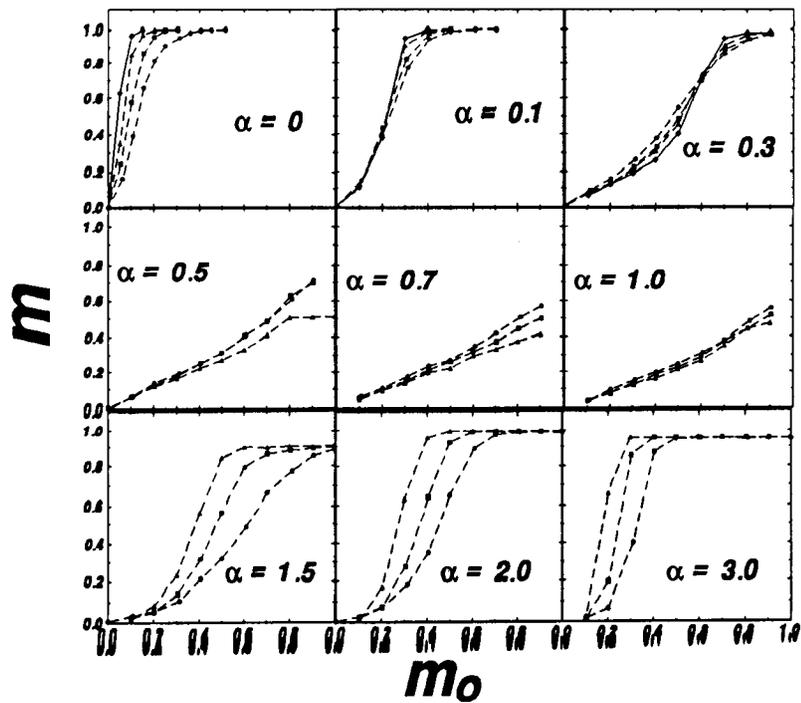


Figura 4.11: Superposição final contra a superposição inicial para $\epsilon = 0.3$ e vários valores de α .

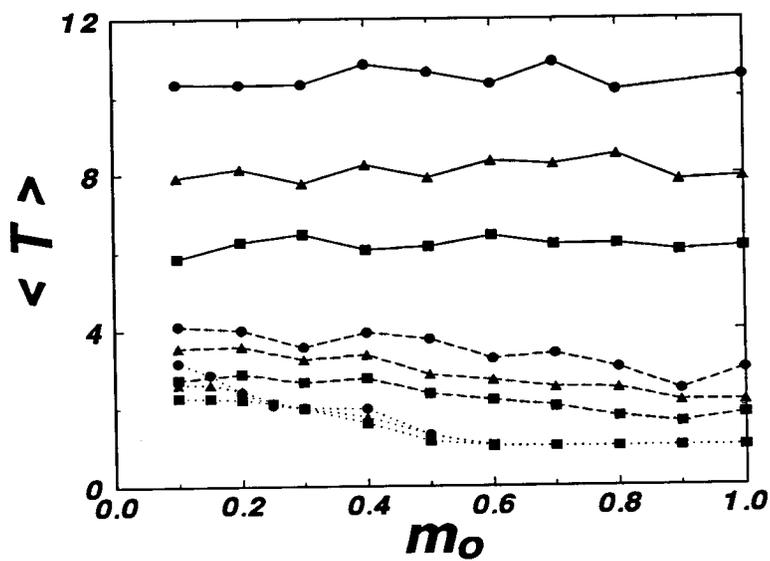


Figura 4.12: Tempo médio de convergência versus a superposição inicial para $\alpha = 0.1, 0.5$ e 1 (linhas pontilhada, tracejada e sólida, respectivamente) e $N = 128(\blacksquare)$, $256(\blacktriangle)$ e $512(\bullet)$. Note que não há dependência sensível no valor inicial da superposição, exceto quando α é pequeno.

No modelo truncado, mostrado na fig. 4.12, o tempo médio de convergência é independente da superposição inicial: para qualquer valor de m_0 , o número necessário de passos para atingir um estado estacionário é aproximadamente o mesmo. Para entendermos este comportamento atípico é preciso lembrar que para valores não muito pequenos de α o estado no fundo da bacia de atração pode ser essencialmente diferente da memória armazenada uma vez que a transição é contínua: a superposição final m diminui continuamente desde 1 até 0 à medida que α cresce. Neste caso, quando estados iniciais com superposição m_0 com uma dada memória são gerados, estão sobre a superfície de uma hiperesfera cujo raio é proporcional a $1 - m_0$ e está centrada no padrão (veja a ilustração, fig. 4.13). Em média, a distância no espaço de fase entre estes estados iniciais e o fundo da bacia de atração é igual à distância entre a memória (ou seja, o centro da hiperesfera) e o fundo da bacia, sendo, portanto, independente de m_0 . Conseqüentemente, o tempo de convergência médio não depende de m_0 . Para pequenos valores de α , por outro lado, a qualidade de recuperação é boa ($m \approx 1$) e o argumento acima não vale mais: de fato, um pequeno decréscimo em $\langle T \rangle$ é observado à medida que m_0 cresce.

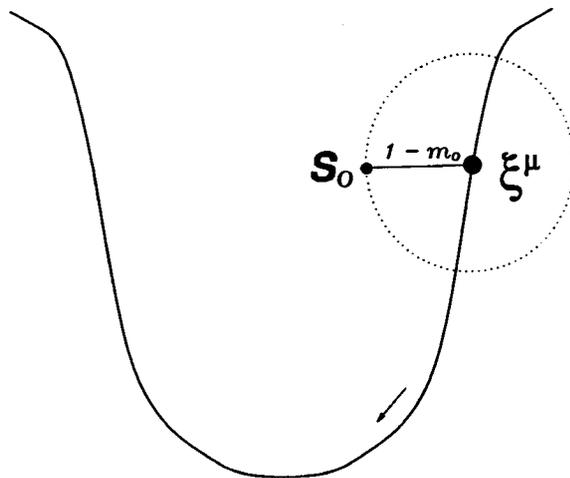


Figura 4.13: Forma esquemática da bacia de atração de uma memória e a superfície formada pelos estados iniciais S_0 que possuem superposição m_0 com ela para grandes valores de α .

Esta geometria também implica que as dispersões não nulas ao redor dos tempos médios de convergência nem sempre assinalam a existência de estados espúrios: sempre que as memórias estejam razoavelmente distantes do fundo das bacias de atração, $m \ll 1$, devido às diferentes distâncias entre os estados iniciais e o mínimo de energia, haverá uma dispersão

não nula no tempo de convergência.

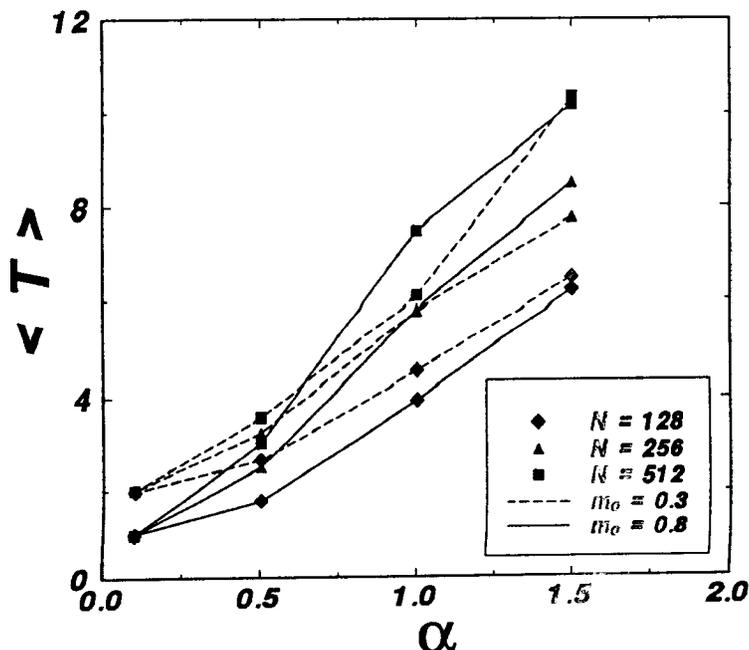


Figura 4.14: Tempo médio de convergência contra α . As curvas para $m_0 = 0.3$ e 0.8 colapsam assinalando independência das condições iniciais.

O tempo médio de convergência é também mostrado em função do parâmetro α , fig. 4.14, para dois valores da superposição inicial m_0 . As diferenças entre as duas curvas existem somente para pequenos valores de α e, à medida que α aumenta, ambas colapsam confirmando a independência de $\langle T \rangle$ na superposição inicial.

A quantidade \mathcal{M} , descrita no capítulo 1 como indicadora de uma diminuição ou não no número de estados espúrios do sistema, foi aqui medida para tentar detectar o efeito da desaparecimento dos estados de vidro de spin. Os resultados são apresentados na figura 4.15: obtivemos $\mathcal{M} = 0$ para $\alpha = 0$ e para α ligeiramente superior a 0.5 ($\alpha/\alpha_c \simeq 0.11$). Este resultado corrobora a predição analítica de que os estados de vidro de spin desaparecem acima de um certo valor α_g . Entretanto, a concordância é apenas qualitativa uma vez que o valor previsto de α_g para $\varepsilon = 0.5$ é 1. Acreditamos que esta diferença possa ser atribuída à instabilidade da solução com simetria de réplicas à baixas temperaturas. Outra possibilidade é de que as bacias de atração dos estados espúrios sejam muito pequenas comparadas às das memórias de modo que os estados iniciais aleatórios teriam pouca probabilidade de atingir tais estados.

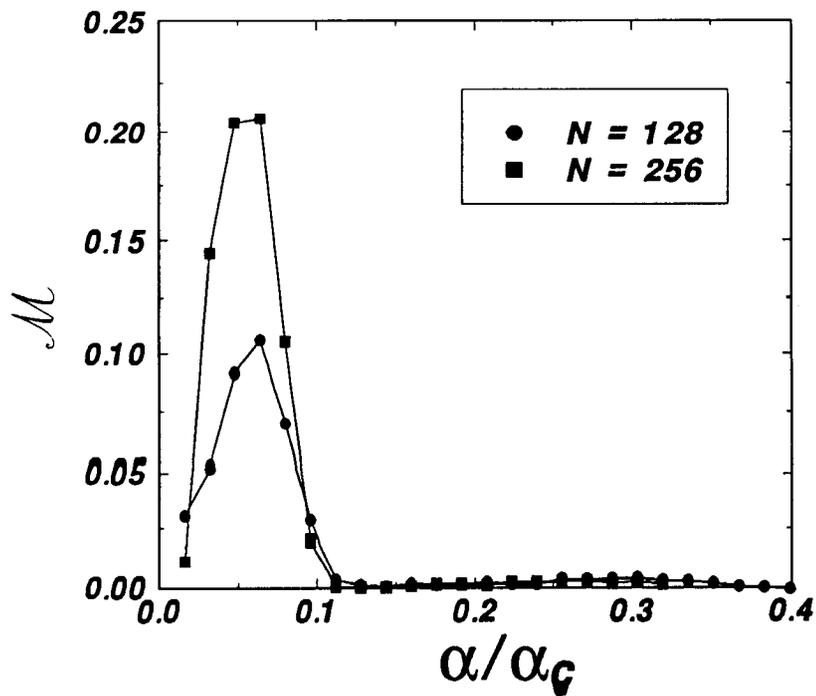


Figura 4.15: A quantidade \mathcal{M} versus α/α_c para $\varepsilon = 0.5$. O máximo aumenta com o tamanho da rede, indicando o efeito dos estados espúrios, mas para $0.5 \lesssim \alpha < \alpha_c \simeq 4.893$ temos $\mathcal{M} \simeq 0$, assinalando a baixa ocupação do espaço de fase pelos estados espúrios.

4.4 Conclusões

Neste capítulo apresentamos os resultados para os estados de equilíbrio do modelo truncado com sinapses de segunda e quarta ordem. O comportamento obtido é muito diverso do encontrado para as generalizações do modelo de Hopfield. Por exemplo, o parâmetro que mede a qualidade de recuperação, m , não é monotonicamente decrescente com α , mas pode apresentar até duas regiões distintas de recuperação, separadas por uma lacuna. Além disso, para altos valores de α , os estados de vidro de spin não são mais estáveis, resultado que recebe forte suporte das simulações numéricas. A lacuna existe para valores de ε menores do que um certo valor crítico ($\varepsilon_c \simeq 0.3587$) e, quando $\varepsilon \simeq \varepsilon_c$, o comportamento é muito complexo na vizinhança da lacuna.

Como a transição é contínua, os valores de α_c devem ser considerados com cuidado. Por exemplo, para $\varepsilon = 1$ temos que $\alpha_c \simeq 3.232$, mas já para $\alpha > 1$ a superposição final, m , é menor do que 0.9 e a recuperação não é satisfatória. Apesar disso, as bacias de atração são grandes e independem de α .

O tempo médio de convergência $\langle T \rangle$, para valores não muito pequenos de α , não depende do valor inicial da superposição, o que é uma consequência geral de termos uma transição contínua. Outra consequência é que existe uma dispersão considerável ao redor

destes valores médios, mas cuja causa não são os estados espúrios que possam existir ao redor das memórias. Em resumo, para transições contínuas, a existência de estados espúrios metaestáveis não está assegurada pela presença de dispersões não nulas ao redor dos tempos médios de convergência. Para obtermos tal informação seria preciso criar estados iniciais com uma superposição efetiva com o estado que se encontra no fundo da bacia de atração, e não com a memória originalmente armazenada.

A existência de um pico na segunda região de recuperação nos permite introduzir um valor ótimo de ϵ , agora uma função de α , que faz com que a recuperação seja sempre perfeita. Tal modelo ótimo é interessante uma vez que quanto maior é a quantidade de informação aprendida, menor é o peso dos termos de quarta ordem e, como aparentemente ocorre nos sistemas nervosos reais, tais conexões de alta ordem podem ser encaradas apenas como correções às muito mais abundantes sinapses de segunda ordem. Este valor ótimo de ϵ somente é válido para $T = 0$, e o modelo provavelmente estará próximo do comportamento ótimo para baixas temperaturas. Seria interessante verificar o que ocorre para valores maiores da temperatura.

Capítulo 5

O Modelo Truncado - A Dinâmica

Neste capítulo estudamos a dinâmica do modelo truncado no limite de alta diluição [85], isto é, quando somente uma pequena fração das sinapses é mantida. Como no caso estudado no capítulo anterior em que a conectividade é total, as sinapses de ordem mais alta introduzem efeitos novos no comportamento do sistema, em particular uma fase caótica surge dependendo dos valores dos parâmetros ϵ , T e α . Por outro lado, como as conexões são assimetricamente cortadas, agora não mais existe uma função energia que decresce à medida que o sistema evolui, o que impede o uso das técnicas do capítulo anterior. Apesar disso, a dinâmica do sistema e seus atratores (pontos fixos ou não) são facilmente obtidos, como veremos.

5.1 O Modelo Truncado Diluído

Como já vimos no capítulo 1, quando as conexões sinápticas são severa e assimetricamente cortadas, a dinâmica da rede pode ser avaliada exatamente [29]. Para introduzirmos assimetria no sistema, as sinapses J_{ij} e J_{ji} são independentemente cortadas com probabilidade $1 - C/N$, onde C é a conectividade média de cada neurônio, sendo que o limite de extrema diluição é obtido quando $C \ll \ln N$. O mesmo vale para as conexões de quarta ordem, mas com probabilidade de serem cortadas de $1 - C/N^3$.

A dinâmica em paralelo do sistema é governada por

$$S_i(t+1) = \begin{cases} +1 & \text{com probabilidade } (1 + \exp[-2\beta_o h_i(t)])^{-1} \\ -1 & \text{com probabilidade } (1 + \exp[+2\beta_o h_i(t)])^{-1} \end{cases}, \quad (5.1)$$

onde o parâmetro $\beta_o \equiv T_o^{-1}$ mede o nível de ruído na rede e $h_i(t)$ é o campo local atuando

sobre o i -ésimo neurônio no instante t :

$$h_i = \sum_j J_{ij} S_j - \varepsilon \sum_{j,k,l} J'_{ijkl} S_j S_k S_l \quad . \quad (5.2)$$

Os acoplamentos são:

$$J_{ij} = C_{ij} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \quad (5.3)$$

$$J'_{ijkl} = \frac{1}{3} (J_{ijkl} + J_{ljk i} + J_{kjil}) \quad (5.4)$$

$$J_{ijkl} = C_{ijkl} \sum_{\mu \neq \nu} \xi_i^{\mu} \xi_j^{\mu} \xi_k^{\nu} \xi_l^{\nu} \quad (5.5)$$

e C_{ij} e C_{ijkl} são variáveis aleatórias distribuídas de acordo com as probabilidades ρ e $\tilde{\rho}$:

$$\rho(C_{ij}) = \frac{C}{N} \delta(C_{ij} - 1) + \left(1 - \frac{C}{N}\right) \delta(C_{ij}) \quad (5.6)$$

$$\tilde{\rho}(C_{ijkl}) = \frac{C}{N^3} \delta(C_{ijkl} - 1) + \left(1 - \frac{C}{N^3}\right) \delta(C_{ijkl})$$

onde $C \ll \log N$.

No capítulo anterior vimos que as auto-interações de quarta ordem do tipo J_{iikl} e J_{ijkk} desempenham um papel importante. Estas conexões, em princípio, poderiam criar correlações entre os estados do neurônio i em diferentes tempos, o que impediria o cálculo da dinâmica como é feita na ref. 29. Na verdade, no caso diluído estas interações são desprezíveis no limite termodinâmico pois

$$J_{iikl} \simeq \alpha \frac{C}{N} J_{kl} \quad , \quad (5.7)$$

e, como $C/N \rightarrow 0$ uma vez que $C \ll \ln N$, não introduzem correlações.

Assumindo que o estado inicial está correlacionado com somente uma das memórias, queremos obter uma equação de recorrência do tipo $m(t+1) = f(m(t))$ para a superposição $m(t)$ entre o estado da rede e a memória,

$$m(t) = \frac{1}{N} \sum_i \langle \xi_i^1 S_i(t) \rangle \quad , \quad (5.8)$$

onde $\langle \dots \rangle$ significa as médias térmica e sobre as possíveis configurações da rede. Usando o método desenvolvido em 29 (ver apêndice D), após tomar o limite $C \rightarrow \infty$, a equação que governa a evolução em paralelo do sistema é:

$$m(t+1) = \int_{-\infty}^{+\infty} \mathcal{D}y \tanh \beta \left[m(t) - \sqrt{2\alpha} y (1 - \varepsilon m^2(t)) \right] \quad , \quad (5.9)$$

onde $C\alpha = P - 1$, $\beta = C/T_o$ e $\mathcal{D}y$ vale:

$$\mathcal{D}y = \frac{dy}{\sqrt{\pi}} e^{-y^2} . \quad (5.10)$$

Também estudamos o caso onde o estado inicial tem superposição macroscópica com as duas primeiras memórias, as quais têm superposição fixa entre si,

$$\kappa = \frac{1}{N} \sum_i \xi_i^1 \xi_i^2 , \quad (5.11)$$

e as $P - 2$ remanescentes são descorrelacionadas. As quantidades de interesse são as superposições com estas duas memórias:

$$m_1(t) = \frac{1}{N} \sum_i \langle \xi_i^1 S_i(t) \rangle$$

$$m_2(t) = \frac{1}{N} \sum_i \langle \xi_i^2 S_i(t) \rangle$$
(5.12)

e as equações para a evolução temporal são:

$$M(t+1) = (1 + \kappa) \int \mathcal{D}y \tanh \beta \left[M(1 - \varepsilon K_-) - y\sqrt{2\alpha}(1 - 2\varepsilon K_+) \right]$$

$$m(t+1) = (1 - \kappa) \int \mathcal{D}y \tanh \beta \left[m(1 + \varepsilon K_-) - y\sqrt{2\alpha}(1 - 2\varepsilon K_+) \right]$$
(5.13)

onde:

$$K_{\pm} = \frac{M^2 \pm m^2}{4} \quad (5.14a)$$

$$M = m_1 + m_2 \quad (5.14b)$$

$$m = m_1 - m_2 . \quad (5.14c)$$

Nas próximas subseções estas equações serão numericamente resolvidas e os vários comportamentos do sistema apresentados.

5.2 O Limite $T = 0$

No limite $T = 0$, o mapa $f(m)$, eq.(5.9), fica

$$f(m) = \text{erf} \left[\frac{m(t)}{\sqrt{2\alpha}(1 - \varepsilon m^2(t))} \right] . \quad (5.15)$$

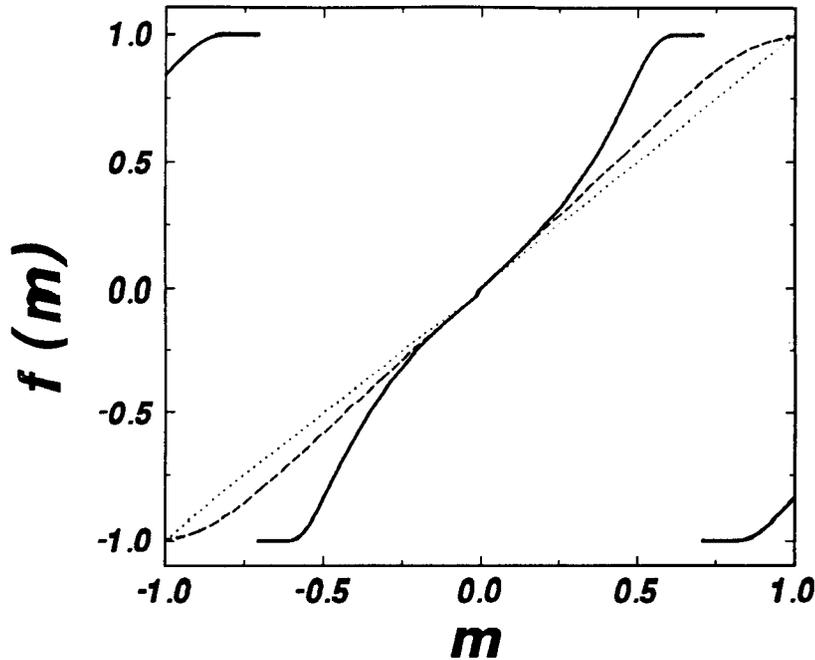


Figura 5.1: Lado direito da eq.(5.15) para $\alpha = 0.5$ e dois valores de ϵ : 0.5 (linha tracejada) e 2 (linha cheia). Para $\epsilon > 1$, $f(m)$ apresenta descontinuidades em $m = \pm\epsilon^{-1/2}$.

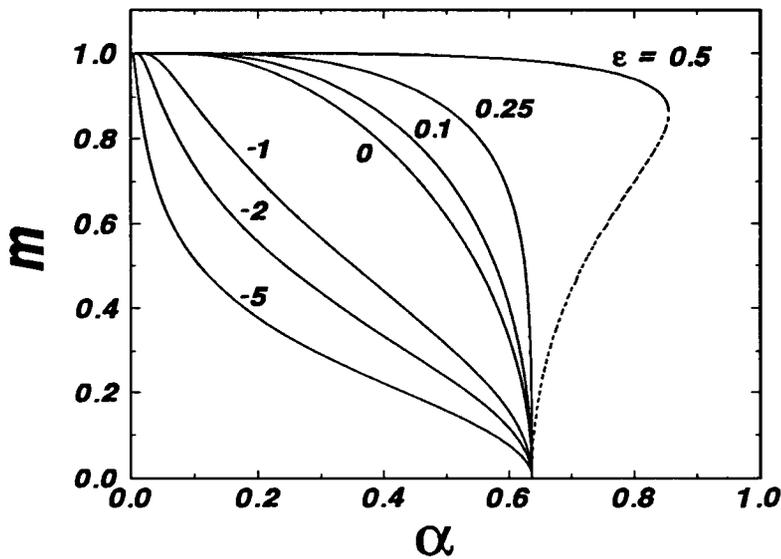


Figura 5.2: Pontos fixos m versus α para vários valores de ϵ . A linha cheia indica as soluções estáveis e a tracejada, as instáveis. Para $\alpha > 2/\pi$ a solução $m = 0$ é sempre estável. A linha tracejada também delimita as bacias de atração das duas soluções estáveis: se m_0 for menor do que a solução instável, o ponto fixo será $m = 0$, caso contrário o sistema converge para a solução com $m \neq 0$.

Note que quando $\varepsilon > 1$, o argumento da função erro muda de sinal toda vez que o denominador se anula, o que ocorre em $m = \pm\varepsilon^{-1/2}$, introduzindo descontinuidades no mapa. A fig. 5.1 mostra o mapa para dois valores diferentes de ε sendo que $f(m)$ é contínua para $\varepsilon < 1$ e descontínua para os outros valores.

É preciso agora encontrar os pontos fixos da equação (5.15), $m = f(m)$, os quais fornecem o comportamento assintótico ($t \rightarrow \infty$) da superposição. Estes pontos são mostrados na fig. 5.2 como uma função de α para vários valores de $\varepsilon < 1$. O valor crítico para a capacidade da rede, $\alpha_c(\varepsilon)$, é definido como o maior valor de α para o qual existe uma solução estável não trivial ($m \neq 0$). Dois casos precisam ser ressaltados: em primeiro lugar, para $\varepsilon < 0.266$ o sistema tem somente um ponto fixo estável que diminui continuamente até 0 em $\alpha_c = 2/\pi$. Em segundo, para $0.266 < \varepsilon < 1$ a transição é descontínua e α_c é maior do que $2/\pi$, crescendo com ε . Neste último caso há três diferentes regimes: para $\alpha < 2/\pi$ há somente a solução de recuperação; para $2/\pi < \alpha < \alpha_c$ esta solução coexiste com a solução $m = 0$ e para $\alpha > \alpha_c$ há somente a solução $m = 0$. No regime intermediário, aparece uma solução instável que separa as bacias de atração das duas soluções estáveis e a bacia da solução de recuperação diminui com α à medida que nos aproximamos de α_c . Além disso, para um dado valor de α , quanto mais próximo ε estiver de 1, melhor é o reconhecimento. A região de recuperação abaixo da linha α_c , onde existe o ponto fixo não nulo, é denominada R_1 , enquanto que acima de α_c a região com $m = 0$ é denominada D (veja a figura 5.3).

Quando $\varepsilon = 1$, como o argumento da função erro apresenta uma divergência se $m = \pm 1$, existe sempre esta solução de recuperação, a qual coexiste com a solução instável encontrada para $\alpha > 2/\pi$ ($0 < m < 1$). Esta solução instável, que delimita a fronteira entre as bacias de atração das duas soluções coexistentes, tende a zero à medida que nos aproximamos de $\varepsilon = 1$. Portanto, a bacia de atração da solução de recuperação aumenta enquanto que a bacia da solução $m = 0$ desaparece neste limite. Isto significa que a capacidade diverge e que a recuperação é perfeita para todos estados iniciais com superposição não nula. Este comportamento é o análogo de ε_{opt} encontrado no sistema completamente conectado onde $\varepsilon_{opt} = (1 + P/N)^{-1}$ ao passo que aqui temos

$$\varepsilon_{opt} = \frac{1}{1 + P/N} \simeq \frac{1}{1 + \alpha C/N} \rightarrow 1 \quad , \quad (5.16)$$

uma vez que $C \ll \ln N$. Também determinamos que, quando $\varepsilon \rightarrow 1$,

$$\alpha_c \sim |1 - \varepsilon|^{-2} \quad . \quad (5.17)$$

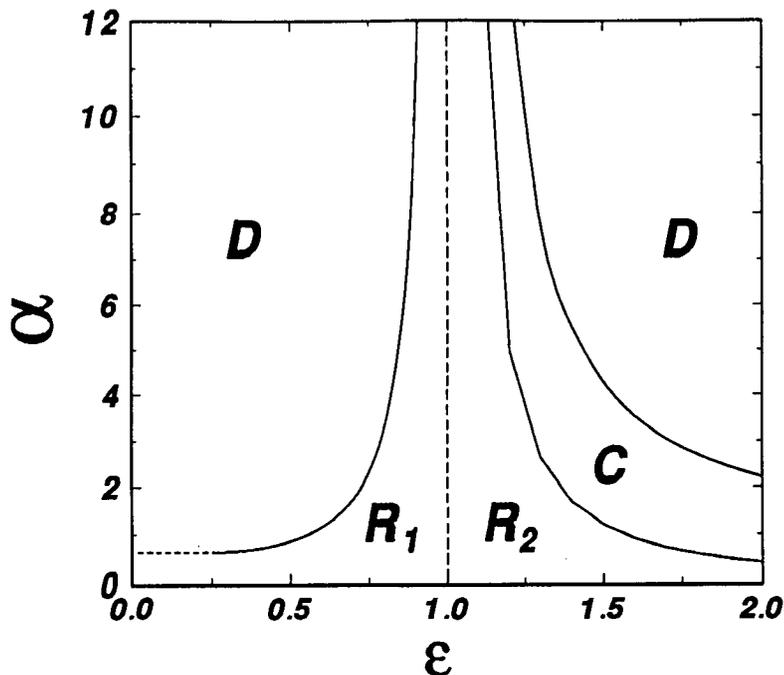


Figura 5.3: Diagrama α versus ε em $T = 0$. A transição é contínua para $\varepsilon < 0.266$. O valor ótimo de ε é 1 onde $\alpha_c \rightarrow \infty$ como $|1 - \varepsilon|^{-2}$. Para $\varepsilon > 1$ o sistema apresenta uma fase periódica ou caótica (C), uma de recuperação (R_2) e uma desordenada (D). A fronteira entre as fases R_2 e C é o ponto onde ocorre a primeira bifurcação.

Para valores de $\varepsilon > 1$ surge um rico comportamento e o sistema apresenta um novo tipo de recuperação. O comportamento geral do sistema é apresentado na fig. 5.3, mostrando os vários regimes. Para ε fixo e baixos valores de α (> 0), o sistema está numa fase de recuperação R_2 , diferente de R_1 pois aqui a superposição oscila entre os valores m e $-m$.

À medida que α cresce o sistema entra em uma nova fase, onde começa a oscilar entre dois valores distintos de m , digamos m_1 e m_2 ($|m_1| \neq |m_2|$), ao passo que períodos maiores aparecem para valores crescentes de α levando ao regime caótico. A rota para o caos seguida pelo sistema não é exatamente a de duplicação de período uma vez que ele sofre *split bifurcations* [86]: antes de dobrar o período do atrator, o sistema dobra o número de atratores estáveis. Por exemplo, uma órbita de período 2 se desdobra em duas de período 2 antes de se tornar uma órbita de período 4. A transição do duplo atrator para o de período dobrado ocorre quando o sistema entra em uma órbita superestável (aquela que contém o ponto crítico). Este comportamento é muito parecido ao encontrado no mapa cúbico $f(m) = \alpha m(1 - m^2)$ onde somente uma *split bifurcation* é permitida para a região periódica (um sistema com dois pontos críticos pode somente ter dois atratores) [86]. A figura 5.4 exemplifica o comportamento do sistema.

Como existem regiões onde mais de um atrator coexistem, é interessante obter alguma

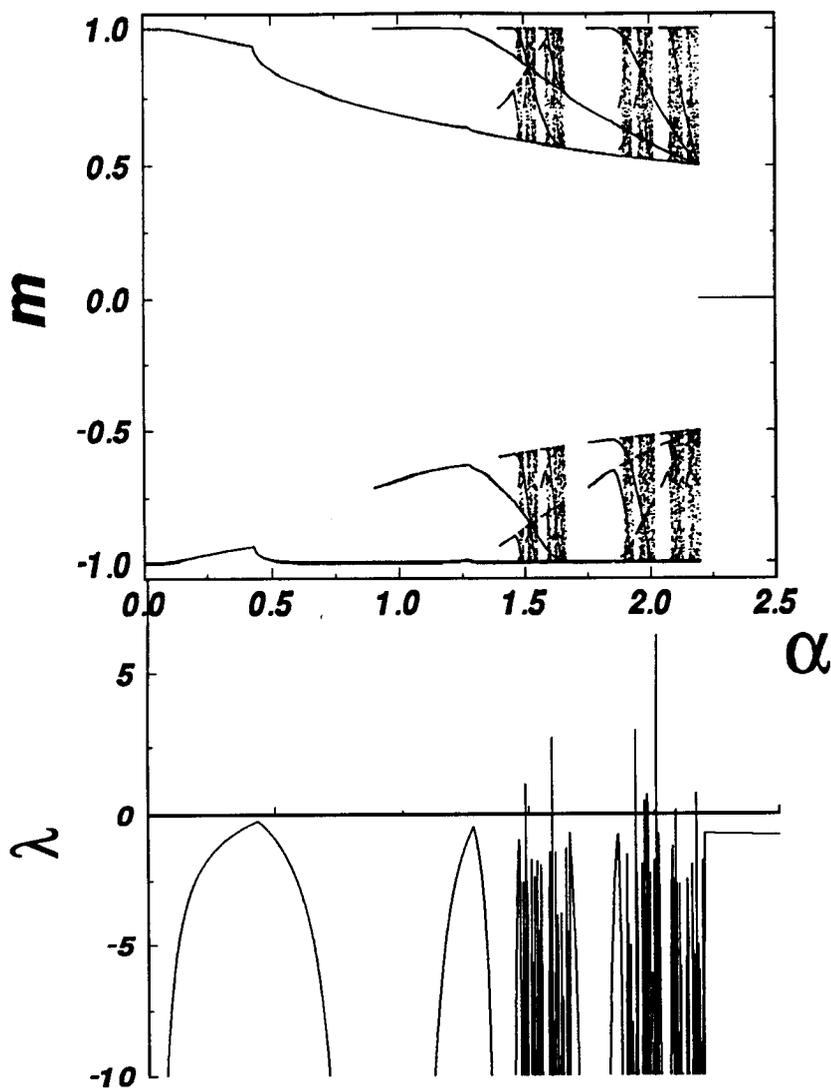


Figura 5.4: A superposição m e o respectivo expoente de Lyapunov versus α em $T = 0$ para $\varepsilon = 2$. Para grandes valores de α há somente a solução $m = 0$ enquanto que para valores menores o comportamento do sistema é complexo. A superposição inicial é $m_0 = 0.5$ e 500 (1000) pontos foram utilizados após um transiente de 1000 (5000) passos para a figura superior (inferior).

informação sobre a forma das bacias de atrações destes atratores, ou seja, aqueles valores de m que conduzem a um ou outro atrator. Na figura 5.5 mostramos a função $f(m)$ iterada várias vezes (e.g., 16) contra m para $\alpha = 0.6$ e $\varepsilon = 2$ e observamos que a função $f^{(16)}$ é formada por um conjunto de platôs em alguns valores fixos de m (neste caso, quatro valores), os quais são os quatro pontos dos dois ciclos de período dois que coexistem nesta região. Para um valor inicial de m perto de 0 os domínios são extremamente misturados, apesar de que para 16 iterações o sistema ainda não tenha atingido o comportamento assintótico porque o tempo de convergência é maior quanto menor for a superposição inicial. Portanto, as bacias de atração não assumem formas simples (embora não sejam fractais), e uma pequena variação na condição inicial pode levar tanto a um quanto a outro atrator.

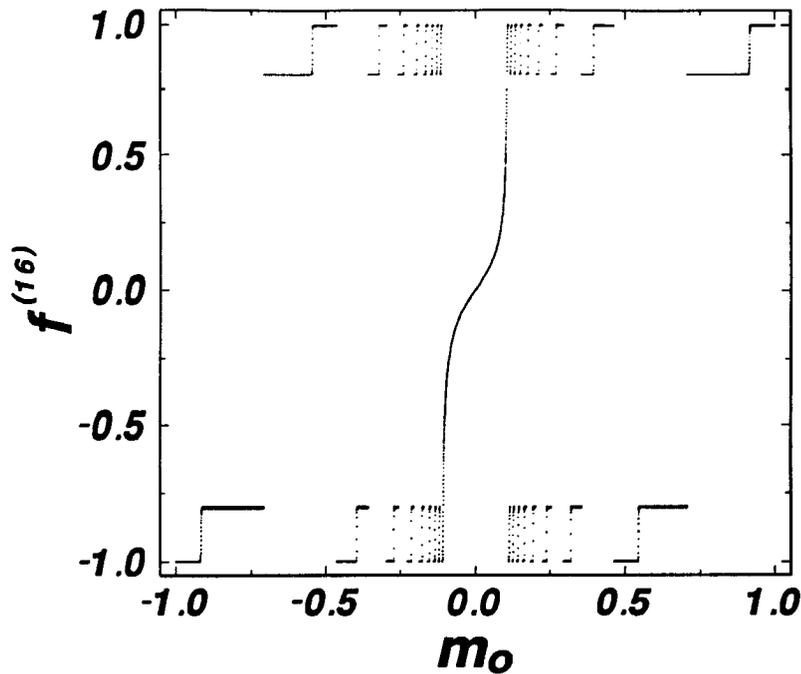


Figura 5.5: Mapa eq.(5.9) iterado 16 vezes mostrando a existência de vários platôs para $\epsilon = 2$ e $\alpha = 0.6$.

Para decidir se uma dada seqüência temporal é caótica ou não calculamos o expoente de Lyapunov através de sua definição:

$$\lambda \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \ln |f'(m_i)| \quad , \quad (5.18)$$

onde n é o comprimento da seqüência temporal utilizada. Este expoente nos fornece a taxa com que duas órbitas, inicialmente infinitesimalmente próximas, se separam com o tempo: no regime caótico temos $\lambda > 0$ ao passo que se o comportamento é periódico, $\lambda < 0$. Para $f(m)$ dada pela eq.(5.15), o expoente é

$$\lambda = \ln \frac{2}{\sqrt{\pi}} - \frac{1}{2\alpha} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \frac{m_i^2}{(1 - \epsilon m_i^2)^2} + \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \ln \frac{|1 + \epsilon m_i^2|}{\sqrt{2\alpha}(1 - \epsilon m_i^2)^2} \quad . \quad (5.19)$$

Quando $\epsilon = 0$ pode-se mostrar que $\lambda < 0$ para todo valor de $\alpha < \alpha_c$, ou seja, o modelo de Hopfield somente apresenta órbitas periódicas, como se sabe. É interessante notar que tanto no regime periódico quanto no caótico, o sistema usualmente passa muito próximo das memórias armazenadas, o que pode ser visto na fig. 5.4 para $\epsilon = 2$. Isto permite interpretar esta fase cíclica ou caótica como uma forma diversa de recuperação, na qual o sistema reconhece não por alcançar um ponto fixo, mas por estar vagando ao redor das memórias.

Para valores grandes de α o sistema entra em uma fase de não recuperação (o único ponto fixo estável da eq.(5.15) agora é $m = 0$) como pode ser observado no diagrama de fase fig. 5.3. Note também as similaridades entre o diagrama de fase em $T = 0$ para o modelo no caso diluído e no não diluído (figura 4.3): ambos têm uma fase de recuperação ao redor do valor ótimo de ε para todos os valores de α devido ao cancelamento do termo de ruído. A diferença é que no caso não diluído, ε_{opt} é uma função de α , $\varepsilon_{opt} = \varepsilon_{opt}(\alpha)$, enquanto que no diluído, é uma constante, $\varepsilon_{opt} = 1$. Além disso, no caso conectado não há uma fase caótica (somente pontos fixos) porque as conexões são simétricas permitindo a introdução de uma função de Lyapunov e o sistema sempre atinge um ponto fixo.

5.3 O Caso $T \neq 0$

Quando ruído térmico é introduzido no sistema, algumas modificações ocorrem no seu comportamento. A figura 5.6 mostra o diagrama T contra α para $\varepsilon \leq 1$ onde temos somente duas fases cujos pontos fixos são $m = 0$ (D) e $m \neq 0$ (R_1). Em particular, para valores de ε onde a transição é contínua ($\varepsilon < 0.266$), a linha T_c é independente de ε e satisfaz

$$\beta_c \int_{-\infty}^{+\infty} \mathcal{D}y \operatorname{sech}^2 [\beta_c y \sqrt{2\alpha}] = 1 \quad (5.20)$$

onde $\mathcal{D}y$ é dado por (5.10). Note que para $\varepsilon = 1$ temos sempre uma fase de recuperação a baixas temperaturas.

Quando $\varepsilon > 1$, o sistema apresenta uma fase desordenada para baixas temperaturas e valores grandes de α , como pode ser visto nas figuras 5.3 e 5.7 (para $\varepsilon = 2$), sendo que à medida que a temperatura aumenta, surge uma fase de recuperação em seguida a uma fase cíclica/caótica. Para pequenos valores de α , o sistema se encontra em uma fase de recuperação a baixas temperaturas e passa diretamente à fase desordenada à medida que T cresce. Existe uma certa reentrância perto da fronteira entre a fase cíclica/caótica e a fase desordenada onde, partindo da fase cíclica/caótica a baixas temperaturas (com α fixo), o sistema entra no regime desordenado para logo retornar à fase cíclica/caótica. O intervalo de temperatura onde existe a solução de recuperação torna-se cada vez mais estreito à medida que α cresce. Esta região existe ao redor da temperatura T^* onde o mapa eq.(5.9) apresenta a solução $m = \pm\varepsilon^{-1/2}$ para todos os valores de α :

$$\frac{1}{T^*} = \frac{\sqrt{\varepsilon}}{2} \ln \frac{\sqrt{\varepsilon} + 1}{\sqrt{\varepsilon} - 1} \quad (5.21)$$

No limite $\varepsilon \rightarrow 1^+$, $T^* = 0$ e $m = \pm 1$, como é mostrado na fig. 5.3 e, quando $\varepsilon \rightarrow \infty$, $T^* \rightarrow 1$ ($m = 0$). Note que, além das três fases que aparecem em $T = 0$, agora encontramos

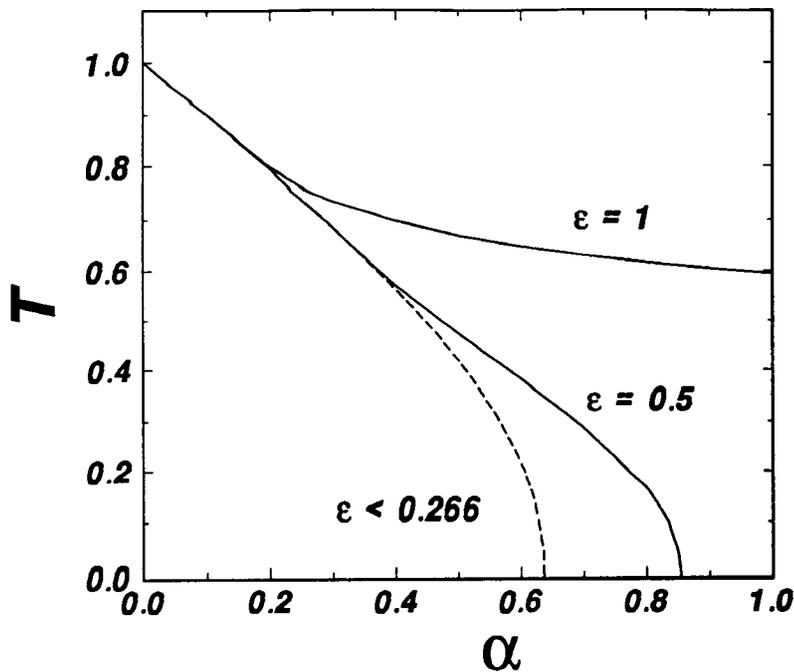


Figura 5.6: Diagrama de fase para valores de $\epsilon \leq 1$ mostrando a existência de somente duas fases: uma de recuperação ($T < T_c$) e uma com $m = 0$ ($T > T_c$). Para $\epsilon < 0.266$, T_c (linha pontilhada) é dada pela eq.(5.20).

também uma fase R_1 mesmo para $\epsilon > 1$. Aumentando o valor de α a região $m = 0$ domina exceto para pontos próximos a T^* . Na fig. 5.8 o comportamento de m para $\epsilon = 2$ e $\alpha = 2$ é mostrado como uma função de T com o correspondente expoente de Lyapunov. Note a duplicação de período reversa levando para uma região de recuperação à altas temperaturas: o sistema deixa a fase caótica e se torna periódico à medida que a temperatura aumenta até uma certa temperatura onde existe um ponto fixo e a informação armazenada pode ser recuperada (embora nem sempre com boa qualidade).

5.4 Recuperação de duas memórias

Também analisamos o caso no qual duas memórias possuem superposição macroscópica κ entre si, sendo a dinâmica governada pela eq.(5.13), iniciando sempre da condição inicial $m_1(0) = 1$ e $m_2(0) = \kappa$. Os atratores destas equações foram numericamente estudados e cinco diferentes fases foram distinguidas: uma fase de recuperação (R_1) a qual corresponde a um ponto fixo com $|m_1| > |m_2| > 0$, outra fase de recuperação (R_2) na qual o sistema oscila em um ciclo de período dois entre os estados com $|m_i| = \text{cte}$, $|m_1| > |m_2|$, uma fase mista (M) com $m_1 = m_2 \neq 0$, uma fase desordenada (D) com $m_1 = m_2 = 0$ e uma cíclica ou caótica (C), com m_1 e m_2 descrevendo órbitas caóticas ou com período maior ou igual

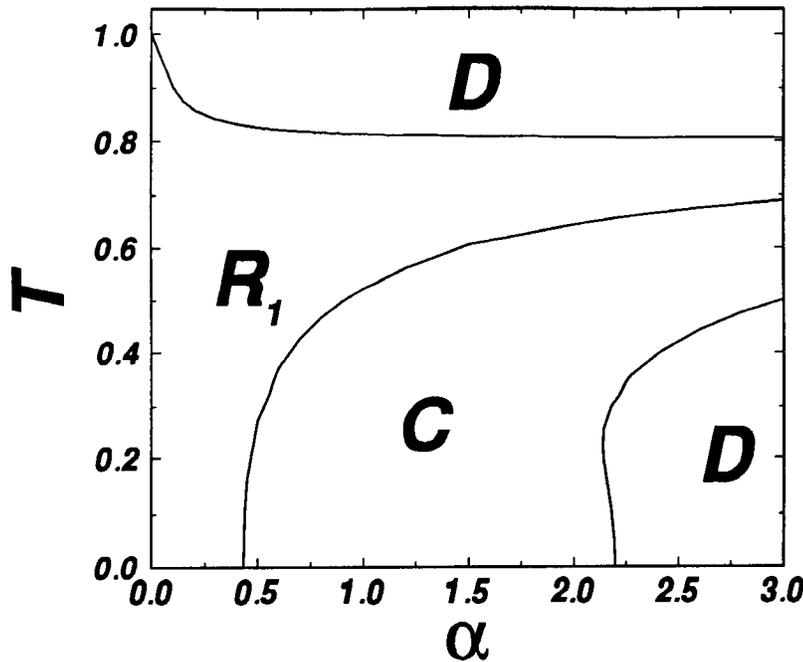


Figura 5.7: Diagrama de fase para $\epsilon = 2$. A fase de recuperação (R_1 para $T \neq 0$ e R_2 para $T = 0$) e a fase periódica ou caótica (C) são rodeadas pela fase $m = 0$ (D). Note que para $T^* \simeq 0.8$ o sistema possui a solução $m = \pm 1/\sqrt{2} \simeq \pm 0.707$ para todos valores de α .

a 2.

Para $\alpha = 0$, quando consideramos superposição macroscópica com somente um padrão, a equação da referência 29 é recuperada. Mas, na situação em que a superposição macroscópica é com duas memórias, não recuperamos as mesmas equações, pois a dependência em ϵ se mantém. Surpreendentemente, isto se mantém válido mesmo quando $\kappa = 0$: o sistema pode apresentar um regime cíclico ou caótico, ausente se procuramos por uma solução de recuperação. Na figura 5.9 apresentamos o diagrama de fase T contra ϵ para $\kappa = 0.2$ e $\alpha = 0$. Para $T = 0$ o sistema apresenta somente duas fases: a de recuperação para $|\epsilon| < \kappa^{-1}$ (R) e a cíclica e caótica (C) para $|\epsilon| > \kappa^{-1}$, enquanto que a fase mista (M) aparece para uma temperatura que depende do valor de κ . É importante enfatizar que a fronteira entre a fase de recuperação e a cíclica ou caótica para valores positivos de ϵ mostra o primeiro aparecimento de órbitas cíclicas (bifurcações) pois dentro da fase C ainda existem ilhas de recuperação. Além disso, na fronteira entre a fase mista e a desordenada, há um ponto, ϵ^* , acima do qual a transição é contínua e abaixo é descontínua. Por exemplo, para $\kappa = 0.2$, $\epsilon^* \simeq -0.94$. Pode-se mostrar que a linha de transição no caso contínuo é dada por $T_c = 1 + \kappa$.

O diagrama α versus ϵ , mostrado na fig. 5.10 para $T = 0$, é qualitativamente similar

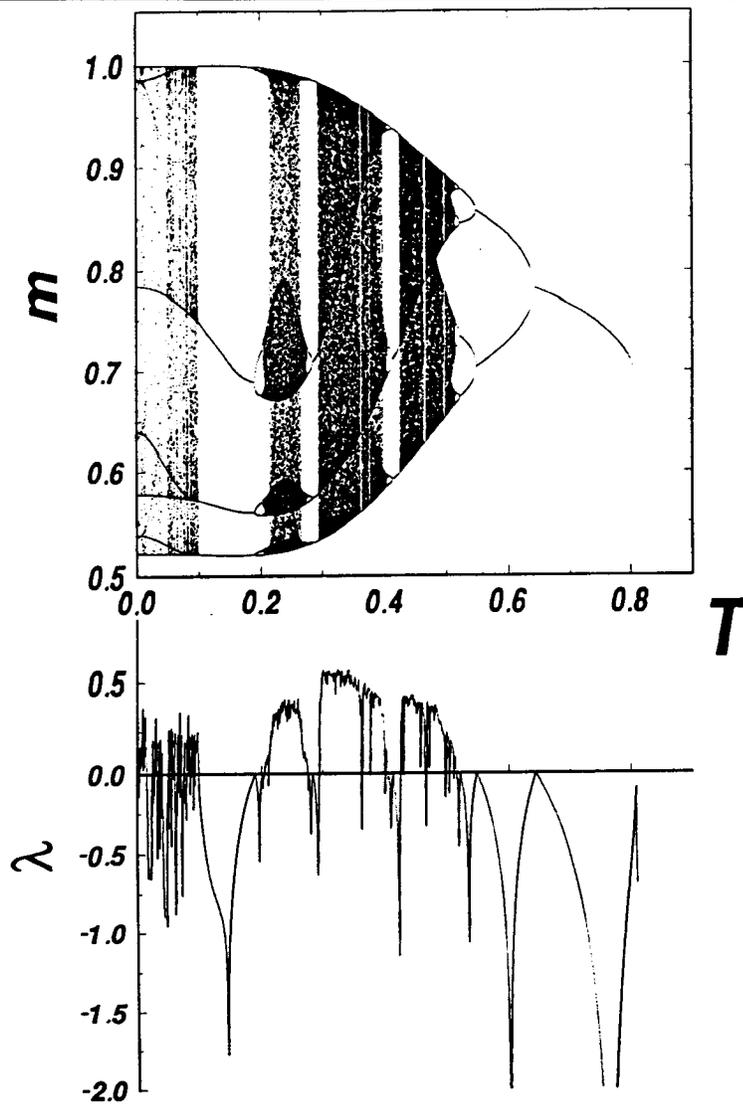


Figura 5.8: A superposição m contra T para $\varepsilon = 2$ e $\alpha = 2$ com o correspondente expoente de Lyapunov. O sistema somente tem um regime de recuperação a altas temperaturas.

àquele onde há somente superposição com uma memória (fig. 5.3). Pode-se também notar um valor de ε onde a capacidade diverge que aqui é dado por

$$\varepsilon_{opt} = \frac{1}{1 + \kappa^2} \quad . \quad (5.22)$$

Além disso, a fronteira entre a fase desordenada e a mista é

$$\alpha = \frac{2}{\pi} (1 + \kappa)^2 \quad . \quad (5.23)$$

É interessante observar que mesmo para $\kappa = 0$, ainda existe a fase mista, o que implica que podem existir outras soluções envolvendo misturas de um número maior de memórias. Em $\alpha = 0$, mesmo quando $\kappa = 0$, ainda existe uma dependência em ε e o sistema pode apresentar um regime cíclico ou caótico, ausente se procuramos por uma solução de recuperação. Portanto, mesmo quando as fontes de ruído não estão presentes (α e T), o sistema pode

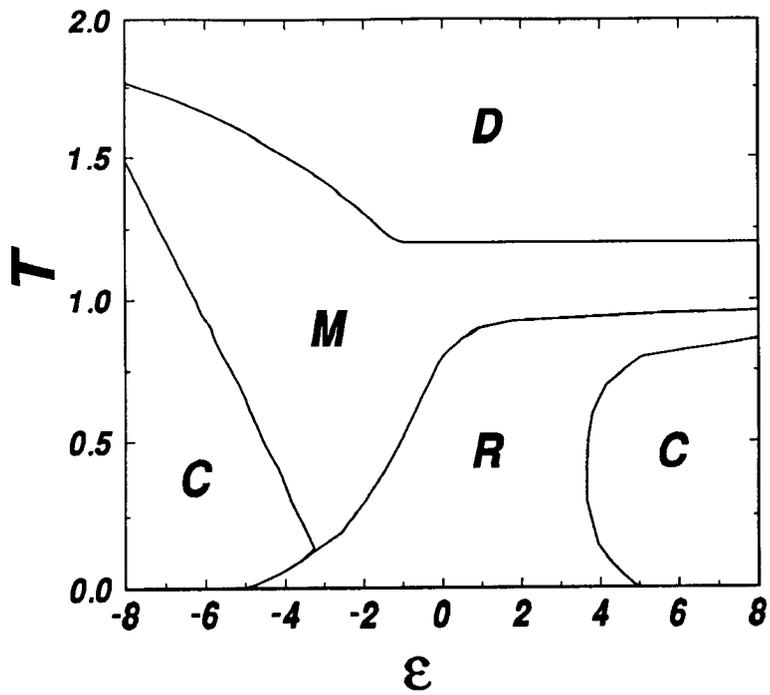


Figura 5.9: Diagrama de fase para $\alpha = 0$ e $\kappa = 0.2$.

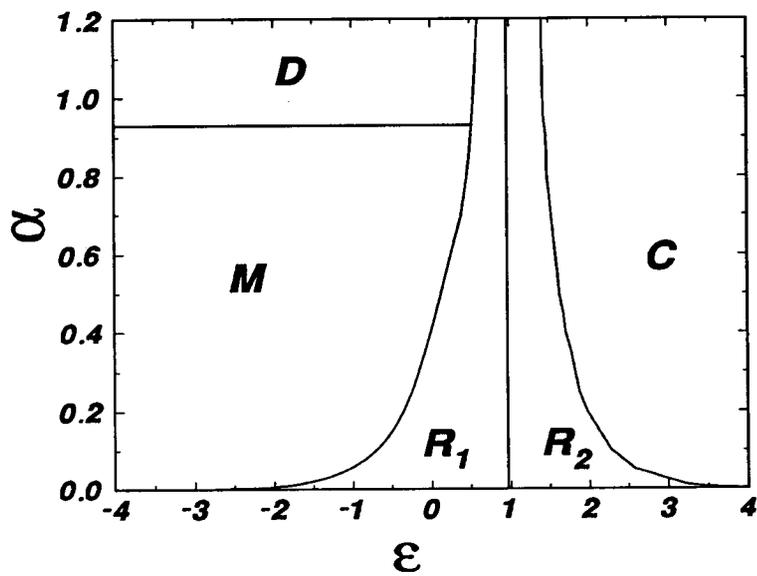


Figura 5.10: Diagrama de fase para $T = 0$ e $\kappa = 0.2$. Abaixo da fase R_2 , em $\alpha = 0$, o sistema está em uma fase R_1 , e abaixo de M , em uma fase C . Isto compatibiliza este diagrama com o anterior (fig. 5.9).

apresentar um comportamento mais complicado do que pontos fixos. Na linha $\alpha = 0$ no diagrama da figura 5.10, existe a solução de recuperação R_1 no intervalo $[-\kappa^{-1}, \kappa^{-1}]$.

5.5 Conclusão

Neste capítulo apresentamos a solução da dinâmica do modelo truncado diluído, a qual apresenta um rico comportamento dependendo dos parâmetros α, ε e T . Em analogia com o correspondente modelo não diluído, existe um valor ótimo de ε para o qual o sistema sempre recupera as memórias armazenadas. É importante salientar que para alguns valores de ε a fase de recuperação se encontra em regiões de alta temperatura: à medida que a temperatura baixa, o sistema passa de uma fase desordenada (D) para uma ordenada (R_1) e, após passar através de uma região onde apresenta órbitas periódicas e caóticas, ele novamente entra em uma região desordenada. Neste caso, a presença de ruído térmico melhora a capacidade de recuperação do sistema.

O comportamento periódico (com períodos maiores do que um) ou caótico somente está presente para $\varepsilon > 1$. Aparentemente, a razão para que isso ocorra é que abaixo de $\varepsilon = 1$ o termo de segunda ordem (Hopfield) sobrepassa o de quarta ordem, ao passo que para $\varepsilon > 1$ ocorre o contrário. Neste caso, devido à assimetria nas conexões (a presença de μ e ν na regra de aprendizado), ciclos são introduzidos. Isto não ocorre quando todas as sinapses estão presentes porque podemos simetrizá-las, o que permite a introdução de uma energia para o sistema, a qual sempre decresce ao longo da dinâmica. Coerentemente, se $\varepsilon < 1$ o mapa eq.(5.9) é inversível e contínuo, o que não permite comportamento caótico para sistemas dinâmicos unidimensionais [87].

Um comportamento interessante do sistema, talvez mais próximo da realidade, é que existem regiões no espaço dos parâmetros onde a recuperação não se dá atingindo um ponto fixo estável, mas sim pelo fato do sistema estar restrito a um atrator em cuja vizinhança está a memória.

Capítulo 6

Conclusões

Os métodos da mecânica estatística são extremamente úteis para o estudo do comportamento cooperativo existente em uma grande gama de sistemas compostos de muitas unidades simples. Nesta tese utilizamos tais técnicas para avaliar os efeitos da introdução de interações de ordem maior do que dois entre neurônios, obtendo informações sobre as propriedades destes sistemas para conjuntos típicos de padrões. Nos capítulos iniciais introduzimos o tema e brevemente revisamos os modelos existentes contendo tais interações, salientando os principais resultados obtidos. Finalmente, nos capítulos 4 e 5, os quais constituem o corpo principal do trabalho, estudamos o efeito de uma nova regra de aprendizado (o modelo truncado). Basicamente existem três modelos para redes atratoras com multi-interações, sendo que a distinção entre eles é feita através das suas regras de aprendizagem: o modelo de Hopfield generalizado [35, 58, 62–64, 66, 67], a regra da pseudo-inversa generalizada [78, 79] e um modelo recentemente introduzido [8–10] que incorpora várias ordens de interações. O modelo truncado é um caso particular deste último onde apenas algumas das ordens de interação são mantidas (aqui tratamos somente de termos de segunda e quarta ordem).

O modelo de Hopfield generalizado pode ser formado por uma ou várias ordens de interação simultaneamente. No primeiro caso, para interações de ordem ℓ , o número de memórias que o sistema é capaz de armazenar escala com $N^{\ell-1}$, ao passo que se tivermos vários termos, $P \sim N^{\ell_{min}-1}$, onde ℓ_{min} é a menor ordem de interação. Em ambos os casos simulações numéricas podem ser feitas e comparadas com os resultados analíticos, com resultados bastante bons. A pequena discrepância que existe é provavelmente devida à instabilidade da solução com simetria de réplicas. Os diagramas de fase são obtidos [35, 62, 63] (também para o caso em que $\ell \rightarrow \infty$ [62]) e a evolução dinâmica é calculada para o caso de extrema diluição [66, 67], ou seja, quando somente uma pequena fração das sinapses é mantida. Já para a generalização da pseudo-inversa, poucos resultados foram

obtidos, sendo apenas que a capacidade de armazenamento é da mesma ordem que a do modelo de Hopfield generalizado [78], embora a recuperação seja perfeita.

O terceiro modelo, ao longo do texto também chamado de modelo completo, foi criado tendo-se em mente um sistema onde as memórias nunca deixassem de ser mínimos absolutos de energia, mesmo quando a carga da rede fosse muito grande, e foi estudado em dois limites específicos: quando temos um conjunto de P padrões armazenados (caso OPS) e quando, além destes, as correspondentes anti-memórias são também guardadas (caso PAS). O comportamento do modelo, na configuração PAS, como memória associativa é bastante eficiente: possui alta capacidade de memória e o tempo necessário para recuperá-las é muito pequeno ao passo que a configuração OPS apresenta propriedades de associatividade somente no limite em que a rede não está muito carregada.

Já o modelo truncado é obtido do modelo completo mantendo-se apenas algumas ordens de interação na energia (no caso estudado, segunda e quarta ordem), o que pode ser tomado como o modelo de Hopfield mais termos de correção de ordem mais alta. Ao contrário de outros modelos, aqui não podemos ter somente os termos de ordem mais alta porque estes envolvem misturas de várias memórias e nenhum padrão pode ser recuperado somente com eles. Portanto, a regra de aprendizagem contém termos de misturas entre as memórias, o que faz com que o comportamento seja muito diverso do encontrado para os modelos anteriores. Por exemplo, os auto-acoplamentos são importantes na versão não diluída do modelo: as sinapses J_{iikl} e J_{ijkk} dão origem a contribuições para a energia da mesma ordem que as sinapses J_{ij} , o que não acontece no modelo de Hopfield generalizado. No caso com diluição, estas interações são desprezíveis no limite termodinâmico. As principais propriedades e conclusões para o modelo são:

- A dependência em ϵ , o peso relativo do termo de quarta ordem, ao contrário do modelo de Hopfield generalizado, não é apenas quantitativa mas qualitativa. Por exemplo, o parâmetro que mede a qualidade de recuperação, m , não é monotonicamente decrescente com α , mas pode apresentar até duas regiões distintas de recuperação, separadas por uma lacuna. A lacuna existe para valores de $\epsilon < \epsilon_c \simeq 0.3587$ e, quando $\epsilon \simeq \epsilon_c$, o comportamento é muito complexo na vizinhança da lacuna sendo que os efeitos de quebra de simetria de réplicas (RSB) devem ser levados em conta para decidir que tipo de pontos críticos realmente existem.
- Devido à existência de um pico na segunda região de recuperação é possível introduzir uma regra de aprendizado ótima, variável no tempo, escolhendo o peso ϵ tal que $\epsilon_{opt} = (1 + \alpha)^{-1}$, o qual pode ser reintroduzido novamente na função energia.

Quanto maior for a quantidade de informação aprendida, menor é o peso dos termos de quarta ordem e, como aparentemente ocorre nos sistemas nervosos reais, o aprendizado torna-se mais fácil e tais conexões de alta ordem podem ser encaradas apenas como correções às muito mais abundantes sinapses de segunda ordem. Aqui aparece uma das diferenças fundamentais entre este modelo e as generalizações do modelo de Hopfield: lá, quanto maior fosse o peso do termo de ordem mais alta, mais informação poderia ser armazenada, ao passo que aqui, com ε_{opt} , este peso torna-se cada vez menor à medida que mais e mais informação é acrescentada ao sistema.

- A solução com simetria de réplicas é instável em $T = 0$, uma vez que a entropia S_0 é sempre negativa, com exceção de alguns valores de α onde $S_0 = 0$: $\alpha = 0$ e $\alpha = (1 - \varepsilon_{opt})/\varepsilon_{opt}$. Para estes valores a quebra de simetria de réplica não trará nenhuma modificação, o que explica o fato das simulações numéricas concordarem tão bem com os resultados analíticos nestes pontos.
- A linha T_g , para $\varepsilon \neq 0$, vai a zero em α_g em analogia ao que ocorre no modelo da pseudo-inversa [56]. Como consequência, para $\alpha > \alpha_g$, o efeito dos estados espúrios na dinâmica é severamente reduzido e a quantidade \mathcal{M} foi introduzida para tentar detectar tal efeito. Os resultados corroboram a predição analítica, embora a concordância seja apenas qualitativa. É possível que a diferença quantitativa seja devida à instabilidade da solução com simetria de réplicas a baixas temperaturas. Outra possibilidade é que as bacias de atração dos estados espúrios sejam muito pequenas comparadas às das memórias de modo que os estados iniciais aleatórios teriam pouca probabilidade de atingir tais estados.
- As bacias de atração são grandes e independentes de α e, para superposições iniciais acima de um valor crítico, o qual vale aproximadamente 0.1 para todo $\alpha < \alpha_c$, as memórias são recuperadas. Por outro lado, a qualidade da recuperação diminui com α uma vez que a transição é contínua.
- O tempo médio de convergência $\langle T \rangle$, para valores não muito pequenos de α , não depende do valor inicial da superposição, o que é uma consequência geral de termos uma transição contínua. Outra consequência é que existe uma dispersão considerável ao redor destes valores médios, mas cuja causa não são os estados espúrios que possam existir ao redor das memórias. Em resumo, para transições contínuas, a existência de estados espúrios metaestáveis não está assegurada pela presença de dispersões não nulas ao redor dos tempos médios de convergência. Para obtermos tal informação

seria preciso criar estados iniciais com uma superposição efetiva com o estado que se encontra no fundo da bacia de atração, e não com a memória originalmente armazenada.

Quando diluímos o sistema, isto é, quando as conexões são severa e assimetricamente cortadas, é possível resolver exatamente a dinâmica [29]. Os principais resultados obtidos neste limite são [85]:

- Em $T = 0$, quando $\varepsilon > 1$, o mapa é descontínuo e não inversível. Para $\varepsilon < 0.266$, existe somente um ponto fixo estável para a dinâmica de $m(t)$, o qual diminui continuamente até 0 em $\alpha_c = 2/\pi$. Para $0.266 < \varepsilon < 1$ a transição para a fase desordenada é descontínua e $\alpha_c > 2/\pi$, crescendo com ε . O reconhecimento, para um valor fixo de α , é tanto melhor quanto mais próximo ε estiver de 1.
- Quando $\varepsilon = 1$, sempre existe a solução de recuperação, analogamente ao que ocorre no caso não diluído. A capacidade de armazenamento diverge e a recuperação é perfeita para todos estados iniciais com superposição não nula.
- Quando $\varepsilon > 1$, o sistema pode ser periódico ou caótico, mas como o sistema usualmente passa muito próximo das memórias, isto pode ser interpretado como uma nova forma de recuperação, na qual o reconhecimento ocorre não quando um ponto fixo é alcançado, mas por que o sistema está vagando ao redor das memórias, restrito a um atrator em sua vizinhança. Para ε fixo e baixos valores de α , o sistema encontra-se numa fase de recuperação e à medida que α cresce, o sistema segue uma rota para o caos ligeiramente diferente da rota de duplicação de período [86] (antes de dobrar o período do atrator, o sistema dobra o número de atratores estáveis).
- Existem várias semelhanças entre os diagrama de fase em $T = 0$ para o modelo no caso diluído e no não diluído (figuras 4.3 e 5.3): ambos têm uma fase de recuperação ao redor do valor ótimo de ε para todos os valores de α devido ao cancelamento do termo de ruído. A diferença é que no caso não diluído, ε_{opt} é uma função de α , $\varepsilon_{opt} = \varepsilon_{opt}(\alpha)$, enquanto que no diluído, é uma constante, $\varepsilon_{opt} = 1$. Além disso, no caso conectado não há uma fase caótica (somente pontos fixos) porque a simetria das conexões permite a introdução de uma função de Lyapunov e o sistema sempre atinge um ponto fixo.
- Quando ruído térmico é introduzido, algumas novas características surgem, embora para $\varepsilon \leq 1$ somente existam duas fases cujos pontos fixos são $m = 0$ e $m \neq 0$, tal

qual no modelo de Hopfield. Quando $\varepsilon \geq 1$, existe uma temperatura T^* ao redor da qual sempre existe uma fase de recuperação. Para $\varepsilon = 1$, $T^* = 0$, ou seja, sempre há uma fase de recuperação para baixas temperaturas. O sistema apresenta uma fase desordenada para baixas temperaturas e valores grandes de α , sendo que à medida que a temperatura aumenta, surge uma fase de recuperação em seguida a uma fase cíclica/caótica. Para pequenos valores de α , o sistema se encontra em uma fase de recuperação a baixas temperaturas e passa diretamente à fase desordenada à medida que T cresce. O intervalo de temperatura onde existe a solução de recuperação torna-se cada vez mais estreito à medida que α cresce e, dependendo do caso, a presença de ruído térmico melhora as capacidades de recuperação do sistema.

- No caso em que duas memórias possuem superposição macroscópica κ entre si, existe além das anteriores uma fase mista ($m_1 = m_2 \neq 0$), a qual existe mesmo para $\kappa = 0$, o que implica que podem existir outras soluções envolvendo misturas de um número maior de memórias. Em $\alpha = 0$, mesmo quando $\kappa = 0$, ainda existe uma dependência em ε e o sistema pode apresentar um regime cíclico ou caótico, ausente se procuramos por uma solução de recuperação. Portanto, mesmo quando as fontes de ruído não estão presentes (α e T), o sistema pode apresentar um comportamento mais complicado do que pontos fixos.

Tendo em vista os modelos estudados nesta tese e algumas das técnicas aqui introduzidas, é interessante traçar algumas perspectivas para trabalhos que naturalmente seguem os realizados aqui. Entre os pontos a serem explorados, destacamos os seguintes:

- Obtenção de \mathcal{M} para o modelo de Hopfield utilizando redes grandes para observar se o comportamento linear persiste e obter os coeficientes, assim como suas dependências no tamanho da rede. Pode-se ainda empregar um processo de eliminação dos estados espúrios [82] e observar como isso afeta \mathcal{M} . O processo deve ser rápido no início pois o algoritmo empregado na ref. 82 elimina os estados espúrios com as maiores bacias de atração. Pode-se ainda comparar estes resultados com os obtidos para outros modelos e estudar o efeito da temperatura.
- Para o modelo completo (capítulo 3) com atividade diferente de 0.5, existem resultados de simulação mostrando que um processo de desaprendizagem análogo ao estudado nas refs. 81 e 82 pode diminuir o papel desempenhado pelo estado central. Como isto pode ser implementado diretamente na função energia, este modelo pode ser resolvido analiticamente utilizando-se o método das réplicas [91]. Neste caso, até

onde temos conhecimento, seria o primeiro caso de solução analítica para este tipo de comportamento. Pode-se também tentar eliminar o estado ferromagnético no modelo de Hopfield.

- A generalização da regra da pseudo-inversa delineada no capítulo 2 não foi ainda estudada analiticamente, sendo interessante obter seu diagrama de fases e informações sobre a capacidade de armazenamento, além, é claro, do comportamento quando a ordem de interação é grande [79]. A dinâmica pode também ser estudada com os métodos da ref. 29 (apêndice D).
- O modelo completo pode ainda ser generalizado:

$$E = N \prod_{\mu=1}^P (1 - m_{\mu}^k) \quad . \quad (6.1)$$

Quando $k = 1$ temos o caso OPS e quando $k = 2$ o PAS é obtido. Ao expandirmos a equação acima, o primeiro termo é o modelo de Hopfield generalizado. Este sistema provavelmente apresenta três regimes qualitativamente distintos: quando $k = 1, 2$ e > 2 . No limite em que $k \rightarrow \infty$, provavelmente o modelo torna-se equivalente ao modelo de energias randômicas [73–76].

- Para o modelo truncado, o valor ótimo de ϵ somente é válido para $T = 0$, e o modelo provavelmente estará próximo do comportamento ótimo para baixas temperaturas. Seria interessante verificar o que ocorre para valores maiores da temperatura. O efeito da inclusão do termo de sexta ordem também pode ser considerado. Além de uma aumento na capacidade de armazenamento, provavelmente teremos a presença de mais uma lacuna e pelo menos um par de pesos ($\epsilon_4^{opt}, \epsilon_6^{opt}$) que aumentariam a capacidade do modelo. Para um número pequeno de termos na expansão talvez o problema possa ser encarado como de otimização e técnicas de *simulated annealing* possam ser empregadas para achar os valores ótimos dos pesos.
- Como a versão diluída do modelo truncado apresenta um comportamento bastante diverso da versão sem diluição, seria interessante estudar a termodinâmica do caso com diluição simétrica [88], tentando observar como ocorre a transição entre um comportamento e outro. Além disso, a presença da fase mista é uma indicação de que o sistema pode apresentar propriedades interessantes de generalização.

Em resumo, vimos os efeitos da introdução de multi-sinapses entre os neurônios, os quais, dependendo do modelo, podem ser bastante interessantes e complexos. Em parti-

cular, o modelo truncado satisfaz o requisito biológico de que sinapses de ordem mais alta devem ser tomadas como correções.

Por fim, um último comentário sobre o papel do caos em sistemas nervosos. Foi proposto [89] que a capacidade de um grande grupo de neurônios abruptamente alterar seu padrão de atividade graças a pequenos sinais pode ser responsável pela flexibilidade de sistemas nervosos reais em criar novos padrões ao interagir com o mundo externo, flexibilidade esta que pode permear os processos criativos e de aprendizado.

Apêndice A

O Método Das Réplicas

Vamos aqui mostrar em detalhe os cálculos relativos a utilização do método das réplicas para os modelos discutidos nos capítulos anteriores.

A.1 O Método

O método das réplicas tem sido utilizado com grande sucesso em uma enorme gama de problemas envolvendo desordem congelada (*quenched disorder*), isto é, elementos aleatórios que não variam no tempo (ou pelo menos variam numa escala de tempo muito menor do que as variáveis dinâmicas do problema). Como exemplo podemos citar redes de neurônios, vidros de spin, modelos de evolução, etc. No caso das redes de neurônios, as variáveis aleatórias são os padrões armazenados e sobre elas é que devemos fazer as médias envolvidas. O objetivo é calcular a energia livre do sistema,

$$f = \frac{F}{N} = -\frac{1}{\beta} \ln Z \quad (\text{A.1})$$

onde Z é a função partição do sistema:

$$Z = \sum_{\{S_i\}} e^{-\beta E[S]} \quad (\text{A.2})$$

e $E[S]$ é a energia do sistema para uma dada configuração S . As quantidades mensuráveis são obtidas a partir da energia livre, o que torna o problema bastante complicado já que a média deve ser efetuada sobre $\ln Z$ e $\langle\langle \ln Z \rangle\rangle \neq \ln \langle\langle Z \rangle\rangle$. Para contornar o problema utilizamos a seguinte relação:

$$\langle\langle \ln Z \rangle\rangle = \lim_{n \rightarrow 0} \frac{\langle\langle Z^n \rangle\rangle - 1}{n} \quad , \quad (\text{A.3})$$

onde Z^n pode ser interpretado como a função partição de n réplicas do sistema original. Como veremos nas seções seguintes, $\langle\langle Z^n \rangle\rangle$ é razoavelmente simples de calcular, o que permite obter a energia livre do sistema. Existem alguns problemas, como o limite para $n \rightarrow 0$, o qual deve ser tomado antes do limite termodinâmico, mas na prática é feito o contrário. A justificativa será *a posteriori*, quando verificarmos os resultados assim obtidos com os de simulação numérica.

A.2 A Média Sobre As Memórias Microscópicas

A média é feita sobre as memórias microscópicas, ou seja, aquelas que não se condensam (as memórias macroscópicas se auto-mediam). Vamos encontrar, nos modelos aqui tratados, a seguinte média:

$$\Phi_k \equiv \left\langle\left\langle \exp \left\{ \frac{\beta k!}{2N^{k-1}} \sum_{i_1 < \dots < i_k} \sum_a \xi_{i_1} \dots \xi_{i_k} S_{i_1}^a \dots S_{i_k}^a \right\} \right\rangle\right\rangle . \quad (\text{A.4})$$

É importante ressaltar que o resultado é extremamente diferente se $k = 2$ ou $k > 2$. Vamos tratar separadamente estes dois casos.

Quando $k = 2$, temos

$$\Phi_2 \equiv \left\langle\left\langle \exp \left\{ \frac{\beta}{N} \sum_{i,j} \sum_a \xi_i \xi_j S_i^a S_j^a \right\} \right\rangle\right\rangle . \quad (\text{A.5})$$

Definindo

$$\Lambda_{ij} \equiv \frac{1}{N} \sum_a S_i^a S_j^a , \quad (\text{A.6})$$

podemos escrever:

$$\Phi_2 = \left\langle\left\langle \exp \left\{ \beta \sum_{i,j} \Lambda_{ij} \xi_i \xi_j \right\} \right\rangle\right\rangle . \quad (\text{A.7})$$

Expandindo a equação acima em cumulantes obtemos

$$\ln \Phi_2 = \beta \sum_{i,j} \Lambda_{ij} \langle\langle \xi_i \xi_j \rangle\rangle + \frac{1}{2!} \left[\left\langle\left\langle \left(\beta \sum_{i,j} \Lambda_{ij} \xi_i \xi_j \right)^2 \right\rangle\right\rangle - \left\langle\left\langle \beta \sum_{i,j} \Lambda_{ij} \xi_i \xi_j \right\rangle\right\rangle^2 \right] + \dots , \quad (\text{A.8})$$

onde $\langle\langle \xi_i \xi_j \rangle\rangle = \delta_{ij}$. Logo, o primeiro cumulante fica:

$$C_1 = \beta \sum_i \Lambda_{ii} . \quad (\text{A.9})$$

Note que como $\Lambda_{ij} \sim \mathcal{O}(N^{-1})$, $C_1 \sim \mathcal{O}(1)$. Já o segundo cumulante fica:

$$\begin{aligned}
 C_2 &= \frac{\beta^2}{2} \sum_{i,j,k,l} \Lambda_{ij} \Lambda_{kl} \langle \xi_i \xi_j \xi_k \xi_l \rangle - \frac{\beta^2}{2} \sum_{i,j} \Lambda_{ii} \Lambda_{jj} \\
 &= \frac{\beta^2}{2} \sum_{i,j,k,l} \Lambda_{ij} \Lambda_{kl} (\delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) - \frac{\beta^2}{2} \sum_{i,j} \Lambda_{ii} \Lambda_{jj} \\
 &= \frac{\beta^2}{2} \left(\sum_{i,j} \Lambda_{ii} \Lambda_{jj} + \sum_{i,j} \Lambda_{ij} \Lambda_{ji} + \sum_{i,j} \Lambda_{ij} \Lambda_{ji} - \sum_{i,j} \Lambda_{ii} \Lambda_{jj} \right) \\
 &= \beta^2 \sum_{i,j} \Lambda_{ij} \Lambda_{ji} \quad .
 \end{aligned} \tag{A.10}$$

Voltando então para a expansão:

$$\begin{aligned}
 \ln \Phi_2 &= \beta \sum_i \Lambda_{ii} + \frac{\beta^2}{2} \sum_{i,j} \Lambda_{ij} \Lambda_{ji} + \frac{\beta^3}{3} \sum_{i,j,k} \Lambda_{ij} \Lambda_{jk} \Lambda_{ki} + \dots \\
 &= -\text{Tr}_{\{i\}} \ln(\mathbf{I} - \beta \mathbf{A})
 \end{aligned} \tag{A.11}$$

onde \mathbf{A} é a matriz formada por Λ_{ij} e \mathbf{I} a matriz identidade. Voltando um pouco atrás e utilizando a eq.(A.6), escrevemos

$$\begin{aligned}
 \ln \Phi_2 &= \beta \sum_a Q_{aa} + \frac{\beta^2}{2} \sum_{a,b} Q_{ab} Q_{ba} + \frac{\beta^3}{3} \sum_{a,b,c} Q_{ab} Q_{bc} Q_{ca} + \dots \\
 &= -\text{Tr}_{\{a\}} \ln(\mathbf{I} - \beta \mathbf{Q})
 \end{aligned} \tag{A.12}$$

onde

$$Q_{ab} = \frac{1}{N} \sum_i S_i^a S_i^b \quad . \tag{A.13}$$

Ou seja,

$$\Phi_2 = \exp \left[-\text{Tr}_{\{a\}} \ln(\mathbf{I} - \beta \mathbf{Q}) \right] \quad . \tag{A.14}$$

Este resultado vale tanto para o modelo de Hopfield quanto para sua generalização (com $\varepsilon_2 \neq 0$). Para o modelo truncado e para o completo é preciso fazer uma pequena modificação na definição de Λ_{ij} :

$$\Lambda_{ij} = \frac{1}{N} \sum_a S_i^a S_j^a (1 - \varepsilon y_a) \tag{A.15}$$

e

$$\Lambda_{ij} = \frac{1}{N} \sum_a S_i^a S_j^a \tilde{S}_a \quad , \tag{A.16}$$

respectivamente, onde y_a e \tilde{S}_a são definidos nos capítulos 4 e 3. Nestes casos não conseguimos somar a série final, como em (A.12), e precisamos mantê-la explicitamente:

$$\Phi_2 = \exp \left[\sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} Q_{a_2 a_3} \dots Q_{a_r a_1} (1 - \varepsilon y_{a_1}) \dots (1 - \varepsilon y_{a_r}) \right] \tag{A.17}$$

para o modelo truncado e

$$\Phi_2 = \exp \left[\sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} Q_{a_2 a_3} \dots Q_{a_r a_1} \tilde{S}_{a_1} \dots \tilde{S}_{a_r} \right], \quad (\text{A.18})$$

para o completo.

Quando $k > 2$, nem todos os termos da série vão contribuir, restando somente o termo de segunda ordem. Definimos $\Lambda_{i_1 i_2 \dots i_k}$ a partir de (A.4):

$$\Lambda_{i_1 \dots i_k} = \frac{k!}{2N^{k-1}} \sum_a S_{i_1}^a \dots S_{i_k}^a \quad (\text{A.19})$$

e expandimos em cumulantes. Como os índices são todos diferentes entre si, todos os termos envolvendo cumulantes de ordem ímpar são nulos, restando apenas os pares. O cumulante de segunda ordem é:

$$\begin{aligned} C_2 &= \frac{1}{2} \left\langle \left\langle \left(\beta \sum_{i_1 < \dots < i_k} \Lambda_{i_1 \dots i_k} \xi_{i_1} \dots \xi_{i_k} \right)^2 \right\rangle \right\rangle \\ &= \frac{\beta^2}{2} \left\langle \left\langle \sum_{i_1 < \dots < i_k} \sum_{j_1 < \dots < j_k} \Lambda_{i_1 \dots i_k} \Lambda_{j_1 \dots j_k} \xi_{i_1} \dots \xi_{i_k} \xi_{j_1} \dots \xi_{j_k} \right\rangle \right\rangle \\ &= \frac{\beta^2}{2} \sum_{i_1 < \dots < i_k} \Lambda_{i_1 \dots i_k}^2 \\ &= \frac{\beta^2 k!}{8N^{k-2}} \sum_{a,b} Q_{ab}^k. \end{aligned} \quad (\text{A.20})$$

Os termos pares do cumulante de quarta ordem, $\langle x^4 \rangle - 3\langle x^2 \rangle^2$, darão origem a contribuições cuja ordem é $\mathcal{O}(N^{2(2-k)})$. Portanto, a menos que $k = 2$, estes termos não contribuirão e

$$\Phi_k = \exp \left[\frac{\alpha \beta k! N}{8} \sum_{a,b} Q_{ab}^k \right], \quad k \neq 2. \quad (\text{A.21})$$

A.3 O Modelo Truncado

A energia do sistema replicado pode ser escrita como

$$E = -\frac{N}{2} \sum_{\mu,a} (m_\mu^a)^2 - \frac{N\varepsilon}{4} \sum_{\mu,a} (m_\mu^a)^4 + \frac{N\varepsilon}{4} \sum_a \left[\sum_\mu (m_\mu^a)^2 \right]^2. \quad (\text{A.22})$$

A função partição, portanto, é:

$$\langle \langle Z^n \rangle \rangle = \sum_{\{S_i^a\}} \left\langle \left\langle \exp \left\{ -\frac{\beta N}{2} \sum_{\mu,a} \left[(m_\mu^a)^2 + \frac{\varepsilon}{2} (m_\mu^a)^4 \right] - \frac{\beta N \varepsilon}{4} \sum_a \left[\sum_\mu (m_\mu^a)^2 \right]^2 \right\} \right\rangle \right\rangle. \quad (\text{A.23})$$

A primeira integral gaussiana é introduzida para linearizar o último termo da equação acima:

$$\exp\left(-\frac{\lambda x_a^2}{2}\right) = i\sqrt{\frac{\lambda}{2\pi}} \int_{-i\infty}^{+i\infty} dy_a \exp\left(\frac{1}{2}\lambda y_a^2 - \lambda x_a y_a\right) \quad (\text{A.24})$$

onde, neste caso,

$$\lambda = \frac{\beta N \epsilon}{2}$$

$$x_a^2 = \left[\sum_{\mu} (m_{\mu}^a)^2 \right]^2 .$$

A função partição é então escrita como:

$$\langle\langle Z^n \rangle\rangle = cte \left\langle\left\langle \sum_{\{S_i^a\}} \int dy \exp\left\{-\frac{\beta N}{2} \sum_{\mu,a} \left[(m_{\mu}^a)^2(1 - \epsilon y_a) + \frac{\epsilon}{2}(m_{\mu}^a)^4\right] + \frac{\beta N \epsilon}{4} \sum_a y_a^2\right\} \right\rangle\right\rangle . \quad (\text{A.25})$$

Separamos o conjunto de P memórias em duas partes, uma com ℓ e outra com $P - \ell$ padrões:

$$\langle\langle Z^n \rangle\rangle = cte \sum_{\{S_i^a\}} \int dy \exp\left\{\frac{\beta N}{2} \sum_{\mu=1}^{\ell} \sum_a \left[(m_{\mu}^a)^2(1 - \epsilon y_a) + \frac{\epsilon}{2}(m_{\mu}^a)^4\right] + \frac{\beta N \epsilon}{4} \sum_a y_a^2\right\}$$

$$\left\langle\left\langle \exp\left\{\frac{\beta N}{2} \sum_a \left[(m^a)^2(1 - \epsilon y_a) + \frac{\epsilon}{2}(m^a)^4\right]\right\} \right\rangle\right\rangle^{\alpha N} . \quad (\text{A.26})$$

O último termo envolve a média sobre as memórias microscópicas, a qual foi detalhada na seção anterior. Note que devido ao termo de segunda ordem, o de quarta ordem pode ser desprezado e resta somente um termo de ruído análogo ao modelo de Hopfield. Ficamos então com:

$$\langle\langle Z^n \rangle\rangle = cte \sum_{\{S_i^a\}} \int dy \exp\left\{\frac{\beta N}{2} \sum_{\mu,a} \left[(m_{\mu}^a)^2(1 - \epsilon y_a) + \frac{\epsilon}{2}(m_{\mu}^a)^4\right] + \frac{\beta N \epsilon}{4} \sum_a y_a^2\right.$$

$$\left. + \frac{\alpha N}{2} \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1 \dots a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \epsilon y_{a_1}) \dots (1 - \epsilon y_{a_r})\right\} . \quad (\text{A.27})$$

Passamos agora a linearizar os termos em m_{μ}^a e Q_{ab} :

$$\langle\langle Z^n \rangle\rangle = cte \sum_{\{S_i^a\}} \int dy dm \left[\prod_{\mu,a} \delta\left(m_{\mu}^a - \frac{1}{N} \sum_i \xi_i^{\mu} S_i^a\right) \right]$$

$$\exp\left\{\frac{\beta N}{2} \sum_{\mu,a} \left[(m_{\mu}^a)^2(1 - \epsilon y_a) + \frac{\epsilon}{2}(m_{\mu}^a)^4\right] + \frac{\beta N \epsilon}{4} \sum_a y_a^2\right.$$

$$\left. + \frac{\alpha N}{2} \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1 \dots a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \epsilon y_{a_1}) \dots (1 - \epsilon y_{a_r})\right\}$$

$$\begin{aligned}
 &= cte \sum_{\{S_i^a\}} \int dy dm dt \exp \left\{ \frac{\beta N}{2} \sum_{\mu,a} \left[(m_\mu^a)^2 (1 - \varepsilon y_a) + \frac{\varepsilon}{2} (m_\mu^a)^4 \right] \right. \\
 &\quad + \frac{\alpha N}{2} \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \varepsilon y_{a_1}) \dots (1 - \varepsilon y_{a_r}) \\
 &\quad \left. + \frac{\beta N \varepsilon}{4} \sum_a y_a^2 - \sum_{\mu,a} t_\mu^a \left(m_\mu^a - \frac{1}{N} \sum_i \xi_i^\mu S_i^a \right) \right\} . \quad (A.28)
 \end{aligned}$$

E introduzindo uma função delta para Q_{ab} , juntamente com sua representação de Fourier:

$$\begin{aligned}
 \langle\langle Z^n \rangle\rangle &= cte \sum_{\{S_i^a\}} \int dy dm dt dq dr \exp \left\{ \frac{\beta N}{2} \sum_{\mu,a} \left[(m_\mu^a)^2 (1 - \varepsilon y_a) + \frac{\varepsilon}{2} (m_\mu^a)^4 \right] \right. \\
 &\quad + \frac{\beta N \varepsilon}{4} \sum_a y_a^2 + \frac{\alpha N}{2} \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \varepsilon y_{a_1}) \dots (1 - \varepsilon y_{a_r}) \\
 &\quad \left. - \beta N \sum_{\mu,a} t_\mu^a m_\mu^a + \beta \sum_{\mu,a} t_\mu^a \sum_i \xi_i^\mu S_i^a - \beta^2 N \alpha \sum_{a < b} r_{ab} \left(Q_{ab} - \frac{1}{N} \sum_i S_i^a S_i^b \right) \right\} . \quad (A.29)
 \end{aligned}$$

E, finalmente, podemos escrever tudo na seguinte forma:

$$\langle\langle Z^n \rangle\rangle = cte \int dy dm dt dq dr \exp[-\beta N f(\mathbf{m}, \mathbf{t}, \mathbf{r}, \mathbf{q}, \mathbf{y})] \quad (A.30)$$

onde $f(\mathbf{m}, \mathbf{t}, \mathbf{r}, \mathbf{q}, \mathbf{y})$ é a energia livre por neurônio e vale:

$$\begin{aligned}
 f &= \lim_{n \rightarrow 0} \left\{ -\frac{1}{2} \sum_{\mu,a} \left[(m_\mu^a)^2 (1 - \varepsilon y_a) + \frac{\varepsilon}{2} (m_\mu^a)^4 \right] - \frac{\varepsilon}{4} \sum_a y_a^2 + \sum_{\mu,a} t_\mu^a m_\mu^a \right. \\
 &\quad + \beta \alpha \sum_{a < b} r_{ab} Q_{ab} - \frac{\alpha}{2\beta} \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \varepsilon y_{a_1}) \dots (1 - \varepsilon y_{a_r}) \\
 &\quad \left. - \frac{1}{\beta} \left\langle \left\langle \ln \left[\sum_{\{S^a\}} \exp \left(\beta \sum_{\mu,a} t_\mu^a \xi_i^\mu S_i^a + \beta^2 \alpha \sum_{a < b} r_{ab} S_i^a S_i^b \right) \right] \right\rangle \right\rangle \right\} . \quad (A.31)
 \end{aligned}$$

Supondo simetria das réplicas, ou seja:

$$m_\mu^a = m_\mu \quad (A.32a)$$

$$t_\mu^a = t_\mu \quad (A.32b)$$

$$y_a = y \quad (A.32c)$$

$$Q_{ab} = q, \quad a \neq b \quad (A.32d)$$

$$r_{ab} = r . \quad (A.32e)$$

O limite $n \rightarrow 0$ de alguns termos em f é extremamente simples. Vamos apenas dar os detalhes dos termos mais complicados. Em primeiro lugar:

$$\begin{aligned}
 L_1 &\equiv \sum_{r=1}^{\infty} \frac{\beta^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} (1 - \varepsilon y_{a_1}) \dots (1 - \varepsilon y_{a_r}) \\
 &= \sum_{r=1}^{\infty} \frac{\beta^r (1 - \varepsilon y)^r}{r} \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} \\
 &= \beta(1 - \varepsilon y) \text{Tr} \mathbf{Q} + \frac{1}{2} \beta^2 (1 - \varepsilon y)^2 \text{Tr} \mathbf{Q}^2 + \frac{1}{3} \beta^3 (1 - \varepsilon y)^3 \text{Tr} \mathbf{Q}^3 + \dots \\
 &= -\text{Tr} \ln [\mathbf{I} - \beta(1 - \varepsilon y) \mathbf{Q}] \\
 &= -\ln \left\{ [1 - \beta(1 - \varepsilon y)(1 - q)]^{n-1} [1 - \beta(1 - \varepsilon y)(1 + q(n-1))] \right\} \quad (\text{A.33})
 \end{aligned}$$

onde utilizamos que $\text{Tr} \ln = \ln \det$. Tomando o limite $n \rightarrow 0$:

$$\lim_{n \rightarrow 0} -\frac{\alpha}{2\beta} L_1 = \frac{\alpha}{2\beta} \ln [1 - \beta(1 - \varepsilon y)(1 - q)] - \frac{\alpha}{2} \frac{(1 - \varepsilon y)q}{1 - \beta(1 - \varepsilon y)(1 - q)} \quad (\text{A.34})$$

O segundo termo em f cujo limite requer alguns truques é

$$L_2 = \left\langle \left\langle \ln \sum_{\{S^a\}} \exp \left[\beta \sum_{\mu} t_{\mu} \xi^{\mu} \sum_a S^a + \beta^2 r \alpha \sum_{a < b} S^a S^b \right] \right\rangle \right\rangle \quad (\text{A.35})$$

Usando a igualdade

$$\sum_{a < b} S^a S^b = \frac{1}{2} \left(\sum_a S^a \right)^2 - \frac{n}{2} \quad (\text{A.36})$$

podemos escrever:

$$L_2 = -\frac{n\beta^2 r \alpha}{2} + \left\langle \left\langle \ln \sum_{\{S^a\}} \exp \left[\beta \sum_{\mu} t_{\mu} \xi^{\mu} \sum_a S^a + \frac{\beta^2 r \alpha}{2} \left(\sum_a S^a \right)^2 \right] \right\rangle \right\rangle \quad (\text{A.37})$$

Utilizando a seguinte integral gaussiana,

$$\exp \left[\frac{1}{2} \beta^2 r \alpha \left(\sum_a S^a \right)^2 \right] = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} z^2 + \beta \sqrt{\alpha r} z \sum_a S^a \right] \quad (\text{A.38})$$

podemos escrever:

$$\begin{aligned}
 L_2 &= -\frac{n\beta^2 r \alpha}{2} + \left\langle \left\langle \ln \int \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2) \sum_{\{S^a\}} \exp \left[\beta \left(\sum_{\mu} t_{\mu} \xi^{\mu} + \sqrt{\alpha r} z \right) \sum_a S^a \right] \right\rangle \right\rangle \\
 &= -\frac{n\beta^2 r \alpha}{2} + \left\langle \left\langle \ln \int \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2) \left[2 \cosh \beta \left(\sum_{\mu} t_{\mu} \xi^{\mu} + \sqrt{\alpha r} z \right) \right]^n \right\rangle \right\rangle \\
 &= -\frac{n\beta^2 r \alpha}{2} + \left\langle \left\langle \ln \int \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2) \exp \left[n \ln 2 \cosh \beta \left(\mathbf{t} \cdot \boldsymbol{\xi} + \sqrt{\alpha r} z \right) \right] \right\rangle \right\rangle \quad (\text{A.39})
 \end{aligned}$$

Expandindo para n pequeno e tomando o limite obtemos:

$$\lim_{n \rightarrow 0} \frac{1}{\beta n} L_2 = \frac{\beta \alpha r}{2} + \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \beta (t \cdot \xi + \sqrt{\alpha r z}) \right\rangle \right\rangle \quad (\text{A.40})$$

onde agora $\langle \rangle$ é a média sobre as memórias ξ^μ e sobre a variável gaussiana z :

$$\langle \dots \rangle = \frac{1}{2^\ell} \sum_{\{\xi^\mu\}} \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \dots \quad (\text{A.41})$$

E a energia livre torna-se (após o limite $n \rightarrow 0$):

$$f = -\frac{1}{2}(1 - \varepsilon y) \sum_{\mu} m_{\mu}^2 - \frac{\varepsilon}{4} \sum_{\mu} m_{\mu}^4 - \frac{\varepsilon}{4} y^2 + \sum_{\mu} t_{\mu} m_{\mu} - \frac{\alpha}{2\beta} \ln [1 - \beta(1 - \varepsilon y)(1 - q)] \\ + \frac{\alpha\beta}{2} r(1 - q) - \frac{1}{2} \frac{\alpha q(1 - \varepsilon y)}{1 - \beta(1 - \varepsilon y)(1 - q)} - \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \beta (t \cdot \xi + \sqrt{\alpha r z}) \right\rangle \right\rangle \quad (\text{A.42})$$

As equações de ponto de sela são obtidas derivando-se f em relação às suas variáveis e igualando a zero:

$$m_{\mu} = \left\langle \left\langle \xi^{\mu} \tanh \beta (t \cdot \xi + \sqrt{\alpha r z}) \right\rangle \right\rangle \quad (\text{A.43a})$$

$$t_{\mu} = m_{\mu}(1 - \varepsilon y) + \varepsilon m_{\mu}^3 \quad (\text{A.43b})$$

$$q = \left\langle \left\langle \tanh^2 \beta (t \cdot \xi + \sqrt{\alpha r z}) \right\rangle \right\rangle \quad (\text{A.43c})$$

$$r = \frac{q(1 - \varepsilon y)^2}{[1 - \beta(1 - \varepsilon y)(1 - q)]^2} \quad (\text{A.43d})$$

$$y = \sum_{\mu} m_{\mu}^2 + \alpha \frac{1 - \beta(1 - \varepsilon y)(1 - q)^2}{[1 - \beta(1 - \varepsilon y)(1 - q)]^2} \quad (\text{A.43e})$$

A.4 O Modelo de Hopfield Generalizado

A função partição do modelo de Hopfield generalizado, mediada sobre as memórias microscópicas, é

$$\langle \langle Z^n \rangle \rangle = \sum_{\{S_i^a\}} \left\langle \left\langle \exp \left\{ \frac{\beta k!}{2N^{k-1}} \sum_{i_1 < \dots < i_k} \sum_{\rho, a} \xi_{i_1}^1 \dots \xi_{i_k}^1 S_{i_1}^a \dots S_{i_k}^a \right\} \right\rangle \right\rangle \quad (\text{A.44})$$

Podemos separar o termo que corresponde à superposição com a primeira memória, notando que os termos correspondentes às auto-interações não vão contribuir (isso pode ser facilmente verificado no caso $k = 3$). Então:

$$\langle \langle Z^n \rangle \rangle = \sum_{\{S_i^a\}} \left\langle \left\langle \exp \left\{ \frac{\beta N}{2} \sum_a (m^a)^k + \frac{\beta k!}{2N^{k-1}} \sum_{i_1 < \dots < i_k} \sum_a \sum_{\rho > 1} \xi_{i_1}^1 \dots \xi_{i_k}^1 S_{i_1}^a \dots S_{i_k}^a \right\} \right\rangle \right\rangle \quad (\text{A.45})$$

Como a média atua somente no último membro e usando o resultado da seção anterior, obtemos:

$$\langle\langle Z^n \rangle\rangle = \sum_{\{S_i^a\}} \exp \left\{ \frac{\beta N}{2} \sum_a (m^a)^k + \frac{\alpha \beta^2 k! N}{8} \sum_{a,b} Q_{ab}^k \right\} . \quad (\text{A.46})$$

Linearizando os termos em m^a , introduzimos funções delta e suas transformadas, obtemos

$$\begin{aligned} \langle\langle Z^n \rangle\rangle &= \text{cte} \int \left[\prod_a dm^a dt^a \prod_{b>a} dr_{ab} dQ_{ab} \right] \\ &\exp \left\{ \frac{\beta N}{2} \sum_a (m^a)^k + \frac{\alpha \beta^2 k! N}{8} \sum_{a,b} Q_{ab}^k - \beta N \sum_a t^a m^a - \alpha \beta^2 N \sum_{a<b} r_{ab} Q_{ab} \right. \\ &\left. + \sum_i \ln \sum_{\{S_i^a\}} \exp \left[\beta \sum_a t^a \xi_i S_i^a + \alpha \beta^2 \sum_{a<b} r_{ab} S_i^a S_i^b \right] \right\} . \quad (\text{A.47}) \end{aligned}$$

E, finalmente, a energia livre é

$$\begin{aligned} f &= \lim_{n \rightarrow 0} \frac{1}{n} \left\{ -\frac{1}{2} \sum_a (m^a)^k - \frac{\alpha \beta k}{8} \sum_{a,b} Q_{ab}^k + \sum_a t^a m^a + \alpha \beta \sum_{a<b} r_{ab} Q_{ab} \right. \\ &\left. - \frac{1}{\beta} \left\langle \left\langle \ln \sum_{\{S^a\}} \exp \left[\beta \sum_a t^a \xi S^a + \alpha \beta^2 \sum_{a<b} r_{ab} S^a S^b \right] \right\rangle \right\rangle \right\} . \quad (\text{A.48}) \end{aligned}$$

Supondo que as réplicas sejam simétricas, isto é,

$$m^a = m \quad (\text{A.49a})$$

$$t^a = t \quad (\text{A.49b})$$

$$r_{ab} = r \quad (\text{A.49c})$$

$$Q_{ab} = q \quad , \quad (\text{A.49d})$$

a energia livre fica, já tomando o limite $n \rightarrow 0$:

$$f = -\frac{1}{2} m^k + \frac{\alpha \beta k!}{8} + tm - \frac{\alpha \beta}{2} r q + \frac{1}{2} \alpha \beta r - \frac{1}{\beta} \langle\langle \ln 2 \cosh \beta(t + \sqrt{\alpha r z}) \rangle\rangle . \quad (\text{A.50})$$

E as equações de ponto de sela são:

$$m = \langle\langle \tanh \beta (t + \sqrt{\alpha r z}) \rangle\rangle \quad (\text{A.51a})$$

$$q = \langle\langle \tanh^2 \beta (t + \sqrt{\alpha r z}) \rangle\rangle \quad (\text{A.51b})$$

$$t = \frac{k}{2} m^{k-1} \quad (\text{A.51c})$$

$$r = \frac{k! k q^{k-1}}{2} . \quad (\text{A.51d})$$

A.5 O Modelo Completo

A função partição do modelo completo, quando o sistema é replicado n vezes e mediado sobre as possíveis configurações de memórias fica [68]:

$$\langle\langle Z^n \rangle\rangle = \sum_{\{S_i^a\}} \left\langle\left\langle \exp \left\{ -\frac{\beta N}{2} \sum_{a=1}^n \exp \left[-\sum_{\rho=\ell+1}^P (m_\rho^a)^2 \right] \prod_{\mu=1}^P [1 - (m_\mu^a)^2] \right\} \right\rangle\right\rangle . \quad (\text{A.52})$$

Para efetuarmos as médias envolvidas é preciso linearizar vários termos. Primeiramente introduzimos uma função delta para o argumento da segunda exponencial:

$$\delta \left[S_a - \frac{1}{\alpha} \sum_{\rho=\ell+1}^P (m_\rho^a)^2 \right] = \frac{\beta N}{4\pi i} \int_{-i\infty}^{+i\infty} d\tilde{S}_a \exp \left\{ \frac{\beta N}{2} \tilde{S}_a \left[S_a - \frac{1}{\alpha} \sum_{\rho=\ell+1}^P (m_\rho^a)^2 \right] \right\} , \quad (\text{A.53})$$

e obtemos:

$$\begin{aligned} \langle\langle Z^n \rangle\rangle &= cte \sum_{\{S_i^a\}} \int_{-\infty}^{+\infty} \int_{-i\infty}^{+i\infty} \left(\prod_a dS_a d\tilde{S}_a \right) \\ &\left\langle\left\langle \exp \left\{ -\frac{\beta N}{2} \left[\sum_a S_a \tilde{S}_a + \sum_{a,\rho} \tilde{S}_a (m_\rho^a)^2 - \sum_a e^{-\alpha S_a} \prod_\mu [1 - (m_\mu^a)^2] \right] \right\} \right\rangle\right\rangle . \end{aligned} \quad (\text{A.54})$$

Como fizemos anteriormente com S_a , introduzimos as quantidades \tilde{m}_μ^a e \tilde{Q}_{ab} por meio de funções delta relacionadas às quantidades m_μ^a e Q_{ab} , respectivamente, e obtemos

$$\begin{aligned} \langle\langle Z^n \rangle\rangle &= cte \sum_{\{S_i^a\}} \int_{-\infty}^{+\infty} \int_{-i\infty}^{+i\infty} \left(\prod_a dS_a d\tilde{S}_a \prod_\mu dm_\mu^a d\tilde{m}_\mu^a \prod_{b>a} dQ_{ab} d\tilde{Q}_{ab} \right) \\ &\exp \left\{ -\frac{\beta N}{2} \sum_a S_a \tilde{S}_a + \frac{\alpha N}{2} \sum_{r=1}^{\infty} \frac{1}{r} \left(\frac{\beta}{\alpha} \right)^r \sum_{a_1, \dots, a_r} Q_{a_1 a_2} \dots Q_{a_r a_1} \tilde{S}_{a_1} \dots \tilde{S}_{a_r} + \right. \\ &-\frac{\beta N}{2} \sum_a e^{-\alpha S_a} \prod_\mu [1 - (m_\mu^a)^2] - \frac{\beta N}{2} \sum_{\mu,a} m_\mu^a \tilde{m}_\mu^a + \frac{\beta}{2} \sum_i \sum_{a<b} \tilde{Q}_{ab} S_i^a S_i^b \\ &\left. + \frac{\beta}{2} \sum_{\mu,a,i} \tilde{m}_\mu^a \xi_i^\mu S_i^a - \frac{\beta N}{2} \sum_{a<b} \tilde{Q}_{ab} Q_{ab} \right\} . \end{aligned} \quad (\text{A.55})$$

Supondo que as réplicas sejam simétricas,

$$\begin{aligned} \tilde{m}_\mu^a &= \tilde{m}_\mu & m_\mu^a &= m_\mu \\ \tilde{S}_a &= \tilde{S} & S_a &= S \\ \tilde{Q}_{ab} &= \tilde{q} & Q_{ab} &= q \quad , a \neq b \end{aligned} \quad (\text{A.56})$$

a energia livre fica:

$$\begin{aligned}
 f = & \frac{1}{2}S\tilde{S} + \frac{\alpha}{2\beta} \ln \left[1 - \frac{\beta\tilde{S}}{\alpha}(1-q) \right] - \frac{1}{2} \frac{\tilde{S}q}{1 - \frac{\beta\tilde{S}}{\alpha}(1-q)} + \frac{1}{2} e^{-\alpha S} \prod_{\mu} (1 - m_{\mu}^2) \\
 & + \frac{1}{2} \sum_{\mu} m_{\mu} \tilde{m}_{\mu} - \frac{1}{4} \tilde{q}(1-q) - \frac{1}{\beta} \left\langle \left\langle \ln 2 \cosh \left(\frac{\beta}{2} \sum_{\mu} \tilde{m}_{\mu} \xi^{\mu} + \sqrt{\frac{\beta\tilde{q}}{2}} z \right) \right\rangle \right\rangle . \quad (\text{A.57})
 \end{aligned}$$

As equações de ponto de sela tornam-se:

$$m_{\mu} = \left\langle \left\langle \xi^{\mu} \tanh \left[\frac{\tilde{\beta}}{2} \left(\sum_{\nu} \frac{m_{\nu}}{1 - m_{\nu}^2} \xi^{\nu} + \sqrt{\alpha r z} \right) \right] \right\rangle \right\rangle \quad (\text{A.58a})$$

$$q = \left\langle \left\langle \tanh^2 \left[\frac{\tilde{\beta}}{2} \left(\sum_{\nu} \frac{m_{\nu}}{1 - m_{\nu}^2} \xi^{\nu} + \sqrt{\alpha r z} \right) \right] \right\rangle \right\rangle \quad (\text{A.58b})$$

$$r = \frac{q}{\left[1 - \frac{\tilde{\beta}}{2}(1-q) \right]^2} \quad (\text{A.58c})$$

$$S = \frac{1 - \frac{\tilde{\beta}}{2}(1-q)^2}{\left[1 - \frac{\tilde{\beta}}{2}(1-q) \right]^2} \quad (\text{A.58d})$$

$$\tilde{\beta} \equiv \frac{2\beta\tilde{S}}{\alpha} = 2\beta e^{-\alpha S} \prod_{\mu} (1 - m_{\mu}^2) . \quad (\text{A.58e})$$

Apêndice B

Expansões para o Modelo Truncado

Neste apêndice mostraremos em detalhe como alguns dos resultados do capítulo 4 são obtidos.

B.1 Transições Contínuas em α_c^\pm

Quando $\alpha \rightarrow \alpha_c^\pm(\varepsilon)$ a superposição $m \rightarrow 0$ e podemos obter $\alpha_c^\pm(\varepsilon)$ expandindo as equações de ponto de sela. Em primeiro lugar, como m é pequeno, assim também deve ser o argumento da função erro:

$$m = \operatorname{erf}\left(\frac{t}{\sqrt{2\alpha r}}\right) \simeq \sqrt{\frac{2}{\alpha\pi}} \frac{t}{\sqrt{r}} \quad . \quad (\text{B.1})$$

Além disso:

$$t = (1 - \varepsilon y)m + \varepsilon m^3 = m \left\{ 1 - \frac{\alpha\varepsilon}{[1 - C(1 - \varepsilon y)]^2} \right\} \quad (\text{B.2})$$

e fazemos a hipótese (que será verificada *a posteriori*) de que o argumento entre chaves é pequeno, ou seja:

$$C(1 - \varepsilon y) \simeq 1 \pm \sqrt{\alpha\varepsilon} \quad . \quad (\text{B.3})$$

Com isso, as equações ficam:

$$y \simeq m^2 + \frac{1}{\varepsilon} \quad (\text{B.4})$$

$$r \simeq \frac{\varepsilon m^4}{\alpha} \quad (\text{B.5})$$

$$C \simeq \pm \sqrt{\frac{2}{\varepsilon\pi}} \frac{1}{m^2} \quad . \quad (\text{B.6})$$

Então, de (B.4) e (B.6):

$$C(1 - \varepsilon y) \simeq \pm \sqrt{\frac{2\varepsilon}{\pi}} . \quad (\text{B.7})$$

Comparando as equações (B.3) e (B.7):

$$\alpha_c^\pm(\varepsilon) = \left(\frac{1}{\sqrt{\varepsilon}} \pm \frac{2}{\pi} \right)^2 . \quad (\text{B.8})$$

B.2 O limite $\varepsilon \rightarrow \infty$

Quando $\varepsilon \rightarrow \infty$, as equações para y , t e r ficam:

$$y = m^2 \quad (\text{B.9})$$

$$t = m \quad (\text{B.10})$$

$$r = \frac{1}{C^2} \quad (\text{B.11})$$

e para C :

$$1 = \sqrt{\frac{2}{\alpha\pi}} \exp\left(-\frac{m^2 C^2}{2\alpha}\right) , \quad (\text{B.12})$$

o que implica que

$$m^2 C^2 = \alpha \ln \frac{2}{\alpha\pi} , \quad (\text{B.13})$$

ou seja, usando a equação para m :

$$m = \operatorname{erf}\left(\frac{mC}{\sqrt{2\alpha}}\right) = \operatorname{erf}\left(\sqrt{\frac{1}{2} \ln \frac{2}{\alpha\pi}}\right) . \quad (\text{B.14})$$

Podemos ver da equação acima que $\alpha_c = 2/\pi$ pois para $\alpha \rightarrow \alpha_c$ (por baixo) teríamos $m \sim (\alpha_c - \alpha)^{1/2}$.

B.3 Temperatura de vidro de spin

Expandindo a equação para q quando este é pequeno, obtemos

$$q \simeq \beta^2 \alpha r \simeq \frac{\beta^2 \alpha q (1 - \varepsilon y)^2}{1 - \beta (1 - \varepsilon y)^2} \quad (\text{B.15})$$

ou seja

$$T_g \equiv \beta^{-1} = (1 \pm \sqrt{\alpha})(1 - \varepsilon y) . \quad (\text{B.16})$$

Da equação para y obtém-se:

$$1 - \varepsilon y = \frac{1 + \beta}{2\beta} - \frac{1}{2\beta} \sqrt{(1 - \beta)^2 + 4\beta\alpha\varepsilon} \quad , \quad (\text{B.17})$$

onde escolhemos o sinal negativo a fim de reobtermos a linha $T_g = 1 + \sqrt{\alpha}$ em $\varepsilon = 0$.
Combinando as duas equações, após uma certa álgebra, resta

$$T_g(\varepsilon) = \frac{1 + \sqrt{\alpha}}{\sqrt{\alpha}} \left[(1 - \alpha\varepsilon)(1 + \sqrt{\alpha}) - 1 \right] \quad . \quad (\text{B.18})$$

Apêndice C

O Modelo de Kauffman

Redes booleanas aleatórias, introduzidas por Kauffman [77] para tentar explicar o aparecimento de ordem em vários sistemas biológicos, consistem em N spins $\sigma_i \in \{0, 1\}$, cada um conectado a outros K spins. A evolução do sistema é efetuada em paralelo, e a cada sítio é atribuída uma função booleana aleatória f_i :

$$\sigma_i^{t+1} = f_i(\sigma_{i_1}^t, \dots, \sigma_{i_K}^t) \quad . \quad (\text{C.1})$$

Existem 2^K possíveis combinações para os valores assumidos pelos argumentos de f_i , portanto, o número de possíveis funções booleanas de K argumentos é 2^{2^K} . É importante frisar que uma vez escolhidos os K vizinhos e a função booleana, o sistema não sofre nenhuma mudança estrutural (*quenched*).

Várias questões podem ser colocadas sobre este sistema, como por exemplo: qual o número de atratores? No caso de haver ciclos limites, qual o seu período? Se considerarmos duas configurações diferentes, qual a probabilidade delas se tornarem iguais? Estas questões foram abordadas tanto do ponto de vista analítico quanto de simulação numérica. Vamos aqui apresentar um método bastante poderoso para estudar a dinâmica do sistema [92] e que será a base para resolvermos futuramente a dinâmica de vários modelos de redes de neurônios. O método consiste em desprezar o fato de que tanto as funções quanto os vizinhos são fixos e fazê-los variar arbitrariamente em cada passo (*annealed approximation*). Vamos argumentar mais adiante que no caso em que K é pequeno ($K \ll \ln N$), esta aproximação é exata, como mostra a excelente concordância entre os resultados obtidos com ela e as simulações numéricas [77].

Vamos tomar duas configurações, \mathcal{C}_1 e \mathcal{C}_2 , as quais possuem n spins diferentes em $t = 0$. Queremos calcular a probabilidade $P_1(m, n)$ de que estas configurações possuam m spins diferentes no próximo instante de tempo. Seja A o conjunto de spins idênticos nas duas configurações ($N - n$ spins) e B o conjunto dos spins diferentes (n spins). Chamamos de

$Q(N_0)$ a probabilidade de que N_0 spins recebam todos os sinais de spins pertencentes ao conjunto A . A probabilidade de que os K vizinhos de um dado spin pertençam ao conjunto A é $(1 - n/N)^K$, portanto, para os N_0 spins temos:

$$Q(N_0) = C_N^{N_0} \left[\left(1 - \frac{n}{N}\right)^K \right]^{N_0} \left[1 - \left(1 - \frac{n}{N}\right)^K \right]^{N-N_0} . \quad (C.2)$$

No instante de tempo seguinte teremos garantidamente N_0 spins idênticos nas duas configurações. Para os $N - N_0$ restantes, a probabilidade de serem iguais (ou diferentes) é $1/2$. Para que tenhamos m spins diferentes nas duas configurações é preciso que destes $N - N_0$ spins, $N - N_0 - m$ sejam iguais. A probabilidade disso ocorrer é:

$$C_{N-N_0}^m \left(\frac{1}{2}\right)^{N-N_0-m} \left(\frac{1}{2}\right)^m = C_{N-N_0}^m \left(\frac{1}{2}\right)^{N-N_0} . \quad (C.3)$$

E:

$$P_1(m, n) = \sum_{N_0=0}^{N-m} \frac{Q(N_0) C_{N-N_0}^m}{2^{N-N_0}} . \quad (C.4)$$

Note que o somatório foi truncado em $N - m$ pois se $N_0 > N - m$ não teríamos m spins diferentes. Usando que

$$C_{N-N_0}^m C_N^{N_0} = \frac{N!}{m!(N-m)!} C_{N-m}^{N_0} , \quad (C.5)$$

podemos efetuar a soma em N_0 :

$$\begin{aligned} P_1(m, n) &= \frac{N!}{m!(N-m)!} \frac{1}{2^N} \left[1 - \left(1 - \frac{n}{N}\right)^K \right]^N \sum_{N_0=0}^{N-m} C_{N-m}^{N_0} \left[\frac{2 \left(1 - \frac{n}{N}\right)^K}{1 - \left(1 - \frac{n}{N}\right)^K} \right]^{N_0} \\ &= \frac{N!}{m!(N-m)!} \frac{1}{2^N} \left[1 - \left(1 - \frac{n}{N}\right)^K \right]^N \left[1 + \left(1 - \frac{n}{N}\right)^K \right]^{N-m} . \end{aligned} \quad (C.6)$$

Esta equação é exata para o caso das duas configurações terem sido escolhidas aleatoriamente. Para calcularmos os passos subsequentes é preciso ter em mente que as configurações para $t \geq 1$ são correlacionadas com as funções f_i e com os conjuntos i_1, \dots, i_K . Aqui, portanto, fazemos a aproximação de que não existe tal correlação, o que é equivalente a escolhermos uma nova função f_i e um novo conjunto i_1, \dots, i_K para cada spin em cada passo de tempo. Formalmente, para um certo tempo t , teríamos:

$$P_t(m, n) \simeq \sum_{q_1=0}^N \dots \sum_{q_{t-1}=0}^N P_1(m, q_{t-1}) P_1(q_{t-1}, q_{t-2}) \dots P_1(q_1, n) . \quad (C.7)$$

Quando N é grande, introduzimos as variáveis contínuas x e y :

$$x = \frac{n}{N} \quad y = \frac{m}{N} \quad (\text{C.8})$$

e a forma binomial eq.(C.6), quando $N \rightarrow \infty$, torna-se uma gaussiana cujo máximo localiza-se em

$$y_1 = \frac{1 - (1 - x)^K}{2} \quad (\text{C.9})$$

Do mesmo modo, a equação (C.7) possui um máximo em

$$y_t = \frac{1 - (1 - y_{t-1})^K}{2} \quad (\text{C.10})$$

A equação acima sempre possui a solução $y = 0$ (as configurações convergem), mas esta só é estável para $K \leq 2$. Para valores maiores, $K > 2$, a solução atinge um ponto fixo não nulo (as configurações não convergem).

O valor crítico $K_c = 2$ concorda com o encontrado em simulações numéricas. Como a aproximação feita acima parece ser bastante forte, esta concordância é surpreendente. O que ocorre, na verdade, é que no limite termodinâmico, a aproximação é exata. Isto pode ser entendido através do seguinte argumento. Seja o spin i no instante t , o qual recebe K sinais do instante de tempo anterior, $t - 1$. Cada um destes K spins em $t - 1$ também recebe K sinais de $t - 2$ e assim por diante até $t = 1$. O estado σ_i que o i -ésimo spin assume no instante t depende, portanto, de M outros spins:

$$M = 1 + K + K^2 + \dots + K^{t-1} = \frac{K^t - 1}{K - 1} \quad (\text{C.11})$$

Se estes sítios forem todos diferentes, não teremos correlações entre as funções f_i (pois serão todas diferentes) e o resultado da aproximação será exato. A probabilidade de que isso ocorra é

$$\begin{aligned} P &= \prod_{i=1}^{M-1} \left(1 - \frac{i}{N}\right) \\ &= 1 - \sum_{i=1}^{M-1} \frac{i}{N} + \sum_{i < j} \frac{ij}{N^2} - \dots \\ &= 1 - \frac{M(M-1)}{2N} + \mathcal{O}\left(\frac{M^4}{N^2}\right) \end{aligned} \quad (\text{C.12})$$

Para que $P \rightarrow 1$ quando $N \rightarrow \infty$, isto é, não tenhamos sítios idênticos na árvore de spins cujos sinais chegam ao sítio i no tempo t , é preciso que ou a conectividade K ou o tempo considerado sejam finitos. Em ambos os casos,

$$\frac{M(M-1)}{N} \simeq \frac{M^2}{N} \rightarrow 0 \quad (\text{C.13})$$

Quando K é finito, M^2 precisa escalar com N^γ onde $\gamma < 1$: $M \sim N^{\gamma/2}$. Logo, $K^t \sim N^{\gamma/2}$ e

$$t \sim \frac{\gamma \ln N}{2 \ln K} . \quad (\text{C.14})$$

Para t finito, podemos ver que K^{2t}/N vai a zero se considerarmos que $K \sim \ln N$.

Pode-se também introduzir uma paisagem de adaptação (*fitness*) para o modelo do seguinte modo. Definimos, para cada configuração dos spins, uma função F dada por

$$F = \frac{1}{N} \sum_i f_i . \quad (\text{C.15})$$

Esta paisagem é semelhante à que aparece em redes de neurônios, com a diferença de que aqui o objetivo é maximizar a adaptação, e não minimizar a energia. Quando $K = 0$, F tem somente um máximo (como todos spins estão desconectados, basta maximizar cada f_i) e quando $K = N - 1$ os valores vizinhos de energia (para estados que se distinguem por somente um spin) são completamente descorrelacionados. Neste último limite, quando os níveis de energia tornam-se aleatórios, o modelo de Kauffman recai no modelo de energias randômicas (REM) de Derrida [73–76]. Este, por sua vez, também é o caso limite de uma família de modelos para vidros de spin: quando temos multi-interações cuja ordem tende a infinito.

Apêndice D

Cálculo da Dinâmica

Vamos mostrar em detalhe como obter a superposição $m(t+1)$ em função da superposição no instante anterior, $m(t)$, para o modelo truncado seguindo o método desenvolvido na ref. 29. O caso ligeiramente mais simples, para o modelo de Hopfield, é obtido no limite $\varepsilon = 0$.

Primeiramente separamos os termos que contêm auto-interações:

$$\begin{aligned} J_{ijkl} &= \sum_{\mu \neq \nu} C_{ijkl} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu \\ &= \sum_{\mu \neq \nu} C_{ijkl} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu (1 - \delta_{ij})(1 - \delta_{kl}) + (P-1) \sum_{\mu} C_{ijkl} \xi_i^\mu \xi_j^\mu \delta_{kl} (1 - \delta_{ij}) \\ &\quad + (P-1) \sum_{\nu} C_{ijkl} \xi_i^\nu \xi_j^\nu \delta_{ij} (1 - \delta_{kl}) + 2P(P-1) C_{ijkl} \delta_{ij} \delta_{kl} \quad . \end{aligned}$$

E para J'_{ijkl} :

$$\begin{aligned} J'_{ijkl} &= \frac{1}{3} (J_{ijkl} + J_{lkji} + J_{klij}) \\ &= \frac{1}{3} \sum_{\mu \neq \nu} (C_{ijkl} + C_{lkji} + C_{klij}) \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu (1 - \delta_{ij})(1 - \delta_{kl}) \\ &\quad + \frac{(P-1)}{3} \sum_{\mu} [C_{ijkl} \xi_i^\mu \xi_j^\mu \delta_{kl} (1 - \delta_{ij}) + C_{ljki} \xi_k^\mu \xi_i^\mu \delta_{lj} (1 - \delta_{kl}) + C_{kjil} \xi_i^\mu \xi_l^\mu \delta_{kj} (1 - \delta_{il})] \\ &\quad + \frac{(P-1)}{3} \sum_{\nu} [C_{iljk} \xi_k^\nu \xi_j^\nu \delta_{ik} (1 - \delta_{kj}) + C_{ljk i} \xi_k^\nu \xi_j^\nu \delta_{ik} (1 - \delta_{lj}) + C_{kjil} \xi_k^\nu \xi_j^\nu \delta_{il} (1 - \delta_{kj})] \\ &\quad + \text{cte} \quad . \end{aligned}$$

As auto-interações, fundamentais no modelo completamente conexo, aqui não contribuem pois:

$$J_{iikl} \simeq \frac{P}{N} J_{kl} = \frac{\alpha C}{N} J_{kl} \rightarrow 0 \quad , \quad (\text{D.1})$$

uma vez que $C \ll \ln N$. O campo local pode ser escrito como:

$$h_i = \sum_{\mu, j} C_{ij} \xi_i^\mu \xi_j^\mu S_j - \frac{\varepsilon}{3} \sum_{\substack{j \neq i \\ k \neq l}} \sum_{\mu \neq \nu} A_{ijkl} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu S_j S_k S_l \quad (\text{D.2})$$

onde $A_{ijkl} = C_{ijkl} + C_{lkji} + C_{klij}$. Separando os termos correspondentes à primeira memória ($\mu = 1$ e $\nu = 1$), obtemos:

$$h_i = \sum_j \xi_i^1 \xi_j^1 C_{ij} S_j + \sum_j \sum_{\mu > 1} \xi_i^\mu \xi_j^\mu C_{ij} S_j - \frac{\varepsilon}{3} \sum_{\substack{j \neq i \\ k \neq l}} A_{ijkl} \left[\sum_{\nu > 1} \xi_i^1 \xi_j^1 \xi_k^\nu \xi_l^\nu + \sum_{\mu > 1} \xi_i^\mu \xi_j^\mu \xi_k^1 \xi_l^1 + \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu \right] S_j S_k S_l \quad (\text{D.3})$$

Lembrando que a variável aleatória C_{ij} é 1(ou 0) com probabilidade C/N (ou $1 - C/N$), os dois primeiros termos de h_i podem ser escritos como

$$\sum_j \xi_i^1 \xi_j^1 C_{ij} S_j + \sum_j \sum_{\mu > 1} \xi_i^\mu \xi_j^\mu C_{ij} S_j = \sum_{r=1}^K \xi_i^1 \xi_{j_r}^1 S_{j_r} + \sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} \quad , \quad (\text{D.4})$$

onde a soma em r é sobre os K sítios conectados ao neurônio i . Já a variável aleatória C_{ijkl} é 1(ou 0) com probabilidade C/N^3 (ou $1 - C/N^3$) e os termos de quarta ordem em (D.3) podem ser reescritos como:

$$\begin{aligned} -\frac{\varepsilon}{3} \sum_{\substack{j \neq i \\ k \neq l}} \sum_{\nu > 1} A_{ijkl} \xi_i^1 \xi_j^1 \xi_k^\nu \xi_l^\nu S_j S_k S_l &= -\frac{\varepsilon}{K^2} \sum_{r,s,t=1}^K \sum_{\nu > 1} \xi_i^1 \xi_{j_r}^1 \xi_{k_s}^\nu \xi_{l_t}^\nu S_{j_r} S_{k_s} S_{l_t} \\ &= -\varepsilon \sum_{r=1}^K \xi_i^1 \xi_{j_r}^1 S_{j_r} \left(\frac{1}{K^2} \sum_{s,t=1}^K \sum_{\nu > 1} \xi_{k_s}^\nu \xi_{l_t}^\nu S_{k_s} S_{l_t} \right) \quad (\text{D.5}) \end{aligned}$$

O termo entre parênteses é um caminho aleatório (*random walk*) com média zero e dispersão αK^{-1} (portanto, será desprezado). Como existe uma correlação entre o estado da rede e a primeira memória, o primeiro termo (fora dos parênteses) não é um caminho aleatório. O segundo termo de quarta ordem de (D.3) é:

$$-\frac{\varepsilon}{3} \sum_{\substack{j \neq i \\ k \neq l}} \sum_{\mu > 1} A_{ijkl} \xi_i^\mu \xi_j^\mu \xi_k^1 \xi_l^1 S_j S_k S_l = -\frac{\varepsilon}{K^2} \sum_{s,t=1}^K \xi_{k_s}^1 \xi_{l_t}^1 S_{k_s} S_{l_t} \left(\sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} \right) \quad (\text{D.6})$$

Este termo contribuirá para o campo local pois a dispersão do termo entre parênteses vale PK (e é da mesma ordem do termo de ruído em (D.4)). E, finalmente, o terceiro termo de

(D.3) vale:

$$-\frac{\varepsilon}{3} \sum_{\substack{j \neq i \\ k \neq l}} \sum_{\mu \neq \nu} A_{ijkl} \xi_i^\mu \xi_j^\mu \xi_k^\nu \xi_l^\nu S_j S_k S_l = -\frac{\varepsilon}{K^2} \sum_{\mu \neq \nu} \sum_{\tau, s, t=1}^K \xi_i^\mu \xi_{j_r}^\mu \xi_{k_s}^\nu \xi_{l_t}^\nu S_{j_r} S_{k_s} S_{l_t} \quad (\text{D.7})$$

que é um caminho aleatório de dispersão $\alpha^2 K$, o que é muito pequeno quando comparado ao termo anterior (lembre que $P \sim K$).

O campo local, excluindo os termos que não contribuem, é

$$h_i = \sum_{r=1}^K \xi_i^1 \xi_{j_r}^1 S_{j_r} + \sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} - \frac{\varepsilon}{K^2} \left(\sum_{s=1}^K \xi_{k_s}^1 S_{k_s} \right)^2 \sum_{\mu > 1} \sum_{r=1}^K \xi_i^\mu \xi_{j_r}^\mu S_{j_r} \quad (\text{D.8})$$

Introduzindo a seguinte notação, $\xi_{j_r}^1 S_{j_r} \equiv \tau_{j_r}$, o campo local pode ser reescrito como:

$$h_i = \sum_{r=1}^K \xi_i^1 \tau_{j_r} + \left[1 - \frac{\varepsilon}{K^2} \left(\sum_{s=1}^K \tau_{k_s} \right)^2 \right] \sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} \quad (\text{D.9})$$

O que queremos calcular é a superposição, ou seja

$$m(t+1) = \frac{1}{N} \sum_i m_i(t+1) = \frac{1}{N} \sum_i \langle \langle \xi_i^1(S_i(t)) \rangle \rangle \quad (\text{D.10})$$

onde a média $\langle \langle \dots \rangle \rangle$ é sobre as possíveis configurações da rede, sobre as possíveis conectividades e sobre as memórias armazenadas. Reescrevendo a equação acima:

$$\begin{aligned} m_i(t+1) &= \langle \langle \xi_i^1 \tanh(\beta_0 h_i) \rangle \rangle \\ &= \langle \langle \tanh(\beta_0 \xi_i^1 h_i) \rangle \rangle \\ &= \langle \langle \tanh \left\{ \beta_0 \left[\sum_{r=1}^K \tau_{j_r} + \left(1 - \frac{\varepsilon}{K^2} \left(\sum_{s=1}^K \tau_{k_s} \right)^2 \right) \sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} \right] \right\} \rangle \rangle \end{aligned} \quad (\text{D.11})$$

Vamos agora efetuar as médias acima mencionadas. O termo de ruído pode ser escrito como:

$$\sum_{r=1}^K \sum_{\mu > 1} \xi_i^\mu \xi_{j_r}^\mu S_{j_r} = (P-1)K - 2S \quad (\text{D.12})$$

onde S é o número de termos negativos na soma e $(P-1)K$ o número total de termos. A presença de S termos negativos ocorre com probabilidade igual a

$$P(S) = C_{(P-1)K}^S \left(\frac{1}{2} \right)^{(P-1)K-S} \left(\frac{1}{2} \right)^S = \frac{1}{2^{K(P-1)}} C_{(P-1)K}^S \quad (\text{D.13})$$

sendo que $\sum_{S=0}^{K(P-1)} P(S) = 1$. Agora queremos a probabilidade de que

$$\xi_i^1 h_i = \sum_{r=1}^K \tau_{j_r} + [(P-1)K - 2S] \left[1 - \frac{\varepsilon}{K^2} \left(\sum_{r=1}^K \tau_{j_r} \right)^2 \right] \quad (\text{D.14})$$

a qual vale

$$P(\{\tau_{j_r}\}, S) = P(\{\tau_{j_r}\})P(S) = P(S) \prod_{j_r=1}^K P(\tau_{j_r}) \quad (\text{D.15})$$

sendo que as probabilidades são descorrelacionadas porque no limite em que estamos trabalhando ($C \ll \ln N$), todos os neurônios j_r são descorrelacionados entre si. Sendo n_+ (n_-) o número de neurônios, conectados com i , alinhados (desalinhados) com a primeira memória, os quais satisfazem

$$Nm_{j_r} = n_+ - n_- \quad (\text{D.16})$$

$$N = n_+ + n_- \quad , \quad (\text{D.17})$$

temos que

$$S_{j_r}(t+1) = \begin{cases} +\xi_{j_r}^1 & \text{com probabilidade } \frac{1+m_{j_r}}{2} \\ -\xi_{j_r}^1 & \text{com probabilidade } \frac{1-m_{j_r}}{2} \end{cases} . \quad (\text{D.18})$$

Portanto, $m_i(t+1)$ fica

$$m_i(t+1) = \left\langle \left\langle \sum_{\{\tau_{j_r}\}} \left(\prod_{j_r=1}^K \frac{1 + \tau_{j_r} m_{j_r}}{2} \right) \frac{1}{2^{K(P-1)}} \sum_{S=0}^{K(P-1)} C_{K(P-1)}^S \right. \right. \quad (\text{D.19})$$

$$\left. \left. \tanh \left\{ \beta_0 \left[\sum_{r=1}^K \tau_{j_r} + [K(P-1) - 2S] \left(1 - \frac{\varepsilon}{K^2} \left(\sum_{r=1}^K \tau_{j_r} \right)^2 \right) \right] \right\} \right\rangle \right\rangle_K$$

onde $\langle \dots \rangle_K$ assinala a última média que resta a fazer, sobre os possíveis K . A probabilidade de termos K neurônios conectados a i é dada por

$$P(K) = C_N^K \left(\frac{C}{N} \right)^K \left(1 - \frac{C}{N} \right)^{N-K} \quad (\text{D.20})$$

a qual, quando $N \rightarrow \infty$, torna-se uma distribuição de Poisson pois

$$C_N^K \left(\frac{C}{N} \right)^K \simeq \frac{N^K}{K!} \left(\frac{C}{N} \right)^K = \frac{C^K}{K!} \quad (\text{D.21})$$

e

$$\left(1 - \frac{C}{N}\right)^{N-K} \simeq \left(1 - \frac{1}{N/C}\right)^N \rightarrow e^{-C} \quad . \quad (\text{D.22})$$

Ou seja

$$P(K) \simeq \frac{C^K e^{-C}}{K!} \quad . \quad (\text{D.23})$$

Logo:

$$m_i(t+1) = \sum_{K=0}^{\infty} \sum_{\{\tau_{j_r}\}} \left(\prod_{j_r=1}^K \frac{1 + \tau_{j_r} m_{j_r}}{2} \right) \frac{C^K e^{-C}}{K!} \frac{1}{2^{K(P-1)}} \sum_{S=0}^{K(P-1)} C_{K(P-1)}^S \tanh \left\{ \beta_0 \left[\sum_{r=1}^K \tau_{j_r} + [K(P-1) - 2S] \left(1 - \frac{\varepsilon}{K^2} \left(\sum_{r=1}^K \tau_{j_r} \right)^2 \right) \right] \right\} \quad (\text{D.24})$$

A soma sobre as configurações $\{\tau_{j_r}\}$ é feita notando-se que teremos n neurônios tais que $\tau_{j_r} = -1$ e $K - n$ tais que $\tau_{j_r} = +1$:

$$m_i(t+1) = \sum_{K=0}^{\infty} \sum_{n=0}^K \sum_{S=0}^{K(P-1)} C_K^m C_{K(P-1)}^S \frac{C^K e^{-C}}{K!} \frac{(1+m)^{K-n} (1-m)^n}{2^{KP}} \tanh \left\{ \beta_0 \left[K - 2n + [K(P-1) - 2S] \left(1 - \frac{\varepsilon}{K^2} (K - 2n)^2 \right) \right] \right\} \quad (\text{D.25})$$

Como $m_i(t+1)$ já não depende de i (o que mostra a equivalência entre os sítios), escrevemos $m_i(t) = m(t)$. Para tomar o limite $C \rightarrow \infty$ e $P \rightarrow \infty$, primeiramente definimos $\alpha C = P - 1$ e $\beta = \beta_0/C$. A distribuição de Poisson, cuja média é C e dispersão vale $C^{1/2}$, no limite em que $C \rightarrow \infty$ tende a uma função delta, $\delta(K - C)$ (note que temos um fator C no denominador da tanh):

$$m(t+1) = \sum_{n=0}^C \sum_{S=0}^{\alpha C^2} C_C^m C_{\alpha C^2}^S \frac{(1+m)^{C-n} (1-m)^n}{2^{C(1+\alpha C)}} \tanh \left\{ \frac{\beta}{C} \left[C - 2n + [\alpha C^2 - 2S] \left(1 - \frac{\varepsilon}{C^2} (C - 2n)^2 \right) \right] \right\} \quad . \quad (\text{D.26})$$

Mas

$$\lim_{C \rightarrow \infty} \frac{1}{2^C} \sum_{n=0}^C C_C^m (1+m)^{C-n} (1-m)^n = \delta \left(n - \frac{C(1-m)}{2} \right) \quad , \quad (\text{D.27})$$

pois, novamente, a dispersão (que é proporcional a $C^{1/2}$) é muito pequena comparada ao valor que a função assume no máximo (lembrando novamente que existe um fator C no denominador da tanh). Portanto

$$m(t+1) = \frac{1}{2^{\alpha C^2}} \sum_{S=0}^{\alpha C^2} C_{\alpha C^2}^S \tanh \left\{ \frac{\beta}{C} \left[mC + [\alpha C^2 - 2S](1 - \varepsilon m^2) \right] \right\} \quad . \quad (\text{D.28})$$

Com a última média (sobre S) a coisa é um pouco diferente pois agora tanto a dispersão quanto o valor médio são da ordem de C^2 :

$$\lim_{C \rightarrow \infty} \frac{1}{2^{\alpha C^2}} C_{\alpha C^2}^S \simeq \sqrt{\frac{2}{\pi \alpha C^2}} \exp \left[-2 \frac{\left(S - \frac{\alpha C^2}{2} \right)^2}{\alpha C^2} \right] . \quad (\text{D.29})$$

E obtemos:

$$m(t+1) = \int_{-\infty}^{+\infty} \frac{dy}{\sqrt{\pi}} e^{-y^2} \tanh \left\{ \beta \left[m + y(1 - \epsilon m^2) \sqrt{2\alpha} \right] \right\} . \quad (\text{D.30})$$

A obtenção das equações que descrevem a evolução do sistema quando este possui superposição com as duas primeiras memórias segue os mesmos passos acima, com a diferença que teremos que considerar alguns termos a mais.

Apêndice E

O Modelo de Energia Randômicas (REM)

No modelo de Hopfield generalizado, quando a ordem de interação é muito grande, recaímos no modelo de energias randômicas (REM) [73–76], cujo diagrama de fases pode ser obtido sem o auxílio de réplicas. Nas seções seguintes o modelo é rapidamente formulado e o limite $k \rightarrow \infty$ do modelo de Hopfield generalizado, resolvido.

E.1 REM

O modelo de energias randômicas consiste nos seguintes três pontos:

1. o sistema possui 2^N níveis de energia E_i ,
2. os quais são distribuídos aleatoriamente de acordo com

$$P(E) = \frac{1}{\sqrt{N\pi J^2}} \exp\left(-\frac{E^2}{NJ^2}\right) \quad (\text{E.1})$$

3. e são independentes.

É importante notar que os itens 1 e 2 são comuns a vários modelos de vidros de spin, entre eles o de Sherrington-Kirkpatrick [93]. A terceira suposição é a que possibilita a solução exata do modelo, análoga à que faremos na próxima seção. Este modelo é o mais simples possível para vidros de spin e seu diagrama de fases qualitativamente se assemelha aos obtidos para outros modelos.

E.2 O limite $k \rightarrow \infty$ do modelo GH

Para mostrar que os níveis de energia do sistema são completamente decorrelacionados neste limite precisamos calcular a probabilidade de uma configuração \mathcal{C}_1 possuir energia

por sítio ϵ_c e superposições $\{m_\mu\}$ macroscópicas com um conjunto de padrões Γ . Tal probabilidade é dada por [62]

$$P(\epsilon, \{m_\mu\}) = \langle\langle \delta(\epsilon - \epsilon_c(\{m_\mu\})) \rangle\rangle \quad (\text{E.2})$$

onde a média $\langle\langle \rangle\rangle$ é sobre as memórias $\mu \notin \Gamma$. Usando que

$$\delta(\epsilon - \epsilon_c) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dK e^{iK(\epsilon - \epsilon_c)} \quad , \quad (\text{E.3})$$

obtemos:

$$\begin{aligned} P(\epsilon, \{m_\mu\}) &= \langle\langle \frac{1}{2\pi} \int_{-\infty}^{+\infty} dK e^{iK(\epsilon - \epsilon_c)} \rangle\rangle \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} dK \exp \left[iK \left(\epsilon + \sum_{\mu \in \Gamma} m_\mu^k \right) \right] \\ &= \langle\langle \exp \left[iK \frac{k!}{N^k} \sum_{\mu \in \Gamma} \sum_{i_1 < \dots < i_k} \xi_{i_1}^\mu \dots \xi_{i_k}^\mu S_{i_1} \dots S_{i_k} \right] \rangle\rangle \\ &= \frac{N}{2\pi} \int_{-\infty}^{+\infty} dK \exp \left[iKN \left(\epsilon + \sum_{\mu \in \Gamma} m_\mu^k \right) \right] \\ &= \langle\langle \exp \left[\frac{iKk!}{N^{k-1}} \sum_{i_1 < \dots < i_k} \xi_{i_1} \dots \xi_{i_k} S_{i_1} \dots S_{i_k} \right] \rangle\rangle^{\alpha N^{k-1}} \quad , \quad (\text{E.4}) \end{aligned}$$

onde fizemos $K \rightarrow NK$. Note que como não temos mais a soma em a podemos fazer $\xi_i S_i \rightarrow \xi_i$. Como as variáveis ξ_i são aleatórias, $\xi_i S_i$ também o será. Podemos então fazer uma transformação de calibre (análoga à que fazemos no modelo de Mattis [90]), $\xi_i S_i \rightarrow \xi_i$, que mantém invariante a distribuição de probabilidade. Precisamos então calcular a seguinte média:

$$\Phi = \langle\langle \exp \left(\frac{iKk!}{N^{k-1}} \sum_{i_1 < \dots < i_k} \xi_{i_1} \dots \xi_{i_k} \right) \rangle\rangle \quad , \quad (\text{E.5})$$

a qual é feita de maneira análoga ao apêndice A e vale

$$\Phi = \exp \left(-\frac{1}{2} \frac{K^2 k!}{N^{k-2}} \right) \quad . \quad (\text{E.6})$$

E, portanto:

$$\begin{aligned} P(\epsilon, \{m_\mu\}) &= \frac{N}{2\pi} \int_{-\infty}^{+\infty} dK \exp \left[iKN \left(\epsilon + \sum_{\mu \in \Gamma} m_\mu^k \right) - \alpha NK^2 \right] \\ &= \sqrt{\frac{N}{2\pi\alpha k!}} \exp \left[-\frac{N}{2\alpha k!} \left(\epsilon + \sum_{\mu \in \Gamma} m_\mu^k \right)^2 \right] \quad . \quad (\text{E.7}) \end{aligned}$$

Analogamente, podemos calcular a probabilidade conjunta $P_{12}(\epsilon_1, \epsilon_2)$ de duas configurações, \mathcal{C}_1 e \mathcal{C}_2 , possuírem superposições macroscópicas $\{m_\mu^1\}$, $\mu \in \Gamma_1$, e $\{m_\mu^2\}$, $\mu \in \Gamma_2$ e energias ϵ_1^c e ϵ_2^c . Esta probabilidade vale

$$P_{12}(\epsilon_1, \epsilon_2) = \langle\langle \delta(\epsilon_1 - \epsilon_1^c) \delta(\epsilon_2 - \epsilon_2^c) \rangle\rangle \quad (\text{E.8})$$

Novamente introduzimos representações para as funções delta e a média a ser calculada agora é:

$$\Phi = \left\langle\left\langle \exp \left[\frac{ik!}{N^{k-1}} \sum_{i_1 < \dots < i_k} \xi_{i_1} \dots \xi_{i_k} \left(K_1 S_{i_1}^1 \dots S_{i_k}^1 + K_2 S_{i_1}^2 \dots S_{i_k}^2 \right) \right] \right\rangle\right\rangle \quad (\text{E.9})$$

onde já reescalamos as variáveis de integração, $K_i \rightarrow NK_i$. Definindo

$$\Lambda_{i_1 \dots i_k} \equiv \frac{ik!}{N^{k-1}} \left(K_1 S_{i_1}^1 \dots S_{i_k}^1 + K_2 S_{i_1}^2 \dots S_{i_k}^2 \right) \quad , \quad (\text{E.10})$$

podemos utilizar os resultados do apêndice A e obter:

$$\Phi = \exp \left[-\frac{k!}{2N^{k-2}} \left(K_1^2 + K_2^2 + 2K_1 K_2 q^k \right) \right] \quad (\text{E.11})$$

onde

$$q = \frac{1}{N} \sum_i S_i^1 S_i^2 \quad (\text{E.12})$$

é a superposição entre as duas configurações. Integrando em K_1 e K_2 , finalmente obtemos:

$$P_{12}(\epsilon_1, \epsilon_2) = \frac{N}{2\pi\alpha k! \sqrt{1 - q^{2k}}} \exp \left\{ -\frac{N}{2\alpha k!(1 - q^{2k})} \left[\left(\epsilon_1 + \sum_{\mu \in \Gamma_1} (m_\mu^1)^k \right)^2 + \left(\epsilon_2 + \sum_{\mu \in \Gamma_2} (m_\mu^2)^k \right)^2 - 2q^k \left(\epsilon_1 + \sum_{\mu \in \Gamma_1} (m_\mu^1)^k \right) \left(\epsilon_2 + \sum_{\mu \in \Gamma_2} (m_\mu^2)^k \right) \right] \right\} \quad (\text{E.13})$$

Quando $k \rightarrow \infty$, a probabilidade acima desacopla uma vez que $q^k \rightarrow 0$ se $q \neq 1$, ou seja, $P_{12} = P(\epsilon_1)P(\epsilon_2)$. Portanto, os níveis de energia são totalmente descorrelacionados neste limite e o terreno de energia do sistema é extremamente rugoso, da mesma forma que o modelo de Kauffman quando todos os sítios estão ligados entre si [77].

No limite em que estamos interessados, somente precisamos considerar duas possibilidades para a superposição: ou $m = 1$ com, digamos, a primeira memória, ou $m = 0$. Qualquer outra solução para a superposição (incluindo as misturas) é degenerada com a solução $m = 0$, pois $m^k \rightarrow 0$ quando $k \rightarrow \infty$ e $m \neq 1$.

A função partição do sistema é

$$Z = \sum_{i=1}^{2^N} e^{-\beta E_i} \quad (\text{E.14})$$

e, separando a contribuição das P memórias as quais possuem energia -1 por sítio:

$$Z = P e^{\beta N} + \int_{-\infty}^{+\infty} d\epsilon \Omega(\epsilon) e^{-\beta \epsilon N} \quad (\text{E.15})$$

onde $\Omega(\epsilon)d\epsilon$ é o número de configurações com energia entre ϵ e $\epsilon + d\epsilon$.

Para obtermos o diagrama de fases é preciso conhecer que valores típicos $\Omega(\epsilon)$ assume. A densidade de estados é obtida a partir de

$$\Omega(\epsilon) = \sum_{i=1}^{2^N} \Theta(E_i - \epsilon) \Theta(\epsilon + d\epsilon - E_i) \quad (\text{E.16})$$

ou seja, somente contribuem os níveis de energia entre ϵ e $\epsilon + d\epsilon$. O valor médio de $\Omega(\epsilon)$ é facilmente obtido:

$$\begin{aligned} \langle \Omega(\epsilon) d\epsilon \rangle &= 2^N \sqrt{\frac{N}{2\pi\alpha k!}} \exp\left(\frac{-N\epsilon^2}{2\alpha k!}\right) d\epsilon \\ &= \sqrt{\frac{N}{2\pi\alpha k!}} \exp\left[N\left(\ln 2 - \frac{\epsilon^2}{2\alpha k!}\right)\right] d\epsilon \quad . \end{aligned} \quad (\text{E.17})$$

Podemos ver que $\langle \Omega(\epsilon) \rangle$ será exponencialmente grande quando

$$\ln 2 > \frac{\epsilon^2}{2\alpha k!} \quad , \quad (\text{E.18})$$

ou seja,

$$|\epsilon| < \sqrt{2\alpha k! \ln 2} \quad , \quad (\text{E.19})$$

onde $m_\mu^k = 0$ pois o termo envolvendo os estados correlacionados com as memórias já está evidenciado em (E.15). Como a densidade de estados é exponencialmente grande ao redor do valor médio, significa que existem muitos estados com energias entre ϵ e $\epsilon + d\epsilon$. Como os níveis de energia são variáveis aleatórias independentes, pelo teorema central do limite, a dispersão ao redor do valor médio será muito pequena. Logo:

$$\langle \Omega(\epsilon) \rangle \sim \Omega(\epsilon) \quad . \quad (\text{E.20})$$

Utilizando (E.17), (E.19) e (E.20) em (E.15):

$$Z = P e^{\beta N} + \sqrt{\frac{N}{2\pi\alpha'}} \int_{-\sqrt{2\alpha' \ln 2}}^{+\sqrt{2\alpha' \ln 2}} \exp\left[N\left(\ln 2 - \frac{\epsilon^2}{2\alpha'} - \beta\epsilon\right)\right] d\epsilon \quad (\text{E.21})$$

onde usamos que $\alpha k! = \alpha'$. Note que como o argumento da exponencial em (E.17) é negativo quando $\epsilon < \epsilon_0 = -\sqrt{2\alpha' \ln 2}$, não teremos nenhum estado com tais energias. Logo, o estado fundamental do sistema é ϵ_0 (quando não há correlação com alguma memória). Abaixo da temperatura crítica T_g o sistema encontra-se no estado fundamental, ao passo que acima de T_g existe uma fase paramagnética onde ϵ é dado pelo ponto de sela da integral acima. Existe ainda uma outra fase quando o expoente da exponencial na integral é negativo e o termo dominante passa a ser o proveniente das memórias. No primeiro caso, onde $\epsilon = -\sqrt{2\alpha' \ln 2}$, a função partição fica:

$$Z = P e^{\beta N} + \sqrt{\frac{N}{2\pi\alpha'}} e^{\beta N \sqrt{2\alpha' \ln 2}} \simeq \sqrt{\frac{N}{2\pi\alpha'}} e^{\beta N \sqrt{2\alpha' \ln 2}} \quad (\text{E.22})$$

e a energia livre:

$$f_{SG} = -\sqrt{2\alpha' \ln 2} \quad . \quad (\text{E.23})$$

Quando $T > T_g$, a integral é dominada pelo ponto de sela,

$$\epsilon = -\alpha' \beta \quad , \quad (\text{E.24})$$

ou seja,

$$Z \simeq \sqrt{\frac{N}{2\pi\alpha'}} e^{N(\ln 2 + \alpha' \beta^2 / 2)} \quad (\text{E.25})$$

e

$$f_P = -T \ln 2 + \frac{\alpha'}{2T} \quad . \quad (\text{E.26})$$

Note que quando $\epsilon = \epsilon_0$ obtemos T_g :

$$T_g = \frac{\alpha'}{\epsilon_0} = \sqrt{\frac{\alpha'}{2 \ln 2}} \quad . \quad (\text{E.27})$$

No último caso, quando o termo ferromagnético é dominante,

$$Z = P e^{\beta N} \quad (\text{E.28})$$

e

$$f_F = -1 \quad . \quad (\text{E.29})$$

As fronteiras entre as várias fases são obtidas igualando-se estas energias livres, gerando o diagrama de fases mostrado no capítulo 2.

Referências Bibliográficas

- [1] HOPFIELD J.J. 1982 *Proc.Natl.Acad.Sci. USA* **79** 2554
- [2] AMIT D.J. 1989 *Modeling Brain Function* (NY: Cambridge University Press)
- [3] CAMPBELL P. 1989 *Physics World* **2** 8
- [4] MULLER B. E REINHARDT J. 1991 *Neural Networks – An Introduction* (Springer Verlag)
- [5] AMIT D., GUTFREUND H. E SOMPOLINSKY H. 1985 *Phys.Rev. A* **32** 1007
- [6] AMIT D., GUTFREUND H. E SOMPOLINSKY H. 1985 *Phys.Rev.Lett.* **55** 1530
- [7] AMIT D., GUTFREUND H. E SOMPOLINSKY H. 1987 *Ann.Phys.* **173** 30
- [8] DE ALMEIDA R.M.C. E IGLESIAS J.R. 1990 *Phys.Lett. A* **146** 239
- [9] ARENZON J.J. 1991 O Modelo RS para Redes de Neurônios, dissertação de mestrado, IF-UFRGS.
- [10] ARENZON J.J., DE ALMEIDA R.M.C. E IGLESIAS J.R. 1992 *J.Stat.Phys.* **69** 385
- [11] BOLLÉ D., HUYGHEBAERT J. E SHIM G.M. 1994 *J.Phys. A* **27** 5871
- [12] ARENZON J.J., DE ALMEIDA R.M.C., IGLESIAS J.R., PENNA T.J.P. E DE OLIVEIRA P.M.C. 1992 *J.Physique I* **1** 55
- [13] PENNA T.J.P., DE OLIVEIRA P.M.C., ARENZON, J.J., DE ALMEIDA R.M.C. E IGLESIAS J.R. 1991 *Int.J.Mod.Phys. C* **2** 711
- [14] HEBB D.O. 1949 *The Organization of Behavior* (Wiley, New York)
- [15] JAMES W. 1890 *Principles of Psychology* (Henry Holt, New York) p.566

- [16] QUILLFELDT J.A. 1994 Tese de doutorado, Instituto de Bioquímica – UFRGS
- [17] AMIT D., GUTFREUND H. E SOMPOLINSKY H. 1987 *Phys.Rev. A* **35** 2293
- [18] HERTZ J., KROGH A. E PALMER R.G. 1991 Introduction to the Theory of Neural Computation (Addison-Wesley Publishing Company)
- [19] MCCULLOCH W.S. E PITTS W.A. 1943 *Bull.Math.Biophys.* **5** 115
- [20] LAUTRUP B. 1988 *preprint*, The Niels Bohr Institute, NBI-HE-88-06
- [21] CLARK J.W. 1988 *Phys.Reports* **158** 91
- [22] PASTUR L.A. E FIGOTIN A.L. 1977 *Sov.J.Low.Temp.Phys.* **3**(6) 378
- [23] MEZARD M., PARISI G. E VIRASORO M.A. 1987 Spin Glass Theory and Beyond (World Scientific Publishing Co.)
- [24] BINDER K. E YOUNG A.P. 1986 *Rev.Mod.Phys.* **58** 801
- [25] PETRITI D. 1994 In: FIESTA 94 IV Escuela de Verano en Física Estadística Y Sistemas Cooperativos (Universidad de Chile)
- [26] CRISANTI A., AMIT D. E GUTFREUND H. 1986 *Europhys.Lett.* **2** 337
- [27] STEFFAN H. E KÜHN R. 1994 *Z.Phys. B* **95** 249
- [28] KOHRING G.A. 1990 *J.Stat.Phys.* **59** 1077
- [29] DERRIDA B., GARDNER E. E ZIPPELIUS A. 1987 *Europhys.Lett.* **4** 167
- [30] ARENZON J.J. E LEMKE N. 1994 *J.Phys. A* **27** 5161
- [31] FORREST, B. M. 1988 *J.Phys. A* **21** 245
- [32] KOHRING G.A. 1990 *J.Phys. A* **23** 2237
- [33] KANTER, I. 1989 *Phys.Rev. A* **40**(5) 2611
- [34] GHOSH M., SEN A.K., CHAKRABARTI B.K. E KOHRING G.A. 1990 *J.Stat.Phys.* **61** 501
- [35] ARENZON J.J. E DE ALMEIDA R.M.C. 1993 *Phys.Rev. E* **48** 4060

- [36] PARGA N. E VIRASORO M.A. 1986 *J. Physique* **47** 1857
- [37] GUTFREUND H. 1988 *Phys. Rev. A* **37** 570
- [38] IDIART M.A.P. 1991 Tese de doutorado, IF-UFRGS
- [39] PARISI G. 1986 *J. Phys. A* **19** L617
- [40] NADAL J.P., TOULOUSE G., CHANGEUX J.P. E DEHAENE S. 1986 *Europhys. Lett.* **1** (10) 535
- [41] NADAL J.P., TOULOUSE G., CHANGEUX J.P. E DEHAENE S. *Europhys. Lett.* **2** (4) 346
- [42] MEZARD M., NADAL J.P. E TOULOUSE G. 1986 *J. Physique* **47** 1457
- [43] KINZEL W. 1985 *Z. Phys. B* **60** 205
- [44] PARISI G. 1986 *J. Phys. A* **19** L675
- [45] FORREST B.M. 1989 *J. Phys. France* **50** 2003
- [46] KÜRTEK K.E. 1990 *J. Phys. France* **51** 1585
- [47] GUYON I., PERSONNAZ L., NADAL J.P. E DREYFUS G. 1988 *Phys. Rev. A* **38** 6365
- [48] VIRASORO M.A. 1989 *Phys. Reports* **184** 301
- [49] FONTANARI J.F. 1990 *J. Physique I* **51** 2421
- [50] MIRANDA E. 1991 *J. Physique I* **1** 999
- [51] BRANCHTEIN M.C. E ARENZON J.J. 1992 *J. Physique I* **2** 2019
- [52] KREBS P.R. E THEUMANN W.K. 1993 *J. Phys. A* **26** 3983
- [53] SILVA C.R., TAMARIT F.A., LEMKE N., ARENZON J.J. E CURADO E.M.F. 1995 *J. Phys. A*, a ser publicado
- [54] LEMKE N., ARENZON J.J., DE ALMEIDA R.M.C. E GOULART S. 1995 *J. Phys. A*, a ser publicado
- [55] PERSONNAZ L., GUYON I. E DREYFUS G. 1985 *J. Phys. Lett.* **46** L-359
PERSONNAZ L., GUYON I. E DREYFUS G. 1986 *Phys. Rev. A* **34** 4217

- [56] KANTER I. E SOMPOLINSKY H. 1987 *Phys.Rev. A* **35** 380
- [57] LENG X., MCGRANN J.V., QUILLFELDT J.A., SHAW G.L. E SHENOY K.V. 1994
In: The Memory System of the Brain, J. Delacour Ed. (World Scientific)
- [58] PERETTO P. E NIEZ J.J. 1986 *Biol.Cybern.* **54** 53
- [59] KANDEL E.R. 1991 In: Principles of Neural Science, Kandel E.R., Schwartz J.H. e
Jessell T.M. Eds. (Appleton & Lange, Norwalk)
- [60] CAREW T.J., CASTELLUCCI V.F. E KANDEL E.R. 1971 *Int. J. Neurosci.* **2** 79
- [61] KANDEL E.R. E TAUC L. 1965 *J.Physiol.* **181** 1
- [62] GARDNER E. 1987 *J.Phys. A* **20** 3453
- [63] ABBOTT L.F. E ARIAN Y. 1987 *Phys.Rev. A* **36** 5091
- [64] HORN D. E USHER M. 1988 *J.Physique* **49** 389
- [65] KOHRING G.A. 1990 *J.Physique* **51** 145
- [66] TAMARIT F.A., STARIOLO D.A. E CURADO E.M.F. 1991 *Phys.Rev. A* **43** 7083
- [67] WANG L. E ROSS J. 1991 *Phys. Rev. A* **44** R2259
- [68] BOLLÉ D., DUPONT P. E HUYGHEBAERT J. 1992 *Phys.Rev. A* **45** 4194
- [69] KANTER I. 1987 *J.Phys. C* **20** L257
1988 *Phys.Rev. A* **37** 2739
- [70] NAEF J.P. E CANNING A. 1992 *J.Physique I* **2** 247
- [71] ARENZON J.J., DE ALMEIDA R.M.C., IGLESIAS J.R., PENNA T.J.P. E DE OLIVEIRA P.M.C. 1993 *Physica A* **197** 1
- [72] BENGTTSSON M. 1992 *Int.J.Neur.Sys.* **2** 315
- [73] DERRIDA B. 1980 *Phys.Rev.Lett.* **45**(2) 79
- [74] DERRIDA B. 1981 *Phys.Rev. B* **24**(5) 2613
- [75] DERRIDA B. 1980 *Phys.Rep.* **67**(1) 29

- [76] GROSS D.J. E MEZARD M. 1984 *Nuclear Physics B* **240** [FS12] 431
- [77] KAUFFMAN S. 1993 *Origins Of Order* (Oxford University Press)
- [78] PERSONNAZ L., GUYON I. E DREYFUS G. 1987 *Europhys. Lett.* **4** 863
- [79] ARENZON J.J., em preparação
- [80] VIANA L. 1988 *J.Physique* **49** 167
- [81] HOPFIELD J.J., FEINSTEIN D.I. E PALMER R.G. 1983 *Nature* **304** 158
- [82] VAN HEMMEN J.L., IOFFE L.B., KÜHN R. E VAAS M. 1990 *Physica A* **163** 386
- [83] BARROZO M.C. E PENNA T.J.P. 1994 *Int. J. Mod. Phys. C* **5** 503
- [84] VAN HEMMEM J.L. 1982 *Phys.Rev.Lett.* **49** 409
- [85] LEMKE N., ARENZON J.J. E TAMARIT F.A. 1995 *J.Stat.Phys.*, a ser publicado
- [86] TESTA J. E HELD G.A. 1983 *Phys.Rev. A* **28** 3085
- [87] OTT E. 1993 *Chaos in Dynamical Systems* (Cambridge University Press)
- [88] CANNING A. E GARDNER E. 1988 *J.Phys. A* **21** 3275
- [89] SKARDA C.A. E FREEMAN W.J. 1987 *Behav.Brain Sci.* **10** 161
- [90] MATTIS D.C. 1976 *Phys.Lett. A* **56** 421
- [91] ARENZON J.J. E LEMKE N., em preparação
- [92] DERRIDA B. E POMEAU Y. 1986 *Europhys.Lett.* **1** 45
- [93] SHERRINGTON D. E KIRKPATRICK S. 1975 *Phys.Rev.Lett.* **35** 1792