

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

**REDES NEURAIS E PADRÕES
CORRELACIONADOS ***

Rubem Erichsen Junior

Tese realizada sob
a orientação do Prof. Dr.
Walter Karl Theumann e
apresentada ao Instituto de
Física da UFRGS, em preen-
chimento parcial dos requizi-
tos para a obtenção do título
de doutor em ciências.

Porto Alegre, dezembro de 1991

*Trabalho parcialmente financiado por CNPq, FAPERGS e FINEP

Para Grace, Carolina e Augusto

Agradeço

ao Prof. Dr. Walter Karl Theumann pela dedicada orientação, incentivo e apoio,

à Prof^a Dr^a Alba Theumann pelo incentivo e apoio na etapa final deste trabalho

e aos meus colegas do Instituto de Física da UFRGS, pela amizade e coleguismo em todas os momentos.

Abstract

We studied here the general effects that the ‘correlation’, defined as the overlap among the set of stored patterns causes to the equilibrium properties of attractor neural networks.

Initially, the Hopfield model for associative memory was studied analytically with a finite number of patterns stored according to the Hebb learning rule. The retrieval properties of states that show an asymmetric overlap with more than one stored pattern (asymmetric mixed states) were determined. The phase diagrams for the overlaps demonstrate that these states are relevant if the amount of thermal noise is not too large and the correlations are not too strong. The storage capacity $\alpha = P/N$ is considered in a network with the addition of $P - p$ uncorrelated patterns and results are given for the critical α_c . This corresponds to the value of α above which the retrieval state is no longer present. Numerical simulations are carried out to discuss the retrieval quality and the size of the basins of attraction of the network.

In the second part of the work, we studied the storage properties of correlated patterns, in the saturation regime, in Gardner–Derrida neural networks with optimal, continuous synapses. The average over quenched disorder were performed with the replica method, and the minimal fraction of storage errors ‘per neuron’ and the critical storage capacity in function of the correlations were calculated in the replica symmetry theory. We studied the stability in the space of replicas of these solutions, verifying that the symmetry breaking is relevant. Finally, we show a preliminar study of the replica symmetry breaking effects in the first order of the iterative Parisi’s scheme for spin glasses.

Resumo

Estudamos os efeitos gerais que a ‘correlação’, definida como superposição entre os padrões exerce sobre as propriedades de equilíbrio de redes neurais atratoras.

Inicialmente, estudamos analiticamente o modelo de Hopfield de memória associativa, com um número finito de padrões armazenados de acordo com a regra de Hebb. Determinamos as propriedades de recuperação dos estados que apresentam uma superposição assimétrica com mais de um padrão (estados mistos assimétricos). Os diagramas de fase demonstram que estes estados são relevantes se a quantidade de ruído na rede e a correlação entre os padrões não são ambas demasiado grandes. Estudamos a capacidade de armazenamento $\alpha = P/N$ em uma rede com a inclusão de $P - p$ padrões não correlacionados, e obtivemos resultados para α_c crítico, que corresponde ao valor de α acima do qual não há mais estados de recuperação. Efetuamos simulações numéricas para discutir a qualidade da recuperação e a dimensão das bacias de atração.

Na segunda parte do trabalho, estudamos as propriedades de armazenamento de padrões correlacionados acima do limite de saturação, em redes neurais de Gardner–Derrida com sinapses contínuas ótimas. Utilizando o método de réplicas para calcular médias configuracionais na rede de neurônios, determinamos, na teoria de simetria de réplicas, a fração mínima de padrões erroneamente armazenados por neurônio e a capacidade de armazenamento crítica como função das correlações. Estudamos a estabilidade destas soluções, e verificamos que a quebra de simetria de réplicas é relevante. Finalmente, apresentamos um estudo preliminar dos efeitos de quebra de simetria de réplicas na ordem mais baixa do

esquema iterativo de Parisi para vidros de spin.

Índice

1	Introdução	3
2	Redes neurais atratoras	8
2.1	A dinâmica do sistema	8
2.1.1	Dinâmica síncrona	10
2.1.2	Dinâmica assíncrona	11
2.2	Termodinâmica do modelo assíncrono	13
2.2.1	Regra de aprendizado de Hebb	13
2.2.2	Capacidade microscópica de armazenamento	14
2.2.3	Capacidade macroscópica de armazenamento	19
3	O modelo de Hopfield com padrões correlacionados	33
3.1	Redes biológicas	36
3.1.1	O modelo AGS	36
3.1.2	O modelo de Perez-Vicente e Amit	38
3.2	Capacidade microscópica de armazenamento	41
3.2.1	Solução simétrica	41
3.2.2	Soluções assimétricas	44
3.3	Capacidade macroscópica de armazenamento	53
3.3.1	Solução simétrica	53
3.3.2	Soluções assimétricas	58
3.4	Simulação numérica	62
3.4.1	Simulação numérica de um sistema de neurônios com padrões correlacionados	64

3.4.2	Bacias de atração	66
4	Redes neurais de Gardner–Derrida e padrões correlacionados	68
4.1	Fração mínima de erros	69
4.2	Estabilidade da solução de simetria de réplicas	79
5	Quebra de simetria de réplicas no problema de Gardner–Derrida	85
5.1	O esquema de Parisi	85
5.2	Padrões não correlacionados	88
5.3	Padrões correlacionados	89
6	Conclusões	91
A	Integração das componentes ‘altas’	95
B	Cálculo de $G_1(q_{\alpha\beta}, M_\alpha)$ e $G_2(\phi_{\alpha\beta}, \varepsilon_\alpha)$	99
C	Diagonalização das matrizes A_1 e A_2	101
	Referências Bibliográficas	104

1. Introdução

Evidências biológicas provenientes do estudo do cérebro indicam que a propriedade de processamento computacional de que são dotados os sistemas neurais naturais reside no comportamento cooperativo de um número considerável ($\sim 10^{12}$ no cérebro humano) de neurônios, que interagem de uma forma complexa por um número também considerável ($\sim 10^{16}$ no cérebro humano) de conexões sinápticas.

Simplificadamente, pode-se descrever um neurônio como um dispositivo não linear de estímulo e resposta. Segundo uma descrição morfológica simplificada (ver figura 1.1), é uma célula microscópica constituída por um corpo chamado *soma*, onde se localiza o núcleo, e uma série de pequenas ramificações chamadas *dendritos*. Há ainda uma ramificação maior, denominada de *axônio*. Um neurônio é estimulado por pulsos eletroquímicos emitidos pelos axônios de outros neurônios. Estes pulsos são coletados nos dendritos, por meio de conexões chamadas *sinapses*. As sinapses podem ser excitatórias ou inibitórias, conforme o estímulo que o neurônio recebe resulte no aumento ou redução do seu *potencial de membrana*. O neurônio emite sua resposta na forma de um trem de pulsos ao longo de seu axônio, que irá estimular outros neurônios. A taxa de emissão de pulsos é uma função sigmoïdal (monotonamente crescente e limitada inferiormente e superiormente) do potencial de membrana.

O modelo de neurônios formais de Mc Culloch e Pitts [2] representa uma simplificação extrema da função desempenhada por um neurônio. Entretanto, ela captura a essência da não linearidade ao propor o modelo de dispositivo binário que assume os estados ativo e inativo conforme esteja disparando à taxa máxima ou quiescente, respectivamente.

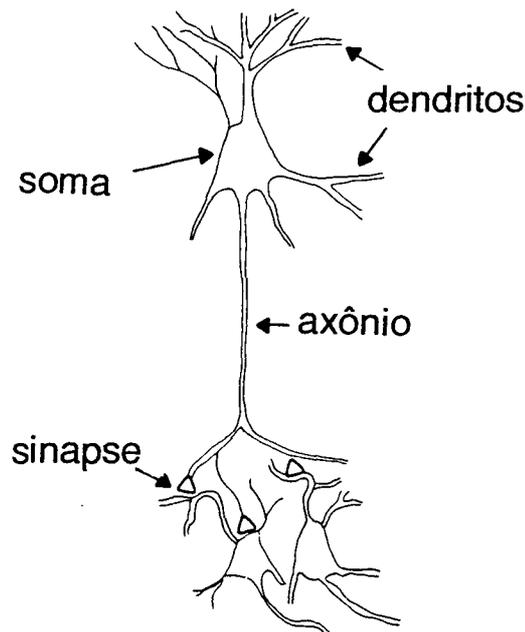


Figura 1.1: Aspecto morfológico de um neurônio, destacando o soma, os dendritos, o axônio e as sinapses (extraído da referência [1])

Devido à óbvia analogia com o modelo de Ising para sistemas magnéticos, o modelo de neurônios formais representou o primeiro passo para despertar o interesse de físicos, principalmente os oriundos da área da mecânica estatística. Em 1949, Hebb [3] relacionou o aprendizado com a modificação das conexões sinápticas. A hipótese de Hebb, que será discutida detalhadamente no capítulo 2, afirma que se dois neurônios estão frequentemente no mesmo estado de atividade, então desencadeia-se um processo metabólico que resulta no fortalecimento da sinapse entre eles. A identificação, ainda que rudimentar, dos componentes elementares de um sistema neural biológico, e o conhecimento básico sobre suas interações mútuas, permitem a construção de modelos físicos.

O raciocínio é uma propriedade de sistemas neurais que exige a evolução temporal através de estados seqüencialmente ordenados. Inspirado nesta hipótese, Little [4] traçou uma analogia entre a matriz transferência que descreve a evolução temporal de uma rede de neurônios formais e a matriz transferência para um sistema de spins de Ising bidimensional. Assim como a existência de ordem de longo alcance no sistema de spins a baixa temperatura, a existência de estados persistentes no cérebro deve estar relacionada com a degenerescência

do maior autovalor da matriz transferência definida para cada problema. Entretanto, devido à assimetria nas interações sinápticas, a matriz transferência para o sistema neural não é diagonalizável

A dinâmica do modelo de Little é síncrona. Todos neurônios atualizam-se em função do potencial de membrana ao mesmo tempo, em múltiplos inteiros do intervalo de tempo elementar τ . Hopfield [5] sugeriu um modelo dotado de uma dinâmica assíncrona, que possui maior fundamentação biológica. De acordo com esta dinâmica, atualiza-se um neurônio por vez. Para que o tempo médio de atualização da rede seja o mesmo da dinâmica síncrona, o intervalo de tempo entre a atualização consecutiva de dois neurônios é τ/N , onde N é o número total de neurônios.

Outra contribuição importante do modelo de Hopfield é a proposta de uma matriz simétrica de interações sinápticas baseada na hipótese de Hebb, denominada de 'regra de aprendizagem de Hebb'. A simetria nas interações sinápticas é pouco realista biologicamente. Entretanto, ela introduz a possibilidade da existência de atratores na dinâmica do sistema. Contemplado com esta possibilidade, o modelo de Hopfield habilita-se a descrever uma importante propriedade dos sistemas neurais naturais: a capacidade de memória associativa. Dada uma informação inicial incompleta, associada com a configuração inicial do sistema, este evolui dinamicamente na direção de uma configuração atratora, correspondente à informação completa ou padrão previamente apresentado à rede neural. Esta idéia dá origem ao conceito de redes neurais atratoras. A regra de aprendizagem de Hebb é uma prescrição que permite tornar um conjunto de padrões os atratores da dinâmica da rede.

Se as interações entre os neurônios são simétricas, a evolução dinâmica do sistema em direção aos atratores pode ser descrita pela minimização de uma função energia convenientemente definida. O modelo de Hopfield foi abordado por métodos de mecânica estatística de equilíbrio por Amit, Gutfreund e Sompolinsky, tanto para um número finito [6] quanto para um número infinito [7, 8] de padrões armazenados. As propriedades de equilíbrio do modelo de Little dotado da regra de aprendizagem de Hebb foram estudadas por Fontanari

e Köberle [10]. Como atratores, além dos pontos fixos como no modelo de Hopfield, há neste caso também os ciclos de período 2.

Uma regra de aprendizagem como a de Hebb implica que as conexões sinápticas são construídas como uma função explícita do conjunto de padrões a serem armazenados. Como alternativa a esta visão aparentemente artificial, foram definidos modelos nos quais a memorização é obtida por meio de *algoritmos* de aprendizagem [11, 12]. Os padrões são apresentados sucessivas vezes à rede até serem memorizados

Nos trabalhos de Amit, Gutfreund e Sompolinsky citados acima estuda-se o armazenamento de padrões nos quais os neurônios possuem atividade média de 50%, ou seja, no estado de equilíbrio metade dos neurônios está ativa e metade permanece quiescente. Entretanto, evidências biológicas [13] atestam que a taxa média de atividade é significativamente menor que 50%. Para aprimorar o modelo de modo a levar em consideração estas evidências, Amit, Gutfreund e Sompolinsky, em um trabalho posterior [9], introduziram o modelo AGS, no qual um parâmetro a , denominado *polarização*, descreve o estado de atividade da rede. Ele é definido tal que $(1 + a)/2$ é a fração de neurônios ativos e $(1 - a)/2$ a fração de neurônios quiescentes. O parâmetro a assume valores dentro do intervalo $[-1, 1]$. Como consequência da polarização, há uma superposição efetiva dos padrões. Relativamente a modelos que descrevem o armazenamento de padrões correlacionados, além do modelo AGS, entre outros que merecem ser citados temos o modelo da pseudo-inversa, de Personaz, Guyon e Dreyfus, [14], cuja termodinâmica foi estudada por Kanter e Sompolinsky [16] e os modelos de Buchmann, Divko e Shulten [17] e de Perez-Vicente e Amit [18].

As redes neurais que armazenam padrões correlacionados representam um passo importante no sentido de ampliar as bases biológicas da teoria. O armazenamento de padrões hierarquicamente correlacionados, ou seja, o caso em que os padrões são gerados a partir de um conjunto de ancestrais também tem sido estudado. Entre importantes contribuições neste sentido, podemos citar os trabalhos de Gutfreund [19] e de Parga e Virasoro [20]. Um modelo de hierarquia no espaço das conexões sinápticas foi proposto por Idiart e Theumann

[21, 22].

Devido aos fatores acima apontados, direcionamos este trabalho para o estudo detalhado das conseqüências do armazenamento de padrões correlacionados em modelos de redes neurais atratoras. No capítulo 2, discutimos detalhadamente o estudo do problema de redes neurais por métodos de mecânica estatística, inclusive a analogia entre temperatura e ruído sináptico. Revisitamos a teoria de Amit, Gutfreund e Sompolinsky, sobre a qual nos baseamos para estudar o modelo de Hopfield com padrões correlacionados.

No capítulo 3 descrevemos em detalhes os aspectos relacionados com o armazenamento de padrões correlacionados pela regra de Hebb. Para o número p de padrões finito, estudamos a estrutura de estados metaestáveis espúrios que infestam a dinâmica da rede. Para o caso em que o número de padrões é da mesma ordem do número de neurônios, estudamos a estrutura de estados espúrios na ausência de ruído sináptico e construímos o diagrama de fases termodinâmicas.

Um importante passo para o desenvolvimento da teoria de redes neurais deve-se a Gardner [23], que permite o estudo de propriedades gerais, independente de regras ou algoritmos de aprendizagem. As conexões sinápticas são escolhidas como variáveis dinâmicas do sistema. O aprendizado é assim representado como a evolução destas variáveis, governada por uma função custo que penaliza os erros de armazenamento. Seguindo este método, Gardner e Derrida [24] calcularam a fração mínima de erros de armazenamento por neurônio quando a rede neural é exigida acima de sua capacidade crítica. Em nosso trabalho estudamos os efeitos que as correlações entre os padrões exercem sobre a fração mínima de erros. Efetuamos a análise da estabilidade da solução de simetria de réplicas [25], segundo a diagonalização de Almeida e Thouless [26]. Constatamos a necessidade de proceder à quebra de simetria de réplicas. No capítulo 5, detalhamos este procedimento. Empregamos para isso o método de Parisi, em seu primeiro passo. No capítulo 6 concluímos com um resumo e com a discussão de nossos resultados e assinalamos o que resta por fazer nesta linha de trabalhos.

2. Redes neurais atratoras

O modelo de redes de neurônios formais representa uma simplificação da complexa natureza dos sistemas biológicos, que se faz necessária para permitir o seu estudo por métodos empregados pela mecânica estatística. Entretanto, apesar da simplificação, a essência destes sistemas, representada pela capacidade computacional que emerge do comportamento coletivo de um número grande de unidades simples que respondem de modo não linear a determinado estímulo externo, é resguardada, como veremos a seguir.

2.1 A dinâmica do sistema

O estado de atividade dos neurônios formais é descrito por uma variável binária. Segundo uma das representações usuais, o estado ativo, em que o neurônio dispara, e o inativo, em que ele permanece quiescente, são representados por $\sigma = 1$ ou $\sigma = 0$, respectivamente [27]. Outra representação, mais atraente em virtude da analogia com sistemas magnéticos, estabelece que a variável S assume os valores $S = +1$ e $S = -1$ para representar os dois estados. As duas representações relacionam-se pela expressão

$$S = 2\sigma - 1 . \tag{2.1}$$

O potencial pós-sináptico acumulado pelo neurônio i após um determinado intervalo de tempo é dado pela expressão (na representação S)

$$U_i = \frac{1}{2} \sum_{j \neq i} J_{ij} (S_j + 1). \tag{2.2}$$

Esta expressão estabelece que o potencial de membrana acumulado no neurônio i é a soma das contribuições sinápticas de todos os neurônios que estiverem ativos, ponderados pelo fator J_{ij} , que representa a eficácia da conexão sináptica entre os neurônios i e j . J_{ij} assume valor positivo ou negativo, conforme a sinapse é excitatória ou inibitória, respectivamente. O neurônio i decide emitir ou não um pulso eletroquímico ao longo de seu axônio após comparar o potencial de membrana acumulado com o seu potencial de limiar T_i . Se o sistema é determinístico então a resposta do neurônio no instante $t + \tau$ fica inequivocamente determinada a partir de U_i e T_i no ciclo anterior t pela expressão

$$S_i(t + \tau) = \text{sgn}[U_i(t) - T_i(t)] , \quad (2.3)$$

onde $\text{sgn}(x)$ é a função sinal de x .

O mecanismo de transmissão dos pulsos sinápticos está sujeito a uma série de fontes de ruído (Amit, 1989), de modo que o valor do potencial pós-sináptico U_i determinado pela equação 2.2 é um valor médio do que é sentido pelo neurônio i . A natureza não determinística do processo pode ser reproduzida introduzindo-se uma função que represente a probabilidade do neurônio i assumir o estado S_i

$$P(S_i) = \frac{\exp[\beta S_i(U_i - T_i)]}{\exp[\beta(U_i - T_i)] + \exp[-\beta(U_i - T_i)]} , \quad (2.4)$$

onde β representa o papel de inverso de temperatura. A probabilidade $P(S_i)$ assume uma forma sigmoideal. No limite $\beta \rightarrow \infty$ recupera-se a descrição determinística.

De modo a reforçar a analogia com um sistema magnético constituído por spins de Ising, é conveniente reescrever

$$\begin{aligned} U_i - T_i &= \frac{1}{2} \sum_{j \neq i} J_{ij} (S_j + 1) - T_i \\ &= h_i - h_i^e , \end{aligned} \quad (2.5)$$

onde

$$h_i = \sum_{j \neq i} J_{ij} S_j \quad (2.6)$$

e

$$h_i^e = T_i - \sum_{j \neq i} J'_{ij}, \quad (2.7)$$

sendo

$$J'_{ij} = \frac{1}{2} J_{ij}. \quad (2.8)$$

Pode-se simplificar o modelo admitindo-se que o potencial de limiar é igual ao valor médio do potencial pós-sináptico, tal que $h_i^e = 0$. A expressão 2.4 assume então a forma familiar

$$P(S_i) = \frac{1}{2} [1 + \tanh(\beta h_i S_i)], \quad (2.9)$$

que é formalmente idêntica à que descreve a dinâmica de Glauber [30] para um sistema de spins de Ising em interação, em contato com um banho térmico à temperatura $T = 1/\beta$.

O estado da rede no instante t é descrito pelo conjunto de N variáveis $S_i, i = 1, 2, \dots, N$ no espaço de configurações, que é um hipercubo de N dimensões. Ocorrem diferentes tempos de propagação de impulsos eletroquímicos, devido à presença de axônios de comprimentos diferentes. Conseqüentemente, cada estado da rede é gerado a partir de uma coleção de estímulos provenientes de uma série de estados precedentes. A dinâmica de um sistema assim é extremamente complexa. A seguir vamos discutir dois modelos de dinâmicas adotados usualmente: as dinâmicas síncrona, ou paralela e assíncrona, ou seqüencial.

2.1.1 Dinâmica síncrona

Na dinâmica síncrona todos neurônios são atualizados simultaneamente em instantes de tempo que são múltiplos inteiros de um determinado intervalo τ . Todos neurônios tomam a decisão de disparar ou não a partir do valor do potencial pós-sináptico acumulado no último ciclo de disparos. São desprezadas as diferenças nos tempos de propagação. Cada estado é gerado inteiramente a partir da configuração imediatamente anterior.

A probabilidade de transição do estado J para o estado I é dada por

$$W(I | J) = \prod_{i=1}^N P(S_i^I) = \frac{\exp(\beta \sum_i h_i^J S_i^I)}{\prod_i [\exp(\beta h_i^J) + \exp(-\beta h_i^J)]}, \quad (2.10)$$

que é simplesmente o produto sobre todos os neurônios da expressão (2.9). Se o sistema é não determinístico, todos os elementos de $W(I | J)$ são não nulos, de modo que o sistema

segue no espaço de configurações uma trajetória em que a cada ciclo ele pode sofrer uma transição para qualquer um dos 2^N vértices do hipercubo N -dimensional.

2.1.2 Dinâmica assíncrona

Do ponto de vista biológico, a dinâmica assíncrona [5] é uma alternativa mais atraente para descrever a evolução temporal de um sistema de neurônios. Para que o ciclo de disparos percorra toda a rede durante o intervalo de tempo correspondente τ da dinâmica síncrona, o intervalo entre o disparo de dois neurônios consecutivos é dado por $\tau' = \tau/N$. A ordem de disparos pode ser seqüencial ou aleatória e a trajetória que o sistema percorre no espaço de configurações é sensível à esta ordem. Dois estados sucessivos diferenciam-se ao máximo pelo estado de um neurônio. Se a ordem de disparos é aleatória, que é mais compatível com um sistema biológico, então a matriz formada pelos elementos $W(I | J)$ possui N elementos em cada linha, além do elemento diagonal, cada um correspondendo à inversão no estado de um neurônio distinto. No espaço de configurações, as transições acontecem somente entre vértices adjacentes do hipercubo.

O desenvolvimento de habilidades computacionais por uma rede neural implica que as trajetórias percorridas no espaço de configurações dependem das condições iniciais, influenciadas por agentes externos. Assim o sistema resulta *não ergódico*. A quebra de ergodicidade em sistemas físicos só ocorre no limite termodinâmico (no caso, o número de neurônios indo ao infinito) e abaixo de uma determinada temperatura crítica (correspondendo a um baixo nível de ruído) [29].

A capacidade de memória associativa é uma propriedade de sistemas computacionais que desperta grande interesse. Tal capacidade assegura a existência de trajetórias na dinâmica do sistema que partindo de um conjunto de estados iniciais o conduzem para pontos fixos atratores, no caso da dinâmica sem ruído, ou para uma distribuição atratora de pontos fixos, no caso da dinâmica com ruído. A informação armazenada na rede corresponde às coordenadas do ponto fixo. Estas coordenadas serão representadas por um conjunto de variáveis ξ_i^μ ; $i = 1, 2, \dots, N$; $\mu = 1, 2, \dots, p$, onde p é o número total de pontos fixos ou

padrões, que compõem a *memória* do sistema.

A evolução do sistema na direção de um ponto fixo pode ser descrita pela minimização de uma determinada *função energia* convenientemente definida. Para a dinâmica assíncrona uma escolha conveniente é

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j \neq i} J_{ij} S_i S_j \quad (2.11)$$

Se o sistema evolui da configuração J , na qual o neurônio i assume o estado S_i , para a configuração I , na qual o neurônio i assume o estado $-S_i$, a variação da energia é dada por

$$\Delta \mathcal{H} = \pm S_i \sum_{j \neq i} J_{ij} \pm \sum_{j \neq i} J_{ji} S_j . \quad (2.12)$$

Em $T = 0$, a regra de atualização definida por 2.2 e 2.3 estabelece que só ocorre transição se o primeiro termo de (2.12) for negativo. Se $J_{ij} = J_{ji}$ então a transição $J \rightarrow I$ implica em $\Delta \mathcal{H} < 0$. Portanto, a simetria nas conexões sinápticas assegura que a dinâmica assíncrona conduz o sistema na direção em que \mathcal{H} é mínimo, onde todos neurônios estão alinhados com o campo local.

Na presença de ruído, a situação é diferente da descrita acima, pois de acordo com (2.9) existe para cada neurônio uma probabilidade não nula de alinhar-se contrariamente ao campo local. Sendo a probabilidade de transição dada por (2.10), a razão entre $W(I | J)$ e $W(J | I)$ é

$$\frac{W(I | J)}{W(J | I)} = \exp \left(-\frac{2h_i S_i}{k_B T} \right) . \quad (2.13)$$

Mas $2h_i S_i / k_B T$ é exatamente a diferença de energia entre as configurações I e J , de modo que a equação (2.13) pode ser reescrita na forma

$$W(I | J) \exp \left(-\frac{\mathcal{H}(J)}{k_B T} \right) = W(J | I) \exp \left(-\frac{\mathcal{H}(I)}{k_B T} \right) , \quad (2.14)$$

que representa a condição de balanço detalhado, satisfeita pela dinâmica de Glauber [30]. Conseqüentemente, em equilíbrio termodinâmico o sistema transita entre configurações que satisfazem a distribuição de Gibbs, isto é, a probabilidade $P(I)$ de encontrar o sistema na configuração I é dada por

$$P(I) \sim \exp \left(-\frac{\mathcal{H}(I)}{k_B T} \right) . \quad (2.15)$$

Neste caso, a média temporal de um determinado observável é igual à média de ensemble. O valor médio do observável \mathcal{O} é dado por

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \sum_{\{s_i\}} \mathcal{O} \exp\left(-\frac{\mathcal{H}\{S_i\}}{k_B T}\right). \quad (2.16)$$

\mathcal{Z} é a função partição, definida por

$$\mathcal{Z}(\mathcal{T}) = \sum_{\{s_i\}} \exp\left(-\frac{\mathcal{H}\{S_i\}}{k_B T}\right). \quad (2.17)$$

Esta discussão remete o problema para o âmbito da termodinâmica, onde as propriedades físicas são obtidas a partir da energia livre por neurônio, definida por

$$f(\beta) = -\frac{1}{\beta N} \ln \mathcal{Z}, \quad (2.18)$$

onde $\beta = 1/k_B T$.

2.2 Termodinâmica do modelo assíncrono

2.2.1 Regra de aprendizado de Hebb

Para estudar a capacidade de memória associativa é necessário partir de alguma suposição a respeito de como a memória (representada por um conjunto de padrões $\{\xi_i^\mu\}$) é impressa nas conexões sinápticas. De acordo com a hipótese de Hebb, se dois neurônios disparam ao mesmo tempo repetidamente, desencadeia-se um processo metabólico que resulta no incremento, ou reforço, da conexão sináptica entre eles. Traduzida para a linguagem matemática, a regra de Hebb determina que para armazenar p padrões $\{\xi_i^\mu, \mu = 1, 2, \dots, p\}$, as conexões sinápticas são dadas por

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (2.19)$$

Esta expressão, que resulta em sinapses explicitamente simétricas, a rigor vai além da hipótese de Hebb. Ela implica que

1. se os estados assumidos pelos neurônios i e j no estado μ são ambos ativos, então aumenta a eficácia sináptica;

2. se os estados de atividade são diferentes (um ativo e outro passivo) então diminui a eficácia sináptica;
3. se ambos neurônios permanecem inativos no padrão μ então também aumenta a eficácia sináptica.

Obviamente as afirmativas 2. e 3. estão ausentes na hipótese original formulada por Hebb.

Um aspecto importante relativo à regra de aprendizagem de Hebb, e que é refletido pela expressão (2.17), é que ela é manifestamente *local*, isto é, a conexão sináptica J_{ij} depende apenas dos estados de atividade dos neurônios i e j .

Para medir a qualidade da recuperação de cada padrão introduz-se a variável denominada ‘overlap’, definida por

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i \rangle_T . \quad (2.20)$$

O overlap m^μ mede a superposição entre o estado da rede e o padrão identificado como μ . O valor máximo de m^μ é igual a 1, e corresponde ao estado em que $\langle S_i \rangle_T = \xi_i^\mu$, para todo i, μ (recuperação perfeita do padrão μ).

2.2.2 Capacidade microscópica de armazenamento

Amit, Gutfreund e Sompolinsky [6] estudaram as propriedades termodinâmicas de um sistema de neurônios cuja dinâmica assíncrona é determinada pelo hamiltoniano (2.11), no limite em que o número de padrões p é pequeno em relação ao número de neurônios. Mais precisamente, p é mantido finito enquanto N é levado ao infinito.

Os p padrões são armazenados de acordo com a regra de aprendizagem de Hebb (equação (2.14)). Propriedades médias do desempenho do sistema como memória associativa foram obtidas do estudo do armazenamento de padrões compostos de variáveis aleatórias independentes que obedecem a distribuição de probabilidades

$$P(\{\xi_i^\mu\}) = \prod_{i,\mu} P(\xi_i^\mu) = \prod_{i,\mu} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)] . \quad (2.21)$$

Cada realização do sistema recozido (‘quenched’) depende de um conjunto de variáveis aleatórias. As propriedades do sistema são determinadas tomando-se a média das quanti-

dades físicas sobre a distribuição de probabilidades (2.19). Por exemplo, a densidade de energia livre é dada por

$$f(\beta) = - \lim_{N \rightarrow \infty} \frac{1}{\beta N} \langle \ln \mathcal{Z} \rangle_{\{\xi_i^\mu\}}, \quad (2.22)$$

onde $\langle \dots \rangle_{\{\xi_i^\mu\}}$ indica a média sobre a distribuição (2.19).

O Hamiltoniano para o sistema é dado por

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j \neq i} J_{ij} S_i S_j - \sum_{\mu} h^{\mu} \sum_i S_i \xi_i^{\mu}, \quad (2.23)$$

onde por conveniência introduzimos os campos h^{μ} conjugados a $\sum_i S_i \xi_i^{\mu}$. A função partição, definida por (??) é dada por

$$\mathcal{Z}(\beta, h) = \sum_{\{S_i\}} \exp \left[\beta \left(\frac{1}{2} \sum_{i,j \neq i} J_{ij} S_i S_j + \sum_{\mu} h^{\mu} \sum_i S_i \xi_i^{\mu} \right) \right]. \quad (2.24)$$

Sendo as conexões sinápticas dadas por (2.17), temos

$$\begin{aligned} \sum_{i,j \neq i} J_{ij} S_i S_j &= \frac{1}{N} \sum_{\mu} \sum_{i,j \neq i} \xi_i^{\mu} \xi_j^{\mu} S_i S_j \\ &= \frac{1}{N} \sum_{\mu} \left[\left(\sum_i \xi_i^{\mu} S_i \right)^2 - N \right]. \end{aligned} \quad (2.25)$$

Introduzindo (2.23) em (2.22), temos

$$\mathcal{Z}(\beta, h) = e^{-\frac{1}{2}\beta p} \sum_{\{S_i\}} \prod_{\mu} \exp \beta \left[\frac{1}{2N} \left(\sum_i \xi_i^{\mu} S_i \right)^2 + h^{\mu} \sum_i \xi_i^{\mu} S_i \right]. \quad (2.26)$$

Usando a propriedade da integral gaussiana

$$e^{\frac{1}{2}\lambda a^2} = \int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}x^2 + \sqrt{\lambda}ax \right), \quad (2.27)$$

pode-se reescrever (2.24) na forma

$$\mathcal{Z}(\beta, h) = e^{-\frac{1}{2}\beta p} \sum_{\{S_i\}} \int \prod_{\mu} \left(\frac{dx^{\mu}}{\sqrt{2\pi}} \right) \exp \left[-\frac{1}{2} \sum_{\mu} (x^{\mu})^2 + \sqrt{\frac{\beta}{N}} \sum_{\mu} \left((x^{\mu} + h^{\mu}) \sum_i \xi_i^{\mu} S_i \right) \right]. \quad (2.28)$$

Para garantir que a energia livre seja uma quantidade extensiva, procede-se a mudança de variáveis

$$x^{\mu} \rightarrow \sqrt{\beta N} x^{\mu}. \quad (2.29)$$

Após inverter a ordem da integração com o traço sobre $\{S_i\}$, temos

$$\begin{aligned} \mathcal{Z}(\beta, h) &= (\beta N)^{\frac{1}{2}p} e^{-\frac{1}{2}\beta p} \int \prod_{\mu} \left(\frac{dx^{\mu}}{\sqrt{2\pi}} \right) \\ &\cdot \exp \left[-\frac{1}{2}\beta N \sum_{\mu} (x^{\mu})^2 \right] \prod_i \sum_{\{S_i\}} \exp \left[\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi_i^{\mu} S_i \right]. \end{aligned} \quad (2.30)$$

Tomando o traço, temos

$$\begin{aligned} \mathcal{Z}(\beta, h) &= (\beta N)^{\frac{1}{2}p} e^{-\frac{1}{2}\beta p} \int \prod_{\mu} \left(\frac{dx^{\mu}}{\sqrt{2\pi}} \right) \\ &\cdot \exp \left\{ -N \left[\frac{1}{2}\beta \sum_{\mu} (x^{\mu})^2 - \frac{1}{N} \sum_i \ln \left[2 \cosh \left(\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi_i^{\mu} \right) \right] \right] \right\} \end{aligned} \quad (2.31)$$

Para $N \gg 1$ a integral em x^{μ} pode ser resolvida pelo método de ponto de sela. Assim temos

$$-\frac{1}{\beta N} \ln \mathcal{Z}(\beta, h) = \frac{1}{2} \sum_{\mu} (x^{\mu})^2 - \frac{1}{\beta N} \sum_i \ln \left[2 \cosh \left(\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi_i^{\mu} \right) \right]. \quad (2.32)$$

Os valores das variáveis x^{μ} que maximizam o lado direito de 2.30) são dados pelas equações de ponto de sela:

$$x^{\mu} = \frac{1}{N} \sum_i \xi_i^{\mu} \tanh \left[\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi_i^{\mu} \right]. \quad (2.33)$$

Se p é finito, então o lado direito das equações (2.30) e (2.31) apresentam a propriedade de automediância, isto é

$$\frac{1}{N} \sum_i (\dots) = \langle \dots \rangle_{\{\xi_i^{\mu}\}}. \quad (2.34)$$

Usando esta propriedade temos

$$\begin{aligned} -\frac{1}{\beta N} \langle \ln \mathcal{Z}(\beta, h) \rangle_{\{\xi_i^{\mu}\}} &= f(\beta, h) \\ &= \frac{1}{2} \sum_{\mu} (x^{\mu})^2 - \frac{1}{\beta} \left\langle \ln \left[2 \cosh \left(\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi^{\mu} \right) \right] \right\rangle_{\{\xi^{\mu}\}} \end{aligned} \quad (2.35)$$

e

$$x^{\mu} = \left\langle \xi^{\mu} \tanh \left[\beta \sum_{\mu} (x^{\mu} + h^{\mu}) \xi^{\mu} \right] \right\rangle_{\{\xi^{\mu}\}}. \quad (2.36)$$

O significado físico do valor que as variáveis x^μ assumem no ponto de sela pode ser identificado calculando-se a média de ensemble dos ‘overlaps’:

$$\begin{aligned} m^\mu &\equiv \left\langle \frac{1}{N} \sum_i \langle S_i \rangle_T \xi_i^\mu \right\rangle_{\{\xi_i^\mu\}} = \left\langle \frac{1}{Z} \sum_{\{S_i\}} \left(\frac{1}{N} \sum_i S_i \xi_i^\mu \right) \exp(-\beta \mathcal{H}) \right\rangle_{\{\xi_i^\mu\}} \\ &= -\frac{\partial f(\beta, h)}{\partial h^\mu}. \end{aligned} \quad (2.37)$$

Derivando a expressão (2.33) com relação a h^μ , obtemos

$$m^\mu = \left\langle \xi^\mu \tanh \left[\beta \sum_\mu (x^\mu + h^\mu) \xi^\mu \right] \right\rangle_{\{\xi^\mu\}}. \quad (2.38)$$

Comparando as expressões (2.36) e (2.34) verificamos que o valor que as variáveis x^μ assumem no ponto de sela é exatamente o ‘overlap’:

$$x^\mu = m^\mu \quad (2.39)$$

Deste ponto em diante substituiremos x^μ por m^μ e assumiremos $h^\mu = 0$. Assim (2.33) e (2.34) podem ser reescritas simplificadamente como

$$f(\beta) = \frac{1}{2} \mathbf{m}^2 - \frac{1}{\beta} \langle \ln[2 \cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi})] \rangle_{\{\boldsymbol{\xi}\}} \quad (2.40)$$

e

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh(\beta \mathbf{m} \cdot \boldsymbol{\xi}) \rangle_{\{\boldsymbol{\xi}\}}, \quad (2.41)$$

Note-se que introduzimos a notação vetorial para o conjunto de ‘overlaps’,

$$\mathbf{m} = (m^1, m^2, \dots, m^p) \quad (2.42)$$

e para o conjunto de padrões

$$\boldsymbol{\xi} = (\xi^1, \xi^2, \dots, \xi^p). \quad (2.43)$$

O modelo pressupõe interações de alcance infinito entre os neurônios. Os estados de equilíbrio, caracterizados macroscopicamente pelo vetor \mathbf{m} são dados pelas soluções das equações de campo médio (2.39), que correspondem às soluções exatas do modelo.

Expandindo-se (2.39) para $m^\mu \rightarrow 0$, obtemos

$$m^\mu = \beta m^\mu + \mathcal{O}(m^3), \quad (2.44)$$

onde m é o módulo de \mathbf{m} . Conseqüentemente, para $T > 1$ a única solução é a paramagnética ($m^\mu = 0$). Soluções com $m^\mu \neq 0$ surgem continuamente quando $T < 1$.

Entre as soluções simétricas,

$$\mathbf{m} = m_n(1, 1, \dots, 1, 0, 0, \dots, 0) \quad (2.45)$$

encontram-se as soluções que correspondem aos estados de recuperação de padrões ($n = 1$).

Nesses casos, as equações de ponto de sela assumem a forma

$$m^\mu = \langle \xi^\mu \tanh(\beta \sum_{\nu=1}^n \xi^\nu) \rangle_{\{\xi\}}. \quad (2.46)$$

Somando sobre μ , temos

$$nm_n = \langle z_n \tanh(\beta m_n z_n) \rangle_{z_n}. \quad (2.47)$$

z_n é uma variável aleatória definida por

$$z_n = \sum_{\mu=1}^n \xi^\mu, \quad (2.48)$$

que obedece a uma distribuição de probabilidades binomial,

$$P(z_n = 2k - n) = \frac{1}{2^n} \binom{n}{k}, \quad (2.49)$$

sendo $0 \leq k \leq n$.

Os estados de recuperação apresentam overlap

$$m_1 = \tanh(\beta m_1), \quad (2.50)$$

com energia livre

$$f(\beta) = \frac{1}{2} m_1^2 - \frac{1}{\beta} \ln[2 \cosh(\beta m_1)], \quad (2.51)$$

que é mínimo absoluto para $0 \leq T \leq 1$ [6]. De (2.48) vemos que $m_1 = 1$ só quando $T = 0$. A recuperação só é perfeita na ausência de ruído sináptico. Para $T \neq 0$, cada neurônio

pode fazer uma transição na direção oposta ao indicado pelo potencial pós-sináptico com uma probabilidade não nula (equação (2.9)). A $T > 0$, o sistema em equilíbrio permanece perambulando sobre uma distribuição de estados em torno de um dos estados fundamentais, que correspondem aos padrões armazenados.

Os estados metaestáveis desempenham um papel importante do ponto de vista dinâmico. Sendo mínimos locais da energia livre, dependendo das condições iniciais podem aprisionar o sistema, impedindo-o de alcançar um mínimo absoluto. De acordo com a análise da estabilidade [6], entre as soluções simétricas apenas a de recuperação é estável em $T \rightarrow 1^-$. As restantes são apenas pontos de sela. A situação é diferente, entretanto, à baixa temperatura. Abaixo de $T \approx 0,461$, a solução simétrica que mistura três padrões é metaestável. À temperaturas inferiores outras soluções simétricas ímpares tornam-se metaestáveis. As soluções simétricas pares permanecem instáveis em qualquer temperatura, inclusive a $T = 0$.

De um modo geral, a baixas temperaturas o espaço de fases da rede neural fica infestado por um grande número de atratores expúrios, inclusive assimétricos (que correspondem a soluções cujo vetor \mathbf{m} tem componentes diferentes), que têm efeitos perversos sobre as bacias de atração dos estados de recuperação. Desse modo, a presença de ruído sináptico pode ter um efeito benéfico sobre o desempenho da rede neural como memória associativa, embora a qualidade de recuperação resulte prejudicada.

2.2.3 Capacidade macroscópica de armazenamento

Quando a rede se encontra no estado de recuperação, com $S_i = \xi_i^\mu$, o campo local sobre o neurônio i é dado por

$$h_i = \sum_{j \neq i} J_{ij} \xi_j^\mu = \xi_i^\mu (1 + \delta_i), \quad (2.52)$$

sendo

$$\delta_i = \frac{1}{N} \sum_{j \neq i} \sum_{\nu \neq \mu} \xi_i^\nu \xi_j^\nu \xi_i^\mu \xi_j^\mu. \quad (2.53)$$

O termo δ_i corresponde ao ruído provocado por correlações ‘microscópicas’ de ordem $1/\sqrt{N}$ com os demais padrões, e é a soma de $(p-1)(N-1)$ termos que assumem valores $+1$ ou -1 com igual probabilidade, multiplicados por $1/N$. A situação é análoga ao problema da

caminhada aleatória em uma dimensão. A média deste termo é

$$\langle \delta_i \rangle_{\{\xi\}} = 0 \quad (2.54)$$

e a sua variância é

$$\langle \delta_i^2 \rangle_{\{\xi\}} = \frac{p-1}{N}, \quad (2.55)$$

onde consideramos $N-1 \approx N$. Quando p é finito, o ruído é desprezível. Os estados de recuperação são estáveis, pois a condição

$$h_i \xi_i^\mu > 0 \quad (2.56)$$

é verificada em todos neurônios, em todos padrões. Entretanto, quando p cresce com N , tal que $p = \alpha N$, sendo α um número finito, o ruído proveniente das correlações microscópicas com os demais padrões não pode ser mais desprezado. Neste caso, para cada neurônio existe uma probabilidade não nula de que a condição (2.54) não se verifique.

A consequência imediata do armazenamento de um número infinito de padrões é que os estados que correspondem à recuperação *sem erros* deixam de ser atratores simples. O conceito de memória associativa deve ser reformulado, vinculando-se agora à existência de *distribuições* de estados atratorae, em torno dos estados de recuperação. No que segue, faremos uma exposição da teoria de campo médio de Amit, Gutfreund e Sompolinsky [8], que analisa a termodinâmica do modelo de Hopfield relativa à capacidade macroscópica de armazenamento.

De (2.11), (2.14) e (2.15), vemos que a energia do estado fundamental da rede neural é dada por

$$E = \frac{1}{2}\alpha - \frac{1}{2} \sum_{\mu=1}^p (m^\mu)^2. \quad (2.57)$$

Para que esta energia seja limitada inferiormente, os estados de equilíbrio da rede neural podem ter ‘overlap’ macroscópico (correspondentes aos padrões que ‘condensam’) apenas com um número finito de padrões. O overlap com os restantes é de ordem $1/\sqrt{N}$. Para assegurar que apenas s (um número finito) padrões condensem macroscopicamente usa-se o artifício de introduzir s campos h^μ , sendo $\mu = 1, 2, \dots, s$, acoplados aos padrões $\{\xi^\mu, \mu = 1, 2, \dots, s\}$.

Quando $p \rightarrow \infty$, a condição $2^p \ll N$, essencial para que as quantidades físicas sejam automediantes, não se verifica. Conseqüentemente, torna-se necessário efetuar explicitamente a média sobre a distribuição de variáveis aleatórias $\{\xi_i^\mu\}$. Utiliza-se o método das réplicas, que se vale da identidade matemática

$$\ln(x) = \lim_{n \rightarrow 0} \frac{x^n - 1}{n} \quad (2.58)$$

para efetuar a média configuracional do logaritmo da função partição:

$$\langle \ln \mathcal{Z}(\beta, h) \rangle = \lim_{n \rightarrow 0} \frac{1}{n} \left(\langle \mathcal{Z}^n(\beta, h) \rangle_{\{\xi\}} - 1 \right). \quad (2.59)$$

A energia livre por neurônio será então

$$f(\beta) = - \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\beta n N} \left(\langle \mathcal{Z}^n(\beta, h) \rangle_{\{\xi\}} - 1 \right). \quad (2.60)$$

A função partição replicada n vezes é o produto de n funções partição, correspondentes a n sistemas independentes:

$$\mathcal{Z}^n(\beta, h) = \prod_{\alpha} \sum_{\{S^\alpha\}} \exp(-\beta \mathcal{H}^\alpha) = \sum_{\{S^\alpha\}} \exp(-\beta \sum_{\alpha} \mathcal{H}^\alpha), \quad (2.61)$$

sendo

$$\mathcal{H}^\alpha = -\frac{1}{2} \sum_{i,j \neq i} \frac{1}{N} \sum_{\mu} \xi_i^\mu \xi_j^\mu S_i^\alpha S_j^\alpha - \sum_{\mu \leq s} h^\mu \sum_i \xi_i^\mu S_i^\alpha. \quad (2.62)$$

Assim temos

$$\langle \mathcal{Z}^n(\beta, h) \rangle = e^{-\frac{1}{2}\beta p n} \left\langle \sum_{\{S^\alpha\}} \exp \left[\frac{\beta}{2N} \sum_{\mu\alpha} \left(\sum_i \xi_i^\mu S_i^\alpha \right)^2 + \beta \sum_{\mu \leq s, \alpha} h^\mu \sum_i \xi_i^\mu S_i^\alpha \right] \right\rangle_{\{\xi\}}. \quad (2.63)$$

O termo quadrático no argumento da exponencial pode ser linearizado com o auxílio da integral gaussiana (equação 2.25):

$$\begin{aligned} \langle \mathcal{Z}^n(\beta, h) \rangle &\approx e^{-\frac{1}{2}\beta p n} \sum_{\{S^\alpha\}} \int \prod_{\mu\alpha} \left(\frac{dm_\alpha^\mu}{\sqrt{2\pi}} \right) \\ &\cdot \left\langle \exp \beta N \left[-\frac{1}{2} \sum_{\mu > s, \alpha} (m_\alpha^\mu)^2 + \sum_{\mu > s, \alpha} m_\alpha^\mu \frac{1}{N} \sum_i \xi_i^\mu S_i^\alpha \right] \right\rangle_{\{\xi\}, \mu > s} \\ &\cdot \left\langle \exp \beta N \left[-\sum_{\mu \leq s, \alpha} (m_\alpha^\mu)^2 + \sum_{\mu \leq s, \alpha} (m_\alpha^\mu + h^\mu) \frac{1}{N} \sum_i \xi_i^\mu S_i^\alpha \right] \right\rangle_{\{\xi\}, \mu \leq s} \end{aligned} \quad (2.64)$$

O Sinal ‘ \approx ’ indica que um fator multiplicativo foi desprezado, por não ser importante no cálculo da energia livre. Na equação (2.62) estão separados os fatores que dependem das componentes ‘altas’ ($\mu > s$) e ‘baixas’ ($\mu \leq s$) dos ‘overlaps’. Após tomar a média sobre os padrões que não condensam macroscopicamente e calcular a integral sobre as componentes ‘altas’ de \mathbf{m} (ver apêndice A) temos

$$\begin{aligned} \langle \mathcal{Z}^n(\beta, h) \rangle &\approx e^{-\frac{1}{2}\beta pn} \int \prod_{\mu \leq s, \alpha} \left(\frac{dm_\alpha^\mu}{\sqrt{2\pi}} \right) \int \prod_{\alpha < \beta} (dq_{\alpha\beta} dr_{\alpha\beta}) \\ &\cdot \exp N \left\{ -\frac{1}{2}\beta \sum_{\mu \leq s, \alpha} (m_\alpha^\mu)^2 - \frac{1}{2}\alpha \text{tr} \ln[(1 - \beta)\mathbf{I} - \beta\mathbf{q}] - \frac{1}{2}\alpha\beta^2 \sum_{\alpha \neq \beta} r_{\alpha\beta} q_{\alpha\beta} \right. \\ &\left. + \left\langle \ln \sum_{\{S^\alpha\}} \exp \left[\frac{1}{2}\alpha\beta^2 \sum_{\alpha\beta} r_{\alpha\beta} S^\alpha S^\beta + \beta \sum_{\mu \leq s, \alpha} (m_\alpha^\mu + h^\mu) \xi^\mu S^\alpha \right] \right\rangle_{\{\xi\}, \mu \leq s} \right\} \end{aligned} \quad (2.65)$$

Quando $N \rightarrow \infty$ a integral é dominada pelos valores das variáveis de integração que maximizam o argumento da exponencial:

$$\begin{aligned} \langle \mathcal{Z}^n(\beta, h) \rangle &\sim \max \exp \beta N n \left[-\frac{1}{2}\alpha - \frac{1}{2n} \sum_{\mu \leq s, \alpha} (m_\alpha^\mu)^2 - \frac{1}{2\beta n} \alpha \ln[(1 - \beta)\mathbf{I} - \beta\mathbf{q}] \right. \\ &\left. - \frac{1}{2n} \alpha\beta \sum_{\alpha \neq \beta} r_{\alpha\beta} q_{\alpha\beta} + \frac{1}{\beta n} \left\langle \ln \sum_{\{S^\alpha\}} \exp \beta \mathcal{H}_{eff} \right\rangle_{\{\xi\}} \right], \end{aligned} \quad (2.66)$$

sendo \mathcal{H}_{eff} o hamiltoniano efetivo

$$\mathcal{H}_{eff} = \frac{1}{2}\alpha\beta \sum_{\alpha\beta} r_{\alpha\beta} S^\alpha S^\beta + \sum_{\mu \leq s, \alpha} (m_\alpha^\mu + h^\mu) \xi^\mu S^\alpha. \quad (2.67)$$

A expressão (2.64) representa o termo dominante de 2.63) para $N \rightarrow \infty$. Estamos interessados no limite $n \rightarrow 0$, com a condição adicional $Nn \rightarrow 0$. De 2.58 e 2.64 obtém-se a energia livre por neurônio, que é o mínimo da expressão

$$\begin{aligned} f(\beta) &= \frac{1}{2}\alpha + \frac{1}{2n} \sum_{\mu \leq s, \alpha} (m_\alpha^\mu)^2 + \frac{1}{2\beta n} \alpha \ln[(1 - \beta)\mathbf{I} - \beta\mathbf{q}] \\ &+ \frac{1}{2n} \alpha\beta \sum_{\alpha\beta} r_{\alpha\beta} q_{\alpha\beta} - \frac{1}{\beta n} \left\langle \ln \sum_{\{S^\alpha\}} \exp \beta \mathcal{H}_{eff} \right\rangle_{\{\xi^\mu\}}, \end{aligned} \quad (2.68)$$

nas variáveis m_α^μ , $r_{\alpha\beta}$ e $q_{\alpha\beta}$. Na expressão acima está implícito que a média é sobre os padrões $\mu \leq s$. As equações de ponto de sela

$$\frac{\partial f(\beta)}{\partial m_\alpha^\mu} = 0, \quad (2.69)$$

$$\frac{\partial f(\beta)}{\partial q_{\alpha\beta}} = 0 \quad (2.70)$$

e

$$\frac{\partial f(\beta)}{\partial r_{\alpha\beta}} = 0 \quad (2.71)$$

representam condições necessárias para que a energia livre por neurônio $f(\beta)$ seja um mínimo.

O significado físico das variáveis m_α^μ , $r_{\alpha\beta}$ e $q_{\alpha\beta}$ é obtido das equações de ponto de sela. Derivando a equação (2.63) em relação a m_α^μ , temos para os valores estacionários de $\langle \mathcal{Z}^n(\beta, h) \rangle$:

$$m_\alpha^\mu = \left\langle \frac{1}{N} \sum_i \xi_i^\mu S_i^\alpha \right\rangle. \quad (2.72)$$

Do mesmo modo, para os valores estacionários de $\langle \mathcal{Z}^n(\beta, h) \rangle$ em relação a $r_{\alpha\beta}$ obtém-se

$$q_{\alpha\beta} = \left\langle \frac{1}{N} \sum_i \langle S_i^\alpha \rangle_T \langle S_i^\beta \rangle_T \right\rangle_{\{\xi\}}. \quad (2.73)$$

Na expressão acima está implícita a independência entre as réplicas. Desta expressão identificamos em $q_{\alpha\beta}$ o parâmetro de ordem de vidro de spin de Edwards e Anderson [31].

Igualmente, do valor estacionário de $\langle \mathcal{Z}^n(\beta, h) \rangle$ em relação a $q_{\alpha\beta}$ temos

$$r_{\alpha\beta} = \left\langle \frac{1}{\alpha} \sum_{\mu>s,\alpha} \left\langle \frac{1}{N} \sum_i \xi_i^\mu S_i^\alpha \right\rangle_T \left\langle \frac{1}{N} \sum_i \xi_i^\mu S_i^\beta \right\rangle_T \right\rangle_{\{\xi\}, \mu>s}, \quad (2.74)$$

de modo que $r_{\alpha\beta}$ é identificado como a média quadrática das componentes 'altas' de \mathbf{m} .

A solução de simetria de réplicas é aquela em que

$$\begin{aligned} m_\alpha^\mu &= m^\mu, \\ q_{\alpha\beta} &= q, \\ r_{\alpha\beta} &= r, \end{aligned} \quad (2.75)$$

para qualquer réplica ou par de réplicas diferentes.

A etapa seguinte deste desenvolvimento é determinar a forma que a expressão 2.66 assume na condição de simetria de réplicas. Para determinar o terceiro termo daquela expressão, nota-se que a matriz

$$[(1 - \beta)\mathbf{I} - \beta\mathbf{q}], \quad (2.76)$$

possei um autovalor

$$\lambda_1 = 1 - \beta + \beta q, \quad (2.77)$$

$(n - 1)$ vezes degenerado, e outro

$$\lambda_2 = 1 - \beta - (n - 1)\beta q, \quad (2.78)$$

não degenerado. Portanto, temos

$$\begin{aligned} \lim_{n \rightarrow 0} \frac{1}{n} \text{tr} \ln[(1 - \beta)\mathbf{I} - \beta] &= \lim_{n \rightarrow 0} \frac{1}{n} \ln [(1 - \beta + \beta q)^{n-1} [1 - \beta - (n - 1)\beta q]] \\ &= \ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q}. \end{aligned} \quad (2.79)$$

O último termo de (2.66) pode ser reescrito como

$$\frac{1}{n} \left\langle \ln \sum_{\{S^\alpha\}} \exp \left[\frac{1}{2} \alpha \beta^2 r (\sum_\alpha S^\alpha)^2 - \frac{1}{2} n \alpha \beta^2 r + \beta \sum_\mu (m^\mu + h^\mu) \xi^\mu \sum_\alpha S^\alpha \right] \right\rangle_{\{\xi\}} \quad (2.80)$$

Linearizando o termo quadrático em S^α no argumento da exponencial com o auxílio da integral gaussiana (equação 2.25) e tomando o traço sobre S^α , temos

$$-\frac{1}{2} \alpha \beta^2 r + \frac{1}{n} \left\langle \ln \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} + n \ln \left[2 \cosh \beta [\sqrt{\alpha r} z + \sum_\mu (m^\mu + h^\mu) \xi^\mu] \right] \right] \right\rangle_{\{\xi^\mu\}} \quad (2.81)$$

No limite $n \rightarrow 0$, após expandir até a primeira ordem em n temos

$$-\frac{1}{2} \alpha \beta^2 r + \int_{-\infty}^{+\infty} \mathcal{D}z \left\langle \ln \left[2 \cosh \beta [\sqrt{\alpha r} z + \sum_\mu (m^\mu + h^\mu) \xi^\mu] \right] \right\rangle_{\{\xi^\mu\}}, \quad (2.82)$$

sendo

$$\int \mathcal{D}z \equiv \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{\frac{1}{2} z^2}. \quad (2.83)$$

Reunindo os termos acima, a energia livre (2.66) resume-se à expressão

$$\begin{aligned} f(\beta) &= \frac{1}{2} \alpha + \frac{1}{2} \sum_\mu (m^\mu)^2 + \frac{\alpha}{2\beta} \left[\ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q} \right] + \frac{1}{2} \alpha \beta r (1 - q) \\ &\quad - \frac{1}{\beta} \int_{-\infty}^{+\infty} \mathcal{D}z \left\langle \ln \left[2 \cosh \beta [\sqrt{\alpha r} z + \sum_\mu (m^\mu + h^\mu) \xi^\mu] \right] \right\rangle_{\{\xi^\mu\}}. \end{aligned} \quad (2.84)$$

Da expressão acima obtemos as equações de ponto de sela

$$m^\mu = \int \mathcal{D}z \left\langle \xi^\mu \tanh \beta [\sqrt{\alpha r} z + \sum_\mu (m^\mu + h^\mu) \xi^\mu] \right\rangle_{\{\xi^\mu\}} \quad (2.85)$$

do valor estacionário de $f(\beta)$ em relação a m^μ

$$q = \int \mathcal{D}z \left\langle \tanh^2 \beta [\sqrt{\alpha r} z + \sum_\mu (m^\mu + h^\mu) \xi^\mu] \right\rangle_{\{\xi^\mu\}} \quad (2.86)$$

do valor estacionário de $f(\beta)$ em relação a r , e

$$r = \frac{q}{(1 - \beta + \beta q)^2} \quad (2.87)$$

do valor estacionário de $f(\beta)$ em relação a q .

A presença do parâmetro de ordem de vidro de spin q na descrição da termodinâmica de redes neurais permite antecipar a existência de uma fase de vidro de spin a baixa temperatura. Esta fase é caracterizada por $m^\mu = 0$ para todo μ e $q \neq 0$, e é candidata a mínimo absoluto da energia livre quando os estados de recuperação são desestabilizados. A variável r representa o efeito das correlações de $\mathcal{O}(1/\sqrt{N})$ com os padrões que não condensam macroscopicamente.

Comportamento do sistema na ausência de ruído ($T = 0$)

Quando $\beta \rightarrow \infty$, o lado direito da equação (2.83) tem a forma

$$\int \mathcal{D}z \operatorname{sgn}(az + x) = \operatorname{erf} \left(\frac{x}{a\sqrt{2}} \right), \quad (2.88)$$

onde

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz \quad (2.89)$$

é a função erro. Portanto, em $T = 0$, (2.83) pode ser reescrita como

$$m^\mu = \left\langle \xi^\mu \operatorname{erf} \left(\frac{\sum_\mu (m^\mu + h^\mu) \xi^\mu}{\sqrt{2\alpha r}} \right) \right\rangle_{\{\xi^\mu\}} \quad (2.90)$$

No limite $\beta \rightarrow \infty$, (2.84) é reescrita como

$$1 - q = 2 \left\langle \int \mathcal{D}z \exp(-2\beta |\sqrt{\alpha r} z + x|) \right\rangle_{\{\xi^\mu\}}, \quad (2.91)$$

sendo

$$x \equiv \sum_{\mu} (m^{\mu} + h^{\mu}) \xi^{\mu} . \quad (2.92)$$

Do termo dominante de (2.89) nesse limite temos

$$\beta(1 - q) \equiv C = \sqrt{\frac{2}{\pi\alpha r}} \left\langle \exp \left[-\frac{1}{2\alpha r} \left(\sum_{\mu} (m^{\mu} + h^{\mu}) \xi^{\mu} \right)^2 \right] \right\rangle_{\{\xi^{\mu}\}} . \quad (2.93)$$

Para que C permaneça finito quando $\beta \rightarrow \infty$, q deve ir a 1. A equação (2.85) é reescrita então como

$$r = \frac{1}{(1 - C)^2} . \quad (2.94)$$

Assim, na ausência de ruído sináptico, o comportamento da rede neural é descrito pelo sistema de equações (2.88) e (2.91), sendo r dado em função de C por (2.92).

Os estados de recuperação são descritos por

$$\mathbf{m} = \pm(m, 0, 0, \dots, 0) . \quad (2.95)$$

Introduzindo (2.93) em (2.88) e (2.91), obtém-se

$$m = \operatorname{erf} \left(\frac{m}{\sqrt{2\alpha r}} \right) \quad (2.96)$$

e

$$C = \sqrt{\frac{2}{\pi\alpha r}} \exp \left(-\frac{m^2}{2\alpha r} \right) . \quad (2.97)$$

A chave para resolver este sistema de equações é definir uma variável auxiliar,

$$y = \frac{m}{\sqrt{2\alpha r}} . \quad (2.98)$$

Assim o sistema reduz-se para uma equação e uma variável:

$$y = \psi(y) , \quad (2.99)$$

onde

$$\psi(y) \equiv \frac{\operatorname{erf}(y)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-y^2}} . \quad (2.100)$$

A equação (2.97) sempre possui uma solução trivial $y = 0$, que corresponde à fase de vidro de spin com $\mathbf{m} = 0$ (lembre que $q \approx 1$). Esta é a única solução para $\alpha > \alpha_c$, onde

$$\left(\frac{d\psi(y)}{dy}\right) = 1 \quad (2.101)$$

em $\alpha = \alpha_c$, sendo y_c o valor de y segundo (2.97) em α_c . Em $\alpha = \alpha_c$ surge descontinuamente uma solução com $m \neq 0$, que corresponde à fase ferromagnética de recuperação. Estas soluções surgem sempre aos pares, pois $-m$, que corresponde ao ‘overlap’ com os anti-padrões, também é solução estável. α_c é determinado pela equação (2.99), que explicitamente é reescrita na forma

$$\frac{2}{\sqrt{\pi}} e^{-y^2} \frac{\sqrt{2\alpha_c} + \frac{2}{\sqrt{\pi}} e^{-y^2} + 2y \operatorname{erf}(y)}{\left(\sqrt{2\alpha_c} + \frac{2}{\sqrt{\pi}} e^{-y^2}\right)^2} = 1. \quad (2.102)$$

A solução simultânea de (2.97) e (2.100) resulta no valor crítico [8]

$$\alpha_c \approx 0,138. \quad (2.103)$$

Usando

$$m = \operatorname{erf}(y) \quad (2.104)$$

no valor de m em α_c ,

$$m \approx 0,967. \quad (2.105)$$

Estes resultados revelam que mesmo quando $p = \alpha N$, se $\alpha \leq \alpha_c$, existe uma distribuição de estados atratora em torno de cada padrão armazenado. A situação é similar ao que se verifica quando p é finito e a rede se encontra na presença de ruído sináptico (Amit, Gutfreund e Sompolinsky, 1985). No caso presente, as correlações microscópicas com os padrões que não condensam são a fonte de ruído. A fração de erros de recuperação *por padrão*, f_e , relaciona-se com m pela expressão

$$f_e = \frac{1}{2}(1 - m). \quad (2.106)$$

Com m dado por (2.103), temos a fração de erros de recuperação em $\alpha = \alpha_c$,

$$f_e|_{\alpha=\alpha_c} = 0,0165. \quad (2.107)$$

Para $\alpha > \alpha_c$ o sistema sofre uma transição descontínua para o estado desordenado, ‘esquecendo’ completamente sua memória, com a fração de erros de recuperação assumindo o valor

$$f_e|_{\alpha > \alpha_c} = 0,5. \quad (2.108)$$

De (2.55) e (2.61), a energia em $T = 0$ é dada por

$$E = -\frac{1}{2}m^2 + \frac{1}{2}\alpha(1-r). \quad (2.109)$$

Em $\alpha = \alpha_c$, para o estado de recuperação temos $E = -0,5014$. Para a solução de vidro de spin, com $\mathbf{m} = 0$, temos de (2.91)

$$C = \sqrt{\frac{2}{\pi\alpha r}}, \quad (2.110)$$

e resolvendo (2.92) para r temos

$$r = \left(1 - \sqrt{\frac{2}{\pi\alpha}}\right)^2, \quad (2.111)$$

que por meio de (2.107) resulta em

$$E_{SG} = -\frac{1}{\pi} - \sqrt{\frac{2\alpha}{\pi}}. \quad (2.112)$$

Em $\alpha = \alpha_c$,

$$E \approx -0,615, \quad (2.113)$$

que é o mínimo global do sistema. Diminuindo α , a energia do estado de vidro de spin aumenta, e em

$$\alpha = 0,051 \quad (2.114)$$

a energia do estado de recuperação passa a ser mínimo absoluto da energia.

Para finalizar a discussão sobre o comportamento da rede neural de Hopfield na ausência de ruído sináptico é interessante analisar os estados mistos simétricos, caracterizados por

$$\mathbf{m} = \pm m_n(1, 1, \dots, 1, 0, 0 \dots). \quad (2.115)$$

Agora as equações (2.88) e (2.91) assumem respectivamente as formas

$$m_n = \frac{1}{n} \left\langle z_n \operatorname{erf} \left(\frac{m_n z_n}{\sqrt{2\alpha r}} \right) \right\rangle_{z_n} \quad (2.116)$$

e

$$C = \sqrt{\frac{2}{\pi\alpha r}} \left\langle \exp \left(\frac{z_n^2 y_n^2}{2\alpha r} \right) \right\rangle_{z_n}, \quad (2.117)$$

sendo z_n a variável aleatória definida por (2.46). Introduz-se a variável auxiliar

$$y_n = \frac{m_n}{\sqrt{2\alpha r}}, \quad (2.118)$$

em termos da qual o sistema reduz-se a

$$y_n = \psi_n(y_n), \quad (2.119)$$

sendo

$$\psi_n(y_n) = \frac{1}{n} \frac{\langle z_n \operatorname{erf}(z_n y_n) \rangle_{z_n}}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} \langle \exp(-z_n^2 y_n^2) \rangle_{z_n}}. \quad (2.120)$$

A equação (2.117) é semelhante à equação (2.97). Para cada valor de n , soluções com $m_n \neq 0$ aparecem descontinuamente abaixo dos valores críticos α_n . Assim como no caso de p finito, apenas as soluções com n ímpar são localmente estáveis. Os valores de α_n são significativamente menores que α_c . Por exemplo, $\alpha_3 \approx 0,03$. Em geral, observa-se que

$$\alpha_n \sim \frac{1}{n}. \quad (2.121)$$

A situação é semelhante com o caso p finito, com $T \neq 0$. O ruído provocado pelas correlações de ordem $1/\sqrt{N}$, assim como o ruído de origem sináptica, desestabilizam os estados simétricos. Posteriormente mostraremos que os estados simétricos são favorecidos quando os padrões que condensam são correlacionados.

Comportamento do sistema na presença de ruído $T > 0$

Raciocinando em termos de um diagrama de fases termodinâmicas nos parâmetros intensivos α e T , até agora temos resultados apenas para $\alpha = 0$, $T \neq 0$ e $\alpha \neq 0$, $T = 0$. Verifica-se no primeiro caso, que para $T < 1$, os estados de recuperação estão presentes, e

quando $T = 1$ o sistema sofre uma transição contínua para a fase paramagnética. Por outro lado, no segundo caso os estados de recuperação existem se $\alpha < \alpha_c \approx 0,138$, enquanto que para $\alpha > \alpha_c$ só a fase de vidro de spin é estável, sendo descontínua a transição desta com a fase ferromagnética de recuperação. Veremos agora o que ocorre para ambos α e T diferentes de zero.

A transição entre as fases paramagnética e de vidro de spin é contínua. De (2.84) temos, para $q \rightarrow 0$, $m = 0 = h$,

$$q \approx \int \mathcal{D}z z^2 \beta^2 \alpha r = \beta^2 \alpha r . \quad (2.122)$$

Introduzindo em (2.120) a expressão (2.85) para r , temos até $\mathcal{O}(q)$,

$$q = \frac{\beta^2 \alpha}{(1 - \beta)^2} . \quad (2.123)$$

Esta equação só tem solução $q \neq 0$ se

$$1 = \frac{\beta^2 \alpha}{(1 - \beta)^2} , \quad (2.124)$$

que resulta em

$$T_G = 1 + \sqrt{\alpha} . \quad (2.125)$$

Para $T < T_G$ temos

$$q \approx \beta^2 \alpha r \approx T_G - T . \quad (2.126)$$

Para os estados de recuperação, as equações 2.83 e 2.84 tomam a forma

$$m = \int \mathcal{D}z \tanh \beta(\sqrt{\alpha r} z + m) \quad (2.127)$$

e

$$q = \int \mathcal{D}z \tanh^2 \beta(\sqrt{\alpha r} z + m) . \quad (2.128)$$

A solução numérica deste sistema de equações resulta no diagrama de fases da figura 2.1. Abaixo da curva T_M a fase de recuperação é meta-estável. A curva assinalada como T_C , obtida comparando a energia livre (equação 2.82) entre as fases de recuperação e vidro de spin, limita a região na qual a fase de recuperação é mínimo absoluto da energia livre.

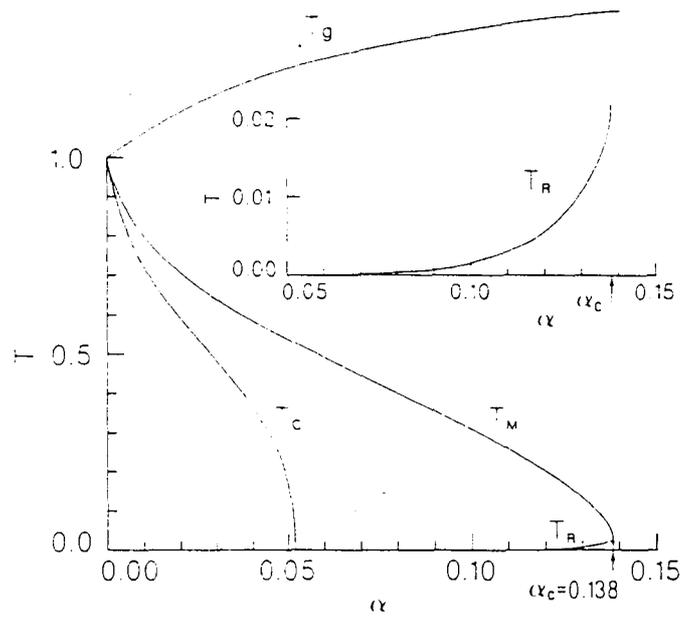


Figura 2.1: Diagrama de fases para o modelo de Hopfield. Extraído da referência [8]. A fase de recuperação é estável abaixo da curva T_M , e é mínimo global abaixo da curva T_c . A fase de vidro de spin é estável abaixo da curva T_g . A inserção mostra a ampliação do eixo vertical. Abaixo da curva T_R a solução de simetria de réplicas é instável.

Abaixo da curva assinalada como T_R a solução de simetria de réplicas é instável [8]. Entretanto, a quebra de simetria de réplicas tem pequeno efeito na determinação da capacidade crítica de armazenamento: em $T = 0$, o valor crítico de α com quebra de simetria é 0,145.

Até o presente discutimos apenas o armazenamento de padrões nos quais as probabilidades de que um determinado neurônio dispare ou não são iguais. Isto está em desacordo com observações em sistemas biológicos, que revelam que o número dos neurônios que disparam é bem menor do que os que permanecem quiescentes. No próximo capítulo iniciaremos a discussão do armazenamento de padrões correlacionados, que aproxima o modelo de redes de neurônios formais das evidências biológicas, e que constituem o assunto deste trabalho.

3. O modelo de Hopfield com padrões correlacionados

A distribuição de probabilidades (2.20) estabelece que durante a recuperação de um padrão a rede transitará entre estados em que 50% dos neurônios estarão ativos, e os outros 50% estarão quiescentes, em média. Neste caso os padrões possuem taxa de atividade média de 50%, e são qualificados como ‘não polarizados’, em virtude do equilíbrio entre o número de neurônios ativos e inativos.

A restrição determinada pela não polarização é incompatível com o que é observado em sistemas biológicos segundo a discussão na introdução deste trabalho. Ela pode ser eliminada, admitindo-se que o conjunto de padrões aleatórios obedeça a distribuição de probabilidades

$$P(\{\xi_i^\mu\}) = \prod_{i,\mu} P(\xi_i^\mu) = \prod_{i,\mu} \left[\frac{1}{2}(1+a)\delta(\xi_i^\mu - 1) + \frac{1}{2}(1-a)\delta(\xi_i^\mu + 1) \right], \quad (3.1)$$

sendo

$$-1 \leq a \leq 1.$$

Denominamos *atividade média* de um padrão a fração de sítios em que este padrão assume o estado +1. O parâmetro a , denominado *polarização*, permite descrever o comportamento de sistemas desde a atividade média igual a zero (quando $a = -1$, $\xi_i^\mu = -1$ para todo i, μ), até 100% (quando $a = 1$, $\xi_i^\mu = 1$, para todo i, μ). Em virtude da simetria perante a inversão do estado de todos neurônios da rede, é suficiente estudar apenas o intervalo $0 \leq a \leq 1$. O aspecto relevante é o quanto a rede está distante da atividade de 50%.

De (3.1), temos

$$\langle \xi_i^\mu \rangle_{\{\xi_i^\mu\}} = a \quad (3.2)$$

e

$$\langle \xi_i^\mu \xi_j^\nu \rangle_{\{\xi_i^\mu\}} = \delta_{\mu\nu} + a^2(1 - \delta_{\mu\nu}) . \quad (3.3)$$

Embora estatisticamente independentes, diz-se que conjuntos de padrões polarizados, que obedecem (3.1), são denominados ‘correlacionados’. Apesar de imprópria (se os padrões são estatisticamente independentes eles não são realmente correlacionados), manteremos esta denominação ao longo deste trabalho.

Se um conjunto de padrões correlacionados é impresso nas conexões sinápticas da rede neural por meio da regra de Hebb (equação 2.11), então o campo local sobre o neurônio i , no estado $S_i = \xi_i^\nu$ é dado por

$$h_i = \frac{1}{N} \sum_{j \neq i} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \approx \xi_i^\nu (1 + \delta_i) , \quad (3.4)$$

onde

$$\delta_i = \sum_{\mu \neq \nu} \frac{1}{N} \sum_{j \neq i} \xi_j^\mu \xi_j^\nu \xi_i^\mu \xi_i^\nu = a^2 \sum_{\mu \neq \nu} \xi_i^\mu \xi_i^\nu . \quad (3.5)$$

é o termo de ruído.

A soma em (3.5) possui $p - 1$ termos iguais a $+1$ ou -1 . Em um número significativo de neurônios o termo de ruído alcança o valor $(p - 1)a^2$. Quando este valor é maior que 1, o termo de ruído no campo local sobre estes neurônios será maior que o sinal, e de acordo com a dinâmica do sistema (equação 2.3), ocorrerá a desestabilização dos padrões. Portanto, para cada p existe um valor crítico da polarização, dado pela equação

$$(p - 1)a_c^2 = 1 , \quad (3.6)$$

acima do qual não existe estado de recuperação. Isto revela que a regra de Hebb não é apropriada para o armazenamento de padrões fortemente correlacionados.

Consideremos a situação $T = 0$, $a < a_c$. Se o sistema se encontra no estado de recuperação do padrão μ , com $S_i = \xi_i^\mu$ para qualquer i , então de acordo com a definição

de ‘overlap’ (equação 2.15), $m^\mu = 1$. Nesta situação, as outras componentes do vetor \mathbf{m} relativas a padrões correlacionados serão, segundo (3.3),

$$m^\nu = \frac{1}{N} \sum_{i=1} \langle \xi_i^\mu \xi_i^\nu \rangle_{\{\xi_i^\mu\}} = a^2, \quad \mu \neq \nu. \quad (3.7)$$

Portanto, os estados de recuperação são caracterizados por vetores ‘overlap’ do tipo

$$\mathbf{m} = (1, a^2, a^2, \dots, a^2). \quad (3.8)$$

Se $a \neq 1$, os estados de recuperação são assimétricos, e misturam todos padrões, embora privilegiando um deles. Isto é uma consequência trivial da correlação entre os padrões. Se $a = 0$, recuperamos o estado de Mattis [32] (equação 2.43, com $n = 1$).

Quando $a > a_c$ o sistema sofre uma transição descontínua para uma fase cujos estados, denominados *espúrios* ou *mistos* são caracterizados por componentes $m^\mu < 1$ e $m^\nu > a^2$, para $\mu \neq \nu$ (segundo Fontanari e Theumann [33]). Embora a capacidade de memória associativa esteja comprometida, acreditamos que esta situação mereça ser examinada em detalhe. A justificativa para isto é que de acordo com um modelo devido a Fontanari [34], determinados estados mistos desempenham um papel importante durante o processo de aprendizado: o sistema inicialmente armazena um conjunto de padrões que misturam diversos conceitos afins, sem individualizá-los, e apenas posteriormente identifica as diferenças entre eles. Diante da relativa simplicidade analítica que a regra de aprendizagem de Hebb oferece, acreditamos que ela é apropriada para o estudo destes estados.

Neste capítulo discutiremos as consequências termodinâmicas do armazenamento de padrões que obedecem a distribuição de probabilidades 3.1. Na seção 3.1 deste capítulo discutiremos as denominadas *redes biológicas*, correspondentes a alterações do modelo de Hopfield, que o habilitam a armazenar padrões fortemente polarizados [9, 18, 17]. Na seção 3.2 estenderemos para o caso $a \neq 0$ [35] o desenvolvimento formal da seção 2.2. Na seção 3.3 estudaremos a capacidade macroscópica de armazenamento oferecida pela regra de Hebb, generalizada para $a \neq 0$. No final deste capítulo apresentaremos resultados de simulações numéricas envolvendo o armazenamento de padrões correlacionados.

3.1 Redes biológicas

O ruído gerado pela correlação microscópica entre os padrões polarizados (equação 3.5) prejudica o seu armazenamento no modelo de Hopfield original. Para cada valor de p existe um valor crítico a_c , dado pela equação 3.6 acima do qual é desestabilizado o estado de recuperação.

O trabalho original de Gardner [23], cuja generalização será abordada no capítulo 4, revela que a capacidade de armazenamento de padrões correlacionados em redes neurais atratoras é monotonamente crescente em $|a|$, e diverge quando $|a| \rightarrow 1$ de acordo com

$$\alpha_c(a) \sim -\frac{1}{(1 - |a|) \ln(1 - |a|)} . \quad (3.9)$$

Note-se que embora a divergência em α_c para $|a| \rightarrow 1$, o conteúdo de informação armazenada na rede é nulo neste caso.

O objetivo desta seção é discutir alterações no modelo de Hopfield que permitem recuperar a capacidade macroscópica de armazenamento de padrões polarizados e inclusive alcançar o limite assintótico ideal representado pela equação (3.9).

3.1.1 O modelo AGS

Uma providência imediata que permite recuperar a capacidade finita de armazenamento para $a \neq 0$ é reduzir a zero o valor médio do ruído 3.5. Isto é alcançado através da regra de aprendizado de Amit, Gutfreund e Sompolinsky [9],

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

Redefinindo os ‘overlaps’ de acordo com

$$m^{\mu} = \frac{1}{N} \sum_i (\xi_i^{\mu} - a) \langle S_i \rangle_T , \quad (3.11)$$

a análise em teoria de campo médio segue os passos da subseção 2.2.3.. Os estados de recuperação são caracterizados por vetores \mathbf{m} do tipo definido pela equação (2.94). Em $T = 0$ a capacidade crítica de armazenamento $\alpha_c = \alpha_c(a)$ é apresentada na figura 3.1.

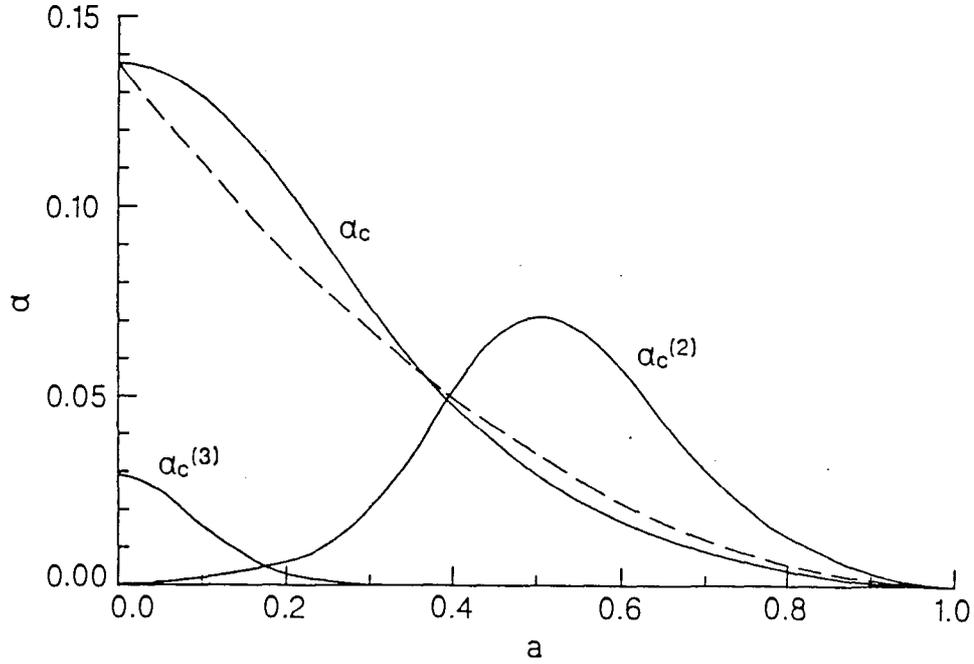


Figura 3.1: Capacidade crítica de armazenamento α_c em função da polarização para o estado de recuperação e estados simétricos que misturam dois ($\alpha_c^{(2)}$) e três ($\alpha_c^{(3)}$) padrões. A curva tracejada representa o resultado da análise sinal-ruído (extraído da referência [9]).

Verifica-se que $\alpha_c(a)$ é uma função monotonamente decrescente de a , com $\alpha_c(a = 1) = 0$. A mesma figura também mostra os valores críticos de α para as soluções que misturam 2 e 3 padrões.

O comportamento insuficiente do modelo AGS fica caracterizado pelo fato de que para $a \geq 0,4$ o estado simétrico que mistura dois padrões permanece estável para valores de α maiores que α_c .

A magnetização média dos estados de recuperação dos padrões que obedecem a distribuição de probabilidades definida por 3.1 é

$$\sum_i \langle \xi_i^\mu \rangle = Na. \quad (3.12)$$

O modelo AGS pode ser aprimorado se a dinâmica do sistema for restrita aos estados com a mesma magnetização dos estados de recuperação. Isto corresponde a introduzir o vínculo

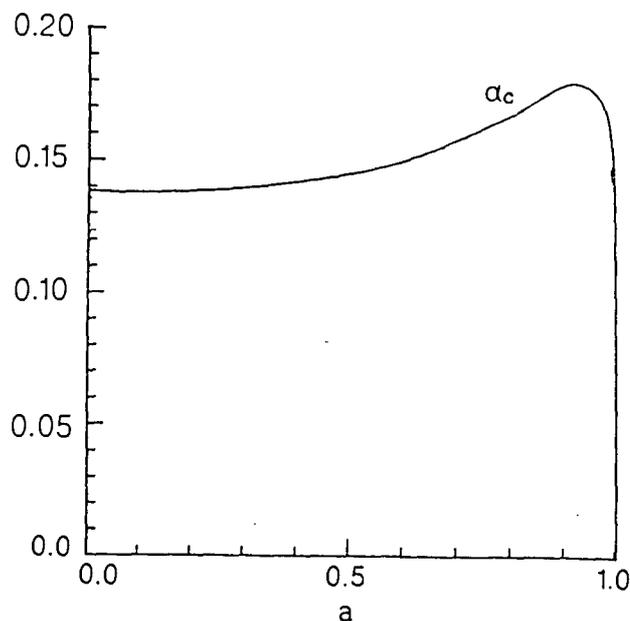


Figura 3.2: Capacidade crítica de armazenamento α_c em função da polarização da rede com vínculo dinâmico (extraído da referência [9]).

dinâmico

$$M \equiv \sum_i \langle S_i \rangle = Na, \quad (3.13)$$

por meio de um campo externo h acoplado a $\sum_i S_i$. O resultado obtido é a eliminação dos estados espúrios que não obedecem o vínculo, e o conseqüente incremento na capacidade de armazenamento, como pode ser verificado na figura 3.2. A capacidade de armazenamento é função monotonamente crescente da polarização para $|a| \leq 0,9$, embora ainda caia abruptamente a zero quando $|a| \rightarrow 1$.

3.1.2 O modelo de Perez-Vicente e Amit

A representação σ é mais conveniente que a representação S para o armazenamento de padrões fortemente correlacionados [17, 36]. Na representação σ verifica-se uma divergência na capacidade de armazenamento para $|a| \rightarrow 1$ [17]. É óbvio que a simples troca de representação é apenas um passo formal. Entretanto, a sua execução é acompanhada da introdução de campos locais que obrigam os neurônios a operarem em torno de seus limiares, onde são mais sensíveis aos estímulos externos [29].

A mudança para a representação σ é um caso particular da transformação mais geral proposta por Perez-Vicente e Amit [18]:

$$S_i \rightarrow \sigma_i = \frac{1}{2}(S_i - b) , \quad (3.14)$$

com $b \in [-1, 1]$. O campo local sobre cada neurônio assume a forma

$$h_i = \sum J_{ij}(S_i - b) - U , \quad (3.15)$$

onde J_{ij} é dado pela expressão (3.10) e U é um campo uniforme a ser ajustado para otimizar a capacidade de armazenamento.

Uma simples análise sinal-ruído permite estimar os valores dos parâmetros b e U que aprimoram o desempenho da rede. Quando a rede se encontra no estado de recuperação do padrão ν o campo local (3.15) é dado por

$$h_i = \frac{1}{N} \sum_j \sum_{\mu=1}^p (\xi_i^\mu - a)(\xi_j^\mu - a)(\xi_j^\nu - b) - U \quad (3.16)$$

separando os termos de sinal e ruído, temos

$$h_i = S_i + R_i , \quad (3.17)$$

sendo

$$S_i = \frac{1}{N} (\xi_i^\nu - a) \sum_j (\xi_j^\nu - a)(\xi_j^\nu - b) - U \quad (3.18)$$

o sinal e

$$R_i = \frac{1}{N} \sum_j (\xi_i^\mu - a)(\xi_j^\mu - a)(\xi_j^\nu - b) \quad (3.19)$$

o ruído. Para $N \rightarrow \infty$,

$$S \approx (1 - a)^2 (\xi_j^\nu - a) - U . \quad (3.20)$$

Se $U = 0$,

$$S = \begin{cases} (1 - a)^2(1 + a) \rightarrow 2(1 - a)^2 & \text{ao } a \rightarrow 1, \text{ se } \xi_i^\nu = 1 , \\ -(1 - a)(1 + a)^2 \rightarrow -4(1 - a) & \text{ao } a \rightarrow 1, \text{ se } \xi_i^\nu = -1 . \end{cases} \quad (3.21)$$

Das expressões acima vemos que

$$|S(\xi_i^\nu = -1)| \gg |S(\xi_i^\nu = +1)| \quad a \rightarrow 1. \quad (3.22)$$

A escolha do campo uniforme

$$U = -a(1 - a^2) \quad (3.23)$$

permite uniformizar o sinal:

$$|S| = 1 - a^2, \quad a \rightarrow 1, \quad (3.24)$$

tanto para $\xi_i^\mu = +1$ quanto para $\xi_i^\mu = -1$.

O termo de ruído (3.19) obedece a uma distribuição gaussiana centrada em zero, com a média quadrática dada por

$$\langle R^2 \rangle \approx \alpha(1 - a^2)(1 - 2ab + b^2). \quad (3.25)$$

A escolha

$$b = a \quad (3.26)$$

permite minimizar o valor de $\langle R \rangle$, o que resulta em

$$\langle R^2 \rangle \approx \alpha(1 - a^2)^3. \quad (3.27)$$

No limite $a \rightarrow 1$ a relação sinal-ruído torna-se

$$\frac{|S|}{\sqrt{\langle R^2 \rangle}} \sim \frac{1}{\sqrt{\alpha(1 - a^2)}}, \quad (3.28)$$

que sugere a capacidade de armazenamento divergente

$$\alpha_c \sim \frac{1}{1 - a^2}. \quad (3.29)$$

Uma análise mais detalhada do modelo de Perez-Vicente e Amit em teoria de campo médio revela que no limite de padrões fortemente correlacionados ($a \rightarrow 1$) a capacidade de armazenamento é similar à prevista por Gardner [23] (equação 3.9), para a escolha ótima das conexões sinápticas J_{ij} .

3.2 Capacidade microscópica de armazenamento

Na discussão que segue, por simplicidade mas sem perda de generalidade, suporemos que todos p padrões armazenados na rede são correlacionados.

3.2.1 Solução simétrica

As soluções simétricas da equação (2.35), estudadas por Fontanari e Theumann [33] são do tipo

$$\mathbf{m} = m_p(1, 1, \dots, 1), \quad (3.30)$$

e representam estados em que o sistema possui overlap igual com todos p padrões correlacionados. De (2.34), (2.35) e 3.30 obtemos a energia livre

$$f(\beta) = \frac{1}{2}p m_p^2 - \frac{1}{\beta} \langle \ln[2 \cosh(\beta m_p z_p)] \rangle_{z_p}, \quad (3.31)$$

e a equação de ponto de sela para m_p ,

$$m_p = \frac{1}{p} \langle z_p \tanh(\beta m_p z_p) \rangle_{z_p}, \quad (3.32)$$

onde a variável aleatória z_p é definida por (2.46). A polarização dos padrões implica, de acordo com (3.1), que

$$P(\xi_i^\mu = \pm 1) = \frac{1}{2}(1 \pm a). \quad (3.33)$$

Conseqüentemente, z_p obedece a uma distribuição de probabilidades binomial,

$$P(z_p = 2k - p) = \binom{p}{k} b_1^k b_2^{p-k}, \quad (3.34)$$

onde k é um inteiro positivo,

$$\binom{p}{k} = \frac{p!}{k!(p-k)!}, \quad (3.35)$$

$$b_1 = \frac{1}{2}(1 + a) \quad (3.36)$$

e

$$b_2 = \frac{1}{2}(1 - a). \quad (3.37)$$

Temperatura zero

De (3.31), temos para a energia livre por neurônio,

$$f(T = 0) = -\frac{1}{2}p m_p^2 . \quad (3.38)$$

No limite de temperatura zero ($\beta \rightarrow \infty$) da equação (3.32) para $m_p (> 0)$

$$m_p = \frac{1}{p} \langle z_p \text{sgn}(z_p) \rangle_{z_p} = \frac{1}{p} \langle |z_p| \rangle_{z_p} . \quad (3.39)$$

obtem-se a forma explícita do 'overlap' para a solução simétrica:

$$m_p = \binom{2l}{l} (b_1 b_2)^l + a \sum_{k=0}^{p-1} \binom{p-1}{k} (b_1^{p-1-k} b_2^k - b_1^k b_2^{p-1-k}) , \quad (3.40)$$

onde

$$l \equiv \text{inteiro} \left(\frac{p}{2} \right) \quad (3.41)$$

é a parte inteira de $p/2$.

Temperatura finita

Expandindo-se (3.32) para $m_p \rightarrow 0$, temos

$$m_p = \frac{1}{p} \beta m_p \langle z_p^2 \rangle_{z_p} + \mathcal{O}(m_p^3) , \quad (3.42)$$

sendo

$$\langle z_p^2 \rangle_{z_p} = p[1 + (p-1)a^2] \quad (3.43)$$

Obtem-se daí a temperatura crítica

$$T_C = 1 + (p-1)a^2 . \quad (3.44)$$

Em $T = T_C$ o sistema sofre uma transição contínua para uma fase paramagnética com a solução única $m_p = 0$.

Estabilidade das soluções simétricas

Para qualquer temperatura entre $T = 0$ e $T = T_C$, o valor de m_p é obtido resolvendo-se numericamente a equação (3.32). As soluções desta equação são extremos da energia livre.

Para que os estados correspondentes sejam estáveis, é necessário que sejam mínimos. A condição que garante a estabilidade é que todos autovalores da matriz \mathbf{A} , cujos elementos são as derivadas segundas da energia livre, sejam positivos. Para as soluções simétricas os elementos de \mathbf{A} são

$$A_{\mu\nu} = \delta_{\mu\nu}(1 - \beta + \beta q) + (1 - \delta_{\mu\nu})\beta(Q - a^2), \quad (3.45)$$

onde

$$q = \langle \tanh^2(\beta m_p z_p) \rangle_{z_p} \quad (3.46)$$

e

$$Q = \langle \xi_1 \xi_2 \tanh^2(\beta m_p z_p) \rangle_{z_p} \quad (3.47)$$

Os autovalores são:

$$\lambda_1 = 1 - \beta(1 - q) + \beta(p - 1)(Q - a^2), \quad (3.48)$$

que é não degenerado, e

$$\lambda_2 = 1 - \beta(1 - q) - \beta(Q - a^2), \quad (3.49)$$

que é $(p - 1)$ vezes degenerado.

Em $T \rightarrow 0$ a estabilidade da solução simétrica depende da paridade de p . De (3.46) e (3.47) vemos que

$$q(T \rightarrow 0) \approx 1 - P(z_p = 0) \quad (3.50)$$

e

$$Q(T \rightarrow 0) \approx a^2 - P(z_p = 0), \quad (3.51)$$

Onde $P(z_p = 0)$ é a probabilidade de z_p ser nulo. Mas $P(z_p = 0) \neq 0$ somente se p é par. Portanto, $\lambda_2 < 0$ se p é par, e as soluções simétricas pares são sempre instáveis (a $T = 0$).

Para temperatura finita, a determinação da estabilidade envolve a análise numérica de (3.32) e o cálculo dos autovalores λ_1 e λ_2 . As figuras 3.3a e 3.3b (de Fontanari e Theumann, ref. [33], resultados inéditos) identificam as regiões de estabilidade no plano (a, T) das soluções simétricas pares e ímpares, respectivamente. A região estável localiza-se sempre à direita de cada curva. Para cada valor de p , a uma determinada temperatura, existe um

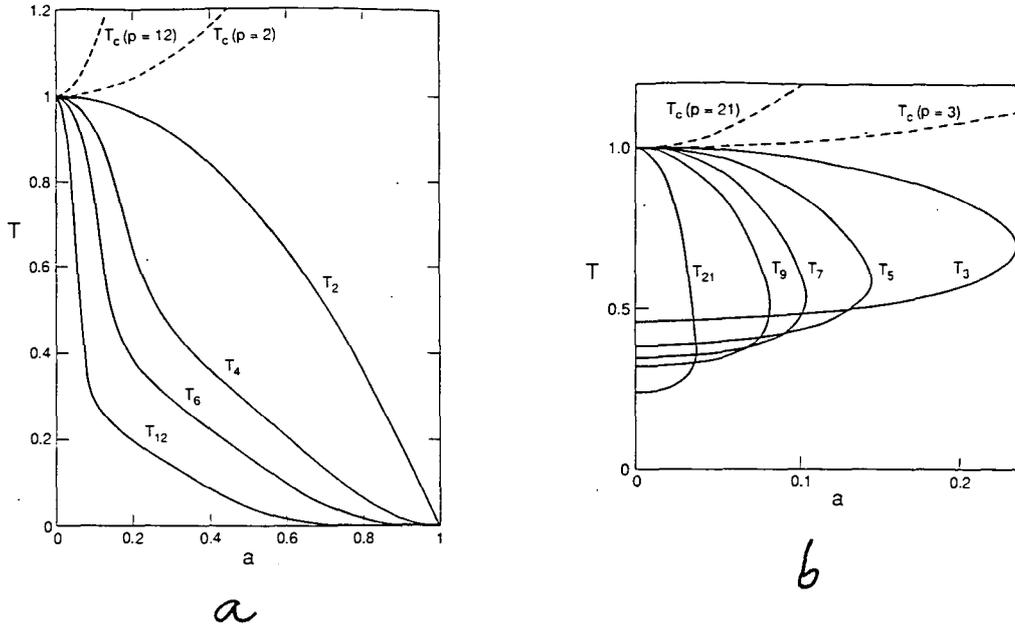


Figura 3.3: Estabilidade das soluções simétricas. (a) pares; (b) ímpares. A solução é estável à esquerda da respectiva curva. (extraído da referência [33])

valor de a acima do qual a solução simétrica é estável. Em outras palavras, a correlação entre os padrões, resultante de suas polarizações, estabiliza as soluções simétricas.

3.2.2 Soluções assimétricas

As soluções assimétricas do sistema de equações (2.35) são todas aquelas cujas componentes de \mathbf{m} assumem valores diferentes entre si. Particularizaremos a discussão em torno de soluções assimétricas dos tipos

$$\mathbf{m} = (m_1, m_{p-1}, m_{p-1}, \dots, m_{p-1}) \quad (3.52)$$

e

$$\mathbf{m} = (m_1, m_1, m_{p-2}, m_{p-2}, \dots, m_{p-2}) . \quad (3.53)$$

A solução de recuperação a $T = 0$, definida pela equação (3.8), é um caso particular de (3.52). Combinando (3.52) e (2.35) temos o sistema formado pelas equações

$$m_1 = \langle b_1 \tanh(\beta\Theta_+) + b_2 \tanh(\beta\Theta_-) \rangle_{z_{p-1}} \quad (3.54)$$

e

$$m_{p-1} = \langle \frac{z_{p-1}}{p-1} [b_1 \tanh(\beta\Theta_+) - b_2 \tanh(\beta\Theta_-)] \rangle_{z_{p-1}}, \quad (3.55)$$

onde

$$\Theta_{\pm} = m_1 \pm m_{p-1} z_{p-1}. \quad (3.56)$$

z_{p-1} é a soma definida por (2.42) mas com $p-1$ termos em lugar de p .

Soluções assimétricas à temperatura zero

Em $T = 0$ o sistema de equações acima tem soluções analíticas [33].

$$m_1 = 1 - 2 \sum_{k=0}^{l-1} \binom{p-1}{k} (b_1^{k+1} b_2^{p-1-k} + b_1^{p-1-k} b_2^{k+1}) \quad (3.57)$$

e

$$\begin{aligned} m_{p-1} = a^2 &+ 2a \sum_{k=0}^{l-2} \binom{p-2}{k} (b_1^{p-2-k} b_2^{k+1} - b_1^{k+1} b_2^{p-2-k}) \\ &+ 2(1 - \delta_{l0}) \binom{p-2}{l-1} (b_1^{p-l} b_2^l + b_1^l b_2^{p-l}), \end{aligned} \quad (3.58)$$

sendo

$$l \equiv \text{inteiro} \left[\frac{1}{2} \left(p + 1 - \frac{m_1}{m_{p-1}} \right) \right]. \quad (3.59)$$

As expressões (3.57) e (3.58) não são soluções explícitas para m_1 e m_{p-1} , pois l depende da razão entre estas duas variáveis. Entretanto, atribuindo-se valores inteiros a l pode-se determinar as soluções para m_1 e m_{p-1} . Cada solução será rotulada pelo valor de l correspondente. O menor valor de l é zero, e o maior é inteiro($p/2$). Resolvendo-se (3.59) para m_1 e m_{p-1} obtemos duas desigualdades que limitam o intervalo em que a solução para cada l está definida:

$$p-1-2l \leq \frac{m_1}{m_{p-1}} \leq p+1-2l. \quad (3.60)$$

Calculando (3.57) e (3.58) com $l = 0$ obtemos a solução de recuperação (3.8). Para p ímpar, a solução com $l = \text{inteiro}(p/2)$ é a solução simétrica. Para p par a solução com $l = p/2$ é assimétrica, com $m_1 < m_{p-1}$. As figuras 3.4a e 3.4b mostram as curvas de m_1 e m_{p-1} em

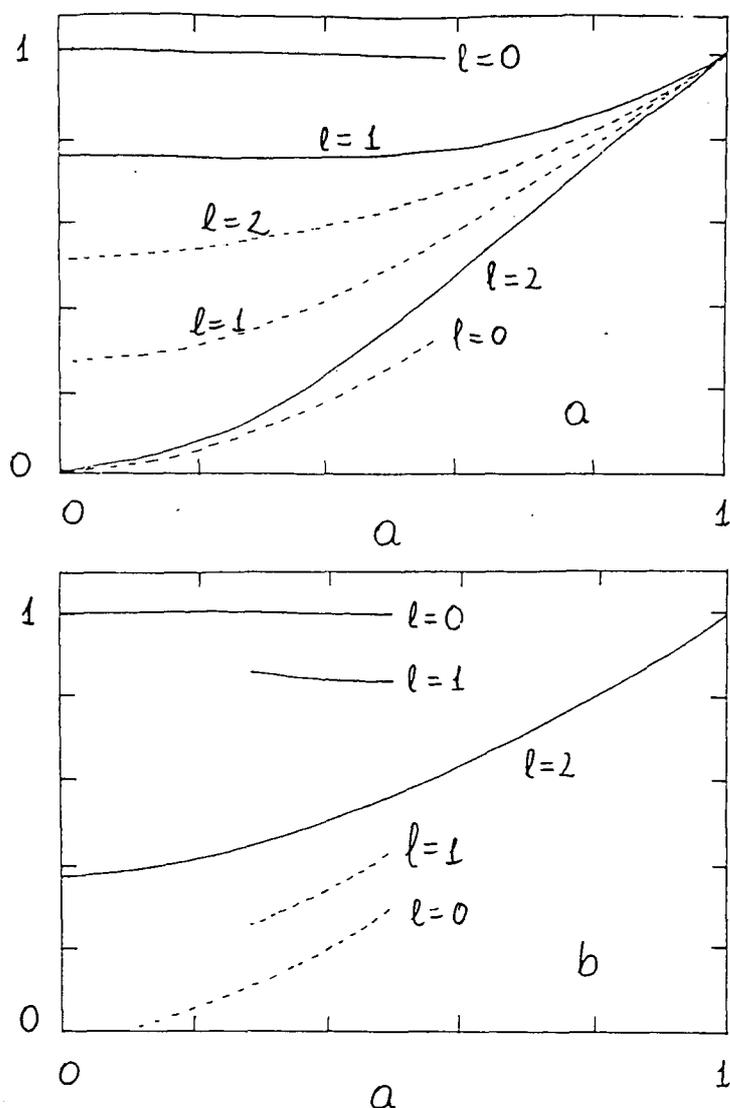


Figura 3.4: Soluções assimétricas para (a) $p = 4$ e (b) $p = 5$. Linha cheia, componente m_1 , linha tracejada, componente m_{p-1} .

função de a para ambos os casos. Na figura 3.4a temos todas soluções do tipo 3.52 para $p = 4$. Na figura 3.4b temos o mesmo, mas para $p = 5$. Uma diferença importante entre os dois casos é a existência, para $p = 4$, de uma solução como a rotulada por $l = 1$. Em geral, para outros valores pares de p ela é rotulada por $l = (p/2) - 1$. Esta solução inexistente para $p = 5$, assim como para qualquer p ímpar. Ela é responsável pela desestabilização das soluções simétricas pares em $T \rightarrow 0$.

É interessante notar que a componente m_{p-1} da solução assimétrica com $l = p/2$, para

p par é igual à solução simétrica m_{p-1} . Isto significa que ao acrescentar mais um padrão correlacionado quando $p-1$ (um número ímpar) já estão armazenados, a solução simétrica ímpar evolui para esta solução com $l = p/2$. Devido à correlação entre os padrões, a componente adicional do vetor \mathbf{m} não pode permanecer nula, e então assume o valor m_1 . Por outro lado, se p é par e for adicionado mais um padrão, a solução simétrica instável m_p evolui para a solução simétrica estável m_{p-1} , com $m_{p-1} = m_p$.

Aumentando o número p de padrões correlacionados, o estado de recuperação é restrito a um intervalo cada vez menor em a . De acordo com (3.6), para $p \gg 1$,

$$a_c \sim \sqrt{p}. \quad (3.61)$$

A mesma conclusão é obtida da análise de (3.60) com $l = 0$. Inspeccionando as equações (3.57) e (3.58) verifica-se que as correções oriundas dos somatórios, para $p \gg 1$ são pequenas quando l é pequeno. Portanto as soluções com l pequeno são pouco diferentes da solução de recuperação (3.8). Para uma análise quantitativa pode-se substituir a distribuição binomial (3.34) para z_p por uma distribuição gaussiana:

$$P(k) = \frac{1}{\sqrt{2\pi p b_1 b_2}} \exp \left[-\frac{(k - p b_1)^2}{2 p b_1 b_2} \right]. \quad (3.62)$$

Isto é justificado apenas quando $p \gg 1$.

Em $T = 0$, as equações (3.54) e (3.55), reescritas com a distribuição gaussiana, assumem então a forma

$$m_1 = \int_{-\infty}^{+\infty} P(k) dk [b_1 \text{sgn}(\Theta_+) + b_2 \text{sgn}(\Theta_-)] \quad (3.63)$$

e

$$m_{p-1} = \int_{-\infty}^{+\infty} P(k) dk \frac{z_{p-1}}{p-1} [b_1 \text{sgn}(\Theta_+) - b_2 \text{sgn}(\Theta_-)]. \quad (3.64)$$

Integrando em k , e notando que $p-1 \approx p$, temos

$$m_1 = b_1 \text{erf}(\phi_+) + b_2 \text{erf}(\phi_-) \quad (3.65)$$

e

$$m_{p-1} = \sqrt{\frac{2(1-a^2)}{p\pi}} [b_1 \exp(\phi_+^2) + b_2 \exp(\phi_-^2)] + a b_1 \text{erf}(\phi_+) - a b_2 \text{erf}(\phi_-), \quad (3.66)$$

onde

$$\phi_{\pm} = \frac{m_1 \pm ap}{2\sqrt{2pb_1b_2}}. \quad (3.67)$$

Usando a expressão assintótica da função erro [37]

$$\operatorname{erf}(x) \approx 1 - \frac{1}{\sqrt{\pi}} \frac{1}{x} \exp(-x^2) \quad (3.68)$$

temos, de (3.65) e (3.66),

$$m_1 = 1 - \mathcal{O}[\exp(-\phi_-)] \quad (3.69)$$

e

$$m_{p-1} = a^2 + \mathcal{O}[\exp(-\phi_-)]. \quad (3.70)$$

A condição necessária para a existência de solução de recuperação, com $m_1 \approx 1$ e $m_{p-1} \approx a^2$ é

$$\frac{m_1}{m_{p-1}} - pa > 0, \quad (3.71)$$

ou seja,

$$a < p^{\frac{1}{3}}, \quad p \gg 1. \quad (3.72)$$

Esta desigualdade representa um acréscimo no intervalo em a em que existe a solução de recuperação, em relação a (3.61), devido à incorporação de soluções com $l \neq 0$. Convém ressaltar, entretanto, que embora possa ser verificado um acréscimo significativo, o que é sugerido por (3.72) é obtido só no limite $p \rightarrow \infty$.

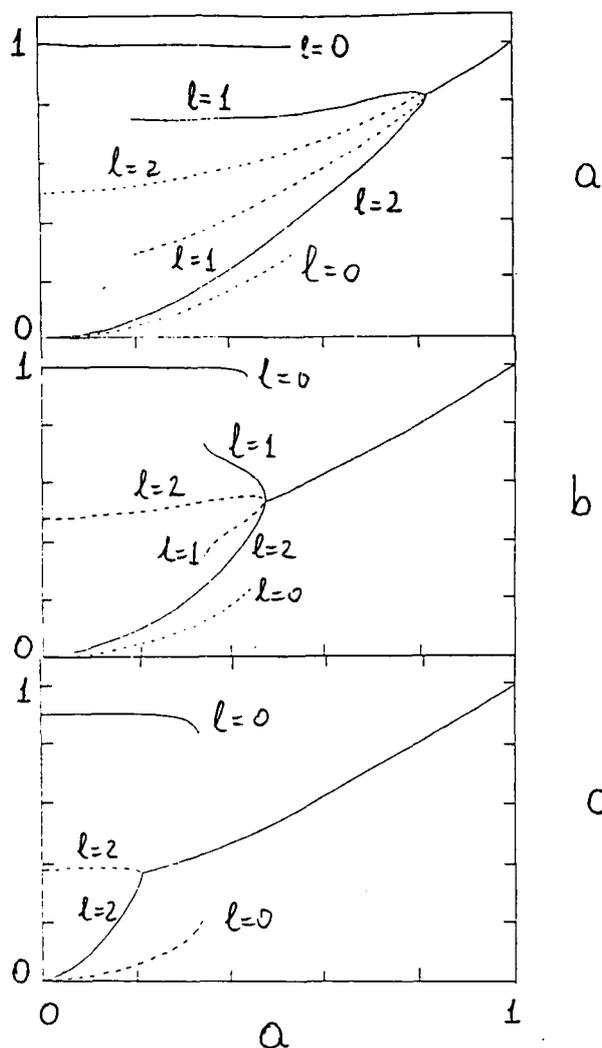


Figura 3.5: Evolução com a temperatura das soluções assimétricas com $p = 4$. Linha cheia, componente m_1 , linha tracejada, componente m_{p-1} . (a) $T = 0,05$; (b) $T = 0,3$; (c) $T = 0,6$.

Temperatura finita

Em $T \neq 0$ as soluções das equações (3.54) e (3.55) são obtidas por meio de análise numérica. Incorporando o parâmetro a como uma terceira variável o resultado desta análise, para cada valor de T , será uma curva no espaço (m_1, m_{p-1}, a) . Na figura 3.3 temos as projeções desta curva sobre os planos (m_1, a) e (m_{p-1}, a) , para $p = 4$ para $T = 0,05$, $T = 0,3$ e $T = 0,6$. Estas curvas representam a evolução com a temperatura das soluções que se observa na figura 3.4a.

A figura 3.5 revela que a tendência dominante com a elevação da temperatura é a supressão das soluções assimétricas e o fortalecimento da solução simétrica. Da mesma forma como ocorre com o caso em que os padrões não são polarizados, a solução de recuperação é entre as soluções assimétricas aquela que persiste até a mais alta temperatura ($T_C = 1$, quando $a = 0$).

As diferentes soluções estáveis de (3.54) e (3.55) representam fases do sistema de neurônios que são pelo menos metaestáveis. De um modo geral, quando o sistema encontra-se em determinada fase assimétrica do tipo (3.52), aumentando-se a polarização a para um valor além daquele em que esta fase existe, o sistema sofre uma transição descontínua para outra fase, simétrica ou assimétrica. A única exceção é a fase rotulada por $l = (p/2) - 1$ para p par, que sofre uma transição contínua para a fase simétrica.

Pode-se determinar a relação entre temperatura e polarização em que a transição descontínua ocorre. Com a definição,

$$\epsilon \equiv m_1 - m_{p-1} , \quad (3.73)$$

obtemos, de (3.54) e (3.55),

$$\begin{aligned} \epsilon = & \frac{2}{p-1} \sum_{k=0}^{p-1} \binom{p-1}{k} \cdot [(p-1-k) (b_1^{k+1} b_2^{p-1-k} \\ & + b_1^{p-1-k} b_2^{k+1}) \tanh\{\beta[m_1 - m_{p-1}(p-1-2k)]\} \\ & + k (b_1^k b_2^{p-k} + b_1^{p-k} b_2^k) \tanh\{\beta[m_1 + m_{p-1}(p-1-2k)]\}] . \end{aligned} \quad (3.74)$$

Se $\beta \gg 1$ e $m_1 \approx m_{p-1}$ então

$$\tanh\{\beta[m_1 + m_{p-1}(p-1-2k)]\} \approx 1 \quad (3.75)$$

e

$$\tanh\{\beta[m_1 - m_{p-1}(p-1-2k)]\} \approx -1 \quad (3.76)$$

para todo $0 \leq k < p/2 - 1$. Assim, para p par temos

$$\epsilon \approx \frac{2p}{p-1} \binom{p-1}{\frac{p}{2}-1} (b_1 b_2)^{\frac{p}{2}} \tanh(\beta\epsilon) . \quad (3.77)$$

Esta equação só tem solução para $\epsilon \neq 0$ ("overlap" assimétrico) se $T < T_S$, sendo

$$T_S = \frac{2p}{p-1} \binom{p-1}{\frac{p}{2}-1} (b_1 b_2)^{\frac{p}{2}}. \quad (3.78)$$

A curva $T_S = T_S(a)$ coincide com a cauda para $a \rightarrow 1$ da curva de estabilidade das soluções simétricas (figura 3.3a). Em outras palavras, as soluções simétricas pares se estabilizam quando são suprimidas as soluções assimétricas com $l = p/2 - 1$. **Isto significa que são estas as soluções assimétricas que desestabilizam as simétricas.**

Expandindo-se a equação (3.77) para $T \rightarrow T_S^-$, obtemos

$$\epsilon \approx \sqrt{3}(T_S - T)^{\frac{1}{2}}. \quad (3.79)$$

Quando p é ímpar, a equação (3.74) resulta em $\epsilon = 0$, de modo que a solução simétrica ímpar não é desestabilizada por uma assimétrica quando $T \rightarrow 0$, como indica a figura 3.3b.

Outras soluções assimétricas

Vamos analisar as soluções assimétricas do tipo (3.53). Combinando as equações (3.53) e (2.35), obtemos um sistema constituído pelas equações

$$m_1 = \left\langle b_1^2 \tanh(\beta\Theta'_+) + b_2^2 \tanh(\beta\Theta'_-) \right\rangle_{z_{p-2}} \quad (3.80)$$

e

$$m_{p-2} = \left\langle \frac{z_{p-2}}{p-2} \left[b_1^2 \tanh(\beta\Theta'_+) + 2b_1 b_2 \tanh(\beta\Theta'_0) - b_2^2 \tanh(\beta\Theta'_-) \right] \right\rangle_{z_{p-2}}, \quad (3.81)$$

onde

$$\Theta'_\pm = 2m_1 \pm m_{p-2} z_{p-2} \quad (3.82)$$

e

$$\Theta'_0 = m_{p-2} z_{p-2}. \quad (3.83)$$

Em $T = 0$ obtemos

$$m_1 = \frac{1}{2}(1 + a^2) + \sum_{k=0}^{l-1} \binom{p-2}{k} (b_1^{k+2} b_2^{p-2-k} + b_1^{p-2-k} b_2^{k+2}) \quad (3.84)$$

e

$$\begin{aligned}
m_{p-2} = & a^3 + 4a \sum_{k=0}^{m-2} \binom{p-3}{k} b_1^{p-2-k} b_2^{k+1} \\
& - 2a \sum_{k=0}^{l-1} \binom{p-3}{k} (b_1^{k+2} b_2^{p-3-k} - b_1^{p-3-k} b_2^{k+2}) \\
& + 2 \binom{p-3}{l-1} (b_1^{l+2} b_2^{p-3-l} - b_1^{p-3-l} b_2^{l+2}) + 4 \binom{p-3}{m-2} (b_1 b_2)^m, \quad (3.85)
\end{aligned}$$

onde

$$l = \text{inteiro} \left(\frac{p}{2} - \frac{m_1}{m_{p-2}} \right) \quad (3.86)$$

e

$$m = \text{inteiro} \left(\frac{p}{2} \right). \quad (3.87)$$

Como exemplo de soluções deste tipo, analisamos os casos $p = 4$ e $p = 5$. Para $p = 4$ temos duas soluções ($l = 0$ e $l = 1$), ambas instáveis. A solução com $l = 1$ corresponde à solução simétrica. Para $p = 5$ existem três soluções, todas localmente estáveis. A solução com $l = 0$ caracteriza-se por $(m_1/m_{p-2}) > 2$. $l = 1$ corresponde à solução simétrica. $l = 2$ caracteriza-se por $(m_1/m_{p-2}) < 1$.

Diagramas de fases

Vimos que existem diversas soluções assimétricas localmente estáveis. Estas soluções dão origem à fases termodinâmicas metaestáveis. A identificação da fase mais estável procede da comparação das energias livres relativas à cada fase. O diagrama completo das fases termodinâmicas, que delimita as regiões de estabilidade absoluta será apresentado posteriormente, quando analisarmos o caso $p \rightarrow \infty$. Por enquanto, achamos importante apenas delimitar no plano (a, T) as regiões em que cada uma das fases originadas pelas soluções assimétricas dos tipos (3.52) e (3.53) são localmente estáveis. Isto é justificado pelo fato de que durante sua evolução dinâmica, o sistema neural pode ser aprisionado pelos mínimos locais, dependendo das condições iniciais. A dimensão da bacia de atração da solução de recuperação é determinada pelos estados metaestáveis espúrios.

Cada uma das figuras 3.6 tem por objetivo delimitar a região de estabilidade de uma fase específica, para uma rede neural com quatro padrões armazenados ($p = 4$). As outras fases que aparecem em cada figura são aquelas que se estabilizam quando a fase especificada é desestabilizada ou suprimida. Cada fase assimétrica é identificada pelo rótulo l correspondente à solução que a origina. A fase simétrica é identificada por S e a paramagnética por P . As transições contínuas são assinaladas por linhas tracejadas, e as descontínuas por linhas cheias. As figuras 3.7 são para uma rede neural com cinco padrões armazenados ($p = 5$). A fase que tem origem da solução assimétrica do tipo (3.53) é identificada por B .

3.3 Capacidade macroscópica de armazenamento

Segundo a discussão no início deste capítulo, a regra de aprendizado de Hebb não permite o armazenamento de um número infinito de padrões correlacionados. O estudo que realizamos (Erichsen e Theumann, ref. [38]) sobre a capacidade macroscópica de armazenamento restringiu-se ao caso de um número finito p de padrões correlacionados, armazenados juntamente com outros $P - p$ padrões não correlacionados, com $P \rightarrow \infty$. Em relação ao desenvolvimento de Amit, Gutfreund e Sompolinsky [8] detalhado na subseção 2.2.3, do qual usaremos as equações gerais, identificamos p com o número finito s de overlaps que ‘condensam’ macroscopicamente. Assim como naquela seção definimos $\alpha \equiv P/N$ como a razão entre o número total de padrões e o número de neurônios. Estudamos em detalhe a solução simétrica (equação 3.30) e entre as soluções assimétricas estudamos apenas as do tipo definido pela equação (3.52), entre as quais se inclui a solução de recuperação.

3.3.1 Solução simétrica

Reescrevendo as equações de ponto de sela na solução de simetria de réplicas, equações (2.83) e (2.84) com o ‘ansatz’ (3.30), temos

$$m_p = \int \mathcal{D}y \langle \tanh[\beta(\sqrt{\alpha}ry + m_p z_p)] \rangle_{z_p} \quad (3.88)$$

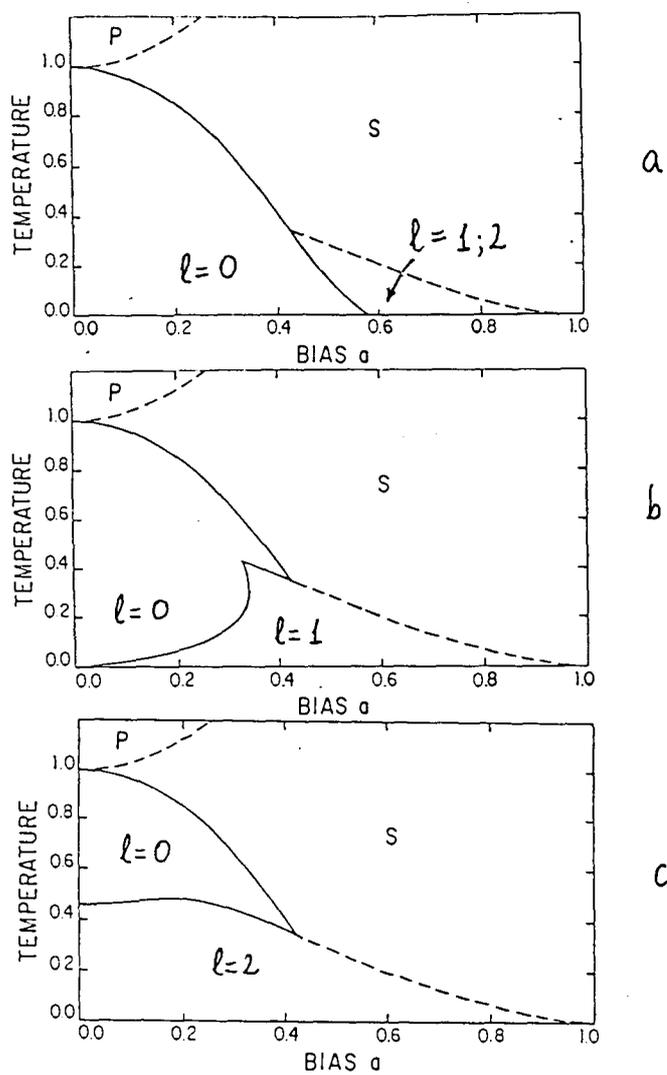


Figura 3.6: Diagrama de fases para $p = 4$, partindo de dentro da bacia de atração da fase identificada por (a) $l = 0$, (b) $l = 1$ e (c) $l = 2$. Curvas cheias e tracejadas representam transições de fase descontínuas e contínuas, respectivamente.

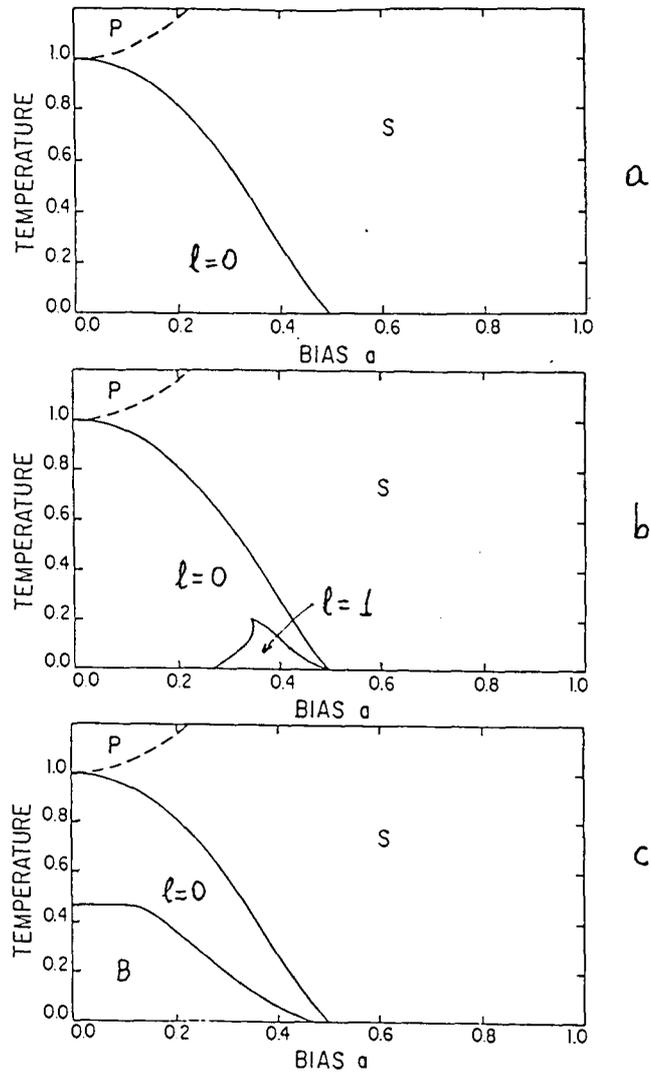


Figura 3.7: Diagrama de fases para $p = 5$, partindo de dentro da bacia de atração da fase identificada por (a) $l = 0$, (b) $l = 1$ e (c) B. Curvas cheias e tracejadas representam transições de fase descontínuas e contínuas, respectivamente.

e

$$q = \int \mathcal{D}y \langle \tanh^2[\beta(\sqrt{\alpha r}y + m_p z_p)] \rangle_{z_p}. \quad (3.89)$$

Temperatura zero

No limite de temperatura zero ($\beta \rightarrow \infty$) o sistema formado pelas equações (3.88) e (3.89) é reescrito na forma

$$m_p = \frac{1}{p} \left\langle z_p \operatorname{erf} \left(\frac{m_p z_p}{\sqrt{2\alpha r}} \right) \right\rangle_{z_p}, \quad (3.90)$$

$$C \equiv \beta(1 - q) = \sqrt{\frac{2}{\pi\alpha r}} \left\langle \exp \left(-\frac{m_p^2 z_p^2}{2\alpha r} \right) \right\rangle \quad (3.91)$$

que devem ser resolvidas utilizando a equação (2.92)

$$r = \frac{1}{(1 - C)^2}. \quad (3.92)$$

Introduz-se novamente uma variável auxiliar como em (2.116)

$$y_p = \frac{m_p}{\sqrt{2\alpha r}}, \quad (3.93)$$

de modo que o sistema composto pelas equações (3.90)–(3.92) reduz-se a uma equação apenas:

$$y_p = \frac{1}{p \sqrt{2\alpha}} \frac{\langle z_p \operatorname{erf}(z_p y_p) \rangle_{z_p}}{\sqrt{\frac{2}{\pi}} \langle \exp(-z_p^2 y_p^2) \rangle_{z_p}}. \quad (3.94)$$

O valor crítico α_S para o surgimento das soluções simétricas é obtido diferenciando a equação acima em relação a y_p :

$$1 = \frac{2}{p\sqrt{\pi}} \frac{\langle z_p^2 \exp(-z_p^2 y_p^2) \rangle_{z_p}}{\sqrt{2\alpha_S} + \frac{2}{\sqrt{\pi}} \langle \exp(-z_p^2 y_p^2) \rangle_{z_p}} (1 + 2p y_p^2). \quad (3.95)$$

Resolvendo para α_S , temos

$$\alpha_S = \frac{1}{\pi} \left[\left\langle \left[\left(\frac{1}{p} + 2y_p^2 \right) z_p^2 - 1 \right] \exp(-z_p^2 y_p^2) \right\rangle \right]^2. \quad (3.96)$$

Em $T = 0$, o parâmetro de ordem de vidro de spin, q converge para 1. Para $\alpha > \alpha_S$ o sistema apresenta-se na fase de vidro de spin. A transição entre esta fase e a simétrica é

descontínua, com os valores de y_p e α_S obtidos resolvendo-se simultaneamente as equações (3.94) e (3.95). Entretanto, a transição é apenas fracamente descontínua (y_p é pequeno sobre a transição), de modo que uma boa estimativa de α_S é obtida calculando-se (3.96) em $y_p = 0$. Obtemos assim

$$\alpha_S = \frac{2}{\pi}(p-1)^2 a^4. \quad (3.97)$$

A dependência com a quarta potência em a está de acordo com um trabalho prévio [10].

Temperatura finita

Expandindo-se (3.88) e (3.89) para m_p e q ambos próximos de zero, obtemos

$$m_p \sim (T_C - T)^{\frac{1}{2}} \quad (3.98)$$

e

$$q \sim T_C - T, \quad (3.99)$$

onde

$$T_C = 1 + (p-1)a^2 \quad (3.100)$$

é a temperatura em que ocorre a transição contínua entre as fases simétrica e paramagnética, que se caracteriza por $m_p = 0 = q$. A temperatura crítica (3.100) independe de α , e portanto é idêntica à determinada para $\alpha = 0$ (equação 3.44).

A temperatura crítica para a transição entre as fases de vidro de spin ($q \neq 0, m_p = 0$) e paramagnética é obtida expandindo-se a equação (3.89) para $q \rightarrow 0$ e $m_p = 0$:

$$q = \frac{\alpha q}{(T-1)^2} + \mathcal{O}(q^2), \quad (3.101)$$

de modo que

$$T_G = 1 + \sqrt{\alpha}. \quad (3.102)$$

A temperatura T_G é necessariamente independente da polarização a , pois esta só se manifesta nos estados que possuem overlap não nulo com os padrões polarizados (o que não ocorre com a fase de vidro de spin).

O comportamento da solução simétrica à temperatura finita é obtido da análise numérica das equações (3.88) e (3.89). Verifica-se que a solução simétrica é suprimida de uma forma contínua quando $T \rightarrow T_S$, como indicam as expressões (3.98) e (3.99).

3.3.2 Soluções assimétricas

As equações (2.91) e (2.93), reescritas com o ‘ansatz’ (3.52), assumem a forma

$$m_1 = \int \mathcal{D}y \langle b_1 \tanh[\beta(\sqrt{\alpha r}z + \Theta_+)] - b_2 \tanh[\beta(\sqrt{\alpha r}z - \Theta_-)] \rangle_{z_{p-1}}, \quad (3.103)$$

$$m_{p-1} = \int \mathcal{D}y \langle \frac{z_{p-1}}{p-1} [b_1 \tanh[\beta(\sqrt{\alpha r}z + \Theta_+)] + b_2 \tanh[\beta(\sqrt{\alpha r}z - \Theta_-)]] \rangle_{z_{p-1}} \quad (3.104)$$

e

$$q = \int \mathcal{D}y \langle b_1 \tanh^2[\beta(\sqrt{\alpha r}z + \Theta_+)] + b_2 \tanh^2[\beta(\sqrt{\alpha r}z - \Theta_-)] \rangle_{z_{p-1}}. \quad (3.105)$$

Temperatura zero

No limite de temperatura zero, o sistema formado pelas equações (3.103)–(3.105) pode ser reescrito na forma

$$m_1 = \left\langle b_1 \operatorname{erf} \left(\frac{\Theta_+}{\sqrt{2\alpha r}} \right) + b_2 \operatorname{erf} \left(\frac{\Theta_-}{\sqrt{2\alpha r}} \right) \right\rangle_{z_{p-1}}, \quad (3.106)$$

$$m_{p-1} = \left\langle \frac{z_{p-1}}{p-1} \left[b_1 \operatorname{erf} \left(\frac{\Theta_+}{\sqrt{2\alpha r}} \right) - b_2 \operatorname{erf} \left(\frac{\Theta_-}{\sqrt{2\alpha r}} \right) \right] \right\rangle_{z_{p-1}} \quad (3.107)$$

e

$$C = \sqrt{\frac{2}{\pi\alpha r}} \left\langle b_1 \exp \left(-\frac{\Theta_+^2}{2\alpha r} \right) + b_2 \exp \left(-\frac{\Theta_-^2}{2\alpha r} \right) \right\rangle_{z_{p-1}}. \quad (3.108)$$

Definimos as variáveis auxiliares

$$y_1 = \frac{m_1}{\sqrt{2\alpha r}} \quad (3.109)$$

e

$$y_{p-1} = \frac{m_{p-1}}{\sqrt{2\alpha r}}. \quad (3.110)$$

O sistema reduz-se a duas equações:

$$y_1 = \frac{1}{D} \langle b_1 \operatorname{erf}(\Phi_+) + b_2 \operatorname{erf}(\Phi_-) \rangle \quad (3.111)$$

e

$$y_{p-1} = \frac{1}{(p-1)D} \langle z_{p-1} [b_1 \operatorname{erf}(\Phi_+) - b_2 \operatorname{erf}(\Phi_-)] \rangle, \quad (3.112)$$

sendo

$$D = \sqrt{2\alpha} + \frac{2}{\pi} \langle b_1 \exp(-\Phi_+^2) + b_2 \exp(-\Phi_-^2) \rangle \quad (3.113)$$

e

$$\Phi_{\pm} = y_1 \pm z_{p-1} y_{p-1} . \quad (3.114)$$

O sistema de equações (3.111) e (3.112) no limite $\alpha \rightarrow 0$ descreve a mesma situação física que o sistema (3.54) e (3.55) no limite $T \rightarrow 0$. Conseqüentemente, ambos devem convergir para as mesmas soluções, que são representadas pelas expressões (3.57) e (3.58). Por outra parte, para $T \neq 0$ as equações 3.54 e 3.55 descrevem o modo como a temperatura, identificada com o ruído sináptico, modifica os estados estáveis de uma rede neural com padrões correlacionados. Analogamente, (3.111) e (3.112) descrevem a influência que exercem sobre os mesmos estados os overlaps microscópicos com os $P - p$ padrões não correlacionados, que não condensam macroscopicamente.

As figuras (3.8) delimitam a região em que as fases que têm origem nas soluções das equações (3.111) e (3.112) são localmente estáveis. Assim como o que foi discutido com relação às figuras (3.6), cada uma destas figuras corresponde a uma da fase específica que é localmente estável. As outras fases que aparecem em cada diagrama são as que permanecem estáveis quando a fase especificada é suprimida ou se desestabiliza. A fase de vidro de spin é identificada por SG, e as outras permanecem com a mesma identificação das figuras (3.6). A curva que limita cada fase pode ser interpretada como a capacidade de armazenamento em função da polarização.

É importante fazer alguns comentários a respeito do que mostram as figuras.

1. A região em que cada solução é estável no plano (a, α) é formalmente semelhante à região de estabilidade no plano (a, T) . Mais uma vez, isto está relacionado ao papel análogo que desempenham na dinâmica do sistema o ruído sináptico e as correlações microscópicas.
2. O valor máximo de α_C para a solução de recuperação ocorre em $a = 0$, e corresponde ao valor conhecido [8] $\alpha_c \approx 0,138$. A curva $\alpha_c(a)$ pode ser comparada com a curva

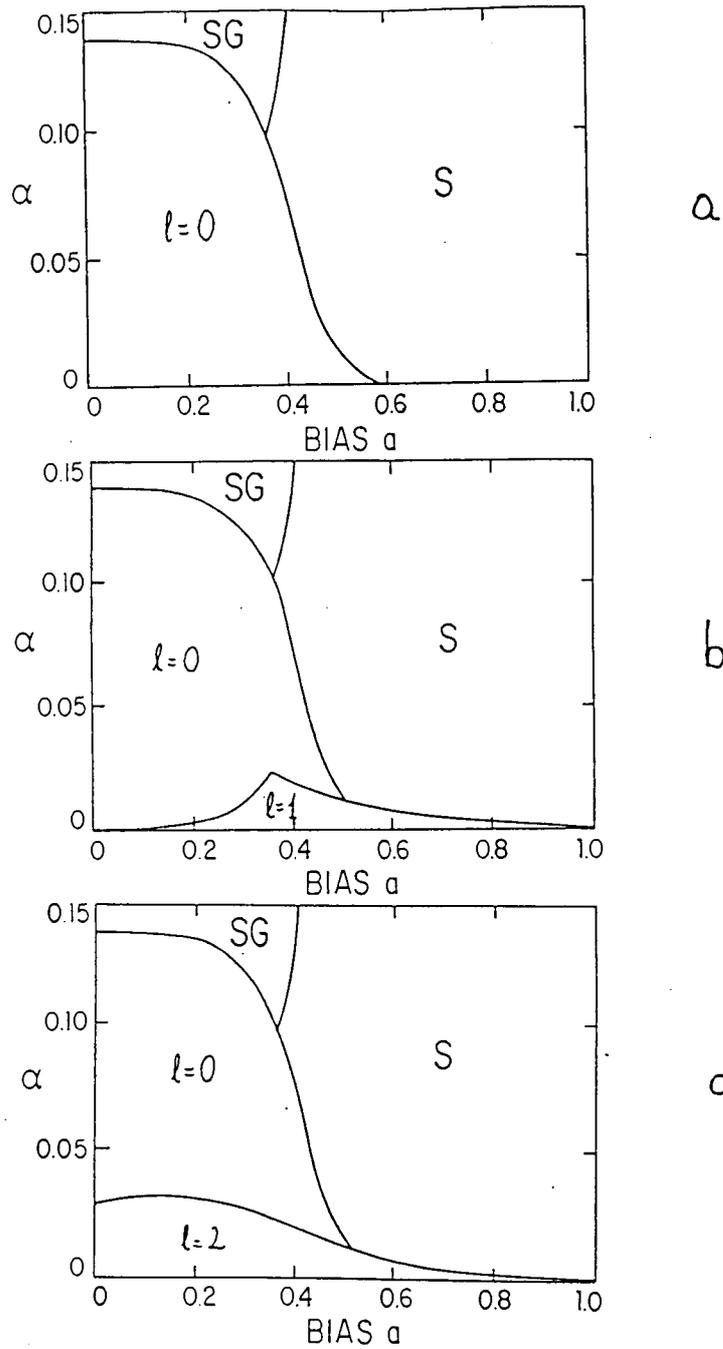


Figura 3.8: Diagrama de fases a $T = 0$ e $p = 4$, partindo da bacia de atração das fases (a) $l = 0$, (b) $l = 1$ e (c) $l = 2$.

análoga obtida para o modelo AGS [9]. A diferença mais marcante é que no modelo AGS esta curva contém uma ‘cauda’ que se estende até $a = 1$, enquanto que no presente caso ela intercepta o eixo horizontal em a_c , dado pela equação (3.6). Isto reflete o ponto mais frágil do modelo original de Hopfield, que é a incapacidade de armazenar padrões fortemente correlacionados. O desempenho do modelo AGS também está longe de ser satisfatório, pois a capacidade de armazenamento cai a zero no limite $a \rightarrow 1$. A capacidade máxima de armazenamento, oferecida pelo conjunto ótimo de conexões sinápticas, determinada por Gardner [23] diverge neste limite, de acordo com a equação (3.9). As redes biológicas de Perez–Vicente e Amit, [18] e Buchmann, Divko e Shulten, [17], que envolvem alterações na regra de aprendizado e na representação dos estados de atividade apresentam melhor desempenho neste limite (ver seção 3.1).

3. Ao contrário da solução de recuperação, que é estável até $\alpha_C \approx 0,138$, as soluções assimétricas expúrias do tipo (3.52) são suprimidas para valores inferiores de α : elas deixam de existir já para $\alpha \approx 0.031$. Estes estados não são mencionados no trabalho de Amit, Gutfreund e Sompolinsky [9].
4. A curva que assinala a transição entre as fases assimétrica e de vidro de spin foi determinada por meio da solução numérica do sistema formado pelas equações (3.94) e (3.95). É importante ressaltar que esta curva é satisfatoriamente reproduzida pela aproximação de transição contínua (3.97). A questão da estabilidade desta solução representa outra sensível diferença com relação ao modelo AGS, que suprime as soluções simétricas ($\alpha_S \rightarrow 0$ quando $a \rightarrow 1$). No presente caso, de acordo com a equação (3.97) temos $\alpha_S \rightarrow 2\sqrt{\pi} (p - 1)^2$ quando $a \rightarrow 1$. Em outras palavras, isto significa que o modelo original de Hopfield, quando armazena padrões fortemente correlacionados, favorece os estados mistos simétricos, em detrimento dos estados de recuperação. O modelo é mais hábil em identificar as semelhanças do que as diferenças entre o conjunto de padrões. Existe na literatura uma regra de aprendizado, devido a Fon-

tanari e Theumann [33] concebida para identificar especialmente as diferenças entre o conjunto de padrões.

Diagrama de fases à temperatura finita

Para concluir esta discussão analítica sobre redes neurais de Hopfield com padrões correlacionados, resta analisar o efeito conjunto sobre os estados de equilíbrio do sistema neural, ruído térmico e correlações microscópicas com padrões correlacionados.

A análise numérica completa das equações de ponto de sela para as soluções simétrica, assimétrica de recuperação, paramagnética e de vidro de spin permite construir o diagrama de fases termodinâmicas em função dos parâmetros T , α e a para o sistema neural. As figuras 3.9a e 3.9b mostram o corte no plano (T, α) do diagrama em $a = 0.2$ e 0.4 respectivamente. Quando mais de uma fase são estáveis, decidimos sobre a mais estável com base no valor mínimo da energia livre. A fase que corresponde à recuperação perfeita de um padrão é identificada pela letra R . A linha cheia limita a região em que esta fase é mínimo absoluto da energia livre. A linha pontilhada limita a região em que ela é metaestável. As linhas tracejadas identificam as transições contínuas $S-P$ (equação 3.100) e $SG-P$ (equação 3.102). As figuras revelam o dramático decréscimo da região dominada pela fase de recuperação em função do aumento da polarização e o não menos dramático acréscimo na área dominada pela solução simétrica. É importante notar que este crescimento ocorre não apenas às custas da fase de recuperação, mas também às custas das fases paramagnética e de vidro de spin. Este último aspecto é positivo, pois revela que a regra de aprendizado de Hebb permite ao sistema neural aprender algo sobre o conjunto de padrões correlacionados, especialmente sobre o que eles têm em comum, embora não consiga identificá-los.

3.4 Simulação numérica

O cálculo do observável \mathcal{O} diretamente da expressão

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{\{S_i\}} \mathcal{O} \exp[-\beta \mathcal{H}(\{S_i\})] \quad (3.115)$$

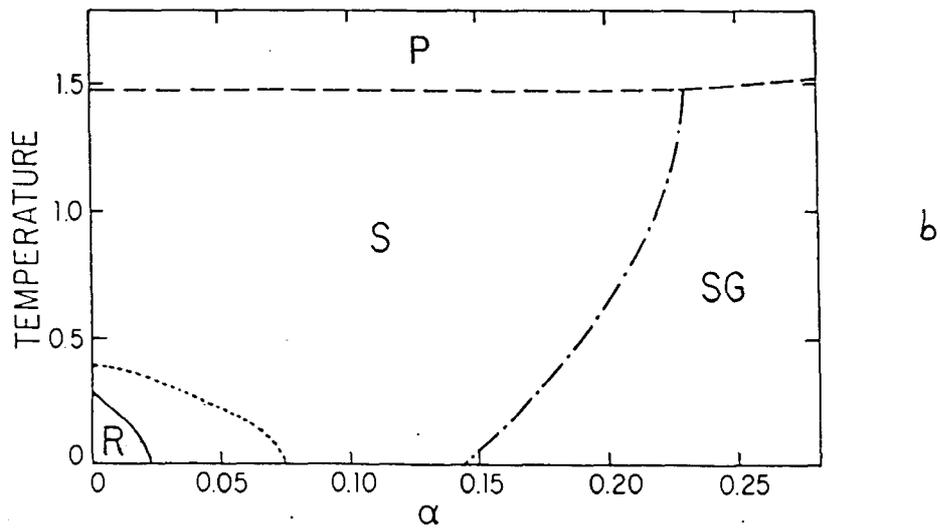
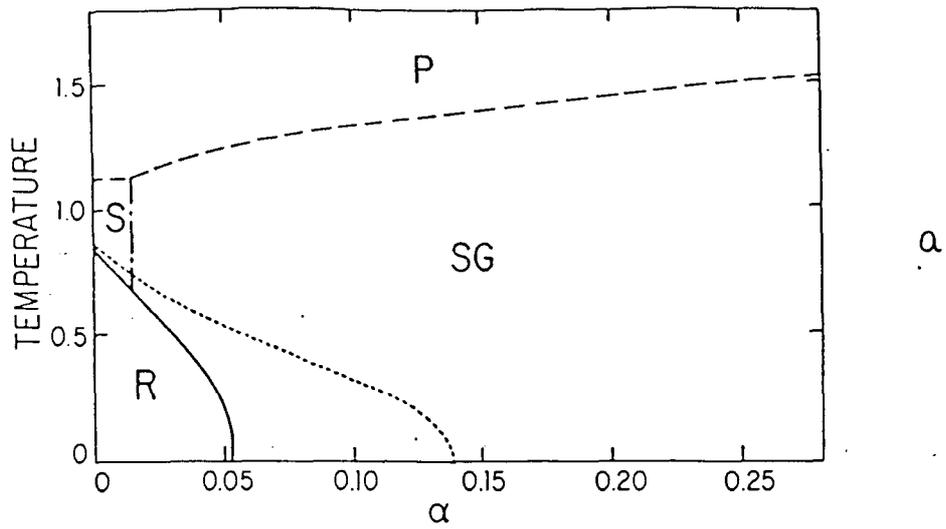


Figura 3.9: Diagrama de fases para $p = 4$ padrões correlacionados e $P - p$ não correlacionados, para (a) $a = 0,2$ e (b) $a = 0,4$. A linha cheia representa transição contínua, enquanto todas as outras representam transições descontínuas. O estado de recuperação é globalmente (R) ou localmente estável dentro das linhas cheia ou pontilhada, respectivamente.

é impraticável quando $N \gg 1$, devido ao enorme número de estados do sistema. Pelo método de Monte Carlo [39], o volume de cálculo numérico é substancialmente reduzido, graças à escolha apropriada de estados. A evolução dinâmica do sistema é encarregada de gerar um conjunto de estados que obedece à distribuição de Gibbs (equação 2.15). No caso de um sistema de N neurônios representado por um conjunto de variáveis de Ising, a dinâmica de Glauber, segundo a qual cada neurônio se atualiza de acordo com a expressão (2.9) é governada pelo hamiltoniano (2.11).

3.4.1 Simulação numérica de um sistema de neurônios com padrões correlacionados

Neste trabalho¹, realizamos simulações numéricas à $T = 0$, que corresponde à evolução dinâmica do sistema de neurônios governada pela equação determinística

$$S_i(t + 1) = \text{sgn}[h_i(t)] , \quad (3.116)$$

onde

$$h_i(t) = \sum_j J_{ij} S_j(t) \quad (3.117)$$

é o campo local sobre cada neurônio no instante t . O objetivo desta simulação numérica foi investigar propriedades de redes neurais de difícil acesso por métodos analíticos, como o tamanho das bacias de atração em torno dos estados de recuperação, além de verificar resultados teóricos.

Conjuntos de $p = 4$ ou $p = 8$ padrões aleatórios, gerados de acordo com a distribuição de probabilidades (3.1) foram armazenados de acordo com as regras de aprendizado de Hebb, (equação 2.14) e do modelo AGS (equação 3.10), em sistemas de 200 e 400 neurônios. As médias configuracionais foram efetuadas sobre conjuntos de 500 relaxações, após o qual determinamos a ‘taxa de recuperação’ f , que corresponde à percentagem de eventos em que o sistema convergiu para o padrão armazenado. Em cada evento, adotamos como critério afirmativo de recuperação se o estado estável que resulta da relaxação difere no máximo em 5% do padrão.

¹Erichsen e Theumann, ref. [38]

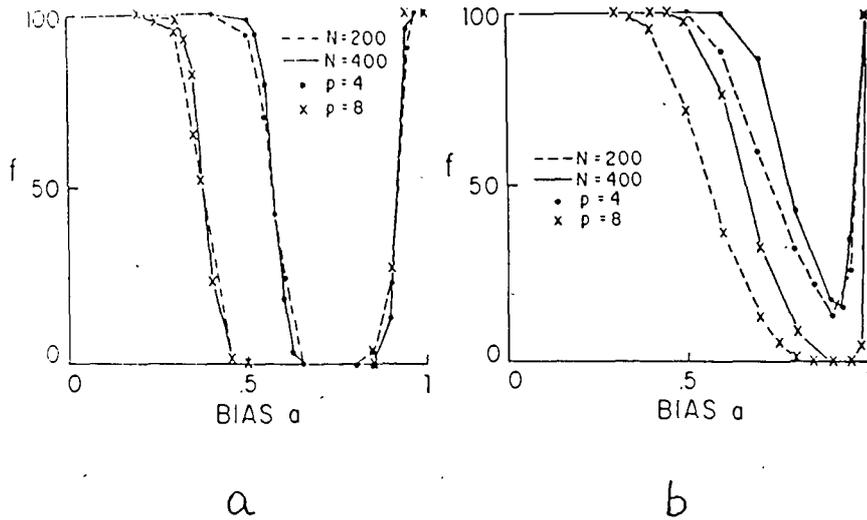


Figura 3.10: Simulação numérica para a fração média de recuperação em função da polarização. (a) modelo de Hopfield com a regra de Hebb; (b) modelo AGS.

A figura 3.10a mostra a estabilidade dos padrões armazenados pela regra de Hebb em função da polarização. O estado inicial de cada relaxação corresponde exatamente a um dos padrões. A queda abrupta de f , com o cruzamento das curvas correspondentes a $N = 200$ e $N = 400$, em $a \approx 0,58$ para $p = 4$ e $a \approx 0.38$ para $p = 8$ sugerindo uma transição descontínua, confirma a desestabilização dos estados de recuperação em $a = a_c$, como previsto pela equação (3.6). Quando $a \rightarrow 1$, $\xi_i^\mu = 1$ para todo i, μ , e de acordo com a regra de Hebb (2.14) $J_{ij} > 0$, para todo i, j . Nesta situação, a rede comporta-se como um sistema puramente ferromagnético, de modo que se estabilizam os estados $S_i = +1$, para todo i e $S_i = -1$, para todo i . Isto explica o comportamento de $f \rightarrow 100\%$ para $a \rightarrow 1$.

Na figura 3.10b são apresentados os resultados relativos à estabilidade dos padrões correlacionados no modelo AGS. A taxa de recuperação neste caso cai continuamente a partir de determinado valor de a , e depois sobe novamente para $f = 100\%$, quando $a = 1$. Verificamos também o aumento de f em função de N , o que sugere que $f = 100\%$ ao $N \rightarrow 1$, para todo a . Isto indica que os padrões correlacionados armazenados por esta regra de aprendizagem são localmente estáveis quando $T = 0$ e $\alpha = 0$.

3.4.2 Bacias de atração

O tamanho da bacia de atração em torno dos estados de recuperação pode ser quantificado em termos do ‘overlap’ mínimo $m_c(\alpha)$ entre o estado inicial do sistema neural e o padrão, de modo que a evolução dinâmica ainda conduz à recuperação [40, 41].

Para investigar a influência da polarização a sobre as bacias de atração, determinamos o comportamento da fração de recuperação em função do ‘overlap’ inicial m_0 . Os estados iniciais foram gerados a partir da inversão aleatória do estado de determinados neurônios. A figura 3.11a mostra a curva $f = f(m_0)$ para $p = 4$ e $N = 200$ e $N = 400$ neurônios, com $a = 0,2$ e $a = 0,5$, com padrões armazenados de acordo com a regra de Hebb. Observa-se que para ambos valores da polarização a , f cresce monotonamente com m_0 , sem nenhum ponto de inflexão que caracterize um valor crítico m_c . Mas isto é natural, pois se $p = 4$, no limite termodinâmico temos $\alpha = 0$. Como $a < a_c$, temos $f = 100\%$ para $0 < m_0 \leq 1$, e os valores de $f < 100\%$ para $m_0 \rightarrow 0$ são efeitos devido ao tamanho finito do sistema.

A figura 3.11b mostra as curvas relativas às mesmas situações, embora com padrões armazenados pela regra de aprendizagem (3.10). Neste caso temos um indício de ponto de inflexão em $a = 0,5$. Para ambas regras de aprendizagem a fração de recuperação é menor para $a = 0,5$ em relação a $a = 0,2$. Isto indica uma redução na bacia de atração em função do aumento da polarização. Quando a cresce, o conjunto de mínimos locais relativos aos padrões fica restrito a uma região cada vez menor no espaço de configurações, ocasionando a redução da dimensão das bacias de atração individuais. Esta constatação permite supor que a regra de aprendizagem de Amit, Gutfreund e Sompolinsky, mesmo mantendo mínimos individuais em torno de cada padrão, não assegura a existência de uma bacia de atração que garanta satisfatórias propriedades de recuperação destes padrões, em particular quando eles são fortemente correlacionados. As propriedades de equilíbrio em relação à capacidade de armazenamento deste modelo são superiores ao modelo original de Hopfield, embora as propriedades dinâmicas não necessariamente o sejam.

Discutimos acima as propriedades de armazenamento e recuperação de padrões correlacionados por meio de uma regra de aprendizagem específica, no caso, a regra de Hebb.

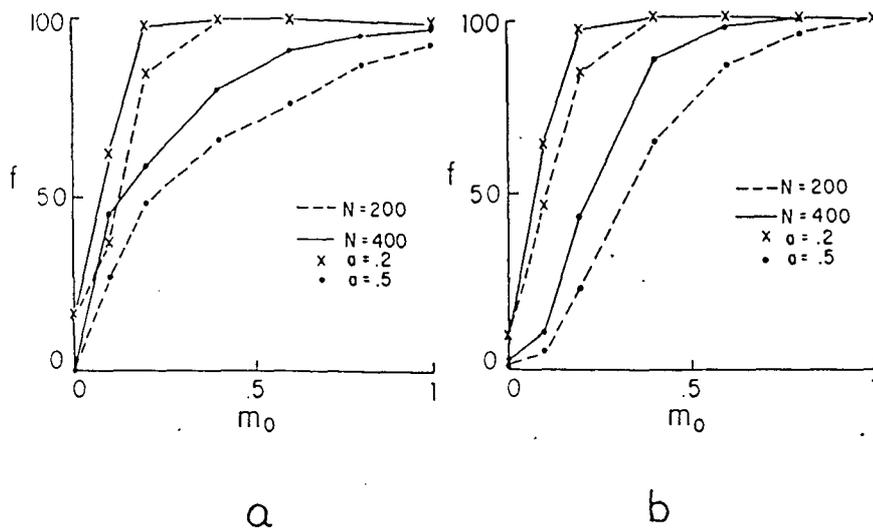


Figura 3.11: Simulação numérica para o tamanho da bacia de atração, baseada na fração média de recuperação. (a) modelo de Hopfield com a regra de Hebb; (b) modelo AGS.

O armazenamento por meio de uma regra de aprendizagem significa que, dado o conjunto de padrões a armazenar, a regra fornece explicitamente o conjunto das conexões sinápticas. No próximo capítulo iniciaremos a discussão de uma abordagem diferente, que consiste em determinar as propriedades ótimas de um sistema neural. Independentemente de regras de aprendizagem específicas, estas propriedades são estudadas por meio da aplicação de métodos da mecânica estatística ao espaço das interações sinápticas.

4. Redes neurais de Gardner–Derrida e padrões correlacionados

As propriedades computacionais que emergem do comportamento coletivo do conjunto de neurônios que constituem um sistema neural são determinadas pelo conjunto $\{J_{ij}\}$ das conexões sinápticas. Cada padrão do conjunto $\{\xi_i^\mu\}$ é ponto fixo da dinâmica representada pela equação

$$S_i(t+1) = \text{sgn}[h_i(t) - T_i] \quad (4.1)$$

se o conjunto das estabilidades locais, definidas pela equação

$$\gamma_i^\mu \equiv \xi_i^\mu [h_i(\{\xi_i^\mu\}) - T_i] = \frac{1}{\sqrt{N}} \sum_j J_{ij} \xi_i^\mu \xi_j^\mu - T_i \xi_i^\mu \quad (4.2)$$

satisfaz a desigualdade

$$\gamma_i^\mu > \kappa \quad (4.3)$$

em todos neurônios da rede. κ é um parâmetro não negativo relacionado com a bacia de atração em torno dos padrões. As conexões sinápticas desempenham o papel de variáveis dinâmicas, e nada sabe-se sobre elas a não ser que durante sua evolução assumem valores que tornam o conjunto de padrões tão estável quanto possível. Cada conjunto $\{J_{ij}\}$ representa um ponto no espaço de configurações. Quando a rede opera abaixo de sua capacidade crítica ($\alpha < \alpha_c$) existe um volume não nulo neste espaço que satisfaz todas as desigualdades (4.3). A capacidade crítica é atingida quando este volume restringe-se a um ponto. O conjunto $\{J_{ij}\}$ correspondente a este ponto é o único que satisfaz todos os vínculos, ou seja, é o *conjunto ótimo de sinapses*.

Esta abordagem original, que assume as sinapses como variáveis dinâmicas, é uma contribuição fundamental de Elizabeth Gardner à teoria geral de redes neurais. Entre muitos resultados importantes devido a Gardner, citamos a capacidade crítica de armazenamento em função do parâmetro κ e da polarização a [23]. Quando estes dois parâmetros são nulos, temos $\alpha_c = 2$, que confirma o resultado prévio de Cover [42]. A capacidade crítica de armazenamento é uma função monótona de a . Quando $|a| \rightarrow 1$ temos o comportamento assintótico representado pela expressão (3.9).

4.1 Fração mínima de erros

É possível o armazenamento de padrões acima da capacidade crítica α_c da rede, desde que seja permitida a violação de uma certa quantidade dos vínculos (4.3). Para um determinado conjunto de sinapses $\{J_{ij}\}$ a função custo de Gardner–Derrida [24],

$$E_i(\{J_{ij}\}) = \sum_{\mu=1}^p \left\{ 1 - \Theta \left[\xi_i^\mu \left(\frac{1}{\sqrt{N}} \sum_j J_{ij} \xi_j^\mu - T_i \right) - \kappa \right] \right\}, \quad (4.4)$$

conta o número de violações no sítio i . $\Theta(x)$ é a função degrau, definida como

$$\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (4.5)$$

A questão fundamental formulada por Gardner–Derrida é ‘qual a fração mínima ϵ de erros de armazenamento por padrão para um determinado valor de $\alpha > \alpha_c$ ’?

A estratégia traçada por Gardner–Derrida para responder a esta questão inicia por considerar que as variáveis $\{J_{ij}\}$ evoluem dinamicamente em equilíbrio com um banho térmico à temperatura β^{-1} , no sentido de minimizar a função custo (4.4), que neste caso desempenha o papel de hamiltoniano do sistema. A função partição para o problema é definida por

$$\mathcal{Z}_i(\beta) = \sum_{\{J_{ij}\}} \exp[-\beta E_i(\{J_{ij}\})]. \quad (4.6)$$

À temperatura finita, o número médio de erros compatível com uma distribuição de equilíbrio

no espaço de configurações é dado pela expressão

$$p\epsilon_i = \langle E_i \rangle = -\frac{d}{d\beta} \ln \mathcal{Z}_i(\beta) . \quad (4.7)$$

A fração mínima de erros de armazenamento no sítio i , compatível com a escolha ótima do conjunto das sinapses corresponde ao estado fundamental do hamiltoniano:

$$p\epsilon_{i,min} = -\lim_{\beta \rightarrow \infty} \frac{d}{d\beta} \ln \mathcal{Z}_i(\beta) . \quad (4.8)$$

Se todos os J_{ij} são independentes (não se estabelece nenhuma relação de simetria entre as sinapses), então a função partição definida sobre toda rede é simplesmente o produto

$$\mathcal{Z}(\beta, h) = \prod_i \mathcal{Z}_i(\beta) . \quad (4.9)$$

As propriedades físicas do sistema seguem de $\ln \mathcal{Z}(\beta)$, que desempenha o papel de energia livre efetiva:

$$\ln \mathcal{Z}(\beta, h) = \sum_i \ln \mathcal{Z}_i(\beta) . \quad (4.10)$$

De (4.8) e (4.10) temos que o número médio total de violações na rede é simplesmente o número médio de violações por neurônio multiplicado pelo número total de neurônios. Deste modo, é suficiente restringirmos o estudo a apenas um neurônio.

A expressão (4.8) dá a média de "ensemble" para uma determinada escolha do conjunto de variáveis aleatórias $\{\xi_i^\mu\}$. As propriedades gerais da rede, independente da escolha das variáveis aleatórias, implicam em calcular a média configuracional de $\ln \mathcal{Z}_i(\beta)$.

Para determinar o comportamento do sistema acima da capacidade crítica de armazenamento, no caso de padrões correlacionados, tomaremos a média configuracional sobre a distribuição de probabilidades (3.1). Isto corresponde a uma generalização do método de Gardner–Derrida [24].

Se o conjunto de sinapses $\{J_{ij}\}$ resulta na fração de erros ϵ_i no neurônio i , então o conjunto $\{\lambda J_{ij}\}$ resulta na mesma fração de erros. Para fixar um conjunto finito no espaço de configurações, adotamos o vínculo esférico

$$\sum_j J_{ij}^2 = N . \quad (4.11)$$

Para tomar a média de $\ln \mathcal{Z}(\beta, h)$ (onde eliminamos o índice de sítio) usamos o método das réplicas:

$$\langle \ln \mathcal{Z}(\beta, h) \rangle = \lim_{n \rightarrow 0} \frac{1}{n} (\langle \mathcal{Z}^n(\beta, h) \rangle - 1). \quad (4.12)$$

Usando a identidade

$$\exp[\beta\Theta(x)] = 1 - (1 - e^\beta)\Theta(x), \quad (4.13)$$

a função partição 4.6 pode ser reescrita como

$$\mathcal{Z}(\beta, h) = \sum_{\{J_{ij}\}} \prod_{\mu} \left\{ e^{-\beta} + (1 - e^{-\beta})\Theta \left[\xi_i^{\mu} \left(\frac{1}{N} \sum_j J_{ij} \xi_j^{\mu} - T_i \right) - \kappa \right] \right\}. \quad (4.14)$$

Replicando a equação (4.14) n vezes, indicando explicitamente o traço sobre as variáveis dinâmicas contínuas e introduzindo a condição de vínculo esférico (equação 4.11), temos

$$\begin{aligned} \mathcal{Z}^n(\beta, h) &= C \prod_{\alpha=1}^n \int \prod_j dJ_{ij}^{\alpha} \delta \left(\sum_j (J_{ij}^{\alpha})^2 - N \right) \\ &\quad \prod_{\mu} \left\{ e^{-\beta} + (1 - e^{-\beta})\Theta \left[\xi_i^{\mu} \left(\frac{1}{N} \sum_j J_{ij}^{\alpha} \xi_j^{\mu} - T_i \right) - \kappa \right] \right\}, \end{aligned} \quad (4.15)$$

sendo

$$C = \left[\int \prod_j dJ_{ij} \delta \left(\sum_j J_{ij}^2 - N \right) \right]^{-n}. \quad (4.16)$$

Usando as representações integrais

$$\delta(x) = \int_{-\infty}^{+\infty} \frac{dy}{2\pi} e^{ixy} \quad (4.17)$$

e

$$\Theta(x - \kappa) = \int_{\kappa}^{\infty} d\lambda \int_{-\infty}^{+\infty} \frac{dy}{2\pi} e^{iy(\lambda - x)}, \quad (4.18)$$

reescrevemos (4.15) e (4.16) na forma

$$\begin{aligned} \mathcal{Z}^n(\beta, h) &= C \int \prod_{\alpha=1}^n \frac{d\varepsilon_{\alpha}}{2\pi} \int \prod_j dJ_{ij}^{\alpha} \exp \left[i\varepsilon_{\alpha} \left(\sum_j (J_{ij}^{\alpha})^2 - N \right) \right] \prod_{\mu} \left[e^{-\beta} \int_{-\infty}^{\kappa} \frac{d\lambda_{\mu}^{\alpha}}{2\pi} \right. \\ &\quad \left. + \int_{\kappa}^{\infty} \frac{d\lambda_{\mu}^{\alpha}}{2\pi} \right] \int_{-\infty}^{+\infty} dx_{\mu}^{\alpha} \exp \left[ix_{\mu}^{\alpha} \left(\lambda_{\mu}^{\alpha} - \frac{1}{\sqrt{N}} \sum_j J_{ij}^{\alpha} \xi_i^{\mu} \xi_j^{\mu} + \xi_i^{\mu} T_i \right) \right] \end{aligned} \quad (4.19)$$

e

$$C = \left\{ \int \frac{d\varepsilon}{2\pi} \int \prod_j dJ_{ij} \exp \left[i\varepsilon \left(\sum_j J_{ij}^2 - N \right) \right] \right\}^{-n}. \quad (4.20)$$

O próximo passo consiste em tomar a média de $\mathcal{Z}^n(\beta, h)$ sobre a distribuição de probabilidades 3.1. Definimos

$$\psi = \left\langle \exp \left(-i \sum_{\alpha} x_{\mu}^{\alpha} J_{ij}^{\alpha} \frac{1}{\sqrt{N}} \xi_i^{\mu} \xi_j^{\mu} \right) \right\rangle_{\{\xi_j^{\mu}\}}. \quad (4.21)$$

ξ_i^{μ} e ξ_j^{μ} são variáveis aleatórias independentes. É conveniente tomar a média sobre os neurônios j em primeiro lugar. Temos então

$$\begin{aligned} \psi &= \frac{1}{2}(1+a) \exp \left(-i \sum_{\alpha} x_{\mu}^{\alpha} J_{ij}^{\alpha} \frac{1}{\sqrt{N}} \xi_i^{\mu} \right) + \frac{1}{2}(1-a) \exp \left(i \sum_{\alpha} x_{\mu}^{\alpha} J_{ij}^{\alpha} \frac{1}{\sqrt{N}} \xi_i^{\mu} \right) \\ &= \cos \left(\sum_{\alpha} x_{\mu}^{\alpha} J_{ij}^{\alpha} \frac{1}{\sqrt{N}} \xi_i^{\mu} \right) - i a \sin \left(\sum_{\alpha} x_{\mu}^{\alpha} J_{ij}^{\alpha} \frac{1}{\sqrt{N}} \xi_i^{\mu} \right). \end{aligned} \quad (4.22)$$

Expandindo até $\mathcal{O}(1/N)$, temos

$$\psi \equiv 1 - \frac{1}{2} \sum_{\alpha\beta} x_{\mu}^{\alpha} x_{\mu}^{\beta} \frac{1}{N} J_{ij}^{\alpha} J_{ij}^{\beta} - i a \sum_{\alpha} x_{\mu}^{\alpha} \frac{1}{\sqrt{N}} J_{ij}^{\alpha} \xi_i^{\mu}. \quad (4.23)$$

Ao tomar o logaritmo e manter apenas os termos dominantes, resulta

$$\ln \psi = -\frac{1}{2}(1-a^2) \sum_{\alpha\beta} x_{\mu}^{\alpha} x_{\mu}^{\beta} \frac{1}{N} J_{ij}^{\alpha} J_{ij}^{\beta} - i a \sum_{\alpha} x_{\mu}^{\alpha} \frac{1}{\sqrt{N}} J_{ij}^{\alpha} \xi_i^{\mu}. \quad (4.24)$$

Somando sobre $j \neq i$ surge o termo

$$\sum_{\alpha\beta} x_{\mu}^{\alpha} x_{\mu}^{\beta} \frac{1}{N} \sum_{j \neq i} J_{ij}^{\alpha} J_{ij}^{\beta} = \sum_{\alpha} (x_{\mu}^{\alpha})^2 + 2 \sum_{\alpha < \beta} q_{\alpha\beta} x_{\mu}^{\alpha} x_{\mu}^{\beta}, \quad (4.25)$$

onde

$$q_{\alpha\beta} = \frac{1}{N} \sum_{j \neq i} J_{ij}^{\alpha} J_{ij}^{\beta} \quad (4.26)$$

De (4.19) e (4.24), ficamos com

$$\begin{aligned}
\langle Z^n(\beta, h) \rangle_{\{\xi_j^\mu\}} &= C \int \prod_{\alpha=1}^n \frac{d\varepsilon_\alpha}{2\pi} \int \prod_j dJ_{ij}^\alpha \exp \left[i\varepsilon_\alpha \left(\sum_j (J_{ij}^\alpha)^2 - N \right) \right] \\
&\cdot \prod \left[e^{-\beta} \int_{-\infty}^{\kappa} \frac{d\lambda_\mu^\alpha}{2\pi} + \int_{\kappa}^{\infty} \frac{d\lambda_\mu^\alpha}{2\pi} \right] \int_{-\infty}^{+\infty} \prod_{\alpha<\beta} dq_{\alpha\beta} \int_{-\infty}^{+\infty} \prod_{\alpha} dM_\alpha \int_{-\infty}^{+\infty} dx_\mu^\alpha \\
&\cdot \exp \left[i \sum_{\mu\alpha} x_\mu^\alpha \lambda_\mu^\alpha - \frac{1}{2}(1-a^2) \left(\sum_{\mu\alpha} (x_\mu^\alpha)^2 + 2 \sum_{\mu} \sum_{\alpha<\beta} x_\mu^\alpha x_\mu^\beta q_{\alpha\beta} \right) - i \sum_{\mu\alpha} (aM_\alpha - T) x_\mu^\alpha \xi_i^\mu \right] \\
&\cdot \prod_{\alpha<\beta} \delta \left(q_{\alpha\beta} - \frac{1}{N} \sum_{j \neq i} J_{ij}^\alpha J_{ij}^\beta \right) \prod_{\alpha} \delta \left(M_\alpha - \frac{1}{\sqrt{N}} \sum_j J_{ij}^\alpha \right), \quad (4.27)
\end{aligned}$$

onde

$$M_\alpha = \frac{1}{\sqrt{N}} \sum_{j \neq i} J_{ij}^\alpha \quad (4.28)$$

é a magnetização no espaço das conexões sinápticas.

Introduzindo nas representações integrais das funções delta (segundo a equação 4.17) os campos $\phi_{\alpha\beta}$ e K_α conjugados a $q_{\alpha\beta}$ e M_α respectivamente, temos uma expressão desacoplada nos índices de neurônio e padrão, onde tanto as variáveis aleatórias quanto as dinâmicas são independentes. O resultado é

$$\begin{aligned}
\langle Z^n(\beta, h) \rangle_{\{\xi_j^\mu\}} &= C \int \prod_{\alpha=1}^n \frac{d\varepsilon_\alpha}{2\pi} \int \prod_{\alpha=1}^n \frac{dK_\alpha dM_\alpha}{2\pi} \int \prod_{\alpha<\beta} \frac{d\phi_{\alpha\beta} dq_{\alpha\beta}}{2\pi} \\
&\cdot \left\langle \exp N \left[\alpha G_1(q_{\alpha\beta}, M_\alpha) + G_2(\phi_{\alpha\beta}, \varepsilon_\alpha, K_\alpha) + i \sum_{\alpha<\beta} \phi_{\alpha\beta} q_{\alpha\beta} + \frac{1}{\sqrt{N}} \sum_{\alpha} K_\alpha M_\alpha \right] \right\rangle_{\xi_i}, \quad (4.29)
\end{aligned}$$

onde $\{\xi^\mu\}$ indica a média sobre ξ_i^μ e ξ_j^μ , e sendo

$$\begin{aligned}
G_1(q_{\alpha\beta}, M_\alpha) &= \ln \left[e^{-\beta} \int_{-\infty}^{\kappa} \frac{d\lambda_\alpha}{2\pi} + \int_{\kappa}^{\infty} \frac{d\lambda_\alpha}{2\pi} \right] \int_{-\infty}^{+\infty} dx_\alpha \\
&\cdot \left\langle \exp \left[i \sum_{\alpha} x_\alpha (\lambda_\alpha - (aM_\alpha - T)) \xi_i - \frac{1}{2}(1-a^2) \left(\sum_{\alpha} (x_\alpha)^2 + 2 \sum_{\alpha<\beta} x_\alpha x_\beta q_{\alpha\beta} \right) \right] \right\rangle_{\xi_i} \quad (4.30)
\end{aligned}$$

e

$$G_2(\phi_{\alpha\beta}, \varepsilon_\alpha, K_\alpha) = \ln \left\{ \int \prod_{\alpha} dJ_\alpha \exp \left[i \sum_{\alpha} \varepsilon_\alpha (J_\alpha^2 - 1) - i \sum_{\alpha<\beta} \phi_{\alpha\beta} \right] \right\}. \quad (4.31)$$

No limite $N \rightarrow \infty$ a integral em (4.29) é dominada pelos valores das variáveis de integração que maximizam o argumento da exponencial. Neste limite, o termo $1/\sqrt{N} \sum_{\alpha} K_{\alpha} M_{\alpha}$ é desprezível em relação aos demais. Integrando (4.29) pelo método do ponto de sela, temos

$$\langle \mathcal{Z}^n(\beta, h) \rangle_{\{\xi^{\mu}\}} \sim \exp \{ nN G(q_{\alpha\beta}, M_{\alpha}, \phi_{\alpha\beta}, \varepsilon_{\alpha}) \} . \quad (4.32)$$

sendo

$$G(q_{\alpha\beta}, M_{\alpha}, \phi_{\alpha\beta}, \varepsilon_{\alpha}) = \frac{\alpha}{n} G_1(q_{\alpha\beta}, M_{\alpha}) + \frac{1}{n} G_2(\phi_{\alpha\beta}, \varepsilon_{\alpha}) + i \frac{1}{n} \sum_{\alpha < \beta} \phi_{\alpha\beta} q_{\alpha\beta} \quad (4.33)$$

No limite $n \rightarrow 0$, a expressão acima resulta em

$$\langle \mathcal{Z}^n(\beta, h) \rangle_{\{\xi^{\mu}\}} \sim 1 + nN G(q_{\alpha\beta}, M_{\alpha}, \phi_{\alpha\beta}, \varepsilon_{\alpha}) . \quad (4.34)$$

O lado direito da equação será calculado no ponto de sela de simetria de réplicas:

$$\begin{aligned} q_{\alpha\beta} &= q \quad \text{todo } \alpha \neq \beta , \\ \varepsilon_{\alpha} &= iE \quad \text{todo } \alpha , \\ \phi_{\alpha\beta} &= iF \quad \text{todo } \alpha \neq \beta , \\ M_{\alpha} &= M \quad \text{todo } \alpha . \end{aligned} \quad (4.35)$$

Mediante o cálculo de $G(q_{\alpha\beta}, M_{\alpha}, \phi_{\alpha\beta}, \varepsilon_{\alpha})$ no apêndice B obtemos

$$\begin{aligned} \frac{1}{N} \langle \mathcal{Z}(\beta, h) \rangle_{\{\xi^{\mu}\}} &= \max \left\{ \alpha \left[\frac{1}{2}(1+a) \int \mathcal{D}y \ln(1 + \gamma_+(q)) \right. \right. \\ &\quad \left. \left. + \frac{1}{2}(1-a) \int \mathcal{D}y \ln(1 + \gamma_-(q)) \right] + E + \frac{1}{2} \ln 2\pi \right. \\ &\quad \left. - \frac{1}{2} \ln(2E + F) + \frac{1}{2} \frac{F}{2E + F} + \frac{1}{2} Fq \right\} . \end{aligned} \quad (4.36)$$

sendo

$$\gamma_{\pm} = (e^{-\beta} - 1) \int_{-\infty}^{\frac{\kappa \mp av}{\sqrt{1-a^2}} - \sqrt{q}y} \mathcal{D}\lambda \quad (4.37)$$

e

$$av = aM - T . \quad (4.38)$$

Equações de ponto de sela

Os pontos estacionários de $1/N \langle \ln \mathcal{Z}(\beta, h) \rangle_{\xi^\mu}$ são obtidos igualando-se a zero as derivadas em relação a E , F , q e v do lado direito de (4.36).

Da derivada em relação a E temos

$$\frac{2(E + F)}{(2E + F)^2} = 1. \quad (4.39)$$

Derivando em relação a F , obtemos

$$q = \frac{F}{(2E + F)^2} = 1. \quad (4.40)$$

Estas equações permitem eliminar E e F algebricamente:

$$E = \frac{1 - 2q}{2(1 - q)^2} \quad (4.41)$$

e

$$F = q(1 - q)^2. \quad (4.42)$$

Combinando (4.41) e (4.42) com (4.36) resulta em

$$\frac{1}{N} \langle \mathcal{Z}(\beta, h) \rangle_{\{\xi^\mu\}} = \max_{q,v} G(q, v), \quad (4.43)$$

sendo

$$G(q, v) = \alpha \left[\frac{1}{2}(1 + a) \int \mathcal{D}y \ln(1 + \gamma_+(q)) \frac{1}{2}(1 - a) \int \mathcal{D}y \ln(1 + \gamma_-(q)) \right] + \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln(1 - q) + \frac{1}{2(1 - q)}. \quad (4.44)$$

A fração de erros é minimizada pela escolha ótima do conjunto de sinapses. Na solução de simetria de réplicas, a evolução dinâmica conduz todas as réplicas do sistema para o mesmo atrator no espaço de configurações. De acordo com (4.26), q é o valor em simetria de réplicas do ‘overlap’ entre os estados de cada par de réplicas. Portanto, para $\beta \rightarrow \infty$,

temos $q \rightarrow 1$. Calculando o lado direito de (4.44) neste limite, temos

$$\begin{aligned}
-\beta\alpha\epsilon_{min} &\equiv \frac{1}{N} \langle \mathcal{Z}(\beta, h) \rangle_{\{\xi^\mu\}} \\
&= \beta \max_{x,v} \left\{ \alpha \left[\frac{1}{2}(1+a) \left(- \int_{\frac{\kappa-av}{\sqrt{1-a^2}}-x}^{\frac{\kappa-av}{\sqrt{1-a^2}}} \mathcal{D}y \frac{1}{x^2} \left(\frac{\kappa-av}{\sqrt{1-a^2}} - y \right)^2 - \int_{-\infty}^{\frac{\kappa-av}{\sqrt{1-a^2}}-x} \mathcal{D}y \right) \right. \right. \\
&\quad \left. \left. + \frac{1}{2}(1-a) \left(- \int_{\frac{\kappa+av}{\sqrt{1-a^2}}-x}^{\frac{\kappa+av}{\sqrt{1-a^2}}} \mathcal{D}y \frac{1}{x^2} \left(\frac{\kappa+av}{\sqrt{1-a^2}} - y \right)^2 - \int_{-\infty}^{\frac{\kappa+av}{\sqrt{1-a^2}}-x} \mathcal{D}y \right) \right] + \frac{1}{x^2} \right\}, \quad (4.45)
\end{aligned}$$

onde

$$x \equiv \sqrt{2\beta(1-q)}. \quad (4.46)$$

Os parâmetros x e v são determinados pelas equações de ponto de sela

$$\frac{\partial}{\partial v} \frac{1}{N} \langle \mathcal{Z}(\beta, h) \rangle_{\{\xi^\mu\}} = 0 \quad (4.47)$$

e

$$\frac{\partial}{\partial x} \frac{1}{N} \langle \mathcal{Z}(\beta, h) \rangle_{\{\xi^\mu\}} = 0. \quad (4.48)$$

Explicitamente, temos

$$\alpha \left[\frac{1}{2}(1+a) \int_{\frac{\kappa-av}{\sqrt{1-a^2}}-x}^{\frac{\kappa-av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa-av}{\sqrt{1-a^2}} - y \right)^2 + \frac{1}{2}(1-a) \int_{\frac{\kappa+av}{\sqrt{1-a^2}}-x}^{\frac{\kappa+av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa+av}{\sqrt{1-a^2}} - y \right)^2 \right] = 1 \quad (4.49)$$

e

$$\frac{1}{2}(1+a) \int_{\frac{\kappa-av}{\sqrt{1-a^2}}-x}^{\frac{\kappa-av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa-av}{\sqrt{1-a^2}} - y \right) = \frac{1}{2}(1-a) \int_{\frac{\kappa+av}{\sqrt{1-a^2}}-x}^{\frac{\kappa+av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa+av}{\sqrt{1-a^2}} - y \right) \quad (4.50)$$

respectivamente.

De (4.45) e (4.49) temos a fração mínima de erros,

$$\epsilon_{min} = \frac{1}{2}(1+a) \int_{-\infty}^{\frac{\kappa-av}{\sqrt{1-a^2}}-x} \mathcal{D}y + \frac{1}{2}(1-a) \int_{-\infty}^{\frac{\kappa+av}{\sqrt{1-a^2}}-x} \mathcal{D}y. \quad (4.51)$$

A equação (4.49) só tem solução para $x < \infty$ se $\alpha > \alpha_c$, sendo

$$\alpha_c = \left[\frac{1}{2}(1+a) \int_{-\infty}^{\frac{\kappa-av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa-av}{\sqrt{1-a^2}} \right)^2 + \frac{1}{2}(1-a) \int_{-\infty}^{\frac{\kappa+av}{\sqrt{1-a^2}}} \mathcal{D}y \left(\frac{\kappa+av}{\sqrt{1-a^2}} \right)^2 \right]^{-1}. \quad (4.52)$$

As equações (4.52) e (4.50) calculada em $x \rightarrow \infty$, reproduzem o resultado de Gardner [23] para a capacidade crítica de armazenamento de padrões correlacionados. Efetivamente, introduzindo $x \rightarrow \infty$ em (4.51) resulta em $\epsilon_{min} = 0$.

Limite $a \rightarrow 0$

A solução do problema em simetria de réplicas para $a \neq 0$ envolve o sistema de equações (4.49) e (4.50) nas variáveis x e v . Quando $a = 0$, a equação (4.50) é trivial. Neste caso v é irrelevante, pois a equação (4.49) torna-se independente de v , resultando na equação de Gardner–Derrida para padrões não correlacionados.

Apesar da irrelevância de v em $a = 0$, é interessante determinar o seu comportamento para a pequeno. Expandindo-se (4.49) e (4.50) até $\mathcal{O}(a)$ obtemos uma expressão explícita para v , que no limite $a = 0$ é

$$v = \frac{1}{D(\kappa, x)} \left[\kappa \operatorname{erf} \left(\frac{\kappa}{\sqrt{2}} \right) - \kappa \operatorname{erf} \left(\frac{\kappa - x}{\sqrt{2}} \right) + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\kappa^2} - \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}(\kappa-x)^2} \right], \quad (4.53)$$

onde

$$D(\kappa, x) = \operatorname{erf} \left(\frac{\kappa}{\sqrt{2}} \right) - \operatorname{erf} \left(\frac{\kappa - x}{\sqrt{2}} \right) - \sqrt{\frac{2}{\pi}} x e^{-\frac{1}{2}(\kappa-x)^2}, \quad (4.54)$$

e sendo x obtido de (4.49) com $a = 0$.

Quando $\kappa = 0$, de (4.49) e (4.53) temos

$$v = \frac{\alpha}{\sqrt{2\pi}} \left(1 - e^{-\frac{1}{2}x^2} \right). \quad (4.55)$$

Portanto, quando $\kappa = 0$, $v(a = 0) > 0$ para qualquer α . Em particular, quando $\alpha < \alpha_c = 2$, então $x \rightarrow \infty$ e $v = \alpha/\sqrt{2\pi}$, em concordância com o resultado de Gardner [23].

Neste trabalho estamos interessados em estudar a situação $\alpha > \alpha_c$. Nesta situação x é finito, e existe um valor crítico κ para o qual $D(\kappa, x)$ é nulo. Na figura 4.3 temos a curva $\kappa_c(\alpha)$. Sobre esta curva $v(a = 0)$ diverge, e a solução do sistema de equações (4.49) e (4.50) não converge para a mesma solução do problema sem correlações (ref. [24]). A função $D(\kappa, x)$ troca de sinal em κ_c : $v(a = 0) \rightarrow +\infty$ para $\kappa \rightarrow \kappa_c^-$ e $v(a = 0) \rightarrow -\infty$ para $\kappa \rightarrow \kappa_c^+$ (ver as curvas para a pequeno na figura 4.1).

A figura 4.1 mostra que para $\kappa < \kappa_c$ só existe uma solução para v para cada valor de κ e a . Para $\kappa > \kappa_c$ e $a \neq 0$ existem três soluções. Uma solução em que $v \rightarrow +\infty$ ao

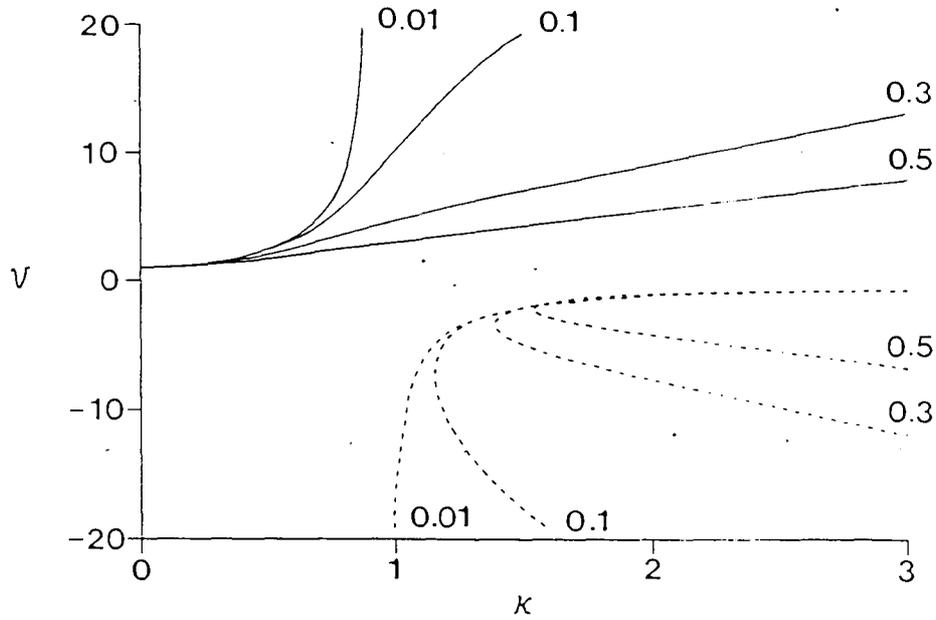


Figura 4.1: *Parâmetro de ordem v como função de κ para $\alpha = 3$ e a entre 0,01 e 0,5.*

$\kappa \rightarrow \kappa_c$, outra em que $v \rightarrow -\infty$ e uma terceira em que $v \rightarrow 0$. O comportamento de $\epsilon_{min}(\kappa)$ correspondente a cada solução é visto na figura 4.2. O comportamento assintótico, obtido da equação (4.51) é o seguinte:

$$\epsilon_{min} = \begin{cases} \frac{1}{2}(1 - a) & \text{se } v \rightarrow +\infty \\ \frac{1}{2} & \text{se } v \rightarrow 0 \\ \frac{1}{2}(1 + a) & \text{se } v \rightarrow -\infty . \end{cases} \quad (4.56)$$

Fisicamente só deve existir uma solução que corresponde à fração mínima de erros de armazenamento. A escolha mais atraente necessariamente recairia sobre a primeira das equações acima, pois ela é realmente o mínimo entre as três. Entretanto, a situação exige a investigação da estabilidade local do ponto de sela de simetria de réplicas correspondente a cada solução.

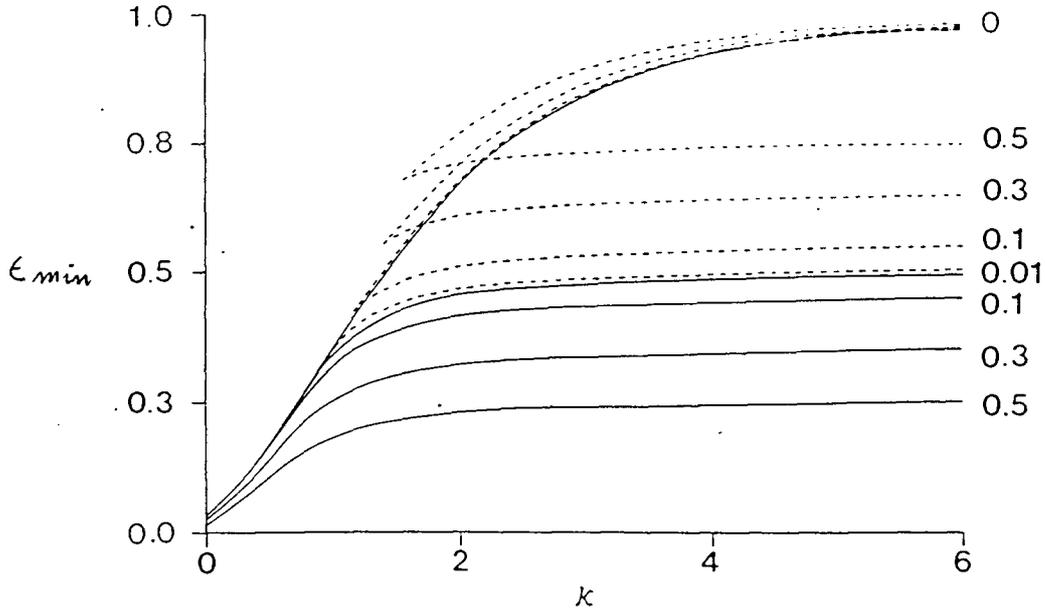


Figura 4.2: Fração mínima de erros como função de κ para $\alpha = 3$ e a entre 0,01 e 0,5. As curvas cheias e pontilhadas são relativas à figura 4.1

4.2 Estabilidade da solução de simetria de réplicas

As soluções de simetria de réplicas que estudamos na seção 4.1 são estáveis quando são máximos locais da função $G(q_{\alpha\beta}, M_\alpha, \phi_{\alpha\beta}, \varepsilon_\alpha)$ definida pela equação (4.33). Uma solução troca de estabilidade quando pelo menos um dos autovalores das flutuações quadráticas de $G(q_{\alpha\beta}, M_\alpha, \phi_{\alpha\beta}, \varepsilon_\alpha)$ troca de sinal. A matriz \mathbf{A} que descreve estas flutuações possui a forma geral

$$\mathbf{A} = \begin{pmatrix} \alpha \mathbf{A}_1 & \mathbf{I} & 0 \\ & 0 & 0 \\ \mathbf{I} & 0 & \mathbf{A}_2 \\ 0 & 0 & \end{pmatrix}, \quad (4.57)$$

sendo

$$\mathbf{A}_1 = \begin{pmatrix} \frac{\partial^2 G_1}{\partial q_{\alpha\beta} \partial q_{\gamma\delta}} & \frac{\partial^2 G_1}{\partial q_{\alpha\beta} \partial M_\gamma} \\ \frac{\partial^2 G_1}{\partial q_{\alpha\beta} \partial M_\gamma} & \frac{\partial^2 G_1}{\partial M_\alpha \partial M_\beta} \end{pmatrix}, \quad (4.58)$$

$$\mathbf{A}_2 = \begin{pmatrix} \frac{\partial^2 G_2}{\partial i\phi_{\alpha\beta}\partial i\phi_{\gamma\delta}} & \frac{\partial^2 G_2}{\partial i\phi_{\alpha\beta}\partial i\epsilon_\gamma} \\ \frac{\partial^2 G_2}{\partial i\phi_{\alpha\beta}\partial i\epsilon_\gamma} & \frac{\partial^2 G_2}{\partial i\epsilon_\alpha\partial i\epsilon_\beta} \end{pmatrix} \quad (4.59)$$

e \mathbf{I} é a matriz identidade de ordem $n(n-1)/2$.

Comparando as equações (4.58) e (4.59) com de Almeida e Thouless [26] observa-se que as matrizes \mathbf{A}_1 e \mathbf{A}_2 no ponto de sela de simetria de réplicas apresentam as mesmas propriedades de simetria da matriz que descreve as flutuações quadráticas da solução de Sherrington e Kirkpatrick [25] para o modelo de vidros de spin de Edwards e Anderson [31].

O espaço das réplicas é formado por $n(n-1)/2$ variáveis $q_{\alpha\beta}$ e $\phi_{\alpha\beta}$ e n variáveis M_α e ϵ_α . Uma determinada solução no espaço das réplicas é representada por um vetor deste espaço. Denominam-se flutuações *longitudinais* em torno de uma solução àquelas flutuações que não alteram a direção do vetor que representa a solução. Denominam-se flutuações *transversais* àquelas flutuações que envolvem apenas uma das variáveis do espaço das réplicas.

Os elementos das matrizes \mathbf{A}_1 e \mathbf{A}_2 são os seguintes:

$$P_i \equiv \frac{\partial^2 G_i}{\partial q_{\alpha\beta}\partial q_{\alpha\beta}}, \quad (4.60)$$

$$Q_i \equiv \frac{\partial^2 G_i}{\partial q_{\alpha\beta}\partial q_{\alpha\gamma}}, \quad (4.61)$$

$$R_i \equiv \frac{\partial^2 G_i}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}}, \quad (4.62)$$

$$C_i \equiv \frac{\partial^2 G_i}{\partial q_{\alpha\beta}\partial M_\alpha}, \quad (4.63)$$

$$D_i \equiv \frac{\partial^2 G_i}{\partial q_{\alpha\beta}\partial M_\gamma}, \quad (4.64)$$

$$A_i \equiv \frac{\partial^2 G_i}{\partial M_\alpha\partial M_\alpha}, \quad (4.65)$$

$$B_i \equiv \frac{\partial^2 G_i}{\partial M_\alpha\partial M_\beta}, \quad (4.66)$$

sendo $i = 1, 2$. Cada matriz tem os seguintes autovalores [26]:

1. Dois autovalores não degenerados que descrevem flutuações longitudinais [26] no espaço de réplicas nas variáveis $q_{\alpha\beta}$ e M_α ou $i\phi_{\alpha\beta}$ e $i\varepsilon_\alpha$:

$$\begin{aligned} \lambda_{1,i}^\pm &= \frac{1}{2} \left[A_i + (n-1)B_i + P_i + 2(n-2)Q_i + \frac{1}{2}(n-2)(n-3)R_i \right] \\ &\pm \left\{ \left[A_i + (n-1)B_i - P_i - 2(n-2)Q_i - \frac{1}{2}(n-2)(n-3)R_i \right]^2 \right. \\ &\left. + 2(n-1) \left[2C_i + (n-2)D_i \right]^2 \right\}^{\frac{1}{2}}. \end{aligned} \quad (4.67)$$

2. Dois autovalores com degenerescência $n-1$, correspondentes a autovetores que descrevem flutuações em direções que misturam as variáveis $q_{\alpha\beta}$ e M_α ou $i\phi_{\alpha\beta}$ e $i\varepsilon_\alpha$:

$$\begin{aligned} \lambda_{2,i}^\pm &= \frac{1}{2} \left[A_i - B_i + P_i + (n-4)Q_i - (n-3)R_i \right] \\ &\pm \left\{ \left[A_i - B_i - P_i - (n-4)Q_i - (n-3)R_i \right]^2 + 4(n-2)(C_i - D_i)^2 \right\}^{\frac{1}{2}}. \end{aligned} \quad (4.68)$$

3. Um autovalor com degenerescência $n(n-3)/2$, que corresponde a autovetores que descrevem flutuações transversais [26] nas variáveis $q_{\alpha\beta}$ ou $i\phi_{\alpha\beta}$:

$$\lambda_{3,i} = P_i - 2Q_i + R_i. \quad (4.69)$$

As flutuações longitudinais não afastam o sistema da solução de simetria de réplicas. A questão então é se esta solução é estável em um sub-espaço de quatro dimensões varrido pelas variáveis q, M, E e F . Se a solução de simetria de réplicas é única, então estas flutuações não provocam sua instabilidade [24]. De (4.67) e (4.68) temos, para $n=0$,

$$\lambda_{1,i}^\pm = \lambda_{2,i}^\pm = \frac{1}{2}(A_i - B_i + P_i - 4Q_i + 3R_i) \pm \frac{1}{2} \left[(A_i - B_i - P_i - 4Q_i - 3R_i) - 8(C_i - D_i) \right]^{\frac{1}{2}}, \quad (4.70)$$

de onde conclui-se que as flutuações em direções que misturam as variáveis $q_{\alpha\beta}$ e M_α ou $i\phi_{\alpha\beta}$ e $i\varepsilon_\alpha$ são degeneradas com as longitudinais quando $n=0$, e portanto impõem as mesmas condições de estabilidade que aquelas.

Investigamos a estabilidade da solução de simetria de réplicas no subespaço das flutuações transversais de $q_{\alpha\beta}$ e $i\phi_{\alpha\beta}$. A estabilidade quanto a estas flutuações é determinada

pelos autovalores da matriz

$$\begin{pmatrix} \alpha\lambda_{3,1} & 1 \\ 1 & \lambda_{3,2} \end{pmatrix}. \quad (4.71)$$

Quando $\alpha = 0$ qualquer conjunto $\{J_{ij}\}$ satisfaz o conjunto de vínculos 4.3. Neste caso, a equação (4.15) é simplesmente o volume do espaço de configurações, e a solução de simetria de réplicas é estável. O produto dos autovalores de (4.71) é -1 quando $\alpha \rightarrow 0$. Partindo-se de $\alpha = 0$, que é uma situação de estabilidade da solução de simetria de réplicas, a transição para uma fase que exige quebra de simetria é marcada pela troca de sinal do produto dos autovalores de (4.71), ou seja

$$1 - \alpha\lambda_{3,1}\lambda_{3,2} > 0 \quad (4.72)$$

é a condição de estabilidade. Os dois autovalores de (4.71) tem sinais opostos na fase de estabilidade da solução de simetria de réplicas devido à escolha da variável F segundo $\phi_{\alpha\beta} = iF$. Sem isto os autovalores teriam o mesmo sinal, que é condição para que $G(q_{\alpha\beta}, M_{\alpha}, \phi_{\alpha\beta}, \varepsilon_{\alpha})$ seja um máximo.

Diagonalizando as matrizes \mathbf{A}_1 e \mathbf{A}_2 (apêndice C), temos respectivamente os autovalores

$$\begin{aligned} \lambda_{3,1} = & \frac{1}{(1-q)^2} \left\{ \frac{1}{4}(1+a) \left[\operatorname{erf} \left(\frac{\kappa - av}{\sqrt{2(1-a^2)}} \right) - \operatorname{erf} \left(\frac{\kappa - av}{\sqrt{2(1-a^2)} - x} \right) \right] \right. \\ & \left. + \frac{1}{4}(1-a) \left[\operatorname{erf} \left(\frac{\kappa + av}{\sqrt{2(1-a^2)}} \right) - \operatorname{erf} \left(\frac{\kappa + av}{\sqrt{2(1-a^2)}} \right) \right] \right\} \end{aligned} \quad (4.73)$$

e

$$\lambda_{3,1} = (1-q)^2. \quad (4.74)$$

Introduzindo as duas equações acima em (4.72) temos a condição

$$\begin{aligned} 1 - \frac{1}{2}\alpha \left\{ \frac{1}{2}(1+a) \left[\operatorname{erf} \left(\frac{\kappa - av}{\sqrt{2(1-a^2)}} \right) - \operatorname{erf} \left(\frac{\kappa - av}{\sqrt{2(1-a^2)} - x} \right) \right] \right. \\ \left. + \frac{1}{2}(1-a) \left[\operatorname{erf} \left(\frac{\kappa + av}{\sqrt{2(1-a^2)}} \right) - \operatorname{erf} \left(\frac{\kappa + av}{\sqrt{2(1-a^2)}} \right) \right] \right\} > 0, \end{aligned} \quad (4.75)$$

que garante a estabilidade da solução de simetria de réplicas contra as flutuações transversais. A figura 4.3 ilustra as curvas que assinalam a igualdade na expressão (4.75) sobre

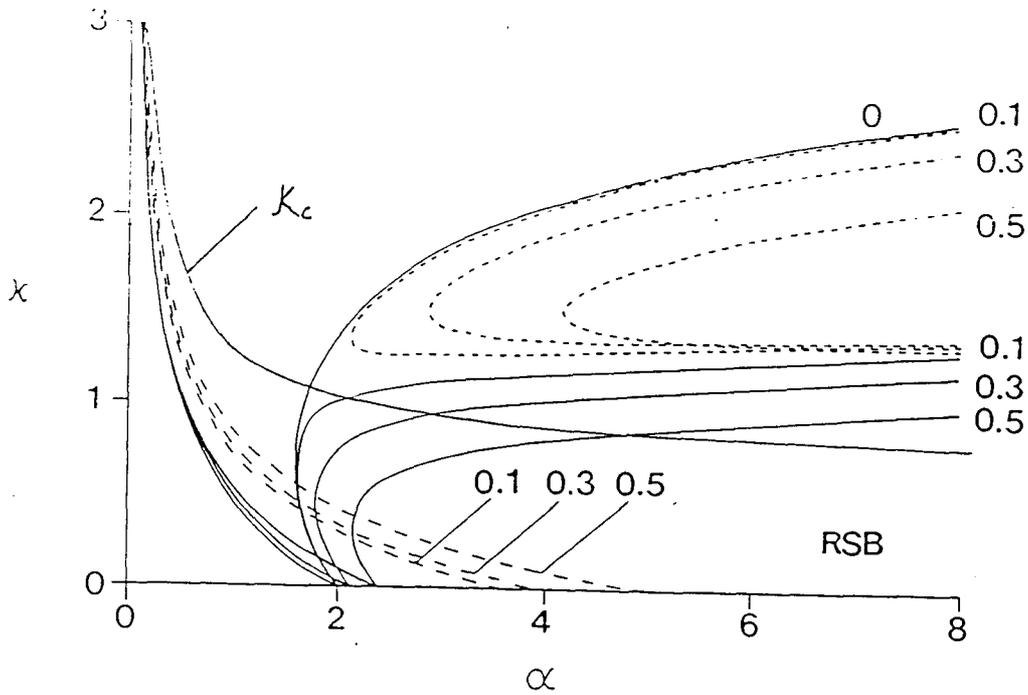


Figura 4.3: Capacidade crítica de armazenamento para a rede ótima, e linhas críticas para a quebra de simetria de réplicas para a entre 0 e 0,5. A figura mostra também a curva $\kappa = \kappa_c(\alpha)$ e a capacidade crítica de armazenamento para $\epsilon_{min} = 0,05$, extrapolando os resultados de simetria de réplicas.

diversos planos (κ, α) de polarização a constante do diagrama de fases. A curva $\kappa = \kappa_c$ também aparece na figura. Só há transição contínua entre $a = 0$ e $a \neq 0$ quando $\kappa < \kappa_c$. A figura 4.3 mostra também a capacidade crítica de armazenamento, que corresponde a $\epsilon_{min} = 0$, além de curvas que correspondem a $\epsilon_{min} = 0,05$, para diversos valores de a . Estas últimas curvas penetram na fase que exige quebra de simetria de réplicas.

Embora a polarização dos padrões resulte no acréscimo importante da fase de simetria de réplica estável, em grande parte do diagrama de fases ela permanece instável. Em particular, quando $\kappa \approx 0$ e $\alpha > \alpha_c$. Esta constatação evidencia a necessidade de procurar soluções para as equações de ponto de sela não restritas às condições de simetria de réplicas representadas pelas equações (4.35).

Vemos na figura 4.3 que a solução de simetria de réplicas não é válida em uma região importante do plano (α, κ) . É necessário, portanto, procurar outra solução. No próximo capítulo discutiremos o primeiro passo da quebra de simetria de réplicas, do esquema

proposto por Parisi [43]

5. Quebra de simetria de réplicas no problema de Gardner–Derrida

A solução de simetria de réplicas para o problema Gardner–Derrida é instável sobre uma importante região no plano (α, κ) , como indica a figura 4.3. Em particular, ela é instável para $\alpha > \alpha_c$, quando κ é pequeno. Nesta situação, diferentes conjuntos de conexões sinápticas, correspondendo a diferentes pontos no espaço de configurações, podem resultar em frações mínimas de erros iguais, embora os erros de armazenamento localizem-se em padrões diferentes. As curvas de estabilidade sobre o diagrama de fases (κ, α) , que vimos no capítulo anterior, mostram o limite da região em que a solução que se restringe a um ponto apenas no espaço de configurações é localmente estável. Esta solução é caracterizada por $q_{\alpha\beta} \rightarrow 1$, para todo par (α, β) .

Neste capítulo descreveremos os nossos cálculos num procedimento de Parisi [43], que permite determinar soluções das equações de ponto de sela não restritas às condições de simetria de réplicas (equações 4.34).

5.1 O esquema de Parisi

O primeiro passo do esquema de Parisi para a quebra de simetria de réplicas [44, 45, 46] consiste em dividir o conjunto de n réplicas em n/m grupos de m réplicas, onde m é um submúltiplo de n . Os elementos de matriz \mathbf{q} assumem o valor $q_{\alpha\beta} = q_1$ se α e β estão no mesmo grupo de réplicas e $q_{\alpha\beta} = q_0$ caso contrário. Os elementos diagonais permanecem

5.2 Padrões não correlacionados

A partir da equação (4.32), a energia livre efetiva de Gardner–Derrida é dada por

$$G(q_{\alpha\beta}, \phi_{\alpha\beta}, \varepsilon_\alpha) = \lim_{n \rightarrow 0} \left[\sum_{\alpha < \beta} \phi_{\alpha\beta} q_{\alpha\beta} + \frac{\alpha}{n} G_1(q_{\alpha\beta}) + \frac{1}{n} G_2(\phi_{\alpha\beta}, \varepsilon_\alpha) \right], \quad (5.6)$$

onde fizemos as trocas de variáveis $\phi_{\alpha\beta} \rightarrow i\phi_{\alpha\beta}$ e $\varepsilon_\alpha \rightarrow i\varepsilon_\alpha$. As funções $G_1(q_{\alpha\beta})$ e $G_2(\phi_{\alpha\beta}, \varepsilon_\alpha)$ são dadas respectivamente por 4.30 e 4.31, calculadas em $a = 0$. No primeiro passo da quebra de simetria, após um desenvolvimento similar ao do apêndice B, obtém-se que a energia livre é dada pelo extremo da função.

$$\begin{aligned} G_{rsb}^{(1)}(q_0, q_1, F_0, F_1, E, m) &= \frac{1}{2} [mF_0q_0 + (1-m)F_1q_1] + E + \frac{1}{2} \ln 2\pi \\ &+ \frac{1-m}{2m} \ln(2E + F_1) - \frac{1}{2m} \ln[2E + F_1 - m(F_1 - F_0)] \\ &+ \frac{1}{2} \frac{F_0}{2E + F_1 - m(F_1 - F_0)} + \frac{\alpha}{m} \int \mathcal{D}z_0 \ln \Psi(q_0, q_1, m), \end{aligned} \quad (5.7)$$

sendo

$$\Psi(q_0, q_1, m) = \int \mathcal{D}z_1 \Phi^m(q_0, q_1), \quad (5.8)$$

e

$$\Phi(q_0, q_1) = e^{-\beta} + \frac{1}{2}(1 - e^{-\beta}) \left[1 - \operatorname{erf} \left(\frac{\kappa - z_0\sqrt{q_0} - z_1\sqrt{q_1 - q_0}}{\sqrt{2(1 - q_1)}} \right) \right]. \quad (5.9)$$

F_0 e F_1 são os valores que $\phi_{\alpha\beta}$ assume neste primeiro passo de quebra de simetria. Estas variáveis, juntamente com E , são eliminadas algebricamente pelas equações de ponto de sela

$$\frac{\partial G_{rsb}^{(1)}}{\partial F_0} = \frac{\partial G_{rsb}^{(1)}}{\partial F_1} = \frac{\partial G_{rsb}^{(1)}}{\partial E} = 0. \quad (5.10)$$

A equação (5.7) depende então de apenas três parâmetros, q_0 , q_1 e m :

$$\begin{aligned} G_{rsb}^{(1)}(q_0, q_1, m) &= \frac{1}{2} + \frac{1}{2} \ln 2\pi - \frac{1-m}{2m} \ln(1 - q_1) \\ &+ \frac{1}{2m} \ln[1 - q_1 + m(q_1 - q_0)] \\ &+ \frac{1}{2} \frac{q_0}{1 - q_1 + m(q_1 - q_0)} + \frac{\alpha}{m} \int \mathcal{D}z_0 \ln \Psi(q_0, q_1, m). \end{aligned} \quad (5.11)$$

Os valores de q_0 , q_1 e m são dados pelas equações de ponto de sela:

$$\frac{\partial G_{rsb}^{(1)}}{\partial q_0} = \frac{1}{2} \frac{mq_0}{[1 - q_1 + m(q_1 - q_0)]^2} + \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi} \frac{\partial \Psi}{\partial q_0} = 0, \quad (5.12)$$

$$\begin{aligned} \frac{\partial G_{rsb}^{(1)}}{\partial q_1} = & \frac{1 - m}{2m} \left\{ \frac{1}{1 - q_1} - \frac{1}{1 - q_1 + m(q_1 - q_0)} \right. \\ & \left. + \frac{mq_0}{[1 - q_1 + m(q_1 - q_0)]^2} \right\} + \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi} \frac{\partial \Psi}{\partial q_1} = 0 \end{aligned} \quad (5.13)$$

e

$$\begin{aligned} \frac{\partial G_{rsb}^{(1)}}{\partial m} = & \frac{1}{2m^2} \{ \ln(1 - q_1) - \ln[1 - q_1 + m(q_1 - q_0)] \} \\ & + \frac{1}{2m} \frac{q_1 - q_0}{1 - q_1 + m(q_1 - q_0)} \left[1 - \frac{mq_0}{1 - q_1 + m(q_1 - q_0)} \right] \\ & - \frac{\alpha}{m^2} \int \mathcal{D}z_0 \ln \Psi + \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi} \frac{\partial \Psi}{\partial m} = 0. \end{aligned} \quad (5.14)$$

5.3 Padrões correlacionados

O único termo de (5.6) que se altera com a introdução de correlações entre os padrões é $G_1(q_{\alpha\beta}, M_\alpha)$. A energia livre depende também da variável $v_\alpha = (aM_\alpha - T)/a$. Esta variável que depende de apenas um índice de réplica, foi mantida simétrica. Após eliminar algebricamente as variáveis E , F_0 e F_1 , temos

$$\begin{aligned} G_{rsb}^{(1)}(q_0, q_1, m, v) = & \frac{1}{2} + \frac{1}{2} \ln 2\pi - \frac{1 - m}{2m} \ln(1 - q_1) \\ & + \frac{1}{2m} \ln[1 - q_1 + m(q_1 - q_0)] + \frac{1}{2} \frac{q_0}{1 - q_1 + m(q_1 - q_0)} \\ & + \frac{1}{2} (1 + a) \frac{\alpha}{m} \int \mathcal{D}z_0 \ln \Psi_+(q_0, q_1, m, v) + \frac{1}{2} (1 - a) \frac{\alpha}{m} \int \mathcal{D}z_0 \ln \Psi_-(q_0, q_1, m, v), \end{aligned} \quad (5.15)$$

onde

$$\Psi_\pm(q_0, q_1, m, v) = \int \mathcal{D}z_1 \Phi_\pm^m(q_0, q_1, v), \quad (5.16)$$

$$\Phi_\pm(q_0, q_1, v) = e^{-\beta} + \frac{1}{2} (1 - e^{-\beta}) \left[1 - \operatorname{erf} \left(\frac{K_\pm(q_0, q_1, v)}{\sqrt{2}} \right) \right], \quad (5.17)$$

e

$$K_{\pm} = \frac{\frac{\kappa \mp av}{\sqrt{1-a^2}} + z_0 \sqrt{q_0} + z_1 \sqrt{q_1 - q_0}}{\sqrt{1 - q_1}}. \quad (5.18)$$

Com a presença da variável v , temos para resolver agora quatro equações de ponto de sela:

$$\begin{aligned} \frac{\partial G_{rsb}^{(1)}}{\partial q_0} &= \frac{1}{2} \frac{mq_0}{[1 - q_1 + m(q_1 - q_0)]^2} + \frac{1}{2}(1+a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_+} \frac{\partial \Psi_+}{\partial q_0} \\ &\quad + \frac{1}{2}(1-a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_-} \frac{\partial \Psi_-}{\partial q_0} = 0, \end{aligned} \quad (5.19)$$

$$\begin{aligned} \frac{\partial G_{rsb}^{(1)}}{\partial q_1} &= \frac{1-m}{2m} \left\{ \frac{1}{1-q_1} - \frac{1}{1-q_1 + m(q_1 - q_0)} + \frac{mq_0}{[1 - q_1 + m(q_1 - q_0)]^2} \right\} \\ &\quad + \frac{1}{2}(1+a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_+} \frac{\partial \Psi_+}{\partial q_1} + \frac{1}{2}(1-a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_-} \frac{\partial \Psi_-}{\partial q_1} = 0, \end{aligned} \quad (5.20)$$

$$\begin{aligned} \frac{\partial G_{rsb}^{(1)}}{\partial m} &= \frac{1}{2m^2} \{ \ln(1 - q_1) - \ln[1 - q_1 + m(q_1 - q_0)] \} \\ &\quad + \frac{1}{2m} \frac{q_1 - q_0}{1 - q_1 + m(q_1 - q_0)} \left[1 - \frac{mq_0}{1 - q_1 + m(q_1 - q_0)} \right] \end{aligned} \quad (5.21)$$

$$\begin{aligned} &- \frac{1}{2}(1+a) \left[\frac{\alpha}{m^2} \int \mathcal{D}z_0 \ln \Psi_+ + \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_+} \frac{\partial \Psi_+}{\partial m} \right] \\ &- \frac{1}{2}(1-a) \left[\frac{\alpha}{m^2} \int \mathcal{D}z_0 \ln \Psi_- + \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_-} \frac{\partial \Psi_-}{\partial m} \right] = 0 \end{aligned} \quad (5.22)$$

e

$$\frac{\partial G_{rsb}^{(1)}}{\partial q_1} = +\frac{1}{2}(1+a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_+} \frac{\partial \Psi_+}{\partial v} + \frac{1}{2}(1-a) \frac{\alpha}{m} \int \mathcal{D}z_0 \frac{1}{\Psi_-} \frac{\partial \Psi_-}{\partial v} = 0. \quad (5.23)$$

O comportamento ótimo da rede é obtido no limite $\beta \rightarrow \infty$. Neste limite, m e $1 - q_1$ se reescalam ambos com $1/\beta$, de modo que a razão $m/(1 - q_1)$ permanece finita.

6. Conclusões

Neste trabalho estudamos o armazenamento de padrões correlacionados em modelos de redes neurais atratoras sob dois enfoques distintos: termodinâmica de equilíbrio com interações sinápticas simétricas e o método de Gardner, que envolve cálculos no espaço das conexões sinápticas, sem impor restrições sobre a simetria destas.

No capítulo 2 fizemos uma rápida revisão sobre os aspectos gerais do problema de sistemas neurais do ponto de vista da mecânica estatística. Discutimos a respeito das dinâmicas síncrona e assíncrona. Discutimos também os trabalhos de Amit, Gutfreund e Sompolinsky [7, 8], nos quais nos apoiamos para estudar o armazenamento de padrões correlacionados no modelo de Hopfield.

Iniciamos o capítulo 3 com uma descrição geral de duas redes biológicas, o modelo AGS [9] e o modelo de Perez-Vicente e Amit [18]. A seguir, estudamos o armazenamento de padrões correlacionados no modelo de Hopfield, cuja simplicidade permitiu uma descrição analítica dos estados espúrios simétricos e assimétricos. Na seção 3.2 discutimos a evolução de alguns destes estados em função da temperatura e do parâmetro a no caso do armazenamento de um número finito de padrões correlacionados. Obtivemos como principais resultados a constatação de que os estados assimétricos espúrios são desestabilizados com o aumento da temperatura e a verificação de que há uma tendência dominante para o fortalecimento dos estados simétricos com o aumento de a . Esta seção corresponde ao detalhamento de um trabalho original [35].

Na seção 3.3 estudamos também o armazenamento de finitos padrões correlacionados, mas juntamente com infinitos padrões não correlacionados, o que permite introduzir o

parâmetro intensivo α . Verificamos a correspondência entre o ruído sináptico e o ruído provocado por correlações microscópicas com os padrões que não condensam [8]. Como resultado mais representativo, temos a dependência do diagrama de fases termodinâmicas com a polarização a . No plano (α, T) , observa-se um grande incremento da superfície ocupada pela fase simétrica com o aumento de a . Esta seção foi publicada na forma de uma versão condensada [38]. Traçamos um paralelo entre as propriedades de armazenamento de padrões correlacionados e os modelos AGS e de Perez-Vicente e Amit. Embora mais eficiente que o modelo de Hopfield com a regra de aprendizagem de Hebb, que estudamos, o modelo AGS também é inapropriado para o armazenamento de padrões fortemente correlacionados ($a \rightarrow 1$). Por outro lado, os modelos de Perez-Vicente e Amit, e de Buchmann, Divko e Shulten [17], que envolvem mudanças na representação dos estados dos neurônios, têm desempenhos próximos ao ótimo previsto por Gardner [23].

O capítulo 3 contém os resultados de simulações numéricas sobre redes neurais de 200 e 400 neurônios dos modelos de Hopfield e AGS com padrões correlacionados. Verificamos que o valor limite de a para a presença de estados de recuperação concorda com o resultado analítico obtido na seção 3.2. Os resultados sugerem que no modelo AGS com finitos padrões armazenados, o estado de recuperação está presente para todo valor de a no intervalo $[-1, 1]$. O tamanho reduzido das redes com as quais trabalhamos não permitiu conclusões definitivas a respeito de bacias de atração em torno dos estados de recuperação.

No capítulo 4, passamos para o estudo de redes neurais segundo o enfoque de Gardner. O conteúdo deste artigo, corresponde a uma parte de uma publicação original [54]. Utilizando o método de Gardner e Derrida [24] calculamos, em simetria de réplicas, a fração mínima de erros de armazenamento por sítio da rede, quando o número de padrões apresentados à ela é maior que a sua capacidade de armazenamento. Também nesta etapa do trabalho obtivemos resultados importantes. Um destes resultados dá conta da existência de uma linha sobre o plano (κ, α) , sobre a qual a magnetização no espaço das interações sinápticas é divergente no limite $a \rightarrow 0$. A origem desta divergência merece uma investigação detalhada. Ela é uma demonstração de que a introdução de correlações entre os padrões produz efeitos

não triviais na teoria de Gardner e Derrida.

Efetuamos a análise da estabilidade da solução de simetria de réplicas, e constatamos que sobre uma região importante do plano (κ, α) ela é instável. Esta instabilidade, associada com a investigação da origem da divergência na magnetização v sugeriram a procura de soluções de quebra de simetria de réplicas. Deduzimos as equações formais de quebra de simetria, até o primeiro passo do esquema de Parisi.

A fração mínima de erros, e mesmo a capacidade crítica de armazenamento sem erros, são parcelas pequenas da informação que pode ser obtida do conjunto de técnicas analíticas que compõem o chamado ‘programa Gardner’. Ambas quantidades são funções das estabilidades locais $\{\gamma_i^\mu\}$, definidas pela equação 4.2. De acordo com Kepler e Abott [49, 50], o método de Gardner permite calcular qualquer função de $\{\gamma_i^\mu\}$, inclusive a sua distribuição $P(\{\gamma_i^\mu\})$, que também segundo Kepler e Abott desempenha importante papel no estudo das propriedades dinâmicas da rede construída com o conjunto ótimo de conexões sinápticas.

Como vimos neste trabalho para o caso de padrões correlacionados, e no trabalho de Gardner e Derrida para padrões sem correlações, a aceitação de uma fração de padrões erroneamente armazenados em cada sítio resulta no aumento da capacidade de armazenamento da rede. Entretanto, é importante que a rede que opera acima da saturação tenha preservada sua capacidade de recuperação de padrões. Em outras palavras, é necessário que a evolução dinâmica do sistema o conduza para atratores próximos dos padrões armazenados. A relação entre as propriedades de armazenamento e recuperação no regime de saturação foi investigada por Amit, Evans, Horner e Wong [51], para uma rede diluída, na qual o número de coordenação de cada neurônio é muito menor que N . Neste regime, a integração das equações dinâmicas para os overlaps revela que só há atratores onde a solução de simetria de réplicas é estável. Entretanto, as propriedades de redes totalmente conexas podem ser completamente distintas da rede diluída correspondente, como pôde ser verificado no estudo da dinâmica do modelo de Hopfield por Derrida, Gardner e Zippelius [52]. Simulações numéricas [53] sugerem que o aprendizado no regime de saturação favorece as propriedades de recuperação. O treinamento com padrões contendo erros permite

alargar a dimensão das bacias de atração em torno dos estados de recuperação. O estudo analítico das propriedades dinâmicas de redes totalmente conexas, tanto dentro do regime de saturação, quanto fora dele, ainda está por ser realizado.

Finalizamos ressaltando a dedução das equações formais de quebra de simetria de réplicas. O encontro de soluções para estas equações poderá estender a investigação das propriedades de armazenamento e recuperação acima da saturação na direção da região em que a solução de simetria de réplicas é instável.

A. Integração das componentes ‘altas’

Definimos

$$\Psi = \left\langle \exp \left\{ \beta N \left[-\frac{1}{2} \sum_{\mu>s} \sum_{\alpha} (m_{\alpha}^{\mu})^2 + \sum_{\mu>s} \sum_{\alpha} m_{\alpha}^{\mu} \frac{1}{N} \sum_i \xi_i^{\mu} S_i^{\alpha} \right] \right\} \right\rangle_{\{\xi_i^{\mu}\}, \mu>s} \quad (\text{A.1})$$

Calculando explicitamente a média indicada obtemos

$$\Psi = \exp \left\{ -\frac{1}{2} \beta N \sum_{\mu>s} \sum_{\alpha} (m_{\alpha}^{\mu})^2 + \sum_{\mu>s} \sum_i \ln \left[\cosh \left(\beta \sum_{\alpha} m_{\alpha}^{\mu} S_i^{\alpha} \right) \right] \right\}. \quad (\text{A.2})$$

As componentes m_{α}^{μ} da expressão acima são de ordem $1/\sqrt{N}$. Com a troca de variáveis

$$m_{\alpha}^{\mu} \rightarrow \frac{m_{\alpha}^{\mu}}{\sqrt{\beta N}}, \quad (\text{A.3})$$

temos $m_{\alpha}^{\mu} \sim \mathcal{O}(1)$. A equação A.2 é reescrita como

$$\Psi = \exp \left\{ -\frac{1}{2} \sum_{\mu>s} \sum_{\alpha} (m_{\alpha}^{\mu})^2 + \sum_{\mu>s} \sum_i \ln \left[\cosh \left(\sqrt{\frac{\beta}{N}} \sum_{\alpha} m_{\alpha}^{\mu} S_i^{\alpha} \right) \right] \right\}. \quad (\text{A.4})$$

Para $N \rightarrow \infty$, podemos expandir o argumento do coseno-hiperbólico até ordem $1/N$, obtendo

$$\Psi = \exp \left\{ -\frac{1}{2} \sum_{\mu>s} \sum_{\alpha} (m_{\alpha}^{\mu})^2 + \frac{\beta}{2N} \sum_{\mu>s} \sum_i \sum_{\alpha\beta} m_{\alpha}^{\mu} m_{\beta}^{\mu} S_i^{\alpha} S_i^{\beta} \right\}. \quad (\text{A.5})$$

O argumento da exponencial em (A.5) pode ser reescrito como

$$-\frac{1}{2} \sum_{\mu} \mathbf{m}_{\mu} \cdot [(1 - \beta)\mathbf{I} - \beta\mathbf{q}] \mathbf{m}_{\mu}, \quad (\text{A.6})$$

onde \mathbf{I} é a matriz identidade e \mathbf{q} é uma matriz simétrica com elementos

$$q_{\alpha\beta} = \frac{1}{N} \sum_i S_i^\alpha S_i^\beta, \quad \alpha \neq \beta \quad (\text{A.7})$$

e

$$q_{\alpha\alpha} = 0. \quad (\text{A.8})$$

Utilizando a forma (A.6), introduzindo as variáveis $q_{\alpha\beta}$ em (A.7) e integrando sobre todas componentes altas dos ‘overlaps’ temos

$$\begin{aligned} \int \prod_{\mu>s} \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \Psi &= \int \prod_{\mu>s} \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \prod_{\alpha<\beta} \left[dq_{\alpha\beta} \delta \left(q_{\alpha\beta} - \frac{1}{N} \sum_i S_i^\alpha S_i^\beta \right) \right] \\ &\quad \exp \left\{ -\frac{1}{2} \sum_{\mu} \mathbf{m}_{\mu} \cdot [(1-\beta)\mathbf{I} - \beta\mathbf{q}] \mathbf{m}_{\mu} \right\}. \end{aligned} \quad (\text{A.9})$$

A exponencial só contém termos independentes nas componentes dos ‘overlaps’ no espaço dos padrões μ , de modo que podemos reescrever a equação (A.9) na forma

$$\begin{aligned} \int \prod_{\mu>s} \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \Psi &= \prod_{\alpha<\beta} \left[dq_{\alpha\beta} \delta \left(q_{\alpha\beta} - \frac{1}{N} \sum_i S_i^\alpha S_i^\beta \right) \right] \\ &\quad \left(\int \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \mathbf{m} \cdot [(1-\beta)\mathbf{I} - \beta\mathbf{q}] \mathbf{m} \right\} \right)^{p-s}. \end{aligned} \quad (\text{A.10})$$

A integral entre parênteses pode ser avaliada por meio da identidade de Hubbard-Stratonowich ([48]),

$$\left(\frac{1}{\sqrt{2\pi}} \right)^n \int d\mathbf{x} \exp \left(-\frac{1}{2} \mathbf{x} \cdot \mathbf{A} \mathbf{x} \right) = \frac{1}{\sqrt{|\mathbf{A}|}}, \quad (\text{A.11})$$

onde $|\mathbf{A}|$ indica o determinante da matriz \mathbf{A} no espaço das n réplicas. Aplicando esta identidade e introduzindo a representação integral para a $\delta(x)$,

$$\delta(x) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} dz e^{-zx}, \quad (\text{A.12})$$

obtemos, a menos de constantes multiplicativas,

$$\begin{aligned} \int \prod_{\mu>s} \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \Psi &= \int \prod_{\alpha<\beta} dq_{\alpha\beta} dr_{\alpha\beta} \exp N \left\{ -\frac{1}{2} \alpha \text{tr} \ln [(1-\beta)\mathbf{I} - \beta\mathbf{q}] \right. \\ &\quad \left. -\frac{1}{2} \alpha \beta^2 \sum_{\alpha, \beta \neq \alpha} r_{\alpha\beta} q_{\alpha\beta} + \frac{1}{2} \alpha \beta^2 \sum_{\alpha, \beta \neq \alpha} r_{\alpha\beta} \frac{1}{N} \sum_i S_i^\alpha S_i^\beta \right\}. \end{aligned} \quad (\text{A.13})$$

É importante esclarecer a origem do primeiro termo da exponencial na expressão acima. Considerando que $p \gg s$ e utilizando a identidade (A.11), o termo de (A.10) que é elevado à potência $p - s$ resulta

$$\begin{aligned} |(1 - \beta)\mathbf{I} - \beta\mathbf{q}|^{-\frac{1}{2}p} &= \exp N \left(-\frac{1}{2}\alpha \ln |(1 - \beta)\mathbf{I} - \beta\mathbf{q}| \right) \\ &= \exp N \left(-\frac{1}{2}\alpha \text{tr} \ln [(1 - \beta)\mathbf{I} - \beta\mathbf{q}] \right), \end{aligned} \quad (\text{A.14})$$

onde usamos a identidade

$$\ln |\mathbf{A}| = \text{tr} \ln \mathbf{A}. \quad (\text{A.15})$$

Introduzindo (A.13) em (2.62), obtém-se a média sobre as componentes altas da função partição replicada,

$$\begin{aligned} \langle \mathcal{Z}^n(\beta, h) \rangle_{\{\xi_i^\mu\}} &\sim e^{-\frac{1}{2}\beta pn} \int \prod_{\mu \leq s} \prod_{\alpha} \frac{dm_\alpha^\mu}{\sqrt{2\pi}} \int \prod_{\alpha < \beta} dq_{\alpha\beta} dr_{\alpha\beta} \\ &\cdot \exp N \left\{ -\frac{1}{2}\beta \sum_{\mu \leq s} \sum_{\alpha} (m_\alpha^\mu)^2 - \frac{1}{2}\alpha \text{tr} \ln [(1 - \beta)\mathbf{I} - \beta\mathbf{q}] - \frac{1}{2}\alpha\beta^2 \sum_{\alpha \neq \beta} r_{\alpha\beta} q_{\alpha\beta} \right\} \\ &\cdot \left\langle \sum_{\{S_i^\alpha\}} \exp \left\{ \frac{1}{2}\alpha\beta^2 \sum_{i\alpha\beta} r_{\alpha\beta} S_i^\alpha S_i^\beta + \beta \sum_{\mu \leq s} \sum_{\alpha} (m_\alpha^\mu + h^\mu) \sum_i \xi_i^\mu S_i^\alpha \right\} \right\rangle_{\{\xi_i^\mu\}}, \end{aligned} \quad (\text{A.16})$$

onde resta tomar a média sobre as componentes baixas $\{\xi_i^\mu\}$, $\mu \leq s$. O fator que depende desta média e do traço sobre as variáveis de estado $\{S_i^\alpha\}$ pode ser reescrito na forma

$$\left\langle \exp \sum_i \ln \sum_{\{S^\alpha\}} \exp \left\{ \frac{1}{2}\alpha\beta^2 \sum_{\alpha\beta} r_{\alpha\beta} S^\alpha S^\beta + \beta \sum_{\mu \leq s} \sum_{\alpha} (m_\alpha^\mu + h^\mu) \xi_i^\mu S^\alpha \right\} \right\rangle_{\{\xi_i^\mu\}}. \quad (\text{A.17})$$

Mas s é um número finito e, portanto, pode-se usar a propriedade da autome-diância para tomar a média, resultando

$$\exp N \left\langle \ln \sum_{\{S^\alpha\}} \exp \left\{ \frac{1}{2}\alpha\beta^2 \sum_{\alpha\beta} r_{\alpha\beta} S^\alpha S^\beta + \beta \sum_{\mu \leq s} \sum_{\alpha} (m_\alpha^\mu + h^\mu) \xi^\mu S^\alpha \right\} \right\rangle_{\{\xi^\mu\}}. \quad (\text{A.18})$$

Assim temos

$$\begin{aligned}
\langle Z^n(\beta, h) \rangle &\sim e^{-\frac{1}{2}\beta Nn} \int \prod_{\mu \leq s} \prod_{\alpha} \frac{dm_{\alpha}^{\mu}}{\sqrt{2\pi}} \int \prod_{\alpha < \beta} dq_{\alpha\beta} dr_{\alpha\beta} \\
&\cdot \exp N \left\{ -\frac{1}{2}\beta \sum_{\mu \leq s} \sum_{\alpha} (m_{\alpha}^{\mu})^2 - \frac{1}{2}\alpha \ln[(1-\beta)\mathbf{I} - \beta\mathbf{q}] - \frac{1}{2}\alpha\beta^2 \sum_{\alpha \neq \beta} r_{\alpha\beta} q_{\alpha\beta} \right. \\
&+ \left. \left\langle \ln \sum_{\{S^{\alpha}\}} \exp \left[\frac{1}{2}\alpha\beta^2 \sum_{\alpha\beta} r_{\alpha\beta} S^{\alpha} S^{\beta} + \beta \sum_{\mu \leq s} \sum_{\alpha} (m_{\alpha}^{\mu} + h^{\mu}) \xi^{\mu} S^{\alpha} \right] \right\rangle_{\{\xi^{\mu}\}} \right\}, \quad (\text{A.19})
\end{aligned}$$

que é a expressão (2.63).

B. Cálculo de $G_1(q_{\alpha\beta}, M_\alpha)$ e $G_2(\phi_{\alpha\beta}, \varepsilon_\alpha)$

Vamos mostrar primeiro o cálculo de $G_1(q_{\alpha\beta}, M_\alpha)$. Tomando a média sobre ξ_i em (4.30) temos

$$G_1(q_{\alpha\beta}, M_\alpha) = \ln[\chi_+(q_{\alpha\beta}, M_\alpha) + \chi_-(q_{\alpha\beta}, M_\alpha)], \quad (\text{B.1})$$

onde

$$\begin{aligned} \chi_\pm = \prod_\alpha \left[e^{-\beta} \int_{-\infty}^\kappa \frac{d\lambda_\alpha}{2\pi} + \int_\kappa^\infty \frac{d\lambda_\alpha}{2\pi} \right] \int_{-\infty}^{+\infty} dx_\alpha \\ \cdot \left\{ \frac{1}{2}(1 \pm a) \exp \left[i \sum_\alpha x_\alpha [\lambda_\alpha \mp (aM_\alpha - T)] - \frac{1}{2}(1 - a^2) \left(\sum_\alpha x_\alpha^2 + \sum_{\alpha < \beta} x_\alpha x_\beta \right) \right] \right\}. \end{aligned} \quad (\text{B.2})$$

Calculando (B.2) em simetria de réplicas, e usando que

$$\sum_\alpha x_\alpha^2 + 2q \sum_{\alpha < \beta} x_\alpha x_\beta = (1 - q) \sum_\alpha x_\alpha^2 + q \left(\sum_\alpha x_\alpha \right)^2, \quad (\text{B.3})$$

resulta em

$$\begin{aligned} \chi_\pm(q, M) = \frac{1}{2}(1 \mp a) \prod_\alpha \left[e^{-\beta} \int_{-\infty}^\kappa \frac{d\lambda_\alpha}{2\pi} + \int_\kappa^\infty \frac{d\lambda_\alpha}{2\pi} \right] \prod_\alpha \int_{-\infty}^{+\infty} dx_\alpha \\ \cdot \exp \left\{ i \sum_\alpha x_\alpha [\lambda_\alpha \mp (aM - T)] - \frac{1}{2}(1 - a^2) \left[(1 - q) \sum_\alpha x_\alpha^2 + q \left(\sum_\alpha x_\alpha \right)^2 \right] \right\}. \end{aligned} \quad (\text{B.4})$$

Linearizando o termo quadrático com o auxílio da integral gaussiana (equação 2.21), integrando sobre x_α , e introduzindo o resultado em (B.4), temos

$$\chi_\pm(q, M) = \frac{1}{2}(1 \pm a) \int \mathcal{D}y \left[\left(e^{-\beta} \int_{-\infty}^\kappa \frac{d\lambda}{2\pi} + \int_\kappa^\infty \frac{d\lambda}{2\pi} \right) \psi_\pm \right]^n, \quad (\text{B.5})$$

sendo

$$\psi_{\pm} = \sqrt{\frac{2\pi}{(1-a^2)(1-q)}} \exp \left\{ -\frac{\left[\frac{\lambda_{\mp}(aM-T)}{\sqrt{1-a^2}} + \sqrt{q}y \right]^2}{2(1-q)} \right\}. \quad (\text{B.6})$$

Fazendo a troca de variáveis

$$\frac{\frac{\lambda_{\mp}v}{\sqrt{1-a^2}} - \sqrt{q}y}{\sqrt{1-q}} \rightarrow \lambda, \quad (\text{B.7})$$

sendo v definido por (4.38), para $n \rightarrow 0$, temos

$$\chi(q, v) = \frac{1}{2}(1 \pm a) \left[1 + n \int \mathcal{D}y \ln[1 + \gamma_{\pm}(q, v)] \right], \quad (\text{B.8})$$

onde $\gamma_{\pm}(q, v)$ definido por 4.37. Introduzindo B.8 em B.1, temos

$$\frac{1}{n}G_1(q, v) = \frac{1}{2}(1+a) \int \mathcal{D}y \ln[1 + \gamma_+(q, v)] + \frac{1}{2}(1-a) \int \mathcal{D}y \ln[1 + \gamma_-(q, v)]. \quad (\text{B.9})$$

Consideramos a seguir o cálculo de $G_2(\phi_{\alpha\beta}, \varepsilon_{\alpha})$. Com a condição de simetria de réplicas (equação 4.35), obtém-se de 4.31

$$G_2(F, E) = \ln \prod_{\alpha} \int_{-\infty}^{+\infty} dJ_{\alpha} \exp \left[-E \sum_{\alpha} (J_{\alpha}^2 - 1) + F \sum_{\alpha < \beta} J_{\alpha} J_{\beta} \right]. \quad (\text{B.10})$$

Linearizando o termo quadrático, temos

$$G_2(F, E) = \ln \int \mathcal{D}y \prod_{\alpha} \int_{-\infty}^{+\infty} dJ_{\alpha} \exp \left[-E(J_{\alpha} - 1) - \frac{1}{2}F J_{\alpha}^2 + 2y\sqrt{f}J_{\alpha} \right]. \quad (\text{B.11})$$

A equação acima resultou desacoplada nos índices de réplica. Após integrar em J e y , para $n \rightarrow 0$ temos como resultado

$$\frac{1}{n}G_2(E, F) = \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln(2E + F) + \frac{F}{2(2E + F)} + E. \quad (\text{B.12})$$

C. Diagonalização das matrizes A_1 e A_2

Reescrevemos (4.30) na forma

$$G_1(q_{\alpha\beta}, M_\alpha) = \ln \prod_\alpha \int \mathcal{D}\lambda_\alpha \int_{-\infty}^{+\infty} dx_\alpha \left[\frac{1}{2}(1+a)e^{\psi_+} + \frac{1}{2}(1-a)e^{\psi_-} \right], \quad (\text{C.1})$$

sendo

$$\psi_\pm = i \sum_\alpha x_\alpha [\lambda_\alpha \mp (aM_\alpha - T)] - \frac{1}{2}(1-a^2) \left(\sum_\alpha x_\alpha^2 + 2 \sum_{\alpha < \beta} x_\alpha x_\beta q_{\alpha\beta} \right) \quad (\text{C.2})$$

e

$$\int \mathcal{D}\lambda_\alpha \equiv e^{-\beta} \int_{-\infty}^{\kappa} \frac{d\lambda_\alpha}{2\pi} + \int_{\kappa}^{\infty} \frac{d\lambda_\alpha}{2\pi}. \quad (\text{C.3})$$

De (C.1) e (C.3) temos

$$\frac{\partial}{\partial q_{\alpha\beta}} G_1(q_{\alpha\beta}, M_\alpha) = -(1-a^2) \langle x_\alpha x_\beta \rangle \quad (\text{C.4})$$

e

$$\frac{\partial^2}{\partial q_{\alpha\beta} \partial q_{\gamma\delta}} G_1(q_{\alpha\beta}, M_\alpha) = (1-a^2) (\langle x_\alpha x_\beta x_\gamma x_\delta \rangle - \langle x_\alpha x_\beta \rangle \langle x_\gamma x_\delta \rangle), \quad (\text{C.5})$$

onde $\langle \dots \rangle$ significa a média definida por

$$\langle \dots \rangle = \frac{\prod_\alpha \int \mathcal{D}\lambda_\alpha \int_{-\infty}^{+\infty} dx_\alpha (\dots) \left[\frac{1}{2}(1+a)e^{\psi_+} + \frac{1}{2}(1-a)e^{\psi_-} \right]}{\prod_\alpha \int \mathcal{D}\lambda_\alpha \int_{-\infty}^{+\infty} dx_\alpha \left[\frac{1}{2}(1+a)e^{\psi_+} + \frac{1}{2}(1-a)e^{\psi_-} \right]}. \quad (\text{C.6})$$

Os elementos diagonais definidos pela equação (4.60) são dados por

$$P_1 = (1-a^2)^2 \left(\langle x_\alpha^2 x_\beta^2 \rangle - \langle x_\alpha x_\beta \rangle^2 \right). \quad (\text{C.7})$$

Para os elementos com um índice de réplica diferente, definidos por (4.61), temos

$$Q_1 = (1-a^2)^2 \left(\langle x_\alpha^2 x_\beta x_\gamma \rangle - \langle x_\alpha x_\beta \rangle \langle x_\alpha x_\gamma \rangle \right). \quad (\text{C.8})$$

Para os elementos com dois índices de réplicas diferentes, definidos pela equação (4.62), a equação (C.5) resulta em

$$R_1 = (1 - a^2)^2 (\langle x_\alpha x_\beta x_\gamma x_\delta \rangle - \langle x_\alpha x_\beta \rangle \langle x_\gamma x_\delta \rangle) . \quad (\text{C.9})$$

Usando a integral gaussiana (equação 2.21) para linearizar o termo quadrático, em simetria de réplicas temos, de C.2

$$\int \prod_\rho dx_\rho e^{\psi_\pm} = \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \int \prod_\rho dx_\rho x_\alpha^2 x_\beta^2 e^{\sum_\rho \varphi_\pm^\rho} , \quad (\text{C.10})$$

sendo

$$\varphi_\pm^\rho = ix_\rho(\lambda \mp av) - \frac{1}{2}(1 - a^2)(1 - q)x_\rho^2 - i\sqrt{(1 - a^2)qt}x_\rho . \quad (\text{C.11})$$

Combinando (C.6) e (C.10) para calcular $\langle x_\alpha x_\beta \rangle$ obtemos

$$\langle x_\alpha^2 x_\beta^2 \rangle = \frac{\int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \prod_\rho \int \mathcal{D}\lambda_\rho \int dx_\rho x_\alpha^2 x_\beta^2 \left(\frac{1}{2}(1 + a)e^{\sum_\rho \varphi_+^\rho} + \frac{1}{2}(1 - a)e^{\sum_\rho \varphi_-^\rho} \right)}{\int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \prod_\rho \int \mathcal{D}\lambda_\rho \int dx_\rho \left(\frac{1}{2}(1 + a)e^{\sum_\rho \varphi_+^\rho} + \frac{1}{2}(1 - a)e^{\sum_\rho \varphi_-^\rho} \right)} . \quad (\text{C.12})$$

Fatorando os os termos correspondentes à cada réplica, e tomando o limite $n \rightarrow 0$ a equação é reescrita como

$$\langle x_\alpha^2 x_\beta^2 \rangle = \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left[\frac{1}{2}(1 + a)\langle x^2(t) \rangle_+^2 + \frac{1}{2}(1 - a)\langle x^2(t) \rangle_-^2 \right] , \quad (\text{C.13})$$

onde as chaves $\langle \dots \rangle_\pm$ indicam a média definida por

$$\langle \dots \rangle_\pm \equiv \frac{\int \mathcal{D}\lambda \int dx (\dots) e^{\varphi_\pm}}{\int \mathcal{D}\lambda \int dx e^{\varphi_\pm}} . \quad (\text{C.14})$$

As demais médias que aparecem nas expressões (C.7)–(C.9) podem ser derivados de forma análoga, obtendo-se no limite $n \rightarrow 0$

$$\langle x_\alpha x_\beta \rangle = \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left[\frac{1}{2}(1 + a)\langle x(t) \rangle_+^2 + \frac{1}{2}(1 - a)\langle x(t) \rangle_-^2 \right] , \quad (\text{C.15})$$

$$\langle x_\alpha^2 x_\beta x_\gamma \rangle = \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left[\frac{1}{2}(1 + a)\langle x^2(t) \rangle_+ \langle x(t) \rangle_+^2 + \frac{1}{2}(1 - a)\langle x^2(t) \rangle_- \langle x(t) \rangle_-^2 \right] \quad (\text{C.16})$$

e

$$\langle x_\alpha x_\beta x_\gamma x_\delta \rangle = \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left[\frac{1}{2}(1 + a)\langle x(t) \rangle_+^4 + \frac{1}{2}(1 - a)\langle x(t) \rangle_-^4 \right] . \quad (\text{C.17})$$

O autovalor $\lambda_{3,1}$ é obtido combinando-se as equações (4.69), (C.7)–(C.9), usando (C.13) e (C.15)–(C.17):

$$\lambda_{3,1} = (1 - a^2)^2 \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} \left[\frac{1}{2}(1 + a) \left(\langle x^2(t) \rangle_+ - \langle x(t) \rangle_+^2 \right)^2 + \frac{1}{2}(1 - a) \left(\langle x^2(t) \rangle_- - \langle x(t) \rangle_-^2 \right)^2 \right]. \quad (\text{C.18})$$

O cálculo explícito dos valores médios definidos por (4.17) é imediato, e resulta em

$$\langle x(t) \rangle_{\pm} = i\chi_{\pm} \quad (\text{C.19})$$

e

$$\langle x^2(t) \rangle_{\pm} = -\frac{\kappa \mp av - \sqrt{(1 - a^2)qt}}{(1 - a^2)(1 - q)} \chi_{\pm}. \quad (\text{C.20})$$

onde definimos

$$\chi_{\pm} = \frac{(1 - e^{\beta})e^{-\phi_{\pm}^2}}{\sqrt{2\pi(1 - a^2)(1 - q)}} \left[e^{-\beta} + \frac{1}{2}(1 - e^{-\beta})(1 - \text{erf}\phi_{\pm}) \right], \quad (\text{C.21})$$

e

$$\phi_{\pm} = \frac{\kappa \mp av - \sqrt{(1 - a^2)qt}}{\sqrt{2(1 - a^2)(1 - q)}}. \quad (\text{C.22})$$

No limite $\beta \rightarrow \infty$ e $q \rightarrow 1$

$$\langle x^2(t) \rangle_{\pm} - \langle x(t) \rangle_{\pm}^2 \sim \frac{1}{(1 - a^2)(1 - q)}. \quad (\text{C.23})$$

introduzindo esta equação em (C.18) resulta na expressão (4.73) para $\lambda_{3,1}$.

Para diagonalizar a matriz \mathbf{A}_2 procederemos como Gardner e Derrida [24], pois a função $G_2(\phi_{\alpha\beta}, \varepsilon_{\alpha})$ independe de a . de (4.31) temos

$$\frac{\partial G_2}{\partial i\phi_{\alpha\beta}} = -\langle J_{\alpha}J_{\beta} \rangle \quad (\text{C.24})$$

e

$$\frac{\partial^2 G_2}{\partial i\phi_{\alpha\beta} \partial i\phi_{\gamma\delta}} = \langle J_{\alpha}J_{\beta}J_{\gamma}J_{\delta} \rangle - \langle J_{\alpha}J_{\beta} \rangle \langle J_{\gamma}J_{\delta} \rangle, \quad (\text{C.25})$$

As chaves $\langle \dots \rangle$ indicam a média definida por

$$\langle \dots \rangle = \frac{\prod_{\alpha} \int dJ_{\alpha}(\dots) \exp \left[i \sum_{\alpha} \varepsilon_{\alpha} (J_{\alpha}^2 - 1) - i \sum_{\alpha < \beta} \phi_{\alpha\beta} J_{\alpha} J_{\beta} \right]}{\prod_{\alpha} \int dJ_{\alpha} \exp \left[i \sum_{\alpha} \varepsilon_{\alpha} (J_{\alpha}^2 - 1) - i \sum_{\alpha < \beta} \phi_{\alpha\beta} J_{\alpha} J_{\beta} \right]}. \quad (\text{C.26})$$

Analogamente às equações (C.7)–(C.9), os elementos de matriz de \mathbf{A}_2 são

$$P_2 = \langle J_\alpha^2 J_\beta^2 \rangle - \langle J_\alpha J_\beta \rangle^2, \quad (\text{C.27})$$

$$Q_2 = \langle J_\alpha^2 J_\beta J_\gamma \rangle - \langle J_\alpha J_\beta \rangle \langle J_\alpha J_\gamma \rangle \quad (\text{C.28})$$

e

$$R_2 = \langle J_\alpha J_\beta J_\gamma J_\delta \rangle - \langle J_\alpha J_\beta \rangle \langle J_\gamma J_\delta \rangle. \quad (\text{C.29})$$

Seguindo passos semelhantes aos que levam de (C.10) a (C.14) e daí a (C.17), temos em simetria de réplicas

$$\langle J_\alpha^2 J_\beta^2 \rangle = \int_{-\infty}^{+\infty} \frac{dt}{2\pi F} \exp\left(\frac{t^2}{2F}\right) \langle J^2(t) \rangle^2, \quad (\text{C.30})$$

$$\langle J_\alpha^2 J_\beta J_\gamma \rangle = \int_{-\infty}^{+\infty} \frac{dt}{2\pi F} \exp\left(\frac{t^2}{2F}\right) \langle J^2(t) \rangle \langle J(t) \rangle^2 \quad (\text{C.31})$$

e

$$\langle J_\alpha J_\beta \rangle = \int_{-\infty}^{+\infty} \frac{dt}{2\pi F} \exp\left(\frac{t^2}{2F}\right) \langle J(t) \rangle^2. \quad (\text{C.32})$$

Agora as chaves indicam a média definida por

$$\langle \dots \rangle = \frac{\int_{-\infty}^{+\infty} dJ(\dots) \exp\left[-\frac{1}{2}(2E + F)J^2 - tJ\right]}{\int_{-\infty}^{+\infty} dJ \exp\left[-\frac{1}{2}(2E + F)J^2 - tJ\right]}. \quad (\text{C.33})$$

Combinando as expressões (C.27)–(C.29) com (C.30)–(4.17), e introduzindo o resultado em (4.69), obtemos o autovalor $\lambda_{3,2}$:

$$\lambda_{3,2} = \int_{-\infty}^{+\infty} \frac{dt}{2\pi F} \exp\left(\frac{t^2}{2F}\right) \left[\langle J^2(t) \rangle - \langle J(t) \rangle^2\right]^2. \quad (\text{C.34})$$

De (C.33) temos

$$\langle J^2(t) \rangle = \frac{t^2}{(2E + F)^2} + \frac{1}{2E + F} \quad (\text{C.35})$$

e

$$\langle J(t) \rangle = -\frac{t}{2E + F}. \quad (\text{C.36})$$

Combinando as expressões acima com (C.34), usando (4.41) e (4.42), resulta na equação (4.74).

Referências Bibliográficas

- [1] DOMANY, E., *J. Stat. Phys.*, **51** (5/6), (1988) 743.
- [2] McCULLOCH, W. S. e PITTS, W. A., *Bull. Math. Biophys.*, **5**, (1943) 115.
- [3] HEBB, D. O., in 'The Organization of Behavior', Wiley, New York (1949).
- [4] LITTLE, W. A., *Math. Biosciences*. **19**, (1974) 101.
- [5] HOPFIELD, J. J., *Proc. Natl. Acad. Sci. USA* **79**, (1982) 2254.
- [6] AMIT, D., GUTFREUND, H. e SOMPOLINSKY, H., *Phys. Rev. A*, **32** (2), (1985) 1007.
- [7] AMIT, D., GUTFREUND, H. e SOMPOLINSKY, H., *Phys. Rev. Lett*, **55**, (1985) 1530.
- [8] AMIT, D., GUTFREUND, H. e SOMPOLINSKY, H., *Ann. Phys.*, **173**, (1987) 30.
- [9] AMIT, D., GUTFREUND, H. e SOMPOLINSKY, H., *Phys. Rev. A*, **35**, (1987) 2293.
- [10] FONTANARI, J. F. e KOBERLE, R., *Phys. Rev. A*, **36** (5), (1988) 2475.
- [11] DIEDERICH, S. e OPPER, M., *Phys. Rev. Lett.*, **58**, (1987) 949.
- [12] PÖPPEL, G. e KREY, U., *Europhys. Lett.*, **4** (9), (1987) 979.
- [13] ABELLES, M., in 'Studies of the Brain Function', Springer-Verlag, New York, (1982).
- [14] PERSONNAZ, L., GUYON, I. e DREYFUS, G., *J. Physique Lett.*, **46**, (1985) L359.

- [15] PERSONNAZ, L., GUYON, I. e DREYFUS, G., *Phys. Rev. A*, **34** (5), (1986) 4217.
- [16] KANTER, I. e SOMPOLINSKY, H. *Phys. Rev. A*, **35** (1), (1987) 380.
- [17] BUCHMANN, J., DIVKO, R. e SHULTEN, K., *Phys. Rev. A*, **39** (5), (1989) 2689.
- [18] PEREZ-VICENTE, C. e AMIT, D. J., *J. Phys. A: Math. Gen.*, **22**, (1989) 559.
- [19] GUTFREUND, H., *Phys. Rev. A*, **37**, (1988) 570.
- [20] PARGA, N. e VIRASORO, M. A., *J. Physique*, **47**, (1986) 1857.
- [21] IDIART, M. A. P. e THEUMANN, A., in Proceed. 'STATPHYS 17 Workshop on Neural Networks and Spin Glass', Ed W. K. Theumann e R. Koberle, World Scientific, (1990).
- [22] IDIART, M. A. P. e THEUMANN, A., *J. Phys. A: Math. Gen.*, **24**, (1991) L649.
- [23] GARDNER, E., *J. Phys. A: Math. Gen.*, **21**, (1988) 257.
- [24] GARDNER, E., DERRIDA, B. *J. Phys. A: Math. Gen.*, **21**, (1988) 271.
- [25] SHERRINGTON, D. e KIRKPATRICK, S., *Phy. Rev. Lett.*, **35** (26), (1975) 1792.
- [26] de ALMEIDA, J. R. L. e THOULESS, D. J. *J. Phys. A: Math. Gen.*, **11** (5), (1978) 983.
- [27] WHILSHAW, D. J., BUNEMAN, O. P. e LONGUET-HIGGINS, H. C., *Nature*, **222**, (1969) 960.
- [28] BRUCE, A. D., GARDNER, E. e WALLACE, D. J., *J. Phys. A: Math. Gen.*, **20**, (1987) 2909.
- [29] AMIT, D., in 'Modelling Brain Function', University Press, Cambridge (1989).
- [30] GLAUBER, R. J., *J. Math. Phys.*, **4**, (1963) 294.

- [31] EDWARDS, S. F. e ANDERSON, P. W., *J. Phys. F*, **5**, (1975) 965.
- [32] MATTIS D. C., *Phys. Lett. A*, **56**, (1976) 421.
- [33] FONTANARI, J. F. e THEUMANN, W. K., *J. Physique*, **51**, (1990) 375.
- [34] FONTANARI, J. F., *J. Phys. France*, **51**, (1990) 2421.
- [35] ERICHSEN Jr., R. e THEUMANN, W. K., in Proceed. 'STATPHYS 17 Workshop on Neural Networks and Spin Glass', ed W. K. Theumann e R. Koberle, World Scientific, 1990.
- [36] TSODIKS, M. V. e FEIGELMAN M. V., *Europhys. Lett.*, **6**, (1988) 101.
- [37] 'Handbook of Mathematical Functions', ed. Abramovitz, M. e Stegun, I., Dover Publications Inc., New York 1984.
- [38] ERICHSEN Jr., R., THEUMANN, W. K., *Intern. Journ. of Neur. Syst.*, **1** (4), (1991) 347.
- [39] BINDER, K., ed. 'Monte Carlo Methods in Statistical Physics', Springer, 1979.
- [40] FORREST, M. B. *J. Phys. A: Math. Gen.*, **21**, (1987) 245.
- [41] FORREST, M. B. *J. Phys. France*, **50**, (1989) 2003.
- [42] COVER, T. M., *IEEE Trans. Eletron. Comput.*, **EC-14**, (1965) 326.
- [43] MEZARD, M., PARISI, G. e VIRASORO, M. A., in 'Spin Glass Theory and Beyond'. World Scientific Notes in Physics V.9, World Scientific Publishing Co., Singapura, 1987.
- [44] PARISI, G., *Phys. Lett.*, **73A** (3), (1979) 203.
- [45] PARISI, G., *J. Phys. A: Math. Gen.*, **13**, (1980) L115.
- [46] PARISI, G., *J. Phys. A: Math. Gen.*, **13**, (1980) 1101.

- [47] RAMMAL, R., TOULOUSE, G. e VIRASORO, M. A., *Rev. of Mod. Phy.*, **58** (3), (1986) 765.
- [48] HUBBARD, J., *Phys. Rev. Lett.*, **3** (2) (1959) 77; STRATONOWICH, R. L., *Soviet Physics Doklady*, **2** (4) (1957) 416.
- [49] KEPLER, T. B. e ABBOTT, L. F., *J. Phys. France*, **49**, (1988) 1657.
- [50] ABBOTT, L. F. e KEPLER, T. B., *J. Phys. A: Math. Gen*, **22**, (1989) 2031.
- [51] AMIT, D. J., EVANS, M. R., HORNER, H. e WONG, K. Y. M., *J. Phys. A: Math. Gen.*, **23**, (1990) 3361.
- [52] DERRIDA, B., GARDNER, E. e ZIPPELIUS, A., *Europhys. Lett.*, **10**, (1987) 687.
- [53] GRINIASTY, M. e GUTFREUND, H., *J. Phys. A: Math. Gen.*, **24**, (1991) 715.
- [54] THEUMANN, W. K. e ERICHSEN Jr., R., *J. Phys. A: Math. Gen.*, **24**, (1991) L565.