UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE FÍSICA

# Categorização no Modelo de Hopfield: Efeitos de Ruído Sináptico e de Diluição Simétrica. *

**Paulo Roberto Krebs**

Tese realizada sob a orientação do Prof. Dr. Walter Karl Theumann e apresentada ao Instituto de Física da UFRGS, em preenchimento parcial dos requisitos para a obtenção do título de Doutor em Ciências.

Porto Alegre

2004

* Trabalho financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS)

para Virgínia e Júlia

# Agradecimentos

Ao Prof. Dr. Walter Karl Theumann, pela orientação, dedicação, estímulo, e compreensão durante a elaboração desta tese.

À minha família, especialmente à Virgínia e à Júlia, por todo apoio, amor e principalmente pela compreensão da minha ausência.

Ao Departamento de Física da Universidade Federal de Pelotas, em especial ao meu colega Dr. Eduardo Fontes Henriques.

Ao Dr. Rubem Erichsen Jr. do IF-UFRGS e ao Dr. Sérgio G. Magalhães da UFSM, pelo interesse, estímulo, discussões, apoio e amizade.

À Profa. Dra. Alba Theumann do IF-UFRGS, ao Dr. José Fernando Fontanari do IFSC-USP e ao Dr. David R. C. Dominguez da Universidad Autónoma de Madrid, Espanha, pelas discussões esclarecedoras.

Aos meus amigos da antiga turma da sala M204.

Aos professores e funcionários do IF-UFRGS, que de alguma forma contribuíram para a minha formação e para a realização desta tese.

# Resumo

Nesta tese estudamos os efeitos de diluição simétrica gradual das conexões entre neurônios e de ruído sináptico sobre a habilidade de categorização de padrões no modelo de Hopfield de redes neurais, mediante a teoria de campo médio com simetria de réplicas e simulações numéricas. Utilizamos generalizações da regra de aprendizagem de Hebb, para uma estrutura hierárquica de padrões correlacionados em dois níveis, representando os ancestrais (conceitos) e descendentes (exemplos dos conceitos). A categorização consiste no reconhecimento dos conceitos por uma rede treinada unicamente com exemplos dos conceitos.

Para a rede completamente conexa, obtivemos os diagramas de fases e as curvas de categorização para vários níveis de ruído sináptico. Observamos dois comportamentos distintos dependendo do parâmetro de armazenamento. A habilidade de categorização é favorecida pelo ruído sináptico para um número finito de conceitos, enquanto que para um número macroscópico de conceitos este favorecimento não é observado. Entretanto a performance da rede permanece robusta contra o ruído sináptico.

No problema de diluição simétrica consideramos apenas um número macroscópico de conceitos, cada um com um número finito de exemplos. Os diagramas de fases obtidos exibem fases de categorização, de vidro de spin e paramagnética, bem como a dependência dos parâmetros de ordem com o número de exemplos, a correlação entre exemplos e conceitos, os ruídos sináptico e estocástico, e a conectividade. A diluição favorece consideravelmente a categorização, particularmente no limite de diluição extrema.

# Abstract

In this thesis we study the effects of gradual symmetric dilution and of synaptic noise on the categorization ability in the Hopfield model of neural networks by means of replica-symmetric mean-field theory and numerical simulations. A generalized Hebbian learning rule is used with hierarchically correlated patterns in a two-level structure of ancestors (concepts) and descendants (examples of the concepts). Categorization consists in recognizing the concepts when the network has been trained only with the examples of the concepts.

Phase diagrams and categorization curves are obtained as a function of the synaptic noise for the completely connected network. Two different behaviours are observed as a function of the load parameter. Although the synaptic noise improves the categorization ability for a finite number of concepts, it does not for a macroscopic number of them. Nevertheless, the network performance is still robust against synaptic noise.

We consider a macroscopic number of concepts, each with a finite number of examples, in the symmetrically dilute network. Phase diagrams are obtained that exhibit a categorization, a spin-glass, and a paramagnetic phase, as well as the dependence of the order parameter on the number of examples, the correlation, the synaptic and stochastic noises, and on the connectivity. It is shown that the dilution improves considerably the categorization ability.

# Índice

# Relação de Figuras

# Relação de Tabelas

# Introdução

O problema mente-cérebro constitui ainda hoje objeto de intensa discussão. O desenvolvimento do cérebro, em mamíferos, a partir de uma coleção de células simples imaturas e não diferenciadas gerando um órgão de elevada complexidade estrutural é algo extraordinário. Basicamente, em todos os sistemas biológicos observa-se que o comportamento de um conjunto de objetos biológicos, por exemplo um animal pluricelular, é completamente diferente daquele observado em um objeto biológico isolado, uma célula. Esse comportamento coletivo é o fenômeno crucial, como enfatiza Parisi [1].

Propor-se a compreender o funcionamento do cérebro é, sem dúvida, uma tarefa bastante audaciosa, fascinante e complexa. Nela estão engajados pesquisadores das mais diversas áreas do conhecimento, movidos pelas mais diversas razões. Essas razões incluem desde o entendimento da fisiologia, da anatomia e do funcionamento do cérebro dentro do contexto da neurobiologia, até o interesse em aplicar as informações obtidas nestas investigações na construção de máquinas capazes de reproduzir, no seu funcionamento autônomo, funções superiores do cérebro.

No contexto da física, o interesse está em explorar as analogias com sistemas complexos que apresentam comportamentos coletivos, o que permite o uso do conjunto de ferramentas e conceitos já desenvolvidos em diversas áreas de pesquisa em física. Nesse sentido, a mecânica estatística contribui de maneira fundamental para as investigações realizadas por físicos. Sistemas em que o comportamento coletivo emerge a partir da interação de um grande número de elementos simples, tais como as transições de fase, têm sido estudados há muito tempo, no contexto da física. Fenômenos como o ferromagnetismo, a supercondutividade

ou sistemas desordenados, como os vidros de spin, são exemplos de fenômenos coletivos que podem ser entendidos a partir da mecânica estatística.

As redes neurais são sistemas constituídos por um grande números de células especiais, os neurônios, que interagem através de processos físico-químicos, dando origem a comportamentos coletivos que se manifestam como memória associativa, reconhecimento de padrões, generalização, categorização, controle motor e todas as demais habilidades cognitivas presentes no cérebro. Nos modelos de redes neurais atratoras, a principal ênfase das investigações tem sido na habilidade de armazenar e reconhecer padrões, funcionando como dispositivos de memória associativa. Porém, essas redes são capazes de realizar tarefas computacionais mais elaboradas como, por exemplo, classificar objetos em diferentes categorias. O problema da categorização é, atualmente, objeto de grande discussão no contexto da ciência cognitiva. É, portanto, interessante investigar este problema no âmbito das redes neurais atratoras.

O objetivo desta tese é estudar o modelo de Hopfield quanto a sua habilidade de categorização, quando são introduzidos ingredientes biológicos como a diluição e o ruído sináptico, utilizando-se a mecânica estatística. Categorização é a capacidade de criar atratores para conceitos a partir de um processo de aprendizagem no qual a rede é exposta apenas a exemplos dos conceitos.

No capítulo 1, fazemos uma breve descrição da biologia do cérebro e, em particular, do funcionamento de um neurônio biológico, visando identificar os ingredientes básicos necessários na construção de modelos de redes neurais artificiais. Para situar o desenvolvimento das redes neurais artificiais ao longo dos anos, apresentamos um pequeno retrospecto histórico dos principais fatos na evolução das redes neurais artificiais, do ponto de vista da física.

No capítulo 2, descrevemos dois dos principais modelos onde se evidenciam analogias com os sistemas estudados na física. Esses modelos são o de Little e o de Hopfield. Descrevemos, também, os principais resultados da solução termodinâmica desenvolvida por Amit, Gutfreund e Sompolinsky para o modelo de Hopfield generalizado. Esses três trabalhos marcaram o ingresso definitivo dos físicos na pesquisa em redes neurais.

No capítulo 3, estudamos, a partir da teoria de campo médio, o efeito do ruído sináptico na habilidade de categorização do modelo de Hopfield tanto para um número finito de padrões, quanto para um número extensivo de padrões. O ruído sináptico, que está presente nos sistemas biológicos, é representado pela temperatura. Os resultados originais são obtidos a partir do cálculo da função de partição na aproximação de simetria de réplicas, que nos permite obter os diagramas de fases e as curvas de categorização.

No capítulo 4, introduzimos a diluição simétrica das conexões sinápticas, visando aproximar o modelo de uma rede real. A manutenção da simetria das conexões sinápticas, embora não sendo biologicamente observada, possibilita a aplicação dos métodos da mecânica estatística de equilíbrio e da teoria de campo médio a este problema. Obtivemos, novamente, resultados originais que estão traduzidos nos diagramas de fases e curvas de categorização em função do grau de diluição da rede.

A seguir, apresentamos as conclusões e comentários gerais, enfatizando os resultados originais, bem como possíveis extensões deste trabalho.

## Capítulo 1

# A Modelagem Biológica

## 1.1 Introdução

O cérebro humano é um sistema complexo composto por um grande número de unidades, os neurônios, de cujo comportamento coletivo emergem as propriedades complexas de aprendizagem, memória, categorização e generalização. O cérebro humano possui, aproximadamente, $10^{10}$ neurônios, sendo que cada um deles se conecta, em média, a $10^4$ outros neurônios. Qualquer modelo de rede neural artificial deve prover uma quantidade mínima de ingredientes que capturem os aspectos essenciais presentes nas redes neurais biológicas. Nesse sentido, os modelos devem ser biologicamente plausíveis, capazes de associatividade, processamento paralelo e apresentar comportamento emergente livre do controle de agentes externos.

Não faremos uma descrição do estado da arte em ciências neurais, a qual pode ser obtida na literatura especializada. Iremos descrever apenas os aspectos essenciais do cérebro necessários para a construção de modelos artificiais de redes neurais [2]. Os elementos fundamentais são os neurônios e as sinapses. Esses elementos foram revelados pelos estudos do médico espanhol Santiago Ramon y Cajal, que, aprimorando a técnica para tingimento de células inventada por Camilo Golgi, puseram fim à discussão entre reticulistas e neuronistas que acontecia no final do século XIX [3][4][5].

## 1.2 O Neurônio Biológico

Existe uma grande variedade de tipos de neurônios em função da estrutura, da função e do tamanho, que pode variar desde aproximadamente 0.01 mm até 1 m. Entretanto, consideraremos que todos eles se comportam de maneira idêntica, não fazendo qualquer diferenciação entre si. Nesse sentido, podemos definir um neurônio canônico como sendo constituído por três partes: a entrada, o centro de processamento e a saída. A entrada representa a árvore dendrítica, o centro de processamento representa o corpo celular (ou soma), e a saída representa o axônio. A figura 1.1 apresenta um desenho esquemático das partes constituintes do neurônio biológico. A função básica dos dendritos é receber os pulsos elétricos provenientes de neurônios adjacentes e transmiti-los ao corpo celular onde serão integrados e comparados a um valor limiar (de referência). Se a soma dos pulsos elétricos recebidos pelo corpo celular for maior que o valor limiar, será emitido um pulso elétrico através do axônio, que é um prolongamento que se projeta a partir do corpo celular e que se ramifica na extremidade para se conectar com neurônios adjacentes. Portanto, a função básica do axônio é transmitir o pulso elétrico emitido pelo corpo celular, chamado de potencial de ação, estabelecendo a comunicação com outros neurônios.

Os neurônios se comunicam através das sinapses ou junções sinápticas, que são os pontos de contato onde o axônio do neurônio pré-sináptico comunica o pulso elétrico aos dendritos ou diretamente ao corpo celular do neurônio pós-sináptico como mostra a figura 1.2. As sinapses de duas células estão separadas por uma pequena fenda sináptica de aproximadamente 200 nm. Usualmente, apenas um axônio se projeta a partir do corpo celular, ramificando-se para se comunicar com muitos neurônios pós-sinápticos.

A transmissão dos sinais entre os neurônios pode ser elétrica ou química. As transmissões elétricas prevalecem no interior dos neurônios, enquanto as transmissões químicas se dão entre neurônios, nas sinapses. As transmissões elétricas iniciam no corpo celular e se propagam pelo axônio para todas as sinapses. Se o neurônio se encontra inativo, seu interior está carregado negativamente em relação ao meio exterior estabelecendo, assim, uma

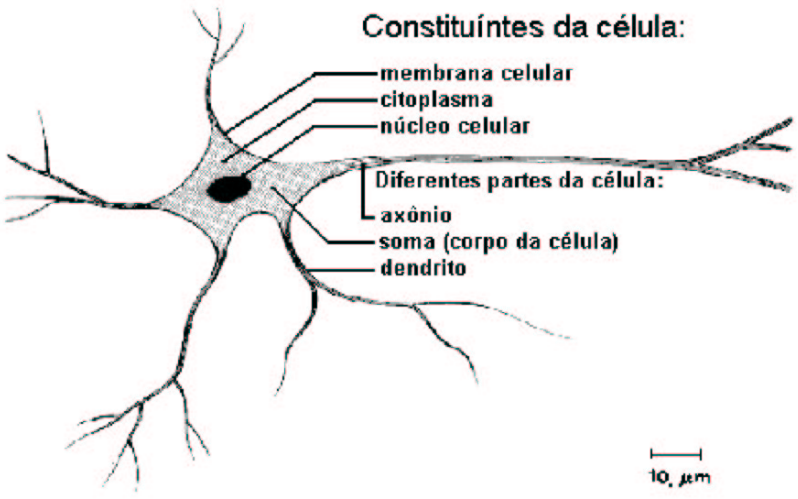



**Fig. 1.1:** *Desenho esquemático do neurônio biológico.*

diferença de potencial entre o interior e o exterior do neurônio. Essa diferença de potencial de aproximadamente -70 mV é chamada de potencial de repouso e é devida a diferentes concentrações de íons de sódio ($Na^+$), potássio ($K^+$) e cloro ($Cl^-$). Na ausência de potencial de ação, a membrana celular é impermeável ao íons de sódio de modo que o interior do neurônio será deficiente em íons positivos.

Quando sinais elétricos provenientes das sinapses estão presentes, ocorre a despolarização do potencial de repouso. Quando esse atinge -60 mV, a membrana celular torna-se permeável aos íons de sódio, que imediatamente entram na célula, neutralizando a diferença de potencial. Neste caso, a sinapse é excitatória. Se os íons de potássio saem da célula, teremos uma hiperpolarização da membrana celular, e a sinapse será inibitória, pois agirá contra a excitação do neurônio. Assim sendo, a atividade do neurônio pré-sináptico será excitar ou inibir a atividade do neurônio pós-sináptico, criando ou não o potencial pós-sináptico.

A transmissão de sinais através da fenda sináptica é fundamentalmente um processo químico. A presença de um sinal elétrico na membrana pré-sináptica causa a liberação

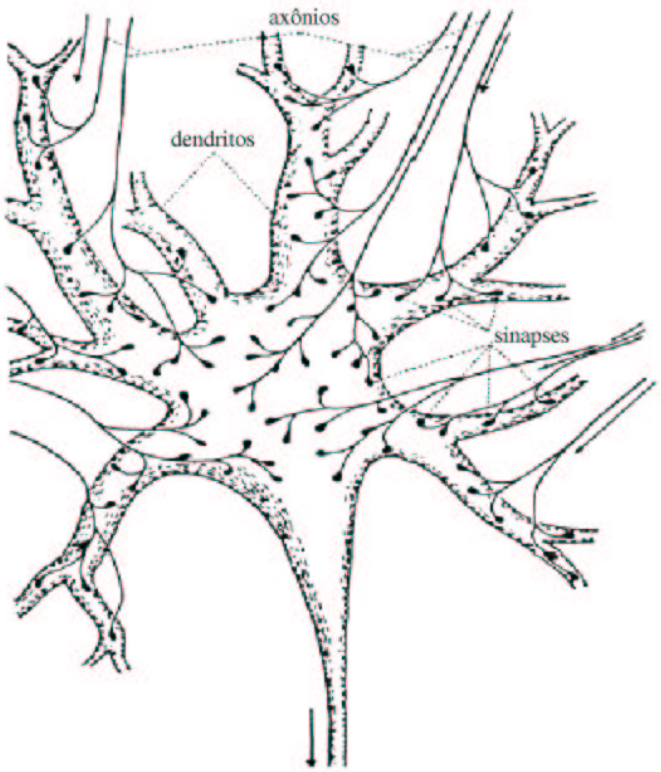
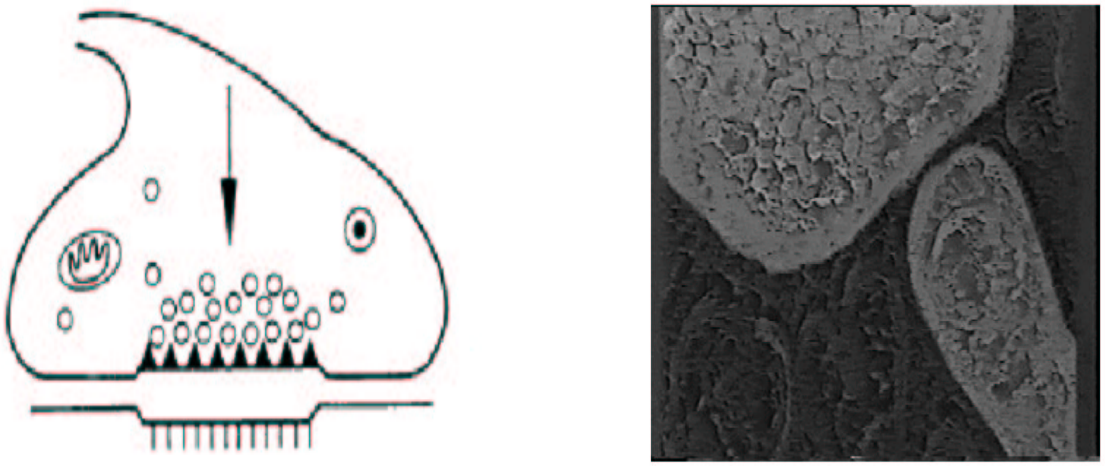



**Fig. 1.2:** *As conexões sinápticas.*

de neuro-transmissores que se difundem através da fenda sináptica e alcançam a membrana pós-sináptica em aproximadamente 0.5 ms (ver figura 1.3). Nela, receptores especiais para os neuro-transmissores alteram a condutância da membrana pós-sináptica para íons de sódio, potássio e cloro, causando uma polarização ou despolarização do potencial pós-sináptico, gerando assim um potencial de ação no neurônio pós-sináptico. Existem evidências de que todas as sinapses de um axônio são ou excitatórias ou inibitórias, fato conhecido como lei de Dale, e de que existem diferenças estruturais significativas entre os dois tipos de sinapses.

Podemos descrever o processo dinâmico fundamental de funcionamento dos neurônios, de maneira simplificada, através da seguinte seqüência de passos: o axônio neural descreve um estado (processo) de tudo ou nada. No primeiro caso, um potencial de ação é transmitido em função da soma realizada no corpo celular. Tanto a forma quanto a amplitude do sinal - da ordem de dezenas de milivolts - são estáveis e são reproduzidas em todos os ramos do axônio.



(a) (b)

**Fig. 1.3:** *(a) Transmissão do sinal na fenda sináptica, (b) fotografia da fenda sináptica*

No outro caso, não haverá propagação de sinal através do axônio, permanecendo o mesmo no seu potencial de repouso. Quando o potencial de ação encontra as sinapses, os neuro-transmissores são liberados na fenda sináptica e encontram a membrana do neurônio pós-sináptico onde causam, por meio dos seus respectivos receptores, a penetração de corrente iônica. A eficácia da sinapse é determinada pela intensidade da corrente iônica que penetra no neurônio pós-sináptico em função do potencial de ação do neurônio pré-sináptico. O potencial pós-sináptico se difunde gradualmente em direção ao corpo celular onde serão somados os potenciais pós-sinápticos provenientes de todos os neurônios pré-sinápticos, que podem ser excitatórios ou inibitórios. Se a soma de todos os potenciais pós-sinápticos for maior que um valor limiar, de algumas dezenas de milivolts, um potencial de ação será então emitido.

O tempo total desde a emissão de um potencial de ação por um neurônio pré-sináptico, liberação dos neuro-transmissores e a emissão de um potencial de ação pelo neurônio pós-sináptico é de aproximadamente 1 a 2 ms. Durante o tempo em que um sinal se propaga pelo axônio, fica impedida a propagação de um segundo sinal, o que limita a taxa de disparos dos neurônios. Existe um período de aproximadamente 2 ms durante o qual um neurônio não

pode emitir um segundo potencial de ação, independentemente da intensidade do potencial de despolarização ao qual o neurônio está submetido. Esse período é chamado de período refratário absoluto. Esse tempo é necessário para que a membrana celular consiga restaurar seu potencial de repouso. Isso limita a freqüência máxima de disparos de um neurônio entre 30 e 150 vezes por segundo.

O fato de um neurônio emitir ou não um potencial de ação através do seu axônio leva naturalmente a uma interpretação do neurônio como um elemento formal, que pode ser encontrado em dois estados: ativo ou inativo. Podemos descrever sua estrutura lógica como sendo constituída de uma unidade de processamento que representa o corpo celular. A ele estão ligados canais de entrada que representam os dendritos e sinapses, por onde flui o sinal emitido pelos neurônios pré-sinápticos. Associamos a cada canal de entrada um parâmetro $J_{ij}$, cujo valor é a eficácia sináptica do canal. O índice $i$ identifica o neurônio em consideração, e o índice $j$ identifica os vários canais de entrada. A eficácia sináptica é quem determina a quantidade do potencial pós-sináptico que entrará na soma realizada pelo corpo celular, para os canais ativados. Existe um canal de saída que representa o estado lógico do neurônio, qual seja, a emissão ou não de um pulso elétrico. Esse canal representa o axônio.

A operação de um neurônio formal consiste, então, na ativação de alguns canais de entrada em um dado momento. Os sinais dos canais ativos são somados no corpo celular por meio de uma soma linear das eficácias sinápticas dos canais ativos. A soma é, então, comparada com o valor limiar e, se superar esse valor, o canal de saída será ativado. Caso contrário, o canal de saída permanece no seu estado inativo.

Todo esse processo pode ser quantificado ao atribuirmos para cada neurônio pré-sináptico uma variável $S_j$ que representa o seu estado ativo ($S_j = 1$) ou inativo ($S_j = 0$), indicando se o canal de entrada do $j$-ésimo neurônio está ativo ou não. O potencial pós-sináptico do neurônio $i$ é a soma das eficácias sinápticas multiplicadas pelo estado $S_j$ de todos os canais de entrada,

$$h_i = \sum_{j=1}^{N} J_{ij} S_j, \tag{1.1}$$

onde $h_i$ representa o potencial pós-sináptico do neurônio $i$, e $N$ é o numero de neurônios

pré-sinápticos. O valor de $h_i$ é, então, comparado com o limiar de ativação do neurônio $i$, $\theta_i$, e o seu estado será determinado pela função lógica

$$S_i' = \psi[h_i > \theta_i]. \tag{1.2}$$

A função $\psi$ assume o valor 1 se o seu argumento for verdadeiro e 0 caso contrário. Dessa forma, a variável $S_i'$ indica a emissão ou não de um pulso elétrico através do canal de saída (axônio).

## 1.3 Retrospectiva Histórica

Essa abordagem foi desenvolvida em 1943 por Warren McCulloch e Walter Pitts [6], em um trabalho que, pela primeira vez, modelava um neurônio real por meio de um neurônio formal bastante simplificado. A rede neural por eles desenvolvida pode implementar qualquer operação do cálculo proposicional. A introdução do tempo foi um elemento essencial, pois os elementos da rede deviam operar de maneira sincronizada para que as funções lógicas pudessem ser efetuadas. Entretanto, as redes de McCulloch e Pitts não são robustas, pois não possuem redundâncias.

Donald Hebb propôs [7], em 1949, que a intensidade das conexões sinápticas não eram fixas. Uma conexão sináptica que é repetidamente ativada por um potencial pós-sináptico terá sua eficiência aumentada, ao passo que uma conexão sináptica que é raramente ativa terá sua eficiência sináptica diminuída. Essa proposição, conhecida como a regra de Hebb, sugere que a aprendizagem se dá pela modificação das eficácias sinápticas como conseqüências das atividades dos potenciais pré e pós-sinápticos.

Por volta dos anos 60, Frank Rosemblatt [8] introduziu um novo tipo de rede neural, chamado de perceptron, pois era um modelo simplificado dos mecanismos biológicos de processamento das informações sensoriais (percepções). Na sua forma mais simples, um perceptron consiste em duas camadas de neurônios interligados, representando a camada de entrada e a camada de saída, respectivamente. Os neurônios da camada de saída recebem sinais dos neurônios da camada de entrada; porém, nenhum sinal é enviado de volta à camada

de entrada. Desse modo, o fluxo de informações propaga-se de maneira unidirecional. Um aspecto importante a ser ressaltado é que as conexões sinápticas nos perceptrons possuem um alto grau de aleatoriedade, ao contrário das redes de McCulloch e Pitts.

Em 1969, Marvin Minsky e Seimour Papert [9] publicaram um livro com um estudo bastante aprofundado sobre as capacidades computacionais do perceptron. Nesse livro, Minsky e Papert criticaram duramente o perceptron a ponto de provocar a quase extinção das pesquisas em redes neurais. Nele, mostraram que funções lógicas simples como a função ou-exclusivo, que requer apenas dois neurônios de entrada conectados a um de saída, não podiam ser implementadas por perceptrons. Entretanto, as críticas de Minsky e Papert não são válidas para perceptrons multi-camadas, os quais são capazes de aprender a classificar mais de duas classes, superando, assim, as limitações dos perceptrons simples.

Um novo desenvolvimento em redes de neurônios formais iniciou-se com o trabalho de Little, que mostrou a similaridade entre redes neurais do tipo proposto por McCulloch e Pitts e sistemas magnéticos de spins do tipo Ising [10]. Essa analogia entre o modelo de Ising e redes neurais é estabelecida observando-se a correspondência entre os estados $S_i = 1$ da rede com o valor $+1$ do spin e $S_i = 0$ da rede com o valor $-1$ do spin do modelo de Ising. Outra contribuição importante foi a introdução de ruído nas conexões sinápticas, o que permitiu o estudo de uma dinâmica estocástica de redes neurais.

A ligação entre redes neurais e sistemas magnéticos foi definitivamente estabelecida a partir do trabalho seminal de John Hopfield em 1982 [11]. Nesse artigo, Hopfield estabelece a relação com sistemas complexos conhecidos como vidros de spin. Nesse modelo, chamado de modelo de Hopfield, o cérebro é visto como um sistema desordenado capaz de modelar funções como memória e aprendizagem, dentre outras que serão exploradas ao longo desta tese. A idéia básica consistiu em perceber que, como as conexões sinápticas podem ser positivas ou negativas, com uma dose razoável de aleatoriedade, tem-se dois ingredientes fundamentais para a constituição de sistemas de vidros de spin, quais sejam, frustração e desordem. Identificando os estados dos neurônios com os spins de Ising e as conexões sinápticas com os acoplamentos magnéticos, evidencia-se a relação entre os dois sistemas. Para

que tal relação fosse possível, Hopfield teve que assumir, corajosamente, que as conexões sinápticas fossem simétricas, ou seja, $J_{ij} = J_{ji}$, o que não é justificável do ponto de vista biológico. Entretanto, esse passo para trás permitiu o uso das ferramentas desenvolvidas pela mecânica estatística para sistemas magnéticos no estudo analítico das redes neurais. Amit, Gutfreund e Sompolinsky resolveram a mecânica estatística do modelo de Hopfield [12][13][14][15].

Buscando uma aproximação com sistemas neurais reais, novos ingredientes são introduzidos nos modelos. A assimetria das conexões sinápticas e a diluição são algumas das modificações que buscam tornar os modelos mais realistas. A assimetria sináptica impede o estudo baseado num Hamiltoniano e, portanto, termodinâmico. Derrida, Gardner e Zippellius resolveram a dinâmica do modelo de Hopfield assimétrico e extremamente diluído [16], analiticamente.

O estudo sistemático das redes neurais não só rendeu possíveis mecanismos para o entendimento do funcionamento do cérebro, como também rendeu novos métodos e técnicas matemáticas. O importante trabalho desenvolvido por Elisabeth Gardner [17][18], hoje conhecido como método de Gardner, que permite estudar as propriedades gerais de redes neurais independentemente de uma regra de aprendizagem explícita, é um marco na mecânica estatística das redes neurais.

## Capítulo 2

# Modelos de Redes Neurais

## 2.1 Introdução

Para que seja possível modelar redes neurais biológicas, é necessário lançar mão de simplificações que permitam o tratamento matemático do modelo. Essas simplificações são levadas até um limite em que o modelo ainda mantém as características biológicas essenciais para reproduzir algumas funções biológicas de interesse. Neste capítulo, faremos uma descrição dos modelos de redes neurais que buscam descrever basicamente a capacidade de memorização (reconhecimento de padrões), introduzidos por Little [10] e por Hopfield [11]. Também serão discutidos alguns aspectos da dinâmica de uma rede neural artificial.

Apesar das simplificações introduzidas, esses modelos mostram-se muito promissores como paradigma para a construção de sistemas computacionais verdadeiramente inteligentes. Novamente, a habilidade de realizar computações paralelas distribuídas (comportamento coletivo) se apresenta como a melhor forma de superar as limitações computacionais que surgem na computação simbólica serial introduzida por von Neumann. Esses modelos exibem um grande número de propriedades desejáveis, que não são encontradas em sistemas de computação simbólica convencionais, tais como robustez, tolerância a falhas (redundância), flexibilidade, tratamento de informações inconsistentes e ruidosas, paralelismo, entre outras. Suas aplicações a tarefas computacionais, tais como reconhecimento de voz e imagens, memória associativa, classificação, compressão de dados, controle adaptativo, filtros de ruídos apresentam um desempenho considerado eficiente, em particular, à

meteorologia [19].

## 2.2 O Modelo de Little

Em 1974, W. A. Little publicou um trabalho [10], introduzindo um modelo de rede neural no qual expressava interesse em investigar a existência de estados persistentes no cérebro. Nesse trabalho, Little partiu da suposição de que os estados do cérebro, em qualquer instante de tempo, podem ser descritos pela configuração definida pelo conjunto dos neurônios que dispararam e que não dispararam um potencial de ação, dentro de um certo intervalo de tempo recente. Desse modo, pôde examinar sob que condições existiriam correlações entre estados separados por um intervalo de tempo muito maior que o período refratário absoluto. A principal motivação para estudar esse problema está na crença de que os estados do cérebro aos quais estamos nos referindo estão relacionados de algum modo com processos mentais ou experiências sensoriais.

Para que o modelo pudesse ser tratado matematicamente, Little teve que introduzir um conjunto de idealizações. A primeira delas foi considerar a rede completamente isolada do mundo exterior, ou seja, a rede não recebe qualquer estímulo externo. A segunda estabelece que os neurônios não podem disparar potenciais de ação em tempos aleatórios, mas apenas de maneira completamente sincronizada, em intervalos de tempo que são múltiplos inteiros de um período $\tau$ que é da ordem do período refratário absoluto. Durante esse período, é realizada a soma dos potenciais pós-sinápticos excitatórios e inibitórios que atuam sobre o neurônio, determinando a emissão ou não de um potencial de ação. A cada período de tempo $\tau$, uma nova soma dos potenciais pós-sinápticos é realizada, desprezando-se as somas em períodos anteriores. Isso significa que a emissão de potenciais pós-sinápticos é um processo de Markov. A terceira e última idealização estabelece que as conexões sinápticas são fixas e não mudam com o tempo, o que significa que o modelo não contempla nenhum processo de aprendizagem, pois este envolve modificações nas conexões sinápticas. Desse modo, a rede contém apenas informações que podem ser interpretadas como informações

hereditárias estabelecidas no processo de desenvolvimento do cérebro.

A partir dessas idealizações, pode-se definir o estado do cérebro pela configuração dos estados dos neurônios. Num determinado instante de tempo $t$, o estado do cérebro é definido pela configuração

$$\mid s_1, s_2, ..., s_N\rangle, \tag{2.1}$$

onde $s_i = +1$ se o $i$-ésimo neurônio disparou e $s_i = -1$ se o $i$-ésimo neurônio não disparou um potencial de ação; $N$ é o número total de neurônios e, como os neurônios podem assumir apenas dois estados, existem $2^N$ estados possíveis.

Na medida em que o neurônio $(j)$ dispara um potencial de ação através de seu axônio, observa-se um aumento ou uma redução no potencial pós-sináptico, representada por $V_{ij}$, do neurônio $(i)$ ao qual está conectado. Se a sinapse for excitatória, $V_{ij}$ será positivo e, se a sinapse for inibitória, $V_{ij}$ será negativo. A ausência de conexão sináptica entre dois neurônios é representado por $V_{ij} = 0$. Dessa forma, o $i$-ésimo neurônio estará sob a ação de um potencial pós-sináptico líquido, $V_i$, que é a soma de todos os potenciais pós-sinápticos dos neurônios que a ele se conectam. Isso pode ser escrito matematicamente como

$$V_i = \sum_j V_{ij}(\frac{s_j + 1}{2}). \tag{2.2}$$

Analisando essa expressão, verifica-se que contribuirão para a soma apenas os termos relativos aos neurônios que dispararam durante o período $\tau$. Se o potencial $V_i$ exceder o limiar de ativação $V_0$, o neurônio $i$ provavelmente emitirá um potencial de ação. Matematicamente, pode-se representar essa probabilidade por

$$p(+1) = \frac{1}{\exp[-\beta(V_i - V_0)] + 1}. \tag{2.3}$$

Se $V_i$ for apreciavelmente maior que $V_0$, o termo da exponencial será pequeno, o que leva a uma probabilidade de disparo $p(+1)$ próxima da unidade. Por outro lado, se $V_i$ for apreciavelmente menor que $V_0$, então a exponencial será muito maior que a unidade, resultando numa probabilidade de disparo $p(+1)$ muito pequena. O parâmetro $\beta$ dá uma medida da suavidade da função $p(+1)$. Embora do ponto de vista biológico tanto $\beta$ quanto $V_0$ devam

variar de neurônio para neurônio, nesse modelo seus valores são mantidos fixos para toda a rede. A probabilidade do neurônio $i$ não disparar é simplesmente $p(-1) = 1 - p(+1)$, o que, após simplificações algébricas, resulta na mesma expressão para $p(+1)$, porém com um sinal positivo na frente do parâmetro $\beta$. Essas duas probabilidades podem ser expressas como

$$p(s_i^{'}) = \frac{1}{\exp[-\beta s_i^{'}(V_i - V_0)] + 1}. \tag{2.4}$$

A partir dessa expressão, pode-se obter a probabilidade de se encontrar a rede num estado $\mid s_1^{'}, s_2^{'}, ..., s_N^{'}\rangle$ num instante de tempo $t^{'} = t + \tau$, estando a rede no estado $\mid s_1, s_2, ..., s_N\rangle$ no instante $t$. Definindo-se um operador $P$ para essa probabilidade e utilizando-se a expressão para $p(s_i^{'})$ obtém-se

$$\langle s_1^{'}, s_2^{'}, ..., s_N^{'} \mid P \mid s_1, s_2, ..., s_N\rangle = \prod_{i=1}^{N} \left( \frac{1}{\exp[-\beta s_i^{'}(V_i - V_0)] + 1} \right) \tag{2.5}$$

em termos de uma matriz de ordem $2^N \times 2^N$, cujos elementos representam a probabilidade de transição de um estado $\mid s_1, s_2, ..., s_N\rangle$ para um estado $\mid s_1^{'}, s_2^{'}, ..., s_N^{'}\rangle$ em um ciclo de atualização $\tau$. Essa probabilidade pode ser escrita de uma maneira mais conveniente como

$$\langle s_1^{'}, s_2^{'}, ..., s_N^{'} \mid P \mid s_1, s_2, ..., s_N\rangle = \prod_{i=1}^{N} \frac{\exp[\frac{\beta}{2} s_i^{'}(V_i - V_0)]}{\sum_{s_i^{'}=\pm 1} \exp[-\frac{\beta}{2} s_i^{'}(V_i - V_0)]}. \tag{2.6}$$

Nesse ponto, o modelo de Little exibe um dos seus aspectos fundamentais, que é a analogia com sistemas de spin tipo Ising. Nesse modelo, a função de partição pode ser expressa em termos de uma matriz de transferência similar àquela definida pela equação (2.6), e a existência de ordem de longo alcance está associada à degenerescência do autovalor máximo da matriz [20]. A analogia fica mais evidente ao observar-se que uma configuração de spins de Ising pode ser representada por um estado análogo a $\mid s_1, s_2, ..., s_N\rangle$, onde $s_i = \pm 1$. A existência de ordem de longo alcance no problema de spins de Ising corresponde à existência de estados persistentes no cérebro. Persistência de estados significa correlação entre duas configurações separadas por um grande intervalo de tempo (da ordem de $10^3$ s).

A probabilidade de transição de se obter um estado $\mid s_1^{'}, s_2^{'}, ..., s_N^{'}\rangle$ após dois ciclos é

dada por

$$\sum_{s_1'',...,s_N''} \langle s_1', s_2', ..., s_N' \mid P \mid s_1'', s_2'', ..., s_N'' \rangle \langle s_1'', s_2'', ..., s_N'' \mid P \mid s_1, s_2, ..., s_N \rangle, \tag{2.7}$$

que pode ser expressa em notação matricial, de forma mais compacta,

$$\langle s_1', s_2', ..., s_N' \mid P^2 \mid s_1, s_2, ..., s_N \rangle. \tag{2.8}$$

Para um número arbitrário de ciclos, digamos $m$, temos

$$\langle s_1', s_2', ..., s_N' \mid P^m \mid s_1, s_2, ..., s_N \rangle. \tag{2.9}$$

Representando por $\psi(\alpha)$ o estado $\mid s_1, s_2, ..., s_N \rangle$ e por $\psi(\alpha')$ o estado $\mid s_1', s_2', ..., s_N' \rangle$, a probabilidade pode ser expressa em função dos auto-vetores do operador $P$, que são supostamente ortonormalizados. Portanto,

$$\langle s_1', s_2', ..., s_N' \mid P \mid s_1, s_2, ..., s_N \rangle = \sum_r \lambda_r \phi_r(\alpha') \phi_r(\alpha), \tag{2.10}$$

sendo os auto-valores do operador $P$ dados por $\lambda_r$. É exatamente a degenerescência desses auto-valores que identifica a existência de estados persistentes no cérebro.

Little calculou a probabilidade $\Gamma(\alpha_1, \alpha_2)$ de se obter um estado $\alpha_2$ após $l$ ciclos de atualização, sabendo-se que a rede encontrava-se no estado $\alpha_1$ após $m$ ciclos de atualização, sendo $m < l$. Na análise dessa probabilidade, Little observou que, na ausência de degenerescência dos auto-valores máximos, a probabilidade $\Gamma(\alpha_1, \alpha_2)$ é simplesmente o produto das probabilidades $\Gamma(\alpha_1)$ e $\Gamma(\alpha_2)$ de cada um dos estados $\alpha_1$ e $\alpha_2$, independentemente. Isso significa que a probabilidade de se obter o estado $\alpha_2$ não é afetada pelo estado $\alpha_1$, não havendo, portanto, correlação entre eles, e, conseqüentemente, não existindo estados persistentes. Entretanto, se os auto-valores máximos forem degenerados, a probabilidade $\Gamma(\alpha_1, \alpha_2)$ não mais será o produto das probabilidades individuais $\Gamma(\alpha_1)$ e $\Gamma(\alpha_2)$. Nesse caso, a probabilidade de se obter o estado $\alpha_2$ passa a depender do estado $\alpha_1$, introduzindo, portanto, uma correlação temporal entre ambos, que persiste mesmo para grandes intervalos de tempo. Isso caracteriza a possibilidade de existirem estados persistentes no cérebro para períodos de tempo arbitrariamente longos.

Outro importante aspecto envolvendo a degenerescência dos auto-valores máximos da matriz $P$ é a possibilidade de existirem correlações entre neurônios separados geograficamente por grandes distâncias. Definindo-se a distância topológica $n_{ij}$ entre os neurônios $i$ e $j$ como o número mínimo de conexões sinápticas entre os neurônios $i$ e $j$, tal que $n_{ij} \neq n_{ji}$, observa-se que são necessários, no mínimo, $n_{ij}$ ciclos de atualização dos neurônios, para que um potencial de ação emitido pelo neurônio $i$ influencie o neurônio $j$. Isso acontecerá se um estado persistir por um intervalo de tempo de, no mínimo, $n_{ij}\tau$. Desse modo, estados persistentes são caracterizados por comportamentos correlacionados entre neurônios distribuídos em diversas regiões do cérebro.

Naturalmente, a existência de estados persistentes e correlações entre neurônios é determinada pelas propriedades da matriz $P$ que dependem dos parâmetros $V_0$, $V_{ij}$ e $\beta$. Little observou, em suas análises numéricas do modelo, que, dependendo dos valores atribuídos a esses parâmetros, obtêm-se ou não estados persistentes. Utilizando uma rede com apenas 4 neurônios, portanto, uma matriz $P$ de ordem $16 \times 16$, $V_0 = 2$ e com $V_{ij} = 1.0$ fixo para todos $i$ e $j$, observou um valor crítico $\beta_0 \approx 1.0$, tal que, para $\beta > \beta_0$, ocorriam estados persistentes, ao passo que, para $\beta < \beta_0$, estes estados não estavam presentes. Alguns aspectos essenciais do modelo podem ser visualizados a partir destes resultados, em particular, a existência de estados persistentes, a existência de um contorno de fases no plano $\beta$, $V_0$ e a existência de um princípio de simetria, obtido pela condição $\sum_j V_{ij}/2 - V_0 = 0$, determinando as regiões de ocorrência de degenerescência dos auto-valores.

Em um trabalho posterior, Little e Shaw [21] mostraram que a probabilidade de transição $W(I \mid J)$ de um estado $I = \mid \{s_i\}\rangle$ no instante de tempo $t$ para um estado $J = \mid \{s'_j\}\rangle$ no instante seguinte de tempo $t+1$ é dada por

$$W(I \mid J) = \frac{\exp[-\beta H(I \mid J)]}{\sum_K \exp[-\beta H(K \mid J)]} \tag{2.11}$$

onde $\beta$ é uma medida do ruído sináptico e

$$H(I \mid J) = -\sum_{ij} V_{ij}\, s_i(I) s'_j(J), \tag{2.12}$$

acoplando os estados $I$ e $J$ em instantes de tempo diferentes.

A evolução dinâmica desse modelo foi estudada por Peretto [22] em 1984, porém utilizando a simetria das conexões sinápticas $V_{ij} = V_{ji}$, introduzida por Hopfield. Com essa condição, Peretto mostrou que a probabilidade de transição satisfaz a condição de balanço detalhado, levando a uma distribuição de Gibbs para os estados dados por $\exp[-\beta H]$, com um Hamiltoniano efetivo

$$H(I \mid I) = -\frac{1}{\beta} \sum_i \ln[2\cosh[\beta \sum_j V_{ij} s_j(I)]] \tag{2.13}$$

e

$$\beta = \frac{1}{T}, \tag{2.14}$$

onde $T$ é a "temperatura" que representa o ruído sináptico, em unidades de $k_B = 1$. Dessa forma, ocorre o desacoplamento dos estados $I$ e $J$ em tempos diferentes, estabelecendo-se uma mecânica estatística para o modelo de Little.

A partir do desenvolvimento das técnicas da mecânica estatística utilizadas no estudo de sistemas magnéticos, em particular dos vidros de spin [23], Amit, Gutfreund e Sompolisky [12] realizaram o que podemos chamar de o primeiro estudo das propriedades termodinâmicas de equilíbrio do modelo de Little. Nesse trabalho, Amit, Gutfreund e Sompolisky restringiram o estudo ao caso em que o número de memórias armazenadas na rede era finito no limite termodinâmico. O principal resultado foi mostrar que o comportamento do modelo de Little para tempos muito longos, que é essencial para a recuperação de memórias, é idêntico ao comportamento do modelo de Hopfield [11], que será discutido mais adiante neste capítulo.

O estudo da mecânica estatística de equilíbrio do modelo de Little, no regime em que o número de memórias armazenadas na rede cresce linearmente com o número de neurônios, $p = \alpha N$, foi realizado por Fontanari e Köberle [25] [26]. Esses autores obtiveram o diagrama de fases incluindo um parâmetro $J_0$, que controla a ocorrência de ciclos, utilizando o método das réplicas, em particular no regime de réplicas simétricas. Observaram a existências de fases paramagnética, ferromagnética (recuperação de memórias) e vidros de spin, bem como uma linha de pontos tri-críticos e ciclos de período 2.

## 2.3 O Modelo de Hopfield

A entrada dos físicos na área de redes neurais foi estabelecida definitivamente com o trabalho seminal de J. J. Hopfield "Neural networks and physical systems with emergent collective computational abilities" [11]. Nesse estudo, Hopfield questionou se a habilidade de grandes conjuntos de neurônios em realizar computações poderia ser uma conseqüência do comportamento coletivo espontâneo devido às interações simples de um grande número de neurônios. Essa indagação foi motivada pelo fato há muito conhecido de que sistemas físicos compostos por um grande número de elementos simples produziam uma série de fenômenos coletivos dos quais o melhor exemplo é o ferromagnetismo. Além disso, seriam as redes neurais, como sistemas análogos aos sistemas magnéticos, capazes de realizar computações úteis, tais como memórias endereçáveis por conteúdo, categorização, etc., que emergiriam como conseqüência do comportamento coletivo? Hopfield responde a essas questões propondo um novo modelo que exibe essas importantes propriedades computacionais espontaneamente.

Dada a imensa complexidade do cérebro e o conhecimento acumulado até o presente momento, muitos dos mecanismos e detalhes da anatomia e do funcionamento dos neurônios são ainda desconhecidos. Tal qual em sistemas físicos, onde a natureza das propriedades que emergem do comportamento coletivo são relativamente independentes de detalhes do modelo, no modelo de Hopfield, observam-se propriedades emergentes do comportamento coletivo que independem do conhecimento de informações detalhadas dos ingredientes introduzidos na modelagem do sistema.

Uma das habilidades mais interessantes do modelo de Hopfield é, sem dúvida, a memória endereçada por conteúdo, que consiste na capacidade do modelo em recuperar informações (memórias) completas a partir do acesso a fragmentos dessa informação. Para conceber o modelo com essa habilidade, Hopfield inspirou-se no comportamento de alguns sistemas físicos. Considerou um sistema descrito por um conjunto de coordenadas $X_1, ..., X_N$ que são as componentes de um vetor de estado $\vec{X}$. Esse sistema possui um conjunto de pontos limites localmente estáveis $X_a, X_b, ...,$ que são fundamentalmente mínimos da energia do


**Fig. 2.1:** *Relevo da função energia com seus mínimos representando as memórias.*

sistema. Quando o sistema é inicialmente colocado em um estado próximo a um desses pontos fixos, sua dinâmica leva-o para o ponto fixo correspondente. Portanto, esses pontos fixos são atratores da dinâmica. Do ponto de vista da recuperação de informações, pode-se pensar esses pontos fixos $X_a, X_b, ...$ como sendo memórias armazenadas no sistema e o estado inicial próximo a $X_a$, por exemplo $X_a + \Delta$, representando o conhecimento parcial dessa memória. A dinâmica do sistema leva-o do estado inicial $X_a + \Delta$ para o estado final $X_a$, e isso é interpretado como a recuperação da informação completa $X_a$. Esse processo pode ser visualizado na figura 2.1.

As unidades processadoras da rede são os neurônios, tendo cada um deles dois estados, $V_i = 0$ e $V_i = 1$, representando o estado inativo e ativo, respectivamente, do $i$-ésimo neurônio. Os neurônios são ligados por conexões sinápticas $T_{ij}$. A ausência de ligações entre dois neurônios, em particular $i$ e $j$ é representada por $T_{ij} = 0$, e o estado do sistema num instante de tempo $t$ é determinado pela totalidade dos estados dos neurônios e é representado por uma palavra binária de $N$ bits. A evolução dinâmica da rede é governada pela regra de atualização

$$V_i = \begin{cases} 1 & \text{se} \quad \sum_{j \neq i} T_{ij} V_j > U_i \\ 0 & \text{se} \quad \sum_{j \neq i} T_{ij} V_j < U_i \end{cases} \tag{2.15}$$

onde $U_i$ é o limiar de ativação do $i$-ésimo neurônio, que é fixo no tempo, e cada neurônio a ser atualizado é escolhido aleatoriamente. Em particular, é atualizado um neurônio a cada ciclo, pois se trata de um processo assíncrono, e $U_i = 0$. Quanto à arquitetura da rede, nesse modelo todos os neurônios estão conectados entre si, e todos os resultados relevantes são

conseqüência dessa realimentação entre os neurônios. A questão agora passa a ser a busca de um mecanismo que implemente as memórias no sistema.

Como o objetivo é armazenar e recuperar informações, essas são codificadas em seqüências de bits. Seja, pois, um conjunto de $n$ memórias $V^s, s = 1, ..., n$ a serem armazenadas nas conexões sinápticas $T_{ij}$ por meio do seguinte algoritmo

$$T_{ij} = \sum_s (2V_i^s - 1)(2V_j^s - 1), \tag{2.16}$$

com $T_{ii} = 0$. Essa expressão é a implementação matemática da regra de Hebb. Com essa prescrição de armazenamento das memórias, observa-se a estabilidade das memórias frente à evolução dinâmica. Supondo que a rede esteja numa configuração que é a memória $s'$ por exemplo, obtém-se

$$\sum_j T_{ij} V_j^{s'} \approx (2V_i^{s'} - 1)\frac{N}{2} \tag{2.17}$$

que será positivo se $V_i^{s'} = 1$ e negativo se $V_i^{s'} = 0$ de acordo com a evolução dinâmica definida na equação (2.15). Desse modo, pode-se observar que as memórias implementadas nas conexões sinápticas serão pontos fixos da dinâmica.

Um aspecto essencial do modelo de Hopfield é a não linearidade, na medida em que assume os neurônios como dispositivos cuja relação entrada-saída (potencial de membrana) é determinada por uma função degrau. Também são considerados atrasos estocásticos na transmissão sináptica, na transmissão dos potenciais de ação através dos axônios e na propagação dos pulsos elétricos pelos dendritos, o que provoca atrasos na entrada de outros neurônios. Esses atrasos são modelados por um único parâmetro, que é o tempo médio de processamento estocástico, no qual um neurônio é atualizado.

Em sistemas biológicos reais, um neurônio $j$ estabelece uma conexão sináptica com um neurônio $i$ ($T_{ij}$), porém é muito pouco provável que o neurônio $i$ estabeleça uma conexão sináptica com o neurônio $j$ ($T_{ji}$). Entretanto, se essas conexões entre os neurônios $i$ e $j$ forem estabelecidas, serão de caráter assimétrico $T_{ij} \neq T_{ji}$. Nesse ponto, Hopfield corajosamente deu um passo para trás, do ponto de vista biológico, e considerou as conexões sinápticas como sendo simétricas $T_{ij} = T_{ji}$. Isso permitiu a introdução de uma função energia definida

como

$$E = -\frac{1}{2}\sum_{i \neq j} T_{ij} V_i V_j. \tag{2.18}$$

A aplicação do algoritmo de evolução dinâmica, definido pela equação (2.15), faz com que a função energia definida na equação (2.18) seja uma função monotonicamente decrescente. Toda inversão do estado de um neurônio ($V_i \rightarrow -V_i$) representado por $\Delta V_i$ implica uma variação de energia

$$\Delta E = -\Delta V_i \sum_{j \neq i} T_{ij} V_j, \tag{2.19}$$

que persiste até que a função energia encontre um mínimo.

A analogia com o modelo de Ising é imediata, bastando que se identifiquem as eficiências sinápticas $T_{ij}$ com o acoplamento de troca $J_{ij}$ e o estado do neurônio $V_i$ com o spin $S_i$. Quando $T_{ij}$ é simétrico e aleatório, identifica-se facilmente a analogia com vidros de spin de Sherrington e Kirkpatrick (SK) [23], posto que o modelo de Hopfield é de longo alcance pois todos neurônios interagem entre si. Essa analogia permitiu que todo o arsenal matemático desenvolvido para os modelos de vidros de spin, em particular no modelo SK com o método das réplicas e sua infinidade de mínimos locais, pode ser aplicado imediatamente ao problema de redes neurais, permitindo o estudo de suas propriedades termodinâmicas de equilíbrio.

Para estudar quantitativamente o modelo, Hopfield realizou uma série de simulações de Monte Carlo à temperatura nula. Dadas as facilidades computacionais da época, as redes simuladas eram compostas por $N = 30$ e $N = 100$ neurônios. Claramente são redes extremamente pequenas se comparadas com sistemas biológicos, entretanto passíveis de fornecer indicações sobre os comportamentos do modelo em diversas situações.

Inicialmente, Hopfield examinou o comportamento da rede quando a condição de simetria $T_{ij} = T_{ji}$ era relaxada, escolhendo cada elemento da matriz sináptica $T_{ij}$ aleatoriamente entre $-1$ e 1. Escolhendo configurações iniciais aleatoriamente, verificou que os estados finais da evolução dinâmica do algoritmo resultavam em aproximadamente três estados atratores. Observou, também, a existência de ciclos simples, nos quais dois estados intercalavam-se sucessivamente, bem como um comportamento caótico em algumas situações.

Ao utilizar a prescrição sináptica dada pela expressão (2.16) para armazenar $n$ memórias aleatórias na matriz sináptica $T_{ij}$, que nesse caso é simétrica, observou que aproximadamente $0.15\ N$ estados puderam ser relembrados com pequeno erro. Nessas condições, a rede funciona como um dispositivo capaz de armazenar e recuperar informações. Para um número de memórias $n > 0.15\ N$, a rede mostrou-se incapaz de recuperar informações mesmo quando o estado inicial era exatamente uma das memórias armazenadas. Outro comportamento observado foi o colapso de memórias, ou seja, duas ou mais memórias muito parecidas levam a um mesmo atrator da dinâmica, categorizando as memórias de acordo com suas similaridades. Para que a rede recupere as informações, é necessário que as mesmas não sejam correlacionadas, caso contrário ocorrerá o colapso de memórias.

O modelo de Hopfield mostrou-se, então, um dispositivo capaz de armazenar informações endereçáveis por conteúdo, robusto, capaz de categorizar. Essas propriedades emergem do comportamento coletivo de um grande número de neurônios que interagem através dos acoplamentos sinápticos, onde as informações estão armazenadas. A simetria das interações sinápticas e a definição de uma função energia permitiram a análise do modelo a partir do ponto de vista da mecânica estatística.

## 2.4 O Modelo de Hopfield Generalizado

A evolução dinâmica proposta originalmente por Hopfield [11] é uma dinâmica serial determinística, em que o estado de um neurônio no instante de tempo $t+1$ é determinado pelo sinal do campo local,

$$S_i(t+1) = \mathrm{sgn}[h_i(t)], \tag{2.20}$$

onde

$$h_i(t) = \sum_j J_{ij} S_i(t) \tag{2.21}$$

é o campo local, $J_{ij}$ é o acoplamento sináptico entre os neurônios $i$ e $j$, e sgn(x) é a função sinal. No entanto, o processo de transmissão dos sinais elétricos no cérebro é acompanhado de uma grande quantidade de ruído, introduzindo um caráter estocástico no processo. Esta

estocasticidade é introduzida na dinâmica atribuindo-se uma probabilidade $P$ do neurônio $i$ assumir o estado $S_i$,

$$P(S_i) = \frac{1}{1 + \exp[-2\beta h_i(t) S_i]}, \tag{2.22}$$

onde, novamente, $\beta = 1/T$ é o inverso da temperatura, que fornece uma medida da intensidade do ruído. Note que essa expressão descreve, essencialmente, a dinâmica de Glauber [27] para um sistemas de spins Ising interagentes, em contato com um banho térmico à temperatura $T$. No limite em que $\beta \to \infty$, suprime-se o ruído e recupera-se a dinâmica determinística.

No desenvolvimento realizado por Amit, Gutfreund e Sompolinsky para o modelo de Hopfield, os estados dos neurônios são representados por spins de Ising nos quais $S_i = \pm 1$, com $i = 1, ..., N$. A capacidade de armazenamento $\alpha$ da rede é definida pela razão entre o número de memórias, $p$, armazenadas na rede e o número de neurônios $N$

$$\alpha = \frac{p}{N}. \tag{2.23}$$

As memórias ou padrões, representados por $\xi_i^\mu = \pm 1$ com $\mu = 1, ..., p$ e $i = 1, ..., N$, são variáveis aleatórias estatisticamente independentes, dadas pela seguinte distribuição de probabilidade

$$P(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1), \tag{2.24}$$

onde $\delta(\xi_i^\mu \mp 1) = 1$ se $\xi_i^\mu = \pm 1$ ou zero de outro modo.

O processo de memorização desses padrões se dá através da regra de aprendizagem de Hebb (2.16), escrita como

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, \tag{2.25}$$

com $J_{ii} = 0$, pois os neurônios não possuem, essencialmente, auto-conexões. Essa regra apresenta algumas vantagens: é aditiva, de modo que novas memórias são adicionadas à rede, o que caracteriza um processo de aprendizagem; é local, pois a conexão $J_{ij}$ depende apenas dos neurônios $i$ e $j$ para todos os padrões; e é simétrica ($J_{ij} = J_{ji}$), possibilitando o

estabelecimento de uma função energia (Hamiltoniano) $H$ para o sistema

$$H = -\frac{1}{2}\sum_{i \neq j} J_{ij} S_i S_j. \tag{2.26}$$

As propriedades termodinâmicas do modelo são obtidas a partir da mecânica estatística de equilíbrio, seguindo procedimentos utilizados para o modelo de Ising no estudo de sistemas magnéticos desordenados, em particular, de sistemas que apresentam comportamentos ferromagnético, paramagnético e vidros de spin [23] [24].

O comportamento da rede neural depende do número de padrões $p$ que se deseja armazenar. Observam-se, claramente, dois regimes distintos. Em um, o número de padrões $p$ é finito no limite termodinâmico em que $N \to \infty$, sendo a capacidade de armazenamento $\alpha = 0$. Em outro, o número de padrões que se deseja armazenar é proporcional ao tamanho da rede, $p = \alpha N$, de forma que tanto $p$ quanto $N$ tendem a infinito, porém a razão $p/N$ permanece finita.

A recuperação de um padrão, $\mu$ por exemplo, é medida através da superposição entre o estado da rede e o respectivo padrão armazenado, definida como

$$m^\mu = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \langle S_i \rangle_T \tag{2.27}$$

sendo

$$\langle S_i \rangle_T = \frac{Tr_S S_i \exp(-\beta H)}{Z} \tag{2.28}$$

a média térmica e $-1 \leq m^\mu \leq 1$. Esse parâmetro de ordem caracteriza a capacidade de rede em recuperar uma memória ($m^\mu \neq 0$) ou uma situação em que a rede é incapaz de recuperá-la ($m^\mu = 0$).

### 2.4.1 Memorização para $p$ finito

Nesse regime, as propriedades termodinâmicas da rede são determinadas a partir da densidade de energia livre

$$f = -\lim_{N \to \infty} \frac{1}{\beta N} \langle \ln Z \rangle_{\{\xi^\mu\}}, \tag{2.29}$$

onde $\langle \ldots \rangle_{\{\xi^\mu\}}$ representa a média configuracional sobre a distribuição dos padrões $\{\xi^\mu\}$, dada pela equação (2.24), e $Z$ é a função de partição

$$Z = Tr_S \exp(-\beta H), \tag{2.30}$$

sendo o traço $Tr_S$ a soma sobre todas as possíveis configurações $\{S_i\}$ da rede.

Acrescenta-se ao Hamiltoniano um campo externo uniforme $h^\mu$, útil no cálculo de propriedades da rede, conjugado aos estados da rede $S_i$, para todos os padrões

$$H = -\frac{1}{2}\sum_{i \neq j} J_{ij} S_i S_j - \sum_\mu h^\mu \sum_i \xi_i^\mu S_i. \tag{2.31}$$

Introduzindo a regra de Hebb, equação (2.25), nesse Hamiltoniano, pode-se escrever a função de partição como

$$Z = e^{-\frac{1}{2}\beta p} Tr_S \exp[\frac{\beta}{2N}\sum_\mu (\sum_i \xi_i^\mu S_i)^2 + \beta \sum_\mu h^\mu \sum_i \xi_i^\mu S_i]. \tag{2.32}$$

Seria uma tarefa simples calcular o traço se a dependência em $S_i$, no expoente, fosse linear. Entretanto, este não é o caso, pois a exponencial depende quadraticamente de $S_i$, introduzindo uma dificuldade adicional. O uso da integral Gaussiana

$$\exp(\lambda a^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[-\frac{1}{2}y^2 + \sqrt{2\lambda} a y]\ dy, \tag{2.33}$$

permite superar essa dificuldade ao custo de se introduzir $p$ variáveis auxiliares, porém lineariza-se a dependência em $S_i$, permitindo o cálculo do traço. Chamando $\{m^\mu\}$ o conjunto dessas variáveis e tomando o traço, a função de partição pode ser escrita como

$$Z = e^{-\frac{1}{2}\beta p} (\beta N)^{\frac{p}{2}} \int \prod_\mu (\frac{dm^\mu}{\sqrt{2\pi}}) e^{-N\beta G(\{m\})} \tag{2.34}$$

onde

$$G(\{m\}) = \frac{1}{2}\sum_\mu (m^\mu)^2 - \frac{1}{\beta N}\sum_i \ln[2\cosh(\beta \sum_\mu (m^\mu + h^\mu)\xi_i^\mu)]. \tag{2.35}$$

No limite termodinâmico, $N \to \infty$, podemos resolver a integral através do método de ponto de sela, resultando

$$-\frac{1}{\beta N}\ln Z = \min_m\ G(\{m\}). \tag{2.36}$$

Os valores de $\{m\}$, chamados pontos de sela, que minimizam a função $G(\{m\})$ são dados pelas condições de extremo

$$\frac{\partial G(\{m\})}{\partial m^\mu} = 0, \tag{2.37}$$

que fornecem as $p$ equações de ponto de sela

$$m^\mu = \frac{1}{N}\sum_i \xi_i^\mu \tanh[\beta \sum_\mu (m^\mu + h^\mu)\xi_i^\mu]. \tag{2.38}$$

Utilizando a propriedade da automediação, que consiste em substituir a média sobre sítios pela média sobre padrões $\{\xi^\mu\}$, obtém-se

$$m^\mu = \langle \xi^\mu \tanh[\beta \sum_\mu (m^\mu + h^\mu)\xi^\mu]\rangle_{\{\xi^\mu\}}, \tag{2.39}$$

e a densidade de energia livre, avaliada nos pontos de sela,

$$f = -\lim_{N\to\infty} \frac{1}{\beta N}\langle \ln Z\rangle_{\{\xi^\mu\}} = \frac{1}{2}\sum_\mu (m^\mu)^2 - \frac{1}{\beta}\langle \ln[2\cosh(\beta \sum_\mu (m^\mu + h^\mu)\xi^\mu)]\rangle_{\{\xi^\mu\}}. \tag{2.40}$$

A interpretação física dos pontos de sela $m^\mu$, dados pela equação (2.39), fica clara ao se calcular a derivada da energia livre em relação ao campo $h^\mu$, que, após o cálculo, é igualado a zero. Os pontos de sela são, de fato, as superposições dadas pela equação (2.27) e, a partir desse ponto, considera-se o campo externo $h^\mu = 0$.

As equações de ponto de sela e a densidade de energia livre podem ser expressas em notação vetorial para o índice $\mu$ com a introdução dos vetores $\mathbf{m}$ e $\boldsymbol{\xi}$, ambos com $p$ componentes,

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh[\beta(\mathbf{m}\cdot\boldsymbol{\xi})]\rangle_{\{\xi^\mu\}}, \tag{2.41}$$

$$f = \frac{1}{2}\mathbf{m}^2 - \frac{1}{\beta}\langle \ln[2\cosh(\beta(\mathbf{m}\cdot\boldsymbol{\xi})]\rangle_{\{\xi^\mu\}}. \tag{2.42}$$

As soluções das equações de ponto de sela, que determinam a correlação entre os estados da rede $\{\langle S_i\rangle_T\}$ com cada um dos $p$ padrões $\{\xi_i^\mu\}$ armazenados, devem ser tais que a densidade de energia livre seja minimizada. Expandindo as equações (2.41) e (2.42) em potências de $\mathbf{m}$ até segunda ordem, resulta

$$f = -T\ln 2 + \frac{1}{2}(1-\beta)\mathbf{m}^2 + \mathcal{O}(m^4) \tag{2.43}$$

e

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

de onde se pode observar que o mínimo da função $f(m)$ é obtido para $\mathbf{m} = 0$, quando $T > 1$, correspondendo ao estado paramagnético com $f = -T \ln 2$. Isso significa que, em altas temperaturas ($T > 1$), a rede possui apenas o estado de equilíbrio trivial $\langle S_i \rangle_T = 0$. Nessa condição, a rede não recupera nenhum padrão armazenado.

Em baixas temperaturas ($T < 1$), surgem, de maneira contínua, soluções não triviais com $\mathbf{m} \neq 0$, indicando uma transição de fase de segunda ordem em $T_c = 1$. Essas soluções são caracterizadas por vetores $\mathbf{m}$ com $n$ componentes não nulas, chamadas soluções simétricas, da forma

$$\mathbf{m} = m_n(\underbrace{1, \ldots, 1}_{n}, \underbrace{0, \ldots, 0}_{p-n}). \tag{2.45}$$

Para $n = 1$, obtém-se a solução de recuperação, pois apenas um dos padrões, $\mu = 1$ por exemplo, possui superposição não nula. Nesse caso, a equação (2.41) assume a forma

$$m^1 = \tanh(\beta m^1), \tag{2.46}$$

e a densidade de energia livre (2.42) correspondente é

$$f = \frac{1}{2}(m^1)^2 - \frac{1}{\beta} \ln[2 \cosh(\beta m^1)]. \tag{2.47}$$

Esses estados, chamados estados de Mattis, são mínimos absolutos no intervalo de temperatura $0 \leq T \leq 1$. Em $T = 0$, recupera-se a dinâmica determinista original de Hopfield dada pela equação (2.20), e, apenas nessa condição, pode-se recuperar perfeitamente um padrão ($m = 1$). Para $T > 0$, a recuperação jamais será perfeita ($m < 1$).

No caso mais geral ($n > 1$), as $n$ componentes da equação (2.41) devem ser somadas, gerando soluções do tipo simétrica expressas com

$$m_n = \frac{1}{n} \langle z_n \tanh(\beta m_n z_n) \rangle_{z_n}, \tag{2.48}$$

$$f_n = \frac{n}{2} m_n^2 - \frac{1}{\beta} \langle \ln[2 \cosh(\beta m_n z_n)] \rangle_{z_n}, \tag{2.49}$$

sendo $z_n = \sum_{\mu=1}^{n} \xi^{\mu}$ uma variável aleatória, cuja distribuição de probabilidades binomial é

$$P(z_n) = \frac{1}{2^n} \begin{pmatrix} n \\ k \end{pmatrix}, \tag{2.50}$$

onde $k = \frac{1}{2}(z_n + n)$, $0 \leq k \leq n$.

Essas soluções simétricas representam estados que são misturas de vários padrões. Observa-se que apenas estados simétricos com número ímpar de componentes são meta-estáveis, enquanto estados com número par de componentes são instáveis. Estados simétricos $m_3$ surgem para temperaturas abaixo de $T_3 = 0.461$, $m_5$ para temperaturas abaixo de $T_5 = 0.385$ e assim sucessivamente. A tabela 2.1 mostra os primeiros valores de $T_n$.

| $n$ | $T_n$ |
|---|---|
| 1 | 1 |
| 3 | 0.461 |
| 5 | 0.385 |
| 7 | 0.345 |

**Tab. 2.1:** *Temperatura crítica abaixo da qual os estados simétricos de n componentes tornam-se meta-estáveis.*

Além das soluções simétricas, existem também, para baixas temperaturas, soluções cujas componentes não nulas possuem valores diferentes, chamadas soluções assimétricas, que assumem a forma genérica

$$\mathbf{m} = (m, m, \cdots, m, \epsilon, \epsilon, \cdots, \epsilon, \delta, \delta, \cdots, \delta, 0, 0, \cdots, 0) \tag{2.51}$$

sendo que algumas dessas soluções surgem descontinuamente.

Para o modelo de Little, vimos, na seção 2.2, que a dinâmica paralela leva o sistema a uma distribuição de Gibbs para estados estacionários com um Hamiltoniano efetivo dado por [22]

$$\overline{H} = -\frac{1}{\beta} \sum_i \ln[2\cosh(\beta \sum_j J_{ij} S_j)], \tag{2.52}$$

com $J_{ij}$ dado pela equação (2.25). Para esse Hamiltoniano, Amit, Gutfreund e Sompolinsky utilizaram o mesmo procedimento matemático desenvolvido para o modelo de Hopfield, obtendo a mesma equação de campo médio para $\mathbf{m}$ e uma densidade de energia livre que é exatamente o dobro da densidade de energia livre do modelo de Hopfield

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh[\beta(\mathbf{m} \cdot \boldsymbol{\xi})] \rangle_{\{\xi^\mu\}}, \tag{2.53}$$

$$f_L = \mathbf{m}^2 - \frac{2}{\beta} \langle \ln[2\cosh(\beta(\mathbf{m} \cdot \boldsymbol{\xi})] \rangle_{\{\xi^\mu\}}. \tag{2.54}$$

Assim, o estado de Mattis é o único estado de mínimo global no intervalo de temperatura $0 \leq T \leq 1$, e estados de mínimos locais aparecem para temperaturas abaixo de $T = 0.461$.

### 2.4.2 Memorização para $p$ proporcional a $N$

Quando o número de padrões cresce linearmente com o tamanho da rede ($p = \alpha N$), um ou um número finito de padrões podem ter uma superposição macroscópica cuja magnitude permanece finita quando $N \to \infty$. Para levar em conta essa possibilidade, adiciona-se um campo externo conjugado a esse conjunto finito de padrões ($\{\xi_i^\nu\}$, $\nu = 1, \cdots, s$) ao Hamiltoniano

$$H = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j - \sum_{\nu=1}^{s} h^\nu \sum_i \xi_i^\nu S_i. \tag{2.55}$$

Dado que, nesse limite, o sistema não satisfaz a condição $2^p \ll N$, a propriedade da automediação não pode ser aplicada, e as médias terão de ser calculadas explicitamente sobre a distribuição dos $\{\xi^\mu\}$. Utiliza-se, então, o método das réplicas, o qual consiste no uso da identidade matemática

$$\ln x = \lim_{n \to 0} \frac{x^n - 1}{n} \tag{2.56}$$

para calcular a densidade de energia livre

$$f = -\lim_{n \to 0} \lim_{N \to \infty} \frac{1}{\beta n N} (\langle Z^n \rangle_{\{\xi^\mu\}} - 1), \tag{2.57}$$

onde $\langle \cdots \rangle_{\{\xi^\mu\}}$ representa a média sobre os padrões $\{\xi\}$,

$$Z^n = Tr_{S^\rho} \exp(-\beta \sum_{\rho=1}^{n} H^\rho) \tag{2.58}$$

é a função de partição replicada e

$$H^\rho = -\frac{1}{2}\sum_{i\neq j}\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu S_i^\rho S_j^\rho - \sum_{\nu=1}^{s}h^\nu\sum_i \xi_i^\nu S_i^\rho. \tag{2.59}$$

Podemos pensar $Z^n$ como a função de partição de $n$ cópias ou réplicas do sistema, sendo cada cópia identificada pelo índice de réplica $\rho$.

Utilizando-se a integral Gaussiana (2.33) para linearizar a dependência quadrática em $S_i^\rho$, a função de partição pode ser escrita como

$$\begin{aligned}\langle Z^n\rangle_{\{\xi^\mu\}} &= \exp(-\frac{\beta pn}{2})(\beta N)^{\frac{pn}{2}}Tr_{S^\rho} \\ &\left\{\int\prod_{\rho,\nu\leq s}\left(\frac{dm_\rho^\nu}{\sqrt{2\pi}}\right)\left\langle\exp\{\beta N[-\frac{1}{2}\sum_{\rho,\nu\leq s}(m_\rho^\nu)^2+\sum_{\rho,\nu\leq s}(m_\rho^\nu+h^\nu)\frac{1}{N}\sum_i\xi_i^\nu S_i^\rho]\}\right\rangle_{\{\xi^\nu\}}\right. \\ &\left.\int\prod_{\rho,\mu>s}\left(\frac{dm_\rho^\mu}{\sqrt{2\pi}}\right)\left\langle\exp\{\beta N[-\frac{1}{2}\sum_{\rho,\mu>s}(m_\rho^\mu)^2+\sum_{\rho,\mu>s}m_\rho^\mu\frac{1}{N}\sum_i\xi_i^\mu S_i^\rho]\}\right\rangle_{\{\xi^\mu\}}\right\}\end{aligned} \tag{2.60}$$

Note que o campo $h^\nu$ está acoplado aos $s$ primeiros padrões, e $\rho = 1,\cdots,n$ é o índice de réplicas. A média foi separada entre os primeiros $s$ padrões ($\nu \leq s$) e os $p-s$ padrões restantes ($\mu > s$), também chamados de memórias baixas e memórias altas ou padrões condensados e não condensados, respectivamente.

Após calcular-se a média sobre as memórias altas e integrar sobre as variáveis re-escalonadas $m_\rho^\mu \to m_\rho^\mu/\sqrt{N}$, para que o limite termodinâmico esteja bem definido, obtém-se a seguinte expressão para a função de partição

$$\begin{aligned}\langle Z^n\rangle_{\{\xi^\mu\}} &= \exp(-\frac{\beta pn}{2})(\beta N)^{\frac{pn}{2}}\int\prod_{\nu\rho}\left(\frac{dm_\rho^\nu}{\sqrt{2\pi}}\right)\int\prod_{(\rho,\sigma)}dq_{\rho\sigma}dr_{\rho\sigma} \\ &\exp\left\{N\left[-\frac{1}{2}\beta\sum_{\rho,\nu}(m_\rho^\nu)^2-\frac{1}{2}\alpha Tr\ln[(1-\beta)\mathbf{I}-\beta\mathbf{q}]-\frac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma}r_{\rho\sigma}q_{\rho\sigma}\right.\right. \\ &\left.\left.\left\langle\ln Tr_{S^\rho}\exp\left[\frac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma}r_{\rho\sigma}S^\rho S^\sigma+\beta\sum_{\rho,\nu}(m_\rho^\nu+h^\nu)\frac{1}{N}\sum_i\xi^\nu S^\rho\right]\right\rangle_{\{\xi^\nu\}}\right]\right\}.\end{aligned} \tag{2.61}$$

Note que a média na equação acima é sobre os padrões condensados, que o subscrito $(\rho,\sigma)$ significa $\rho < \sigma$, $\mathbf{I}$ é a matriz identidade, e $\mathbf{q}$ é uma matriz cujos elementos são $q_{\rho\sigma}$. Note,

também, que esta é uma integral do tipo $\int dx \exp[-NG(x)]$ que pode ser resolvida pelo método do ponto de sela.

No limite termodinâmico, $N \to \infty$, o integrando da equação (2.61) é dominado pelos seus pontos de sela, levando à seguinte densidade de energia livre avaliada em $n \to 0$

$$\begin{aligned} f &= \frac{1}{2}\alpha + \lim_{n\to 0}\left\{ (\frac{\alpha}{2\beta n}) Tr \ln[(1-\beta)\mathbf{I} - \beta\mathbf{q}] + \frac{1}{2n}\sum_{\rho,\nu}(m_\rho^\nu)^2 \right. \\ &+ \left. \frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} r_{\rho\sigma}q_{\rho\sigma} - \frac{1}{n\beta}\langle \ln Tr_S \exp(\beta H_\xi)\rangle_{\{\xi^\mu\}} \right\}, \end{aligned} \tag{2.62}$$

onde $H_\xi$ é o Hamiltoniano efetivo dado por

$$H_\xi = \frac{\alpha\beta}{2}\sum_{\rho\neq\sigma} r_{\rho\sigma}S^\rho S^\sigma + \sum_{\rho,\nu}(m_\rho^\nu + h^\nu)\frac{1}{N}\sum_i \xi^\nu S^\rho. \tag{2.63}$$

Os estados estacionários são obtidos a partir das equações de ponto de sela que minimizam a densidade de energia livre, dadas por

$$\frac{\partial f}{\partial m_\rho^\mu} = 0, \quad \frac{\partial f}{\partial q_{\rho\sigma}} = 0, \quad \frac{\partial f}{\partial r_{\rho\sigma}} = 0. \tag{2.64}$$

O significado físico dos parâmetros de ordem $m_\rho^\nu$, $q_{\rho\sigma}$ e $r_{\rho\sigma}$ é obtido a partir das equações de ponto de sela, determinando os valores estacionários do integrando da equação (2.61) no limite termodinâmico. Para a superposição de um estado da rede com um padrão armazenado tem-se

$$m_\rho^\nu = \frac{1}{N}\langle \sum_i \xi_i^\nu \langle S_i^\rho\rangle_T\rangle_{\{\xi^\mu\}}. \tag{2.65}$$

O parâmetro $q_{\rho\sigma}$ é identificado como o parâmetro de ordem de Edwards-Anderson para vidros de spin

$$q_{\rho\sigma} = \langle \frac{1}{N}\sum_i \langle S_i^\rho\rangle_T \langle S_i^\sigma\rangle_T\rangle_{\{\xi^\mu\}}, \tag{2.66}$$

e o multiplicador de Lagrange $r_{\rho\sigma}$

$$r_{\rho\sigma} = \frac{1}{\alpha}\sum_{\mu>s}\langle [\frac{1}{N}\sum_i \xi_i^\nu \langle S_i^\rho\rangle_T \frac{1}{N}\sum_i \xi_i^\nu \langle S_i^\sigma\rangle_T]\rangle_{\{\xi^\mu\}}$$

é identificado como a média quadrática das superposições de uma configuração da rede com as memórias altas, ou seja,

$$r_{\rho\sigma} = \frac{1}{\alpha}\sum_{\mu>s}\langle m_\rho^\mu m_\sigma^\mu\rangle_{\{\xi^\mu\}}. \tag{2.68}$$

Observa-se, a partir das equações (2.65), (2.66) e (2.68), que todas as réplicas são, a princípio, equivalentes. Isso sugere que as soluções não dependam do valor do índice de réplica. Essa é a chamada solução de *simetria de réplicas*, em que se supõe que os valores dos pontos de sela para os parâmetros de ordem não dependam dos índices de réplica

$$m_\rho^\nu = m^\nu, \tag{2.69}$$

$$q_{\rho\sigma} = q, \quad \rho \neq \sigma, \tag{2.70}$$

$$r_{\rho\sigma} = r, \quad \rho \neq \sigma. \tag{2.71}$$

Introduzindo essas soluções nas equações (2.62) e (2.63) e tomando o limite em que $n \to 0$, obtém-se a densidade de energia livre para réplicas simétricas

$$\begin{aligned} f &= \frac{1}{2}\alpha + \frac{1}{2}\sum_\nu (m^\nu)^2 + \frac{\alpha}{2\beta}[\ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q}] + \frac{\alpha\beta r}{2}(1-q) \\ &- \frac{1}{\beta}\int \mathcal{D}z \langle \ln 2\cosh\beta[\sqrt{\alpha r}z + \sum_\nu (m^\nu + h^\nu)\xi^\nu]\rangle_{\{\xi^\mu\}} \end{aligned} \tag{2.72}$$

sendo

$$\int \mathcal{D}z \equiv \int \frac{dz}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}). \tag{2.73}$$

Note que a média a ser realizada é sobre os $s$ padrões condensados e que as flutuações causadas pelos $p-s$ padrões não condensados estão contempladas implicitamente na integral sobre o termo Gaussiano (média sobre o ruído Gaussiano $z$), contido no último termo da densidade de energia livre acima.

Derivando-se a densidade de energia livre $f$ em relação aos parâmetros de ordem $m$, $q$ e $r$ e igualando-as a zero, obtêm-se as equações de ponto de sela de forma auto-consistente,

$$\frac{\partial f}{\partial m} = 0 \quad \to \quad m^\nu = \int \mathcal{D}z \langle \xi^\nu \tanh\beta[\sqrt{\alpha r}z + \sum_\nu (m^\nu + h^\nu)\xi^\nu]\rangle_{\{\xi^\mu\}}, \tag{2.74}$$

$$\frac{\partial f}{\partial r} = 0 \quad \to \quad q = \int \mathcal{D}z \langle \tanh^2\beta[\sqrt{\alpha r}z + \sum_\nu (m^\nu + h^\nu)\xi^\nu]\rangle_{\{\xi^\mu\}}, \tag{2.75}$$

e

$$\frac{\partial f}{\partial q} = 0 \quad \to \quad r = \frac{q}{(1-\beta+\beta q)^2}, \tag{2.76}$$

que devem ser resolvidas numericamente, fornecendo o comportamento da rede nas diversas situações.

**I - Limite de ruído nulo ($T = 0$):**

Na medida em que a temperatura se aproxima de zero ($T \to 0$), a tangente hiperbólica se reduz à função degrau, e a equação para o parâmetro de ordem $m^\nu$ torna-se

$$m^\nu = \left\langle \xi^\nu \mathrm{erf}\left[\frac{\sum_\nu (m^\nu + h^\nu)\xi^\nu}{\sqrt{2\alpha r}}\right]\right\rangle_{\{\xi^\mu\}}, \tag{2.77}$$

ao passo que o parâmetro de ordem $q$ se aproxima de 1, levando a

$$C \equiv \beta(1-q) = \sqrt{\frac{2}{\pi\alpha r}}\left\langle \exp\left\{-\frac{[\sum_\nu (m^\nu + h^\nu)\xi^\nu]^2}{2\alpha r}\right\}\right\rangle_{\{\xi^\mu\}}. \tag{2.78}$$

O parâmetro de ordem $r$ torna-se

$$r = \frac{1}{(1-C)^2}, \tag{2.79}$$

sendo $\mathrm{erf}(x)$ a função erro definida como

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-y^2}\, dy. \tag{2.80}$$

**a) solução de recuperação**

Estados de recuperação são descritos por soluções do tipo $m^\nu = m\delta_{\nu,1}$, também chamados de estados ferromagnéticos em analogia com sistemas magnéticos. Na ausência de ruído ($T = 0$) e assumindo $h^\nu = 0$, as equações para os parâmetros de ordem $m$ e $q$ passam a ser escritas, após calcular a média sobre os padrões, como

$$m = \mathrm{erf}\left(\frac{m}{\sqrt{2\alpha r}}\right), \tag{2.81}$$

$$C = \sqrt{\frac{2}{\pi\alpha r}}\exp\left(-\frac{m^2}{2\alpha r}\right), \tag{2.82}$$

e o parâmetro $r$ é dado pela equação (2.79) acima.

Definindo uma nova variável $y = \frac{m}{\sqrt{2\alpha r}}$, pode-se reduzir as três equações (2.81), (2.82) e (2.79) a uma única equação dada por

$$y = \frac{\mathrm{erf}(y)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}}\exp(-y^2)}, \tag{2.83}$$

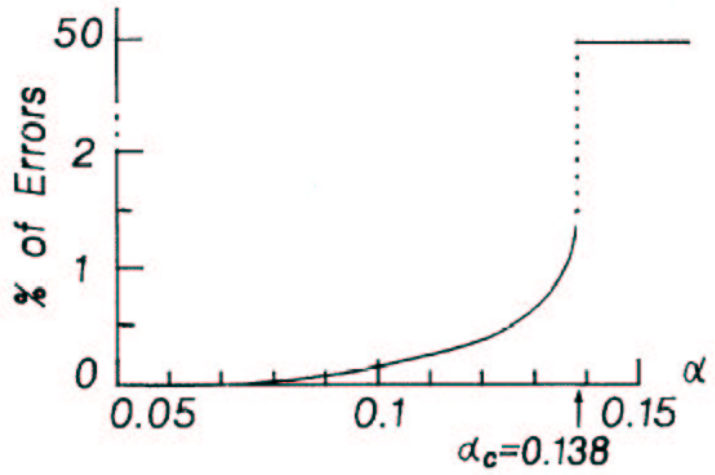



**Fig. 2.2:** *Porcentagem de erro médio,* $\frac{N_e}{N} = \frac{1-m}{2}$*, do estado de recuperação com simetria de réplicas, a* $T = 0$ [13].

sendo a densidade de energia livre

$$f = \frac{1}{2}\mathrm{erf}^2(y) + \frac{1}{\pi}e^{-y^2} - \frac{2}{\pi}\left[e^{-y^2} + \sqrt{\frac{\alpha\pi}{2}}\right]\left[y\sqrt{\pi}\mathrm{erf}(y) + e^{-y^2}\right]. \tag{2.84}$$

Essa equação possui sempre a solução trivial $y = 0$ que corresponde a $m = 0$, indicando a completa ausência de superposição com qualquer um dos padrões armazenados. Essa solução, chamada de solução de vidro de spin, existe para todos os valores de $\alpha$ e é a única solução para $\alpha > \alpha_c = 0.138$. O comportamento de $\frac{1-m}{2}$ em função de $\alpha$ pode ser observado na figura 2.2.

Para $\alpha < \alpha_c$, surgem descontinuamente soluções com $m \neq 0$ (solução de recuperação), sendo que em $\alpha_c$ a superposição vale $m = 0.967$. Embora para $\alpha$ finito jamais tenhamos $m = 1$, observam-se valores muito próximos à unidade indicando que os estados estacionários são muito próximos dos padrões armazenados e, desse modo, são recuperados sem ambigüidade. Acima de $\alpha_M = 0.051$, a solução de vidro de spin possui menor energia que a solução de recuperação, porém, para valores de $\alpha$ menores de $\alpha_M$, a situação se inverte, e a solução de recuperação é o estado de mínimo global, como mostra a figura 2.3.

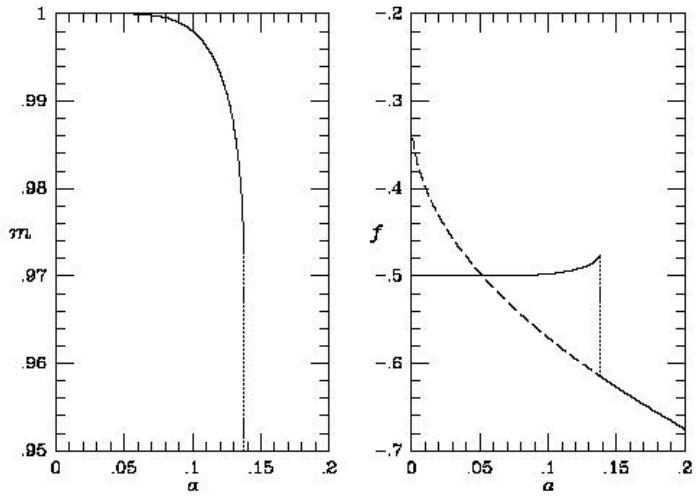



**Fig. 2.3:** *A esquerda: comportamento da superposição $m$ em função de $\alpha$ a $T = 0$. A localização da descontinuidade, onde $m$ desaparece, define a capacidade de armazenamento crítica $\alpha_c \sim 0.138$. A direita: densidade de energia livre a $T = 0$ para estados de recuperação (linha contínua) e para estados de vidro de spin (linha tracejada). O cruzamento de ambas determina $\alpha_M = 0.051$* [28].

### b) solução simétrica

Outra solução observada é aquela em que $n$ componentes do vetor $\mathbf{m}$ são não nulas, indicando uma superposição simétrica com $n$ padrões armazenados. Nesse caso, as equações (2.77) e (2.78) tornam-se

$$m_n = \frac{1}{n}\left\langle z_n \mathrm{erf}\left(\frac{z_n m_n}{\sqrt{2\alpha r}}\right)\right\rangle_{z_n} \tag{2.85}$$

e

$$C = \sqrt{\frac{2}{\pi\alpha r}}\left\langle \exp\left(-\frac{m_n^2 z_n^2}{2\alpha r}\right)\right\rangle_{z_n}, \tag{2.86}$$

onde $z_n = \sum_{\mu=1}^{n} \xi^\mu$ é uma variável aleatória binomial. Definindo $y_n \equiv \frac{m_n}{\sqrt{2\alpha r}}$, obtém-se, novamente, uma única equação

$$n y_n = \frac{\langle z_n \mathrm{erf}(z_n y_n)\rangle_{z_n}}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}}\langle \exp(-z_n^2 y_n^2)\rangle_{z_n}}. \tag{2.87}$$

Na medida em que o número de componentes $n$ aumenta, encontram-se valores críticos $\alpha_n$ acima dos quais a única solução é $y_n = 0$. Abaixo de $\alpha_n$, soluções simétricas $m_n \neq 0$

surgem descontinuamente. Apenas soluções com $n$ ímpar são localmente estáveis. Por exemplo, para $\alpha > \alpha_3 \approx 0.03$ observa-se $m_3 = 0$ e para $\alpha \leq \alpha_3$ observa-se $m_3 \neq 0$, em particular $m_3(\alpha_3) \approx 0.496$.

**II - Soluções a temperatura finita ($T \neq 0$):**

Para que o modelo de Hopfield funcione como um dispositivo de memória, é fundamental analisar as soluções de recuperação que traduzem esse comportamento desejado. Essas soluções são descritas pela condição $m^\mu = m\delta_{\mu,\nu}$. Nesse caso, as equações (2.74) e (2.75) são expressas, com $h^\nu = 0$, da seguinte forma:

$$m = \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh[\beta(\sqrt{\alpha r}z + m)], \tag{2.88}$$

$$q = \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2[\beta(\sqrt{\alpha r}z + m)], \tag{2.89}$$

e

$$r = \frac{q}{(1-\beta+\beta q)^2}. \tag{2.90}$$

A densidade de energia livre é dada por

$$\begin{aligned} f &= \frac{1}{2}\alpha + \frac{1}{2}m^2 + \frac{\alpha}{2\beta}[\ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q}] + \frac{\alpha\beta r}{2}(1-q) \\ &- \frac{1}{\beta}\int \mathcal{D}z \ln 2\cosh\beta[\sqrt{\alpha r}z + m]. \end{aligned} \tag{2.91}$$

Esse sistema de equações não lineares deve ser resolvido numericamente, fornecendo as soluções para os parâmetros de ordem $m$ e $q$. Quando mais de uma solução existir simultaneamente, as energias livres respectivas são comparadas, permitindo, assim, determinar a solução que é mínimo global. Variando-se a temperatura ($T$) e o parâmetro de carga $\alpha$, obtém-se o diagrama de fases apresentado na figura 2.4, onde os diferentes regimes de comportamento da rede podem ser identificados em função dos valores assumidos pelos parâmetros de ordem $m$ e $q$.

Observam-se quatro regiões distintas separadas pelas linhas críticas $T_c$, $T_M$ e $T_g$. Acima da linha $T_g$, que é uma linha de transição de fase de segunda ordem, encontra-se uma região

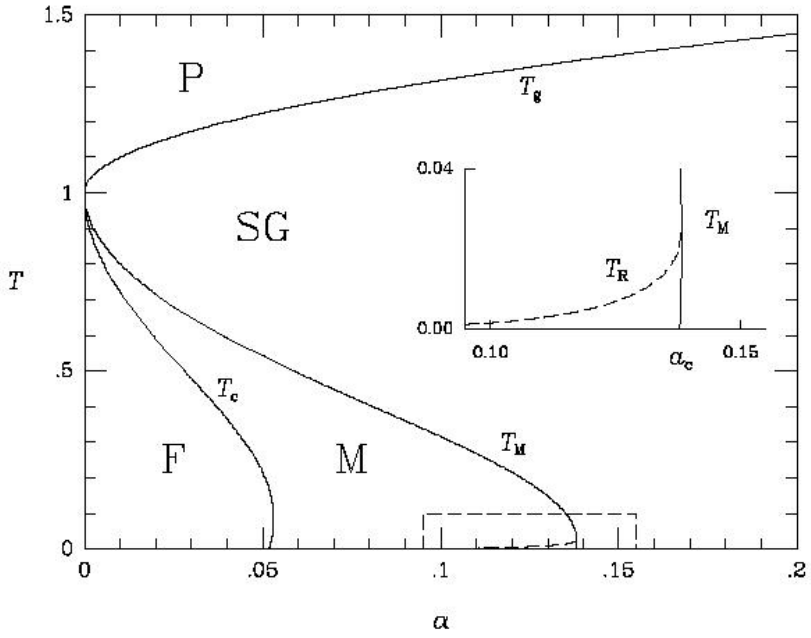



**Fig. 2.4:** *Diagrama de fases $\alpha \times T$ para estados de recuperação com simetria de réplicas* [28].

onde a única solução é a paramagnética (P), na qual $m = 0$ e $q = 0$. Abaixo da linha $T_g$ e acima da linha $T_M$, encontra-se a região onde existe a solução de vidro de spin (SG) com $m = 0$ e $q \neq 0$. Essas soluções não guardam qualquer correlação com as memórias armazenadas.

Na região M, à esquerda da linha $T_M$, surge descontinuamente a solução de recuperação de padrões com $m \neq 0$ e $q \neq 0$, caracterizando uma transição de primeira ordem. Essas soluções são mínimos locais, pois sua densidade de energia livre é maior que aquela dos estados de vidro de spin nessa região. Na medida em que o valor de $\alpha$ decresce, para uma dada temperatura, a densidade de energia livre desses estados de recuperação vai diminuindo, até tornar-se menor que a densidade de energia livre dos estados de vidro de spin, o que acontece à esquerda da linha $T_c$, sobre a qual se observa a igualdade como mostra a figura 2.5. Assim sendo, na região F, a solução de recuperação é mínimo global, e a solução de vidro de spin é mínimo local. No detalhe, observa-se a linha de Almeida-Thouless $T_R$

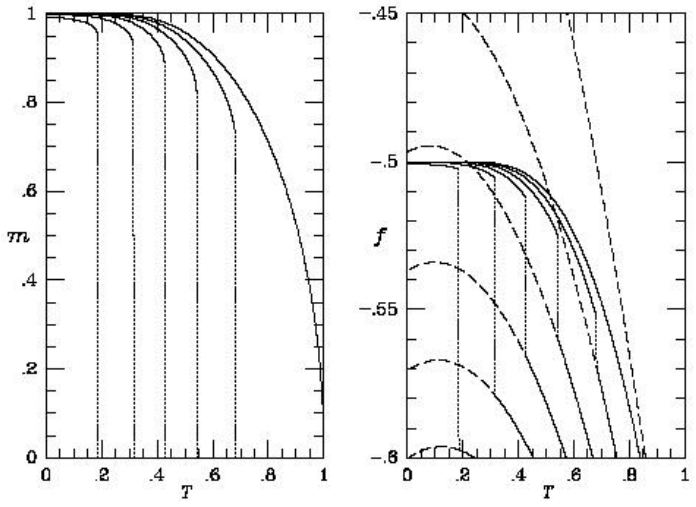



**Fig. 2.5:** *A esquerda: superposição m dos estados de recuperação em função da temperatura para* $\alpha = 0$, 0.025, 0.05, 0.075, 0.1 *e* 0.125, *da direita para a esquerda respectivamente. A direita: densidade de energia livre para os estados de recuperação (linhas contínuas) e estados de vidro de spin (linhas tracejadas) para* $\alpha = 0$, 0.025, 0.05, 0.075, 0.1 *e* 0.125 *da direita para a esquerda, respectivamente* [28].

abaixo da qual as soluções com simetria de réplicas deixam de ser estáveis.

Amit, Gutfreund e Sompolinsky [14] realizaram simulações de Monte Carlo [29] a $T = 0$ para redes de 3000 neurônios e obtiveram um valor para a capacidade crítica de armazenamento da rede $\alpha_c = 0.145 \pm 0.01$, que é levemente maior que o valor obtido analiticamente na aproximação de campo médio com simetria de réplicas $\alpha_c = 0.138$. Simulações de Monte Carlo, realizadas por Kohring [30] em redes com 30000 neurônios, determinaram como capacidade máxima de armazenamento $\alpha_c = 0.143 \pm 0.001$, enquanto que os resultados obtidos por Striefvater, Muller e Kühn [31] indicam $\alpha_c = 0.141 \pm 0.0015$ e Volk [32] $\alpha_c = 0.143 \pm 0.002$. Crisanti, Amit e Gutfreund [33] realizaram o primeiro passo da quebra da simetria de réplicas segundo o método de Parisi [34] para $T = 0$, obtendo $\alpha_c^{1RSB} = 0.144$ e $m(\alpha_c) = 0.983$. Posteriormente, Steffan e Kühn [35] obtiveram para o primeiro passo da quebra de simetria de réplicas $\alpha_c^{1RSB} \simeq 0.138186$ e para o segundo passo $\alpha_c^{2RSB} \simeq 0.138187$, que são ligeiramente superiores a $\alpha_c^{RS} \simeq 0.137905$ para simetria de réplicas. Segundo Györ-

gyi [36], a determinação de $\alpha_c$ com quebra de simetria de réplicas é ainda uma questão em aberto.

## 2.5 Padrões Correlacionados

Possivelmente um dos aspectos mais indesejáveis, se não o mais, é a dificuldade do modelo de Hopfield em armazenar padrões correlacionados. Na análise acima, os padrões são escolhidos com igual probabilidade de terem um neurônio ativo ou inativo. Essa escolha implica uma atividade média de 50%, ou seja, os padrões a serem armazenados possuem a metade dos neurônios ativos e a metade inativos sendo, portanto, ortogonais. Essa situação não é plausível do ponto de vista biológico, pois se observa que, em média, apenas 5% dos neurônios estão ativos em nível de atividade cortical. Além disso, o cérebro armazena uma grande quantidade de informações correlacionadas. A contemplação dessa condição biológica leva à necessidade de armazenar padrões correlacionados. Para tratar desse aspecto, Amit, Gutfreund e Sompolinsky [15] estudaram a capacidade de memória associativa para padrões aleatórios cuja atividade média diferia de 50%, escolhendo-os de acordo com a seguinte distribuição de probabilidade

$$P(\xi_i^\mu) = \frac{1+a}{2}\delta(\xi_i^\mu - 1) + \frac{1-a}{2}\delta(\xi_i^\mu + 1), \tag{2.92}$$

onde $a = \langle \xi_i^\mu \rangle$ é a tendência dos padrões. Os resultados obtidos com a regra de aprendizagem de Hebb (equação (2.25)) indicam que, para uma atividade cortical de 5%, apenas dois padrões poderiam ser armazenados para serem perfeitamente recuperados. Na tentativa de superar essa limitação severa, introduziu-se uma regra de aprendizagem modificada

$$\bar{J}_{ij} = \frac{1}{N}\sum_\mu (\xi_i^\mu - a)(\xi_j^\mu - a). \tag{2.93}$$

Embora essa regra seja não local, ela permite o armazenamento de um número extensivo de padrões, proporcional a $N$. Porém, o número de estados espúrios é o dobro daquele encontrado para o caso de padrões não correlacionados. Além disso, em $T = 0$, os estados simétricos com número par de padrões passam a ser estáveis juntamente com os estados

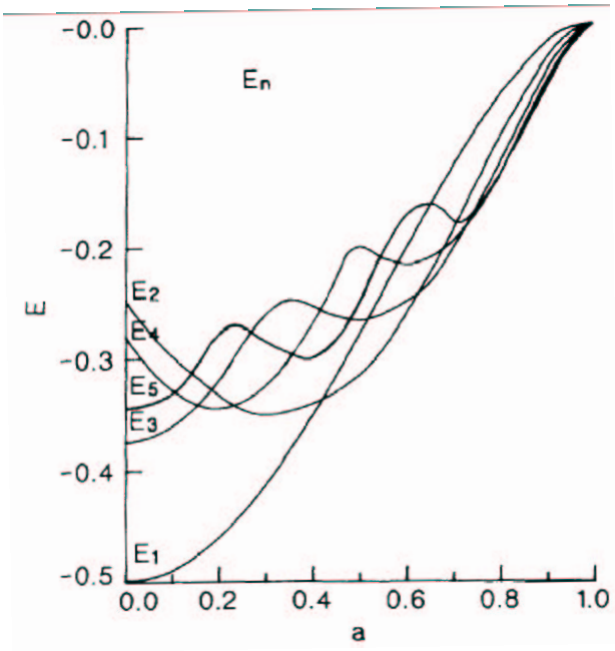



**Fig. 2.6:** *Energia dos primeiros cinco estados simétricos a $T = 0$, para um número finito de padrões armazenados, em função de $a$* [15].

simétricos ímpares. A medida em que $|a|$ aumenta, as energias desses estados começam a se cruzar e, quando $|a| > \sqrt{2} - 1$, tornam-se mínimos globais da densidade de energia livre como mostra a figura 2.6. Para temperatura finita ($T \neq 0$), apenas a baixas temperaturas os estados de recuperação são estáveis, embora os estados espúrios com número par de componentes possuam menor energia.

Quando o número de padrões armazenados aumenta com $N$ como $p = \alpha N$, a análise de sinal-ruído leva à conclusão de que os padrões serão recuperados sem erro se

$$\alpha < \frac{(1 - |a|)^2}{2 \ln N}. \tag{2.94}$$

A solução de campo médio conduz a um valor crítico para a capacidade de armazenamento $\alpha = \alpha_c(a)$. Abaixo de $\alpha_c$, existe um estado de recuperação dinamicamente estável ($m \neq 0$), enquanto que acima de $\alpha_c$, somente o estado de vidro de spin ($m = 0$) é estável, como mostra a figura 2.7.

A principal virtude desse modelo é que ele mantém a simplicidade da regra de aprendizagem. Em particular, a atualização sináptica devido à aprendizagem de novos padrões

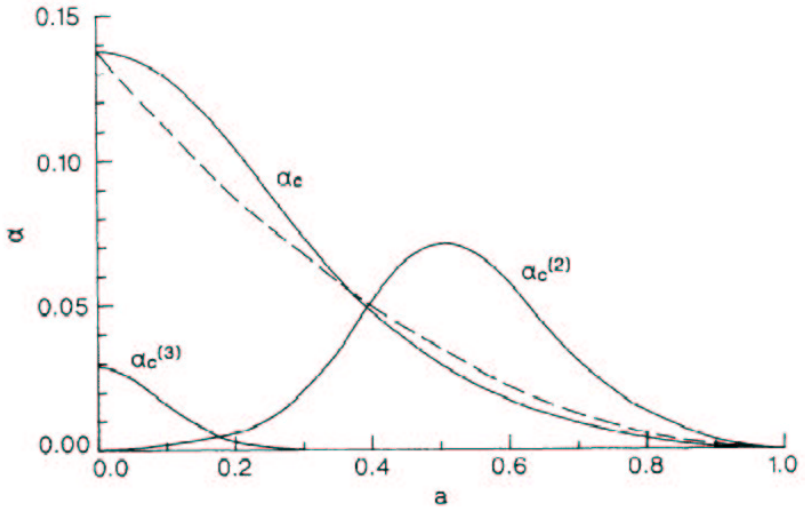



**Fig. 2.7:** *Diagrama $\alpha_c \times a$ para estados de recuperação, estados simétricos com dois e três padrões e previsão da análise sinal-ruído (linha tracejada)* [15].

continua a depender apenas dos neurônios $i$ e $j$, além da tendência $a$ dos padrões.

Capítulo 3

# Categorização com Ruído no Modelo de Hopfield

## 3.1 Introdução

Vivemos em um ambiente que nos bombardeia com uma quantidade imensa de informações, cuja manipulação adequada é um fator extremamente importante na sobrevivência da espécie humana. Classificar as informações de acordo com as categorias a que pertencem, organizando-as de maneira hierárquica, diminuindo assim as limitações da memória associativa, é uma das atividades cognitivas mais básicas. A categorização, que podemos conceituar como o processo através do qual entidades distintas são tratadas como equivalentes, é fundamental no entendimento dos objetos e eventos que acontecem ao nosso redor. Indivíduos capazes de categorizar as informações do ambiente com maior eficiência são capazes de se adaptarem mais facilmente que outros.

O problema da categorização é um dos temas centrais da ciência cognitiva, sendo ainda uma problema em aberto. Entretanto, os estudos já realizados permitem compreender de forma relativamente clara a maneira como as memórias estão organizadas [37]. Por exemplo, um indivíduo trata diferentes visões de um cão da raça poodle como exemplos do mesmo cão. Posteriormente, agrupa todos os cães da raça poodle dentro da categoria poodle. Depois, agrupa as raças poodles e dálmatas na categoria dos cães; cães e gatos na categoria de animais domésticos e assim sucessivamente. Percebe-se claramente uma organização

hierárquica das categorias.

Padrões pertencentes a uma mesma categoria estão correlacionados de modo que, numa situação real, é necessário o armazenamento de informações correlacionadas. O modelo de Hopfield, na sua forma original, possui grande limitação para tratar com padrões correlacionados. Para superar essa limitação, vários trabalhos propuseram regras de aprendizagem modificadas capazes de armazenar padrões estruturados de maneira hierárquica [38][39][40][41].

Uma abordagem diferente foi proposta por Fontanari [42][43]. Nessa, os estados simétricos, considerados espúrios no problema de recuperação de memórias, são vistos como representações para as categorias. Esses estados surgem espontaneamente e possuem uma correlação fixa com o conjunto de memórias pertencentes a uma determinada categoria, que foram armazenadas através da regra de Hebb não modificada. Desse modo, a rede é capaz de criar atratores que contenham as informações comuns às categorias.

Um dos vários fatores que podem influenciar a categorização é o ruído sináptico. Neste capítulo, buscamos compreender em que medida isso se dá. Para tanto, utilizamos a teoria de campo médio.

## 3.2 Organização Hierárquica de Padrões

Para modelar a habilidade de categorizar, construiremos uma estrutura de padrões aleatórios que apresentem uma organização hierárquica com vários níveis. A figura 3.1 ilustra a estrutura hierárquica. O primeiro nível da hierarquia é formado por um conjunto de $p_1$ padrões aleatórios $\{\xi_i^\mu\}$, com $\mu = 1, ..., p_1$, estatisticamente independentes e igualmente distribuídos, no sentido de que todos os padrões satisfazem à mesma distribuição de probabilidade, dada por

$$P(\xi_i^\mu) = \frac{1}{2}(1+a)\delta(\xi_i^\mu - 1) + \frac{1}{2}(1-a)\delta(\xi_i^\mu + 1), \tag{3.1}$$

onde, novamente, a delta de Kronecker $\delta(\xi_i^\mu \mp 1) = 1$ ou 0, se $\xi_i^\mu = \pm 1$ respectivamente, e $-1 < a < 1$ é a tendência ("bias"). Desse modo, a média de cada padrão gerado por meio

dessa distribuição é

$$\langle \xi_i^\mu \rangle = \sum_{\xi_i^\mu \pm 1} \xi_i^\mu P(\xi_i^\mu) = \frac{1+a}{2} - \frac{1-a}{2} = a, \tag{3.2}$$

e a média do produto de dois padrões é

$$\langle \xi_i^\mu \xi_i^\nu \rangle = a^2 + (1-a^2)\delta_{\mu\nu}, \tag{3.3}$$

onde $\delta_{\mu\nu} = 1$ ou 0 se $\mu = \nu$ ou $\mu \neq \nu$, respectivamente.

O segundo nível da hierarquia é gerado a partir do primeiro, introduzindo-se um novo parâmetro de superposição $0 < b < 1$. Geramos $p_2$ padrões estatisticamente independentes e igualmente distribuídos $\{\xi_i^{\mu\nu}\}$, com $\nu = 1, ..., p_2$, para cada padrão $\{\xi_i^\mu\}$, com probabilidade

$$P(\xi_i^{\mu\nu}) = b_1 \delta(\xi_i^{\mu\nu} - 1) + b_2 \delta(\xi_i^{\mu\nu} + 1), \tag{3.4}$$

onde

$$b_1 = \frac{1}{2}(1 + \xi_i^\mu b) \tag{3.5}$$

$$b_2 = \frac{1}{2}(1 - \xi_i^\mu b) \tag{3.6}$$

são as probabilidades que $\xi_i^{\mu\nu} = +1$, ou $\xi_i^{\mu\nu} = -1$, respectivamente.

Assim, a superposição entre os padrões do segundo nível é

$$\langle \xi_i^{\mu\nu} \xi_i^{\mu'\nu'} \rangle = [b^2 + (1-b^2)\delta_{\mu\mu'}\delta_{\nu\nu'}][a^2 + (1-a^2)\delta_{\mu\mu'}], \tag{3.7}$$

onde a média é calculada sobre os padrões $\{\xi_i^{\mu\nu}\}$ e $\{\xi_i^\mu\}$, nesta ordem. Se não há superposição entre dois conceitos, então

$$\langle \xi_i^{\mu\nu} \xi_i^{\mu'\nu'} \rangle = [b^2 + (1-b^2)\delta_{\nu\nu'}]\delta_{\mu\mu'}. \tag{3.8}$$

Pequenos valores de $b$ implicam baixa superposição entre exemplos e conceitos, tornando a criação de uma representação mais difícil, enquanto que valores maiores facilitam a categorização.

A superposição entre padrões do primeiro nível com padrões do segundo nível da estrutura hierárquica é dada por

$$\langle \xi_i^{\mu\nu} \xi_i^{\mu'} \rangle = b[a^2 + (1-a^2)\delta_{\mu\mu'}]. \tag{3.9}$$

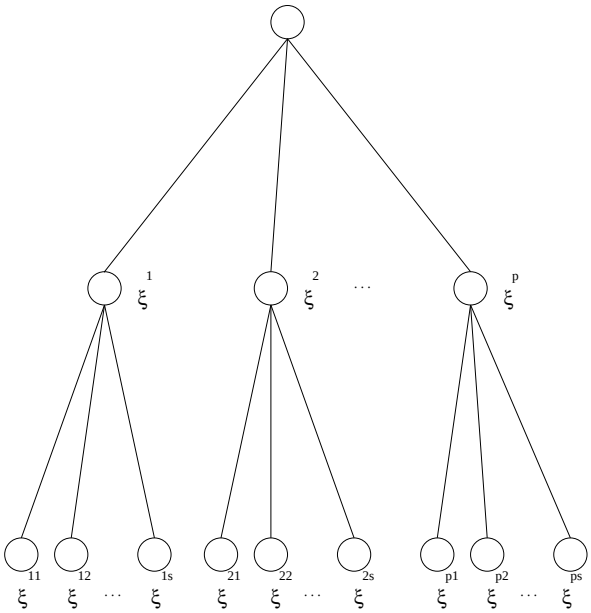



**Fig. 3.1:** *Representação gráfica de uma estrutura hierárquica com dois níveis. No primeiro nível temos p ancestrais* $\{\xi^{\mu}\}$ *e no segundo nível temos s descendentes* $\{\xi^{\mu\nu}\}$.

Essa forma de gerar os padrões agrupa-os em classes de modo que padrões pertencentes à mesma classe têm alta superposição, enquanto padrões pertencentes a classes diferentes têm baixa superposição. Por outra parte, a atividade dos padrões do segundo nível é dada por

$$\langle \xi_i^{\mu\nu} \rangle = \sum_{\xi_i^{\mu} \pm 1} P(\xi_i^{\mu}) [ \sum_{\xi_i^{\mu\nu} \pm 1} \xi_i^{\mu\nu} P(\xi_i^{\mu\nu}) ] = b \langle \xi_i^{\mu} \rangle. \tag{3.10}$$

Utilizando a equação (3.2), obtém-se o resultado

$$\langle \xi_i^{\mu\nu} \rangle = ab. \tag{3.11}$$

A geração de padrões dos próximos níveis segue o mesmo procedimento e encontra-se detalhada no apêndice A.

## 3.3 O Modelo Hierárquico de Dois Níveis

Considera-se, nesta seção, o modelo de Hopfield com estados binários $S_i = \pm 1, i = 1, ..., N$, cuja dinâmica é governada pelo Hamiltoniano

$$H = -\frac{1}{2}\sum_{i \neq j} J_{ij} S_i S_j. \tag{3.12}$$

As informações armazenadas na rede são pontos fixos dessa dinâmica. Treinar a rede com exemplos significa que, durante o estágio de aprendizagem, um conjunto finito de $s$ exemplos $\{\xi_i^{\mu\nu}\}, \nu = 1, ..., s$ de cada conceito $\{\xi_i^{\mu}\}, \mu = 1, ..., p$ é apresentado e armazenado na rede, usando a regra de aprendizagem de Hebb generalizada

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p}\sum_{\nu=1}^{s} \xi_i^{\mu\nu}\xi_j^{\mu\nu} \qquad (i \neq j). \tag{3.13}$$

Assim como a regra de Hebb, essa regra possui duas características importantes que são: ser local, $J_{ij}$ depende apenas dos neurônios $i$ e $j$; ser aditiva, pois cada novo exemplo armazenado independe dos exemplos já armazenados, $J_{ij} \rightarrow J_{ij} + \frac{1}{N}\xi_i^{\mu\nu}\xi_j^{\mu\nu}$.

No contexto da estrutura hierárquica de dois níveis estudada nesta tese, os conceitos representam o primeiro nível. Por simplicidade, o presente estudo trata de conceitos sem tendência, ou seja, $a = 0$. A distribuição de probabilidade se reduz a

$$P(\xi_i^{\mu}) = \frac{1}{2}\delta(\xi_i^{\mu} - 1) + \frac{1}{2}\delta(\xi_i^{\mu} + 1) \tag{3.14}$$

e

$$\langle \xi_i^{\mu}\xi_i^{\nu} \rangle = \delta_{\mu\nu}. \tag{3.15}$$

Os exemplos representam o segundo nível da estrutura hierárquica, dados pela distribuição de probabilidades (3.4).

No âmbito das redes neurais, o problema da categorização consiste em investigar a capacidade da rede em criar uma representação para o conceito, tendo sido exposta apenas aos exemplos, durante o treinamento. Se os conceitos forem pontos fixos da dinâmica do Hamiltoniano $H$, então ela terá sido capaz de criar uma representação para o conceito.

Para caracterizar a performance da rede como um dispositivo capaz de categorizar, é necessário introduzir um parâmetro que informe quantitativamente a qualidade do reconhecimento do conceito. Isso é feito através do parâmetro de ordem $m^\mu$, que descreve a superposição entre o estado da rede $S_i$ e o conceito $\xi_i^\mu$

$$m^\mu = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu S_i, \tag{3.16}$$

para $\mu = 1, ..., p$, com $0 \leq m^\mu \leq 1$.

O reconhecimento de um conceito significa existir uma superposição finita entre o estado da rede e o conceito. Esse reconhecimento emerge espontaneamente na dinâmica da rede treinada com exemplos, sendo que a rede não esteve exposta aos conceitos. A medida do insucesso em reconhecer um conceito é dada pelo erro de categorização, que é definido como a distância de Hamming

$$\epsilon^\mu = \frac{1}{2}(1 - m^\mu), \tag{3.17}$$

entre o estado $S_i$ e o conceito $\xi^\mu$, onde $\mu = 1, ..., p$. Quanto maior $m^\mu$, tanto maior a capacidade de categorizar e, conseqüentemente, menor o erro de categorização.

## 3.4 Teoria de Campo Médio

As propriedades termodinâmicas do modelo são obtidas a partir da densidade de energia livre por neurônio, associada ao Hamiltoniano dado pela equação (3.12),

$$f = -\lim_{N\to\infty} \frac{1}{N\beta} \langle \ln Z \rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}, \tag{3.18}$$

sendo $\langle ... \rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}$ a média configuracional sobre exemplos e conceitos, nesta ordem.

A função de partição $Z$ é dada por

$$Z = Tr_S \exp(-\beta H), \tag{3.19}$$

onde o Hamiltoniano é

$$H = -\frac{1}{2N}\sum_{i\neq j}\sum_{\mu,\nu} \xi_i^{\mu\nu}\xi_j^{\mu\nu} S_i S_j, \tag{3.20}$$

com a regra de aprendizagem de exemplos (3.13).

Podemos reescrever o Hamiltoniano de forma mais conveniente, excluindo os termos em $i = j$, como

$$H = -\frac{1}{2N}\sum_{\mu\nu}(\sum_{i}\xi_i^{\mu\nu}S_i)^2 + \sum_{i\mu}h^{\mu}\xi_i^{\mu}S_i + \frac{ps}{2N} \tag{3.21}$$

onde o termo incluindo o campo auxiliar $h^{\mu}$, acoplado aos conceitos condensados, é introduzido para calcular a superposição $m^{\mu}$ no limite $h^{\mu} \to 0$, de acordo com

$$m^{\mu} = \frac{\partial f}{\partial h^{\mu}} \mid_{h^{\mu}=0} . \tag{3.22}$$

Observam-se dois regimes distintos no comportamento geral do modelo de Hopfield, quanto ao número de conceitos que se pretende representar na rede. O parâmetro que identifica os regimes, se o número de exemplos de cada conceito é finito, é o parâmetro de reconhecimento de conceitos, definido como a razão entre o número de conceitos e o número de neurônios que compõe a rede

$$\alpha = \frac{p}{N}. \tag{3.23}$$

Quando se deseja criar representações para um número finito de conceitos, a rede encontra-se no regime em que $\alpha = 0$, pois no limite termodinâmico, $N \to \infty$, a razão $\frac{p}{N}$ tende a zero. Entretanto, se o número de conceitos for proporcional ao número de neurônios, $p = \alpha N$, então, no limite termodinâmico tem-se $N \to \infty$, $p \to \infty$, porém o parâmetro de reconhecimento é finito e não nulo, $\alpha \neq 0$.

## 3.5 Número Finito de Conceitos

Para investigarmos o comportamento da rede na presença de ruído térmico no regime em que o número de conceitos é finito ($\alpha = 0$) [44], devemos obter a densidade de energia livre $f$ a partir da equação (3.18). Para tanto, introduzimos o Hamiltoniano (3.21) na expressão para a função de partição (3.19), que escrevemos como

$$Z = Tr_S \exp[-\beta(-\frac{1}{2N}\sum_{\mu\nu}(\sum_{i}\xi_i^{\mu\nu}S_i)^2 + \sum_{i\mu}h^{\mu}\xi_i^{\mu}S_i + \frac{ps}{2N})]. \tag{3.24}$$

O cálculo do traço torna-se possível mediante a transformação dada pela equação (2.33). Nesse caso, introduzimos *ps* integrais

$$\prod_{\mu\nu} \exp[\frac{\beta}{2N}(\sum_i \xi_i^{\mu\nu} S_i)^2] = \prod_{\mu\nu} \int_{-\infty}^{\infty} \frac{dm^{\mu\nu}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(m^{\mu\nu})^2 + \sqrt{\frac{\beta}{N}} \sum_i \xi_i^{\mu\nu} S_i m^{\mu\nu}], \quad (3.25)$$

sobre as variáveis auxiliares $m^{\mu\nu}$, definidas como

$$m^{\mu\nu} \equiv \frac{1}{N} \sum_i \xi_i^{\mu\nu} S_i. \quad (3.26)$$

Desste modo, a função de partição passa a ser

$$Z = e^{-\beta ps/2} Tr_S \prod_{\mu\nu} \int_{-\infty}^{\infty} \frac{dm^{\mu\nu}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(m^{\mu\nu})^2 + \sqrt{\frac{\beta}{N}} \sum_i \xi_i^{\mu\nu} S_i m^{\mu\nu} - \beta \sum_{i\mu} h^\mu \xi_i^\mu S_i]. \quad (3.27)$$

No limite termodinâmico, a energia livre deve ser uma quantidade extensiva; para tanto, fazemos a seguinte mudança de variável:

$$m^{\mu\nu} \to \sqrt{N\beta} m^{\mu\nu}, \quad (3.28)$$

o que nos permite reescrever a função de partição como

$$Z = (\frac{N\beta}{2\pi})^{ps/2} e^{-\beta ps/2} \int_{-\infty}^{\infty} \prod_{\mu\nu} dm^{\mu\nu} \exp[-\frac{N\beta}{2} \sum_{\mu\nu} (m^{\mu\nu})^2] Tr_S \exp[\beta S_i (\sum_{i\mu\nu} m^{\mu\nu} \xi_i^{\mu\nu} - \sum_{i\mu} h^\mu \xi^\mu)]. \quad (3.29)$$

Podemos, agora, calcular facilmente o traço da exponencial linearizada,

$$Tr_S \exp[\beta S_i (\sum_{i\mu\nu} m^{\mu\nu} \xi_i^{\mu\nu} - \sum_{i\mu} h^\mu \xi^\mu)] = \exp\{\sum_i \ln[2\cosh[\beta(\sum_{\mu\nu} m^{\mu\nu} \xi_i^{\mu\nu} - \sum_\mu h^\mu \xi^\mu)]]\}, \quad (3.30)$$

e a função de partição toma a seguinte forma

$$Z = (\frac{N\beta}{2\pi})^{ps/2} e^{-\beta ps/2} \int_{-\infty}^{\infty} \prod_{\mu\nu} dm^{\mu\nu} \exp[-N\beta G(\{m\}\{\xi^{\mu\nu}\}\{\xi^\mu\})], \quad (3.31)$$

onde

$$G(\{m\}\{\xi^{\mu\nu}\}\{\xi^\mu\}) = \frac{1}{2} \sum_{\mu\nu} (m^{\mu\nu})^2 - \frac{1}{N\beta} \sum_i \ln[2\cosh[\beta(\sum_{\mu\nu} m^{\mu\nu} \xi_i^{\mu\nu} - \sum_\mu h^\mu \xi^\mu)]]. \quad (3.32)$$

No limite termodinâmico, podemos utilizar o método do ponto de sela para calcular as integrais em $m^{\mu\nu}$, que serão dominadas pelos mínimos da função $G(\{m\}\{\xi^{\mu\nu}\}\{\xi^{\mu}\})$, resultando

$$Z = (\frac{N\beta}{2\pi})^{ps/2} e^{-\beta ps/2} \exp[-N\beta G(\{m\}\{\xi^{\mu\nu}\}\{\xi^{\mu}\})], \tag{3.33}$$

sendo os valores de $m^{\mu\nu}$ aqueles obtidos da condição de extremo

$$\frac{\partial}{\partial m^{\mu\nu}} G(\{m\}\{\xi^{\mu\nu}\}\{\xi^{\mu}\}) = 0. \tag{3.34}$$

O extremo é dado por

$$\min_{m} \; G(\{m\}\{\xi^{\mu\nu}\}\{\xi^{\mu}\}) = \frac{1}{2}\sum_{\mu\nu}(m^{\mu\nu})^2 - \frac{1}{N\beta}\sum_{i} \ln[2\cosh[\beta(\sum_{\mu\nu} m^{\mu\nu}\xi_i^{\mu\nu} - \sum_{\mu} h^{\mu}\xi^{\mu})]]. \tag{3.35}$$

Utilizando a propriedade de automediação, podemos substituir a soma sobre sítios pela média configuracional

$$\frac{1}{N}\sum_{i}(\cdots) = \langle\cdots\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}, \tag{3.36}$$

e a densidade de energia livre passa a ser escrita como

$$f = \frac{1}{2}\sum_{\mu\nu}(m^{\mu\nu})^2 - \frac{1}{\beta}\langle \ln\, 2\, \cosh[\beta(\sum_{\mu\nu} m^{\mu\nu}\xi^{\mu\nu} - \sum_{\mu} h^{\mu}\xi^{\mu})]\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}. \tag{3.37}$$

A superposição com os exemplos, definida na equação (3.26), é obtida a partir das equações de ponto de sela

$$\frac{\partial f}{\partial m^{\mu\nu}}|_{h^{\mu}=0} = 0, \tag{3.38}$$

resultando

$$m^{\mu\nu} = \langle \xi^{\mu\nu} \tanh(\beta\sum_{\rho\sigma} m^{\rho\sigma}\xi^{\rho\sigma})\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}. \tag{3.39}$$

A superposição com os conceitos obtida da equação (3.22) é dada por

$$m^{\mu} = \langle \xi^{\mu} \tanh(\beta\sum_{\rho\sigma} m^{\rho\sigma}\xi^{\rho\sigma})\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}. \tag{3.40}$$

A equação (3.39) admite soluções que podem ser caracterizadas como soluções simétricas ou assimétricas no sentido da discussão apresentada na seção 2.4.1, representadas

pelas equações (2.45) e (2.51). No problema de categorização, estamos interessados em soluções que nos permitam investigar o comportamento da rede como um dispositivo capaz de criar representações para conceitos. Nesse sentido, as soluções de recuperação de exemplos (assimétrica) e de categorização (simétrica) são as soluções de interesse, como veremos a seguir.

### 3.5.1 Soluções de recuperação de exemplos

Consideramos uma classe particular de soluções da forma $m^{\mu\nu} = m^{1\nu}\delta_{\mu 1}$, com $m^{1\nu}$ dada por [43]

$$m^{1\nu} = m^{11}\delta_{1\nu} + (1 - \delta_{1\nu})m_{s-1}, \tag{3.41}$$

que diferencia entre a memória associativa para um exemplo qualquer, aqui o exemplo $\xi^{11}$ e a memória associativa com a combinação simétrica dos $s-1$ exemplos restantes do conceito $\mu = 1$. A superposição simétrica é definida por

$$m_{s-1} = \frac{1}{(s-1)N}\sum_i x_{s-1}S_i, \tag{3.42}$$

com $x_{s-1} = \sum_{\nu>1}\xi^{1\nu}$. Esse tipo de solução assegura que qualquer outro comportamento da rede será uma propriedade espontânea da rede e não uma conseqüência de uma classe particular de soluções escolhidas. Como todos os exemplos são equivalentes, qualquer um pode ser escolhido como o exemplo $\nu = 1$. As equações para $m^{11}$ e $m_{s-1}$ são

$$m^{11} = \langle \xi^{11}\ \tanh[\beta(m^{11}\xi^{11} + m_{s-1}x_{s-1})]\rangle_{\{\xi^{11}\}\{x_{s-1}\}\{\xi^1\}}, \tag{3.43}$$

$$m_{s-1} = \frac{1}{s-1}\langle x_{s-1}\ \tanh[\beta(m^{11}\xi^{11} + m_{s-1}x_{s-1})]\rangle_{\{\xi^{11}\}\{x_{s-1}\}\{\xi^1\}}, \tag{3.44}$$

sendo a média calculada sobre $\xi^{11}$, $x_{s-1}$ e $\xi^1$, nessa ordem. A combinação simétrica dos exemplos, $x_{s-1}$, é uma variável aleatória discreta cuja distribuição de probabilidade é a

distribuição binomial

$$P(x_{s-1}) = \begin{pmatrix} s-1 \\ k \end{pmatrix} b_1^k b_2^{s-1-k}, \tag{3.45}$$

sendo $k = \frac{1}{2}(x_{s-1} + s - 1)$ o número de exemplos $\xi_i^{1\nu} = +1$ e $s - 1 - k = \frac{1}{2}(s - x_{s-1})$ o número de exemplos $\xi_i^{1\nu} = -1$. A equação para a superposição com o conceito $\xi^1$, definida na equação (3.16), para a classe de soluções considerada é

$$m^1 = \langle \xi^1 \tanh[\beta(m^{11}\xi^{11} + m_{s-1}x_{s-1})] \rangle_{\{\xi^{11}\}\{x_{s-1}\}\{\xi^1\}}. \tag{3.46}$$

A razão para termos mantido a distribuição binomial em vez da distribuição Gaussiana, como foi feito na referência [43], é que a distribuição binomial permite investigar o comportamento da rede para um número arbitrário, não necessariamente grande, de exemplos $s$.

Calculando a média das equações (3.43) e (3.44), cujos detalhes encontram-se no apêndice B, obtemos

$$m^{11} = \frac{1}{2}\sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} \left[P_1(k)\tanh(\beta\Lambda_+) + P_1(k-1)\tanh(\beta\Lambda_-)\right], \tag{3.47}$$

$$\begin{aligned} m_{s-1} &= \frac{1}{2(s-1)}\sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} (2k - s + 1) \\ &\quad \times \left[P_1(k)\tanh(\beta\Lambda_+) - P_1(k-1)\tanh(\beta\Lambda_-)\right], \end{aligned} \tag{3.48}$$

onde

$$P_1(k) = \left(\frac{1+b}{2}\right)^{k+1}\left(\frac{1-b}{2}\right)^{s-1-k} + \left(\frac{1+b}{2}\right)^{s-1-k}\left(\frac{1-b}{2}\right)^{k+1} \tag{3.49}$$

e

$$\Lambda_\pm = m^{11} \pm m_{s-1}(2k - s + 1). \tag{3.50}$$

Da mesma forma, calculamos a superposição de categorização $m^1$ a partir da equação (3.46), resultando

$$m^1 = \frac{1}{2}\sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} \left[P_2(k)\tanh(\beta\Lambda_+) + P_2(s-k-1)\tanh(\beta\Lambda_-)\right], \tag{3.51}$$

onde

$$P_2(k) = \left(\frac{1+b}{2}\right)^{k+1}\left(\frac{1-b}{2}\right)^{s-1-k} - \left(\frac{1+b}{2}\right)^{s-1-k}\left(\frac{1-b}{2}\right)^{k+1}. \tag{3.52}$$

A densidade de energia livre para esta classe de soluções é dada por

$$\begin{aligned} f_r &= \frac{1}{2}[(m^{11})^2 + (s-1)m_{s-1}^2] \\ &- \frac{1}{2\beta}\sum_{k=0}^{s-1}\begin{pmatrix} s-1 \\ k \end{pmatrix}[P_1(k)\ln[2\cosh(\beta\Lambda_+)] + P_1(k-1)\ln[2\cosh(\beta\Lambda_-)]]\,. \end{aligned} \tag{3.53}$$

**Limite de ruído nulo ($T = 0$)**

No limite de ruído nulo em que $\beta \to \infty$, a tangente hiperbólica comporta-se assintoticamente como a função sinal, $\lim_{\beta\to\infty}\tanh(\beta\Lambda) = \mathrm{sgn}(\Lambda)$. Podemos, então, escrever as equações (3.47), (3.48) e (3.51) como

$$m^{11} = \frac{1}{2}\sum_{k=0}^{s-1}\begin{pmatrix} s-1 \\ k \end{pmatrix}[P_1(k)\,\mathrm{sgn}(\Lambda_+) + P_1(k-1)\,\mathrm{sgn}(\Lambda_-)]\,, \tag{3.54}$$

$$\begin{aligned} m_{s-1} &= \frac{1}{2(s-1)}\sum_{k=0}^{s-1}\begin{pmatrix} s-1 \\ k \end{pmatrix}(2k-s+1) \\ &\quad \times [P_1(k)\,\mathrm{sgn}(\Lambda_+) - P_1(k-1)\,\mathrm{sgn}(\Lambda_-)] \end{aligned} \tag{3.55}$$

e

$$m^1 = \frac{1}{2}\sum_{k=0}^{s-1}\begin{pmatrix} s-1 \\ k \end{pmatrix}[P_2(k)\,\mathrm{sgn}(\Lambda_+) + P_2(s-k-1)\,\mathrm{sgn}(\Lambda_-)]\,. \tag{3.56}$$

A densidade de energia livre é dada por (ver apêndice B)

$$f_r = -\frac{1}{2}[(m^{11})^2 + (s-1)m_{s-1}^2], \tag{3.57}$$

pois $\lim_{\beta\to\infty}\frac{1}{\beta}\ln[2\cosh(\beta\Lambda)] = |\Lambda|$. Note que, nesse caso, a densidade de energia livre é a energia do sistema por neurônio $H/N$, pois o Hamiltoniano dado pela equação (3.21) pode ser escrito em termos das superposições $m^{\mu\nu}$, sendo igual ã equação acima.

### 3.5.2 Soluções de categorização

Outra classe de soluções consideradas são as superposições simétricas $m^{\mu\nu} = m_s\delta_{\mu 1}, \forall\nu$, com todos os exemplos de um conceito dado, definidas por

$$m_s = \frac{1}{s}\sum_i x_s S_i, \tag{3.58}$$

onde $x_s = \sum_{\nu=1}^{s}\xi^{1\nu}$. Nesse caso, a equação para $m_s$ é

$$m_s = \frac{1}{s}\langle x_s \tanh(\beta m_s x_s)\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}, \tag{3.59}$$

onde $x_s$ é novamente uma variável aleatória cuja distribuição de probabilidades é a distribuição binomial

$$P(x_s) = \left(\begin{array}{c} s \\ k \end{array}\right) b_1^k b_2^{s-k} \tag{3.60}$$

com $k = \frac{1}{2}(x_s + s)$. As equações para as superposições com os exemplos $m_s$ e o conceito $m^1$ são, respectivamente,

$$m_s = \frac{1}{2s}\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right)(2k-s)P_+(k)\tanh[\beta m_s(2k-s)] \tag{3.61}$$

e

$$m^1 = \frac{1}{2}\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right)P_-(k)\tanh[\beta m_s(2k-s)] \tag{3.62}$$

onde

$$P_\pm(k) = \left(\frac{1+b}{2}\right)^k\left(\frac{1-b}{2}\right)^{s-k} \pm \left(\frac{1+b}{2}\right)^{s-k}\left(\frac{1-b}{2}\right)^k. \tag{3.63}$$

Para essa classe de soluções, a densidade de energia livre é dada por

$$f_c = \frac{1}{2}sm_s^2 - \frac{1}{2\beta}\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right)P_+(k)\ln 2\cosh[\beta m_s(2k-s)]. \tag{3.64}$$

**Limite de ruído nulo ($T = 0$)**

Novamente utilizamos o comportamento assintótico da tangente hiperbólica para obtermos as equações para as superposições e a densidade de energia livre nesse limite. Escrevemos, então, as equações (3.61), (3.62) e (3.64) como

$$m_s = \frac{1}{2s} \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} (2k - s) P_+(k) \operatorname{sgn}(m_s x_s), \tag{3.65}$$

$$m^1 = \frac{1}{2s} \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_-(k) \operatorname{sgn}(m_s x_s) \tag{3.66}$$

e

$$f_c = -\frac{1}{2} s m_s^2. \tag{3.67}$$

### 3.5.3 Resultados Numéricos

A partir da equação (3.51), obtemos o erro de categorização

$$\epsilon = (1 - m^1)/2. \tag{3.68}$$

Resolvemos numericamente as equações (3.47), (3.48) e (3.51) para vários valores de temperatura e calculamos a energia livre correspondente em cada caso para analisar a estabilidade das soluções [44].

A figura 3.2 apresenta o diagrama de fases obtido para $b = 0.2$ e $b = 0.25$. A fase de recuperação (R) corresponde a soluções assimétricas em que a rede recupera um exemplo em particular, de modo que $m^{11} \neq m_{s-1}$. Para baixas temperaturas $m^{11} \approx 1$, $m_{s-1} \approx b^2$ e $m_1 \approx b$, correspondendo a um erro de categorização $\epsilon \approx (1 - b)/2$. Na fase de recuperação , existe também uma solução simétrica $m^{11} = m_{s-1}$, que é a solução mais estável na região (S) à direita da linha pontilhada na figura 3.2, para $b = 0.25$, enquanto a solução de recuperação é a solução mais estável na região (R) à esquerda da linha pontilhada. A linha pontilhada é construída a partir da igualdade das densidades de energia livre para as soluções de recuperação (linha tracejada) e categorização (linha contínua), como ilustrado na figura 3.3

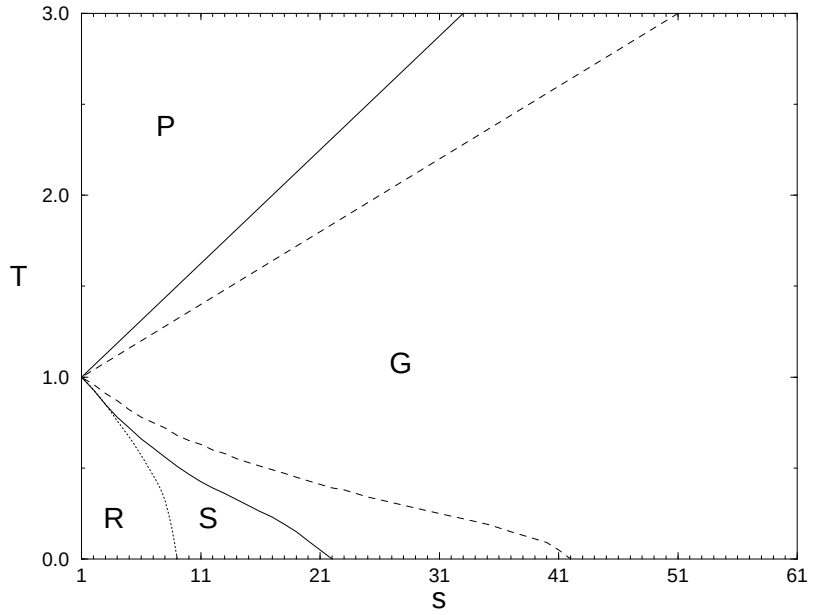



**Fig. 3.2:** *Diagrama de fases ($T \times s$) para $\alpha = 0$ e $b = 0.25$ (linha contínua) e $b = 0.20$ (linha tracejada). A linha pontilhada separa as regiões de estabilidade global para a recuperação de exemplos (R) e estados de mistura simétricos (S) para $b = 0.25$.*

(a). Na figura 3.3 (b), observamos as superposições $m^{11}$ (linha pontilhada), $m_{s-1}$ (linha tracejada), correspondentes à solução de recuperação e a superposição de categorização $m_s$ (linha contínua) para os casos indicados na figura. Existe uma linha de transição de fase de primeira ordem que define um número crítico de exemplos $s_c$, acima do qual a rede passa a categorizar. Na fase de categorização (G), a rede não recupera um exemplo particular mas passa a reconhecer os aspectos comuns dos exemplos na forma de estados de mistura simétricos $m^{11} = m_{s-1} = m_s$.

Aumentando a temperatura, encontramos uma fase paramagnética (P), onde a rede nem categoriza nem recupera os exemplos ($m^{11} = m_{s-1} = m_s = 0$). A transição entre as fases paramagnética e categorização é de segunda ordem, e as soluções encontradas na fase de categorização são soluções simétricas $m_s$. A linha de transição é dada pela temperatura de ordenamento que é obtida a partir da expansão da solução simétrica (3.59) em potências de $m_s$, até primeira ordem

$$m_s = \frac{1}{s}\beta m_s \langle x_s^2 \rangle. \tag{3.69}$$

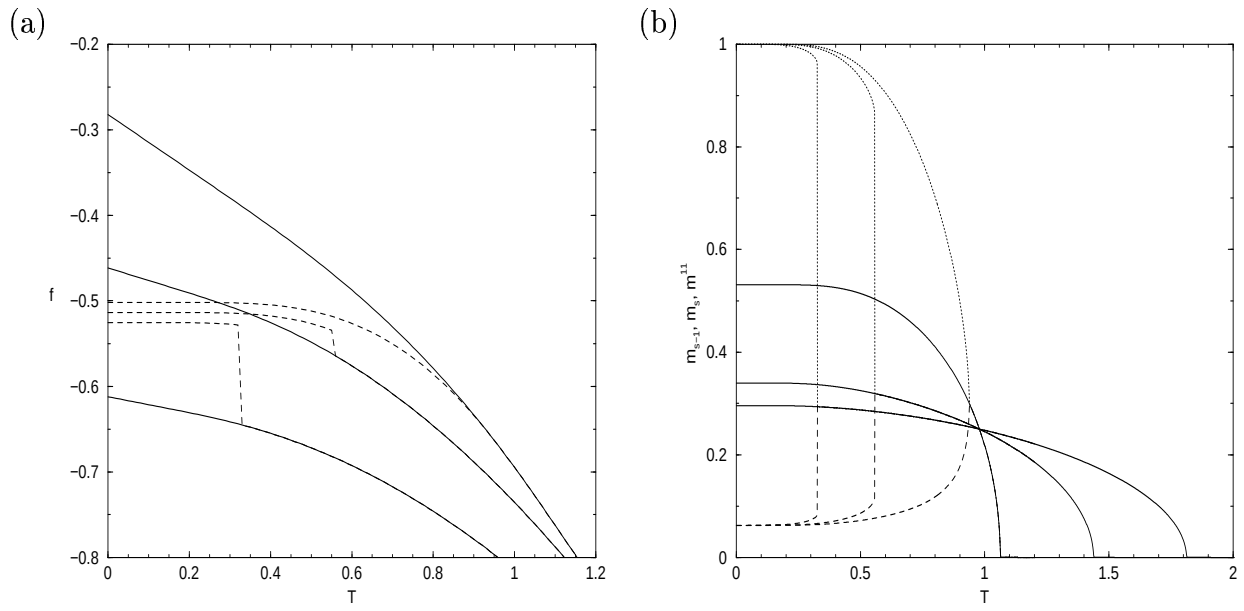



**Fig. 3.3:** *(a) Densidade de energia livre para as soluções de recuperação (linha tracejada) e de categorização (linha contínua) para $s = 2$, $s = 8$ e $s = 14$ da direita para a esquerda respectivamente, e (b) superposições $m_{s-1}$, $m_s$ e $m^{11}$, respectivamente.*

Após o cálculo da média, obtemos a temperatura para a qual observamos o aparecimento da fase de categorização,

$$T_G(b) = 1 + (s-1)b^2. \tag{3.70}$$

Outro aspecto interessante observado é que a fase de categorização cresce enquanto a fase paramagnética decresce, à medida que aumentamos a superposição entre os exemplos e os conceitos, como esperado.

Na figura 3.4, mostramos as curvas de categorização para $b = 0.2$ e vários valores de temperatura. Nela, observamos que a transição para a fase de categorização é descontínua, diferentemente do que se observa na rede de neurônios analógicos [45]. O aumento da temperatura (ruído sináptico) inicialmente favorece a categorização, no sentido de que é necessário um número menor de exemplos, para que a rede entre na fase de categorização. Esse comportamento é observado até a temperatura $T_G = 1$. Ao aumentarmos a temperatura acima desse valor, a categorização deixa de ser favorecida, sendo necessário um número cada vez

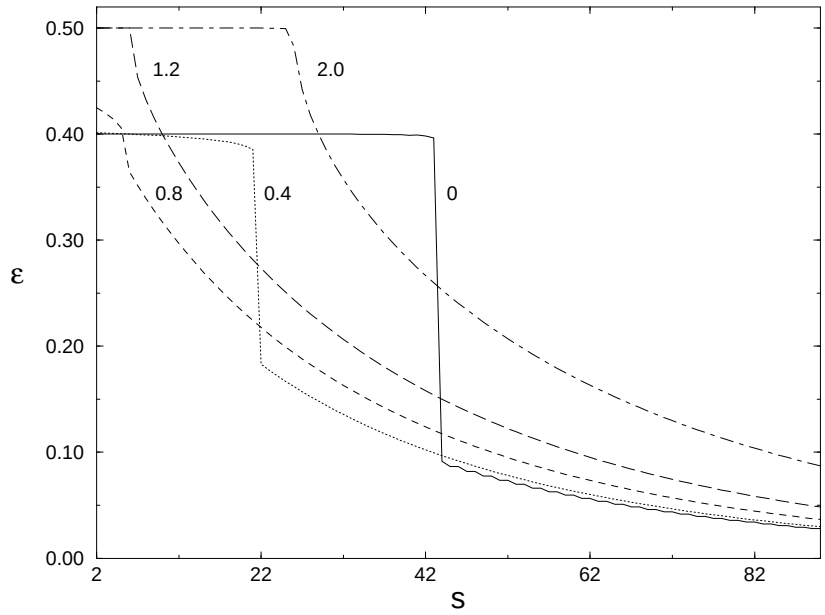



**Fig. 3.4:** *Curvas de categorização em função do número de exemplos para $\alpha = 0$, $b = 0.2$ e para vários valores de temperatura.*

maior de exemplos para se atingir a fase de categorização. Apesar desse desfavorecimento, é sempre possível levar a rede até a fase de categorização, bastando, para isso, que se treine a rede com um número suficientemente grande de exemplos. Outro aspecto interessante é a possibilidade de controlar, com a temperatura, a recuperação do exemplo ou do conceito.

## 3.6 Número Macroscópico de Conceitos

No caso em que $\alpha = p/N$ é finito no limite termodinâmico $N \to \infty$ [44], a rede pode criar um número extensivo de representações (conceitos) tendo acesso apenas a um número finito de exemplos. Isso introduz duas questões que até agora não estavam presentes. Como podemos observar na equação (3.21), expressa em termos das superposições (3.26) na ausência de campo externo

$$H = -\frac{1}{2}\sum_{\mu=1}^{p}\sum_{\nu=1}^{s}(m^{\mu\nu})^2 + \frac{\alpha s}{2}, \tag{3.71}$$

a soma sobre $\mu$ se dá para um número macroscópico de conceitos ($p \to \infty$) . Para que a energia tenha um limite inferior, é necessário que os estados de equilíbrio (atratores) da

rede possuam superposições $m^{\mu\nu}$ da ordem $\mathcal{O}(1)$ com um número finito de conceitos, ditos condensados. Para os demais conceitos, ditos não condensados, as superposições $m^{\mu\nu}$ devem ser da ordem $\mathcal{O}(\frac{1}{\sqrt{N}})$. O outro aspecto importante é que as grandezas físicas de interesse não satisfarão a propriedade da automediação como no caso $\alpha = 0$. Apenas as grandezas correspondentes aos conceitos condensados satisfarão a propriedade da automediação. Isso implica calcular explicitamente a média configuracional da energia livre (3.18), que envolve o logaritmo da função de partição. Para tanto, utilizaremos o método das réplicas [14], que está baseado na seguinte identidade matemática:

$$\ln x = \lim_{n\to 0} \frac{x^n - 1}{n}. \tag{3.72}$$

Escrevemos, assim, a densidade de energia livre em termos das réplicas

$$f = -\lim_{N\to\infty} \lim_{n\to 0} \frac{\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} - 1}{nN\beta}, \tag{3.73}$$

onde $\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}$ é a média térmica e configuracional da função de partição replicada $n$ vezes,

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} = \langle Tr_{S^\rho} \exp(-\beta \sum_{\rho=1}^{n} H^\rho) \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}, \tag{3.74}$$

sendo

$$H^\rho = -\frac{1}{2N} \sum_{\mu\nu} (\sum_i \xi_i^{\mu\nu} S_i^\rho)^2 + \sum_{i\mu} h^\mu \xi_i^\mu S_i^\rho + \frac{\alpha s}{2} \tag{3.75}$$

o Hamiltoniano da $\rho$-ésima réplica. Podemos interpretar fisicamente cada réplica como uma das $n$ cópias idênticas do sistema.

Introduzindo a equação (3.75) na equação para a função de partição replicada (3.74), temos

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} = e^{-\beta p s/2} \langle Tr_{S^\rho} \exp(\frac{\beta}{2N} \sum_{\mu\nu\rho} (\sum_i \xi_i^{\mu\nu} S_i^\rho)^2 - \sum_{i\mu\rho} h^\mu \xi_i^\mu S_i^\rho] \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}. \tag{3.76}$$

Para linearizar o termo quadrático em $S_i^\rho$ utilizamos novamente a integral gaussiana (2.33) e a mudança de variável $m^{\mu\nu} \to \sqrt{N\beta} m^{\mu\nu}$, que nos permite escrever

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} = e^{-\beta p s n/2} (\beta N)^{\frac{psn}{2}} \langle Tr_{S^\rho} \int \prod_{\mu\nu\rho} \frac{dm_\rho^{\mu\nu}}{\sqrt{2\pi}}$$
$$\exp\{\beta N[-\frac{1}{2} \sum_{\mu\nu\rho} (m_\rho^{\mu\nu})^2 + \sum_{\mu\nu\rho} m_\rho^{\mu\nu} \frac{1}{N} \sum_i \xi_i^{\mu\nu} S_i^\rho - \frac{1}{N} \sum_{i\mu\rho} h^\mu \xi^\mu S_i^\rho]\} \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}. \tag{3.77}$$

Apenas os exemplos de um número finito de conceitos condensam, e não há superposição entre conceitos, sendo, portanto, todos equivalentes. Podemos, então, escolher o conceito $\mu = 1$ como o único que condensa, sem perder generalidade. Dessa forma, a função de partição pode ser separada em dois termos

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} = e^{-\beta psn/2}(\beta N)^{\frac{psn}{2}} Tr_{S^\rho} \left\{ \int \prod_{\nu\rho} \frac{dm_\rho^{1\nu}}{\sqrt{2\pi}} \right.$$

$$\left\langle \exp\{\beta N[-\frac{1}{2}\sum_{\nu\rho}(m_\rho^{1\nu})^2 + \sum_{\nu\rho} m_\rho^{1\nu}\frac{1}{N}\sum_i \xi_i^{1\nu}S_i^\rho - \frac{1}{N}\sum_{i\rho} h^1\xi^1 S_i^\rho]\}\right\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}$$

$$\times \quad \left. \int \prod_{\substack{\mu>1\\ \nu\rho}} \frac{dm_\rho^{\mu\nu}}{\sqrt{2\pi}} \left\langle \exp\{\beta N[-\frac{1}{2}\sum_{\substack{\mu>1\\ \nu\rho}}(m_\rho^{\mu\nu})^2 + \sum_{\substack{\mu>1\\ \nu\rho}} m_\rho^{\mu\nu}\frac{1}{N}\sum_i \xi_i^{\mu\nu}S_i^\rho]\}\right\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} \right\}, \quad (3.78)$$

onde a média da primeira exponencial é sobre os exemplos e conceitos condensados, enquanto que a média da segunda exponencial é sobre os exemplos e conceitos não condensados. Podemos reescrever a última linha da seguinte forma

$$\% = \int \prod_{\substack{\mu>1\\ \nu\rho}} \frac{dm_\rho^{\mu\nu}}{\sqrt{2\pi}} \exp[-\frac{\beta N}{2}\sum_{\substack{\mu>1\\ \nu\rho}}(m_\rho^{\mu\nu})^2] \left\langle \exp[\beta N \sum_{\substack{\mu>1\\ \nu\rho}} m_\rho^{\mu\nu}\frac{1}{N}\sum_i \xi_i^{\mu\nu}S_i^\rho]\right\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} . \quad (3.79)$$

Expandindo a segunda exponencial até segunda ordem,

$$\left\langle \exp[\beta N \sum_{\substack{\mu>1\\ \nu\rho}} m_\rho^{\mu\nu}\frac{1}{N}\sum_i \xi_i^{\mu\nu}S_i^\rho]\right\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} =$$

$$\prod_i [1 + \beta\sum_{\mu\nu\rho} m_\rho^{\mu\nu}\langle\xi_i^{\mu\nu}\rangle S_i^\rho + \frac{\beta^2}{2}\sum_{\substack{\mu\mu'\\ \nu\lambda\rho\sigma}} m_\rho^{\mu\nu}m_\sigma^{\mu\nu}\langle\xi_i^{\mu\nu}\xi_i^{\mu'\lambda}\rangle S_i^\rho S_i^\sigma], \quad (3.80)$$

e utilizando-se as equações (3.8) e (3.11), obtém-se

$$\left\langle \exp[\beta N \sum_{\substack{\mu>1\\ \nu\rho}} m_\rho^{\mu\nu}\frac{1}{N}\sum_i \xi_i^{\mu\nu}S_i^\rho]\right\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} = \prod_i \exp[\ln(1 + \frac{\beta^2}{2}\sum_{\substack{\mu>1\\ \nu\lambda\rho\sigma}} m_\rho^{\mu\nu}m_\sigma^{\mu\nu}B_{\nu\lambda}S_i^\rho S_i^\sigma], \quad (3.81)$$

onde

$$B_{\nu\lambda} = b^2 + (1 - b^2)\delta_{\nu\lambda}. \quad (3.82)$$

Após substituir $m^{\mu\nu} \to N^{-\frac{1}{2}}m^{\mu\nu}$ e considerando $\ln(1 + x) \approx x$, resulta

$$\% = \int \prod_{\substack{\mu>1\\ \nu\rho}} \frac{dm_\rho^{\mu\nu}}{\sqrt{2\pi}} \exp[-\frac{\beta}{2}\sum_{\substack{\mu>1\\ \nu\rho}}(m_\rho^{\mu\nu})^2 + \frac{\beta^2}{2}\sum_{\substack{\mu>1\\ \nu\lambda\rho\sigma}} m_\rho^{\mu\nu}m_\sigma^{\mu\lambda}B_{\nu\lambda}\frac{1}{N}\sum_i S_i^\rho S_i^\sigma]. \quad (3.83)$$

O termo $\sum_{\mu>1\,\nu\,\lambda} m_\rho^{\mu\nu} m_\sigma^{\mu\lambda}$ na segunda exponencial é da ordem de 1, embora $m_\rho^{\mu\nu}$ seja da ordem $1/\sqrt{N}$ para os conceitos não condensados. Para tratar desse termo, introduz-se uma variável auxiliar $q_{\rho\sigma}$ através das propriedades da delta de Dirac

$$1 = \int dq_{\rho\sigma}\ \delta(q_{\rho\sigma} - \frac{1}{N}\sum_i S_i^\rho S_i^\sigma). \tag{3.84}$$

Desse modo, podemos escrever

$$\begin{aligned} \% &= \prod_{\mu>1}\int \prod_{\rho<\sigma} dq_{\rho\sigma} \int \prod_{\nu\rho} \frac{dm_\rho^{\mu\nu}}{\sqrt{2\pi}} \prod_{\rho<\sigma} \delta(q_{\rho\sigma} - \frac{1}{N}\sum_i S_i^\rho S_i^\sigma) \\ &\times \exp[-\frac{\beta}{2}\sum_{\nu\rho}(m_\rho^{\mu\nu})^2 + \frac{\beta^2}{2}\sum_{\nu\lambda\rho\sigma} m_\rho^{\mu\nu} m_\sigma^{\mu\lambda} B_{\nu\lambda} Q_{\rho\sigma}], \end{aligned} \tag{3.85}$$

sendo

$$Q_{\rho\sigma} = q_{\rho\sigma} + (1 - q_{\rho\sigma})\delta_{\rho\sigma}. \tag{3.86}$$

Utilizando a representação integral da delta de Dirac

$$\delta(q_{\rho\sigma} - \frac{1}{N}\sum_i S_i^\rho S_i^\sigma) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} dr_{\rho\sigma}\ \ \exp[-r_{\rho\sigma}(q_{\rho\sigma} - \frac{1}{N}\sum_i S_i^\rho S_i^\sigma)], \tag{3.87}$$

e substituindo $m_{\mu\nu} \to y_{\rho\nu}/\sqrt{\beta}$, temos

$$\% = \int \prod_{\rho<\sigma} dq_{\rho\sigma} \int \prod_{\rho<\sigma} \frac{dr_{\rho\sigma}}{\sqrt{2\pi i}} \exp\{\beta N[\frac{\alpha}{\beta}\ln G(q_{\rho\sigma})]\} \exp[-\frac{1}{2}\sum_{\rho\neq\sigma} q_{\rho\sigma} r_{\rho\sigma} + \frac{1}{2N}\sum_i \sum_{\rho\neq\sigma} r_{\rho\sigma} S_i^\rho S_i^\sigma], \tag{3.88}$$

sendo

$$G(q_{\rho\sigma}) = \int \prod_{\rho\nu} \frac{dy_{\rho\nu}}{\sqrt{2\pi}} \exp[-\frac{1}{2}\sum_{\rho\nu} y_{\rho\nu}^2 + \frac{\beta}{2}\sum_{\rho\sigma\nu\lambda} y_{\rho\nu} y_{\sigma\lambda} B_{\nu\lambda} Q_{\rho\sigma}]. \tag{3.89}$$

Fazendo a mudança de variável $r_{\rho\sigma} \to \alpha N\beta^2 r_{\rho\sigma}$, desconsiderando fatores multiplicativos que não irão contribuir para a energia livre, e utilizando a propriedade da automediação para os conceitos condensados, escrevemos a função de partição como

$$\langle Z^n\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} = \int \prod_{\rho\nu} dm_\rho^{1\nu} \int \prod_{\rho<\sigma} dq_{\rho\sigma} dr_{\rho\sigma} \exp\{-N\beta H(\{m_\rho^{\mu\nu}, q_{\rho\sigma}, r_{\rho\sigma}\}) \tag{3.90}$$

onde

$$H(\{m_\rho^{\mu\nu}, q_{\rho\sigma}, r_{\rho\sigma}\}) = \frac{1}{2}\sum_{\rho\nu}(m_\rho^{1\nu})^2 + \frac{\alpha\beta}{2}\sum_{\rho\neq\sigma} q_{\rho\sigma} r_{\rho\sigma} - \frac{\alpha}{\beta}\ln G(q_{\rho\sigma}) - \frac{1}{\beta}\langle \ln Tr_{S^\rho} e^{\beta H_\xi}\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}$$

e

$$H_\xi = \sum_{\rho\nu} m_\rho^{1\nu}\xi^{1\nu}S^\rho + \frac{\alpha\beta}{2}\sum_{\rho\neq\sigma} r_{\rho\sigma}S^\rho S^\sigma - h^1\xi^1\sum_\rho S^\rho. \tag{3.92}$$

No limite termodinâmico ($N\to\infty$), o integrando é dominado pelos seus pontos de sela, de modo que a densidade de energia livre é (ver apêndice C)

$$f = \lim_{n\to 0}\left[\frac{1}{2n}\sum_{\rho\nu}(m_\rho^{1\nu})^2 + \frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} - \frac{\alpha}{n\beta}\ln G(q_{\rho\sigma}) - \frac{1}{n\beta}\langle \ln Tr_{S^\rho} e^{\beta H_\xi}\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}\right]. \tag{3.93}$$

As equações de ponto de sela

$$\frac{\partial f}{\partial m_\rho^{\mu\nu}} = 0, \quad \frac{\partial f}{\partial q_{\rho\sigma}} = 0, \quad \frac{\partial f}{\partial r_{\rho\sigma}} = 0 \tag{3.94}$$

que minimizam a densidade de energia livre, fornecem os estados estacionários e permitem interpretar fisicamente os parâmetros $m_\rho^{\mu\nu}$, $q_{\rho\sigma}$ e $r_{\rho\sigma}$ como

$$m_\rho^{\mu\nu} = \frac{1}{N}\langle\sum_i \xi_i^{\mu\nu}\langle S_i^\rho\rangle_T\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}, \tag{3.95}$$

$$q_{\rho\sigma} = \langle\frac{1}{N}\sum_i \langle S_i^\rho\rangle_T\langle S_i^\sigma\rangle_T\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}, \tag{3.96}$$

$$r_{\rho\sigma} = \frac{1}{\alpha}\sum_{\mu>1}\langle m_\rho^{\mu\nu}m_\sigma^{\mu\nu}\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}. \tag{3.97}$$

A princípio, todas as réplicas são equivalentes, o que sugere que os parâmetros de ordem não dependam dos índices de réplica. Supomos, então, a *simetria de réplica* expressa através das seguintes relações

$$m_\rho^{\mu\nu} = m^{\mu\nu}, \tag{3.98}$$

$$q_{\rho\sigma} = q, \quad \rho\neq\sigma, \tag{3.99}$$

$$r_{\rho\sigma} = r, \quad \rho\neq\sigma. \tag{3.100}$$

Substituindo essas relações na equação (3.93) e utilizando a integral Gaussiana para linearizar a dependência quadrática $S^\rho S^\sigma$, o que permite o cálculo do traço, obtém-se a densidade de energia livre na aproximação de réplicas simétricas (ver apêndice C),

$$f = \frac{1}{2}\sum_\nu (m^{1\nu})^2 + \frac{\alpha r C}{2} - \frac{\alpha}{\beta}\ln G(q) - \frac{1}{\beta}\int_{-\infty}^{\infty} Dz\langle\ln[2\cosh(\beta\Delta)]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}, \tag{3.101}$$

onde

$$\ln G(q) = -\frac{1}{2}[(s-1)\ln(1-C(1-b^2)) + \ln(1-C(1-b^2+sb^2)) - \frac{\beta q s(1-C(1-b^2)(1-b^2+sb^2))}{(1-C(1-b^2))(1-C(1-b^2+sb^2))}] \quad (3.102)$$

e

$$\Delta = z\sqrt{\alpha r} + \sum_\nu m^{1\nu}\xi^{1\nu} - h^1\xi^1, \quad (3.103)$$

$$C = \beta(1-q) \quad (3.104)$$

$$Dz = \frac{dz}{\sqrt{2\pi}}\exp(-z^2/2). \quad (3.105)$$

A partir das equações de ponto de sela (3.94), obtemos as equações para os parâmetros de ordem (tomando $h^\mu = 0$)

$$m^{1\nu} = \langle \int_{-\infty}^{\infty} Dz\,\xi^{1\nu}\tanh[\beta(\sqrt{\alpha r}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} \quad (3.106)$$

$$q = \langle \int_{-\infty}^{\infty} Dz\,\tanh^2[\beta(\sqrt{\alpha r}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} \quad (3.107)$$

$$r = sq\frac{[1-C(1-b^2)(1-b^2+sb^2)]^2 + (s-1)b^4}{[1-C(1-b^2)]^2[1-C(1-b^2+sb^2)]^2} \quad (3.108)$$

e, a partir da equação (3.22), obtemos

$$m^1 = \langle \int_{-\infty}^{\infty} Dz\,\xi^1\,\tanh[\beta(\sqrt{\alpha r}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}. \quad (3.109)$$

### 3.6.1 Soluções de categorização

Visto que a habilidade de categorização é caracterizada por uma solução simétrica ($m^{\mu\nu} = m_s\delta_{\mu 1}$), as equações (3.106) e (3.107) passam a ser escritas como

$$m_s = \frac{1}{s}\langle \int_{-\infty}^{\infty} Dz\,x_s\tanh[\beta(\sqrt{\alpha r}z + m_s x_s)]\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^\mu\}} \quad (3.110)$$

$$q = \langle \int_{-\infty}^{\infty} Dz\,\tanh^2\beta(\sqrt{\alpha r}z + m_s x_s)\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^\mu\}}, \quad (3.111)$$

onde as médias são sobre $x_s = \sum_{\nu=1}^{s} \xi^{1\nu}$, que é uma variável aleatória que satisfaz a distribuição binomial

$$P(x_s) = \begin{pmatrix} s \\ k \end{pmatrix} b_1^k b_2^{s-k} \tag{3.112}$$

com $k = \frac{1}{2}(x_s + s)$.

Calculando-se a média sobre os exemplos e os conceitos e deixando-se a média binomial explícita, as equações anteriores são escritas da seguinte forma:

$$m_s = \frac{1}{2s} \int_{-\infty}^{\infty} Dz \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} (2k - s) P_+(k) \tanh(\beta \Delta_s) \tag{3.113}$$

$$q = \frac{1}{2} \int_{-\infty}^{\infty} Dz \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_+(k) \tanh^2(\beta \Delta_s) \tag{3.114}$$

sendo $P_\pm(k)$ dada pela equação (3.63) e

$$\Delta_s = \sqrt{\alpha r} z + m_s(2k - s). \tag{3.115}$$

A superposição com os conceitos passa a ser escrito como

$$m^1 = \frac{1}{2} \int_{-\infty}^{\infty} Dz \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_-(k) \tanh(\beta \Delta_s), \tag{3.116}$$

e a densidade de energia livre é

$$f = \frac{1}{2} s m_s^2 + \frac{\alpha r C}{2} - \frac{\alpha}{\beta} \ln G(q) - \frac{1}{2\beta} \int_{-\infty}^{\infty} Dz \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_+(k) \ln 2 \cosh(\beta \Delta_s). \tag{3.117}$$

**Limite de ruído nulo ($T = 0$)**

No limite de ruído nulo, em que $\beta \to \infty$, a equação para o parâmetro de ordem $m_s$ passa a ser

$$m_s = \frac{1}{2s} \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} (2k - s) P_+(k) \mathrm{erf}\left(\frac{m_s(2k - s)}{\sqrt{2\alpha r}}\right), \tag{3.118}$$

onde erf($x$) é a função erro definida pela equação (2.80). Nesse limite, o parâmetro $q$ tende a 1, porém de tal forma que $C = \beta(1-q)$ permanece finito, sendo, então, o parâmetro de ordem que deve ser considerado

$$C = \frac{1}{\sqrt{2\pi\alpha r}} \sum_{k=0}^{s} \left( \begin{array}{c} s \\ k \end{array} \right) P_+(k) \exp[-\frac{m_s^2(2k-s)^2}{2\alpha r}]. \tag{3.119}$$

A superposição com os conceitos passa a ser escrita como

$$m^1 = \frac{1}{2} \sum_{k=0}^{s} \left( \begin{array}{c} s \\ k \end{array} \right) P_-(k) \mathrm{erf}(\frac{m_s(2k-s)}{\sqrt{2\alpha r}}). \tag{3.120}$$

A densidade de energia livre é dada por

$$\begin{aligned} f &= \frac{1}{2} s m_s^2 + \frac{\alpha r C}{2} - \frac{\alpha s[1 - C(1-b^2)(1-b^2+sb^2)]}{2[1-C(1-b^2)][1-C(1-b^2+sb^2)]} \\ &- \frac{1}{2} \sum_{k=0}^{s} \left( \begin{array}{c} s \\ k \end{array} \right) P_+(k) \left\{ \sqrt{\frac{2\alpha r}{\pi}} \exp[-\frac{m_s^2(2k-s)^2}{2\alpha r}] \right. \\ &+ \left. m_s(2k-s) \mathrm{erf}(\frac{m_s(2k-s)}{\sqrt{2\alpha r}}) \right\}. \end{aligned} \tag{3.121}$$

### 3.6.2 Resultados Numéricos

A solução numérica das equações para os parâmetros de ordem, (3.113) e (3.114), permite que se construa o diagrama de fases ($\alpha \times T$) apresentado na figura 3.5 para uma superposição $b = 0.4$ e para $s = 10$ exemplos [44].

O diagrama de fases mostra a existência de uma temperatura crítica $T_p$, acima da qual $m_s = 0$ e $q = 0$, caracterizando uma fase paramagnética (P). Nessa fase, a rede perde totalmente sua habilidade de categorização e de recuperação de exemplos. Abaixo dessa temperatura, e acima da curva $T_c$, existe uma fase de vidro de spin (SG), na qual os parâmetros de ordem assumem os valores $m_s = 0$ e $q \neq 0$. Nessa região, as soluções não possuem qualquer correlação com os conceitos ou com os exemplos. Essa fase corresponde à ausência de estados de categorização sobre uma extensão macroscópica da rede, não excluindo uma ordem local. Uma superposição finita $m_s$, caracterizando a fase de categorização (G),

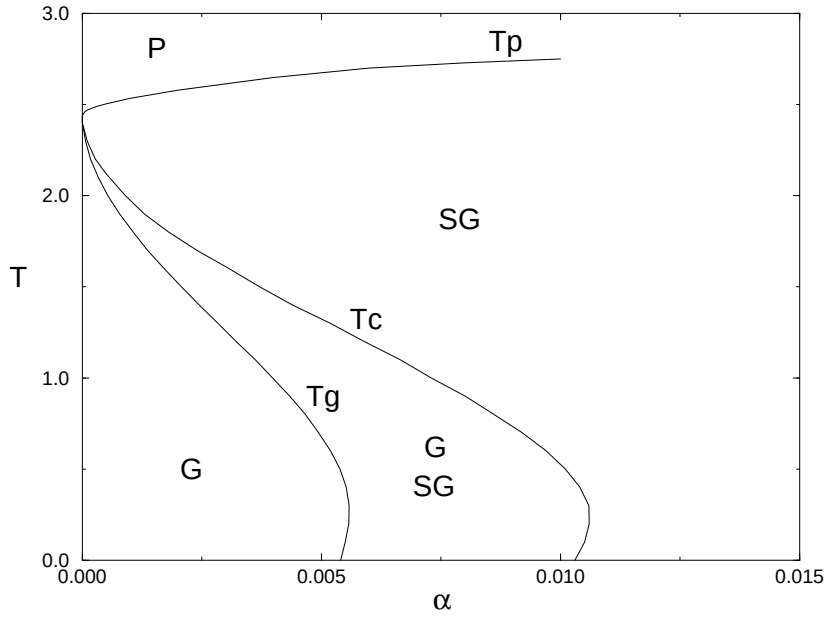



**Fig. 3.5:** *Diagrama de fases para $b = 0.4$ e $s = 10$ exemplos.*

aparece descontinuamente no contorno de fase $T_c$. Contudo, a solução de vidro de spin tem uma energia livre menor na região limitada pelas curvas $T_c$ e $T_g$. Abaixo da curva $T_g$, a fase de categorização é a mais estável, e os estados da rede estão fortemente correlacionados com os conceitos, como podemos observar na figura 3.6. Nessa região, encontramos soluções não nulas para os parâmetros de ordem, $m_s \neq 0$ e $q \neq 0$. Um aspecto muito interessante é a forma reentrante do diagrama de fases para baixas temperaturas. Essa reentrância também foi observada por Naef e Canning [46], para o modelo de Hopfield dentro da aproximação de simetria de réplicas. Outro aspecto interessante é o crescimento da região de categorização com o aumento do número de exemplos, como pode ser verificado na figura 3.6.

A figura 3.7 apresenta curvas de categorização para valores diferentes de temperatura com $\alpha = 0.03125$ e $b = 0.5$. Nela, é possível observar que, para $\alpha$ finito, a temperatura tem um efeito diferente do encontrado no caso $\alpha = 0$. Nesse caso, não é mais possível usar a temperatura para controlar a criação dos conceitos para um número fixo de exemplos. A fase de vidro de spin, que compete com a fase de categorização, é estabilizada pelo ruído Gaussiano devido aos exemplos não condensados $\{\xi^{\mu\nu}\}, \mu > 1$. Apesar disso, a habilidade de categorização da rede permanece robusta contra o ruído sináptico.

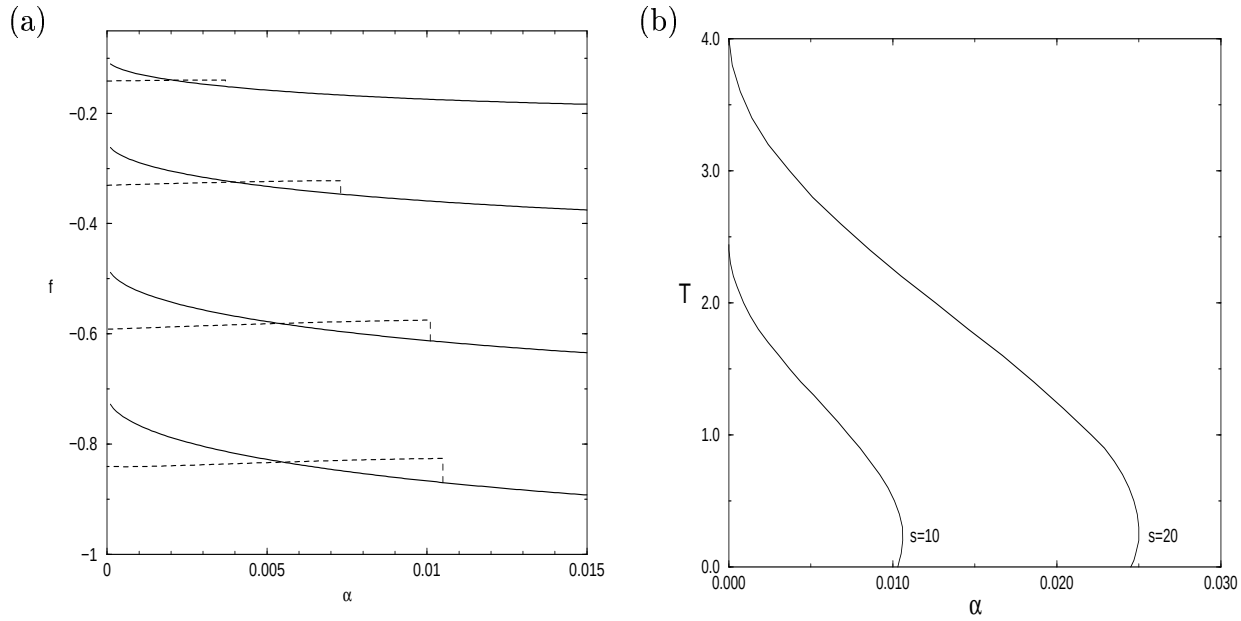



**Fig. 3.6:** *(a) Densidade de energia livre para as soluções de categorização (linha tracejada)e vidros de spin (linha contínua) para $T = 0.1$, $T = 0.5$, $T = 1.0$ e $T = 1.5$ de baixo para cima, com $b = 0.4$ e $s = 10$ exemplos, (b) Diagrama de fases $\alpha \times T$ para $b = 0.4$ e $s = 10$, $s = 20$ exemplos.*

Na figura 3.8, apresentamos o diagrama de fases ($T \times s$) para um valor fixo de $\alpha = 0.03125$ e $b = 0.5$. Nota-se que, para cada temperatura $T$, há um valor crítico de exemplos $s_c$, abaixo do qual não há estados de categorização. Observamos, também, que o aumento da temperatura implica a necessidade de treinar a rede com um número maior de exemplos para que a rede crie representações para os conceitos, ou seja, para que a rede categorize. Também observamos que, a $T = 0$, o número crítico de exemplos aumenta com $\alpha$, como evidenciado na figura 3.9.

## 3.7 Simulações Numéricas

Para verificarmos *qualitativamente* os resultados obtidos pela teoria de campo médio discutidos nas seções anteriores, realizamos simulações de Monte Carlo [29] para os regimes

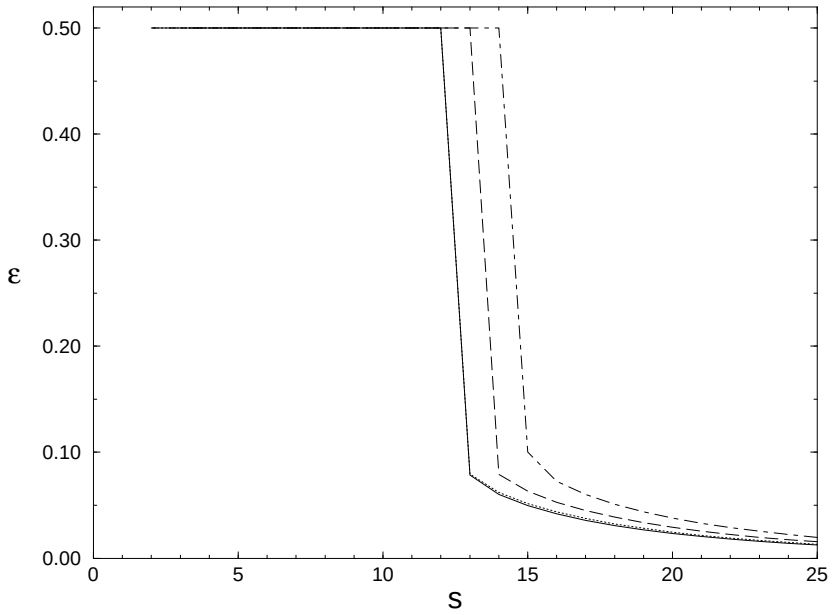



**Fig. 3.7:** *Curvas de categorização para* $\alpha = 0.03125$ *e* $b = 0.5$ *fixos a temperaturas* $T = 0$ *(linha cheia),* $T = 0.4$ *(linha pontilhada),* $T = 0.8$ *(linha tracejada) e* $T = 1.2$ *(linha ponto-tracejada).*

$\alpha \to 0$ e $\alpha \neq 0$, tanto à temperatura finita quanto à temperatura zero [44].

Para temperatura zero, a evolução do sistema consiste em atualizar os neurônios, selecionados aleatoriamente, a cada incremento de tempo de acordo com a equação determinística

$$S_i(t+1) = sgn[h_i(t)], \tag{3.122}$$

onde $h_i(t) = \sum_{j \neq i} J_{ij} S_j(t)$ é o campo local atuando no neurônio $i$. Esse procedimento decresce a energia do sistema sempre que um neurônio é atualizado, resultando em $S_i(t+1)h_i(t+1) > 0$.

Podemos simular a dinâmica do sistema em termos das superposições $m^{\mu\nu}$ que são calculadas com a rede num estado inicial que coincide com um dos exemplos e atualizado a cada passo da dinâmica. Para esse propósito, é conveniente expressar o Hamiltoniano em termos das superposições

$$H = -\frac{N}{2} \sum_{\mu\nu} (m^{\mu\nu})^2 + \frac{1}{2} ps. \tag{3.123}$$

Utilizamos essa equação para realizar as simulações de Monte Carlo à temperatura finita da seguinte forma:

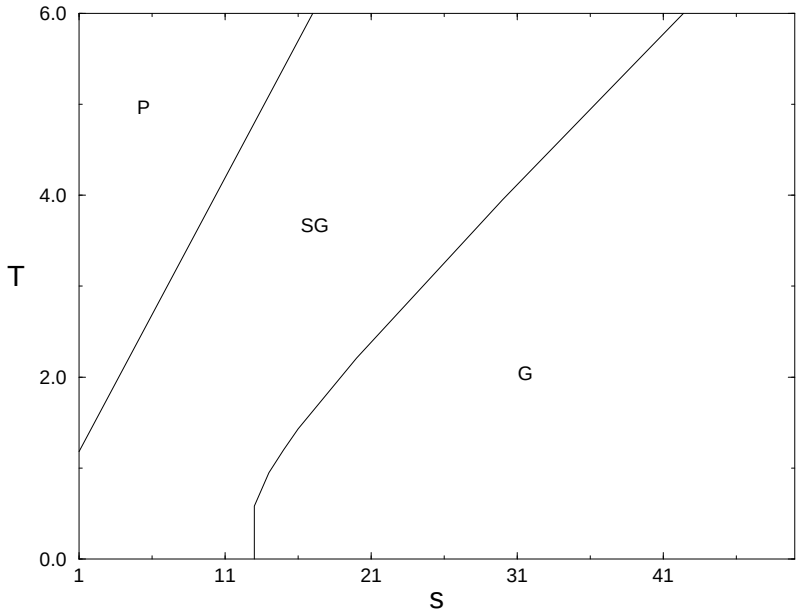



**Fig. 3.8:** *Diagrama de fases para* $\alpha = 0.03125$ *e* $b = 0.5$

1. escolhemos aleatoriamente uma configuração inicial que coincida com um dos exemplos $\xi^{\mu\nu}$;

2. selecionamos um neurônio aleatoriamente e invertemos seu estado;

3. calculamos a variação de energia $\Delta H$ devido a essa inversão;

4. se $\Delta H < 0$, aceitamos a inversão e retornamos ao item 2;

5. se $\Delta H \geq 0$, geramos um numero aleatório $r \in [0, 1]$;

6. se $r < \exp(-\beta \Delta H)$, aceitamos a inversão do estado do neurônio e retornamos ao item 2; caso contrário rejeitamos a inversão e retornamos ao item 2.

Repetimos esse procedimento para vários conjuntos de conceitos e exemplos.

### 3.7.1 Simulações em $\alpha \to 0$

A figura 3.10 mostra os resultados obtidos para o erro de categorização $\epsilon$ em função do número de exemplos $s$ para uma rede com $N = 2000$ neurônios, à temperatura $T = 0$ e para vários valores de superposição $b$. As médias foram calculadas sobre 100 relaxações.

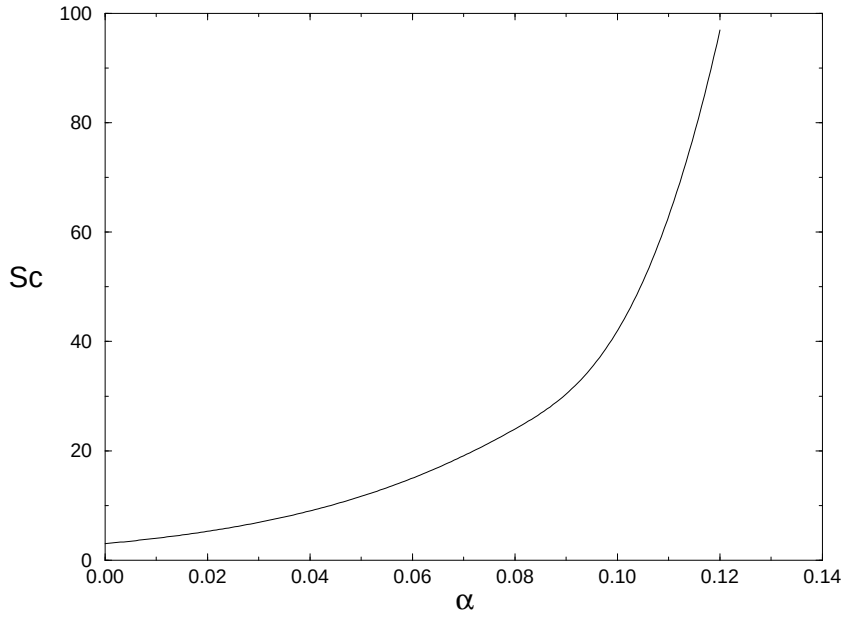



**Fig. 3.9:** *Número crítico de exemplos em função de $\alpha$ para $b = 0.6$.*

Observamos que, com a diminuição da superposição, é necessário um número maior de exemplos para a rede atingir o regime de categorização. Isso pode ser facilmente entendido, pois diminuir a superposição entre exemplos e conceitos significa aumentar as diferenças entre os exemplos, dificultando a extração de aspectos comuns ao conjunto dos exemplos. Além disso, podemos ver claramente o regime de recuperação dos exemplos, caracterizado pelo platô, e a transição para o regime de categorização, caracterizado pelo decaimento da curva.

Na figura 3.11, são apresentados os resultados obtidos na simulação de uma rede com $N = 2000$ neurônios, $b = 0.4$ e 100 relaxações, para temperatura zero e para temperatura finita. Para $T = 0$ e $s < s_c$, a rede recupera o exemplo escolhido como configuração inicial. Com o aumento do número de exemplos $s$, a rede sofre uma transição descontínua em $s_c$, além da qual a rede cria uma representação para o conceito subjacente ao conjunto dos exemplos ao qual pertence o exemplo usado como estado inicial. Para $T = 0.6$, a rede inicia a categorização com um número menor de exemplos que a $T = 0$, em consistência com os resultados da figura 3.4. Para $T = 1.6$, observamos a existência de um platô em $\epsilon = 0.5$, caracterizando a fase paramagnética, até atingir o número crítico de exemplos $s_c$, a partir

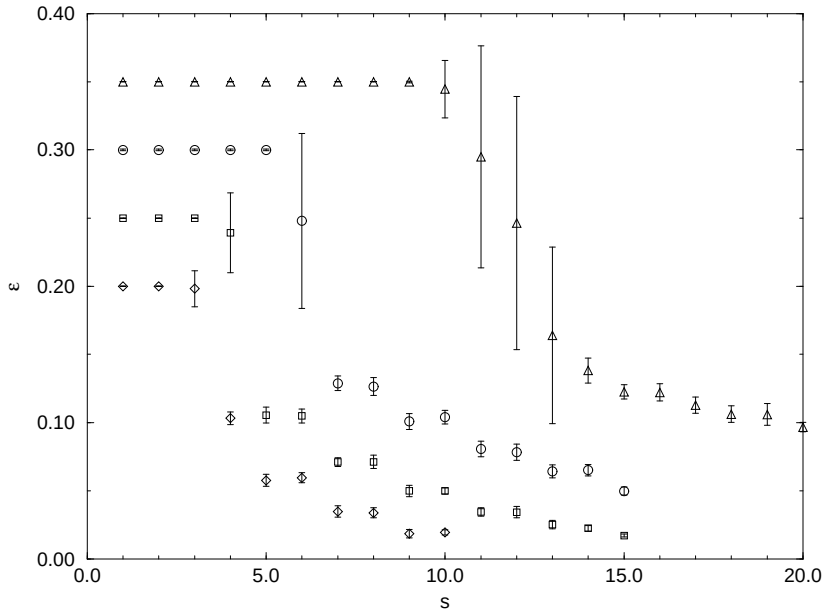



**Fig. 3.10:** *Curvas de categorização para* $b = 0.3$ *(triângulos),* $b = 0.4$ *(círculos),* $b = 0.5$ *(quadrados) e* $b = 0.6$ *(diamantes)*

do qual a rede passa a categorizar continuamente.

### 3.7.2 Simulações em $\alpha \neq 0$

Os resultados numéricos para o erro de categorização, no caso em que $\alpha \neq 0$ a várias temperaturas, são apresentados na figura 3.12, junto com resultados analíticos, para uma rede com $N = 2048$ neurônios, $b = 0.5$ e $\alpha = 0.03125$ para 800 relaxações, partindo de um exemplo como configuração inicial. Observamos que existe um regime de vidro de spin que compete com o regime de categorização na região entre $3 \leq s \leq s_c$, onde $s_c$ é o número crítico de exemplos. Na fase de vidro de spin, existe um número exponencialmente grande de estados meta-estáveis que são estáveis frente à inversão do estado de um neurônio. Esses estados funcionam como armadilhas para o sistema, fazendo com que o sistema fique preso em um desses estados, resultando em superposições $m^{\mu\nu}$ e $m^{\mu}$ não nulas. Essa é a razão pela qual os resultados para o erro de categorização obtidos nas simulações não coincidem com os resultados teóricos de equilíbrio $\epsilon = 0.5$. Outro fator que contribui para essa discrepância é o efeito de tamanho finito do sistema [47] [48]. Em $s \simeq s_c$, o sistema sofre uma transição

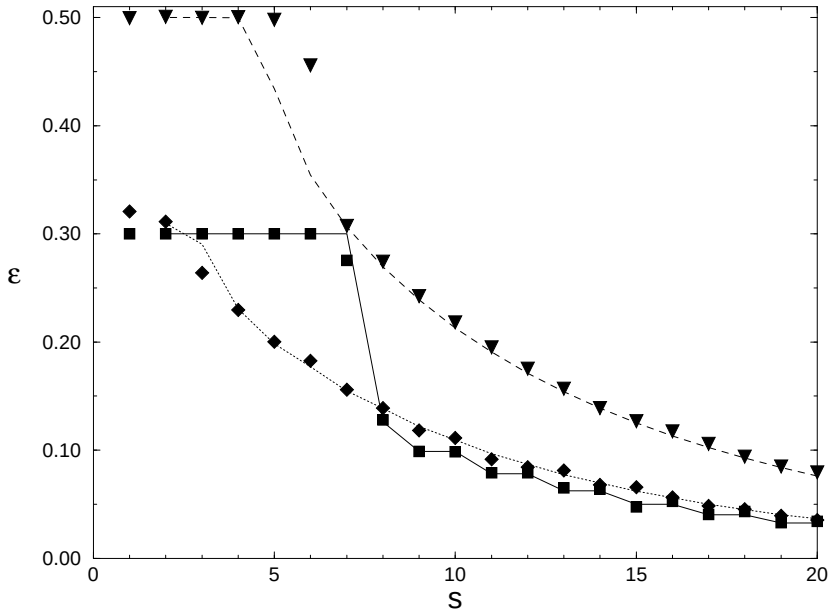



**Fig. 3.11:** *Curvas de categorização para* $\alpha = 0$, $b = 0.4$ *a* $T = 0$ *(quadrados),* $T = 0.6$ *(diamantes) e* $T = 1.6$ *(triângulos). As linhas correspondem aos resultados analíticos.*

para o regime de categorização de forma consistente com os resultados teóricos, exceto para baixos níveis de ruídos. A estimativa precisa do numero crítico de exemplos, a partir de simulações numéricas, depende fortemente do tamanho do sistema.

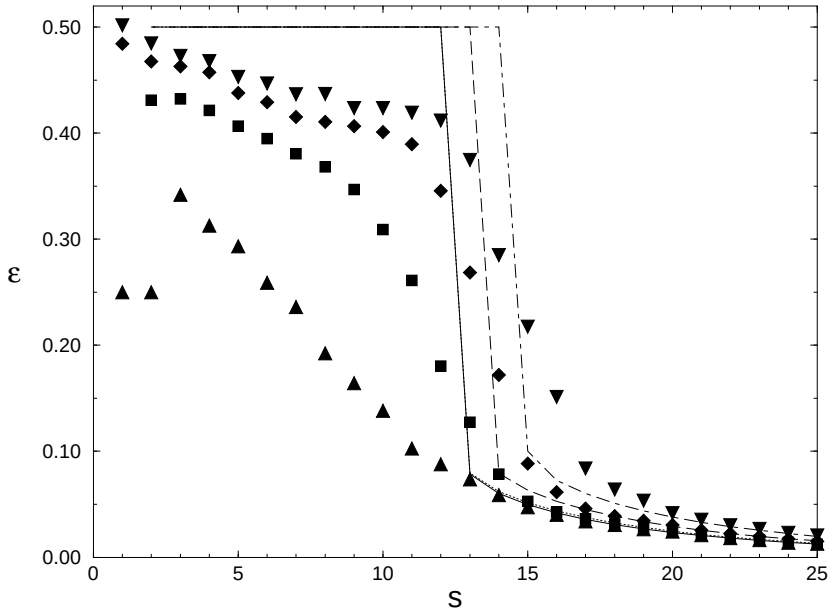



**Fig. 3.12:** *Curvas de categorização para* $\alpha = 0.03125$, $b = 0.5$, *a* $T = 0$ *(triângulos),* $T = 0.4$ *(quadrados),* $T = 0.8$ *(diamantes) e* $T = 1.2$ *(triângulos invertidos). As linhas correspondem aos resultados analíticos para* $T = 0$ *(linha cheia),* $T = 0.4$ *(linha pontilhada),* $T = 0.8$ *(linha linha tracejada) e* $T = 1.2$ *(linhas ponto-tracejadas).*

Capítulo 4

# Categorização no Modelo de Hopfield Simetricamente Diluído

## 4.1 Introdução

Uma das razões para o estudo de redes neurais artificiais diluídas é a observação de que, em sistemas biológicos, os neurônios estão conectados a uma fração de outros neurônios. Essa característica que surge naturalmente durante o desenvolvimento de sistemas biológicos também pode ser introduzida artificialmente nesses sistemas através de lesões. O estudo de redes diluídas permite modelar e compreender os efeitos das lesões sobre a performance de redes capazes de aprendizagem. Isso pode ser feito pela comparação entre o comportamento de redes neurais artificiais e redes biológicas. Dessa forma, os mecanismos de funcionamento de redes biológicas lesadas poderiam ser compreendidos [49].

Nesse capítulo, investigamos o comportamento da rede em função da diluição gradual das conexões sinápticas [50]. Atribui-se, aleatoriamente, o valor zero para uma fração das conexões sinápticas, mantendo-se a sua simetria. Embora em sistemas biológicos as conexões sinápticas sejam assimétricas, essa simetria permite que tratemos o modelo de Hopfield analiticamente com as técnicas da mecânica estatística de equilíbrio, utilizadas no capítulo anterior.

## 4.2 O Modelo

Para introduzirmos a diluição simétrica aleatória no modelo de Hopfield, consideramos $N$ neurônios do tipo Ising, $S_i = \pm 1; i = 1, ..., N$, descritos pelo Hamiltoniano

$$H = -\sum_{i<j} J_{ij}^d S_i S_j + \sum_{i\mu} h^\mu \xi_i^\mu S_i, \tag{4.1}$$

onde a soma é efetuada sobre todos os neurônios $i$, $j$, contando cada par uma vez, sendo a conexão sináptica simétrica ($J_{ij}^d = J_{ji}^d$) e $J_{ii}^d = 0$. A regra de aprendizagem é descrita pela regra de Hebb generalizada

$$J_{i,j}^d = \frac{c_{ij}}{cN} \sum_{\mu\nu} \xi_i^{\mu\nu} \xi_j^{\mu\nu}, \tag{4.2}$$

onde $c_{ij} = c_{ji}$ é uma variável aleatória que pode assumir os valores 1 com probabilidade $c$ e 0 com probabilidade $1 - c$ e $c_{ii} = 0$. A conectividade da rede é determinada pelo parâmetro $c$, que pode variar de 0 a 1, de modo que $cN$ é o número médio de neurônios conectados na rede. Quando $c$ assume o valor 1, a rede será completamente conexa, e recuperamos o modelo de Hopfield descrito no capítulo anterior; entretanto, no limite em que $c \to 0$, a rede encontra-se no regime de diluição extrema. Novamente os exemplos $\xi_i^{\mu\nu} = \pm 1$ são versões correlacionadas dos conceitos $\xi_i^\mu = \pm 1$, dados pela distribuição de probabilidades

$$P(\xi_i^{\mu\nu}) = \frac{1}{2}(1 + b\xi_i^\mu)\delta(\xi_i^{\mu\nu} - 1) + \frac{1}{2}(1 - b\xi_i^\mu)\delta(\xi_i^{\mu\nu} + 1), \tag{4.3}$$

onde $0 \leq b \leq 1$ determina a correlação entre os exemplos e os conceitos. Durante o processo de aprendizagem, a rede é exposta a um conjunto finito de $s$ exemplos $\xi_i^{\mu\nu}$, $\nu = 1, ..., s$ de cada um dos $p = \alpha cN$ conceitos $\xi_i^\mu$, $\mu = 1, ..., p$.

Inspirados no procedimento adotado por Sompolinsky [51] [52], baseado no trabalho de Viana e Bray [53] e generalizado recentemente [54], reescrevemos as conexões sinápticas como

$$J_{ij}^d = J_{ij} + \delta J_{ij}, \tag{4.4}$$

onde

$$\delta J_{ij} = -(1 - c_{ij}/c) J_{ij}, \tag{4.5}$$

sendo

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p}\sum_{\nu=1}^{s}\xi_i^{\mu\nu}\xi_j^{\mu\nu} \tag{4.6}$$

as conexões sinápticas da rede completamente conexa treinada com exemplos.

O termo de diluição $\delta J_{ij}$ pode ser interpretado como um ruído Gaussiano efetivamente independente dos exemplos de treinamento [55], cuja média configuracional, que inclui a média sobre a variável aleatória $c_{ij}$, é

$$\langle \delta J_{ij} \rangle_c = 0 \tag{4.7}$$

e de variância

$$\langle \delta_{ij}^2 \rangle_c = \Delta^2/N, \tag{4.8}$$

onde $\langle ... \rangle_c$ significa média também sobre a diluição, com

$$\Delta^2 \equiv a^2(1-c)/c \tag{4.9}$$

e

$$a^2 = \alpha c s[1 + (s-1)b^4]. \tag{4.10}$$

É importante notar que esse ruído sináptico estático é de natureza totalmente diferente do ruído representado pelo que chamamos de temperatura $T$, ou ruído rápido. Essa separação da matriz sináptica em dois termos leva a um tratamento analítico, em que o primeiro termo consiste no modelo de Hopfield completamente conectado, ao qual se aplica o desenvolvimento teórico realizado no capítulo anterior, e o segundo termo consiste em um termo aleatório Gaussiano que, no limite $N \to \infty$, é tratado de maneira análoga ao vidro de spin de Sherrington-Kirkpatrick. A diluição introduzida dessa maneira caracteriza-se por uma conectividade da mesma ordem do número de neurônios $N$.

Para caracterizar a performance da rede como um dispositivo capaz de categorizar, é necessário introduzir um parâmetro que informe quantitativamente a qualidade do reconhecimento do conceito $\xi^\mu$. Isso é feito, como no capítulo anterior, através do parâmetro de ordem $m^\mu$, que descreve a superposição entre o estado da rede $S_i$ e o conceito $\xi^\mu$

$$m^\mu = \frac{1}{N}\sum_{i=1}^{N}\xi^\mu S_i, \tag{4.11}$$

para $\mu = 1, ..., p$, com $0 \leq m^\mu \leq 1$.

O reconhecimento de um conceito significa existir uma superposição finita entre o estado da rede e o conceito. Esse reconhecimento emerge espontaneamente na dinâmica da rede treinada com exemplos. A medida do insucesso em reconhecer um conceito é dada pelo erro de categorização, que é definido como a distância de Hamming

$$\epsilon^\mu = \frac{1}{2}(1 - m^\mu), \tag{4.12}$$

entre o estado $S_i$ e o conceito $\xi^\mu$, onde $\mu = 1, ..., p$.

## 4.3 Teoria de Campo Médio

Procedemos agora ao estudo da mecânica estatística de equilíbrio do Hamiltoniano (4.1) no limite de $\alpha = p/cN$ finito. A densidade de energia livre por sítio conectado é dada por

$$f = -\lim_{N\to\infty} \frac{1}{cN\beta} \langle \ln Z \rangle_c, \tag{4.13}$$

onde

$$Z = Tr \exp(-\beta H) \tag{4.14}$$

é a função de partição do modelo, $H$ é a soma do Hamiltoniano de Hopfield generalizado e do Hamiltoniano de vidro de spin com interações aleatórias Gaussianas $\delta J_{ij}$, $\beta = T^{-1}$ e $\langle \ldots \rangle_c$ representa a média sobre os exemplos $\{\xi_i^{\mu\nu}\}$, conceitos $\{\xi_i^\mu\}$ e o índice $c$ indica a média sobre as variáveis $\delta J_{ij}$.

A superposição com um determinado conceito é obtida a partir de

$$m^\mu = df/dh^\mu|_{h^\mu=0}, \tag{4.15}$$

onde tomamos o limite do campo $h^\mu \to 0$. Todos os conceitos são equivalentes de modo que nos concentraremos em um único conceito, em particular $\mu = 1$.

Procedendo da maneira padrão, utilizamos o método das réplicas para escrever

$$\langle \ln Z \rangle_c = \lim_{n\to 0} \frac{\langle Z^n \rangle_c - 1}{n}, \tag{4.16}$$

sendo $\langle Z^n \rangle_c$ a função de partição replicada dada por

$$\langle Z^n \rangle_c = \langle Tr_{S^\rho} \exp(\beta \sum_\rho \sum_{i<j} J_{ij} S_i^\rho S_j^\rho + \beta \sum_\rho \sum_{i<j} \delta J_{ij} S_i^\rho S_j^\rho) \rangle_c. \tag{4.17}$$

Substituindo a regra de Hebb (4.6), na equação acima, e como $\sum_{i<j} = \frac{1}{2}\sum_{i,j} - \frac{1}{2}\sum_{i=j}$, podemos escrever

$$\begin{aligned} \langle Z^n \rangle_c &= \langle Tr_{S^\rho} \exp(\frac{\beta}{2N} \sum_{\mu\nu} \sum_\rho \sum_{i,j} (\xi_i^{\mu\nu} S_i^\rho)(\xi_j^{\mu\nu} S_j^\rho) - \beta h^1 \sum_{i\rho} \xi_i^\mu S_i^\rho - \frac{1}{2}\beta nps) \\ &\times \exp(\beta \sum_{i<j} \delta J_{ij} \sum_\rho S_i^\rho S_j^\rho) \rangle_c. \end{aligned} \tag{4.18}$$

Explicitando a média sobre a variável $\delta J_{ij}$, obtemos

$$\begin{aligned} \langle Z^n \rangle_c &= e^{-\frac{1}{2}\beta nps} Tr_{S^\rho} \langle \{ \prod_{\mu\nu\rho} \exp[\frac{\beta}{2N} \sum_i (\xi_i^{\mu\nu} S_i^\rho)^2] \prod_\rho \exp(-\beta h^1 \sum_{i\rho} \xi_i^\mu S_i^\rho) \} \\ &\times \{ \prod_{i<j} \int \frac{d\delta J_{ij}}{\sqrt{2\pi\Delta^2/N}} e^{-\frac{N\delta J_{ij}^2}{2\Delta^2}} \exp(\beta \delta J_{ij} \sum_\rho S_i^\rho S_j^\rho) \} \rangle \end{aligned} \tag{4.19}$$

onde $\langle \ldots \rangle$ representa a média sobre os exemplos e conceitos.

Essa expressão contém o produto de dois termos entre chaves. Podemos tratá-los independentemente até o momento de tomarmos o traço. O primeiro é simplesmente o modelo de Hopfield completamente conexo, estudado no capítulo anterior. Entretanto, o segundo termo é formalmente idêntico ao vidro de spin de Sherrington-Kirkpatrick (SK), porém com a média $J_0$ das interações $J_{ij}$ igual a zero. Vamos nos referir a esse termo como $I_{SK}$. Seguindo o procedimento usual para o modelo SK, completamos o quadrado do argumento da exponencial e obtemos

$$\begin{aligned} I_{SK} &= \prod_{i<j} \int \frac{d\delta J_{ij}}{\sqrt{2\pi\Delta^2/N}} \exp[-\frac{N}{2\Delta^2}(\delta J_{ij}^2 - \frac{2\Delta^2\beta}{N} \delta J_{ij} \sum_\rho S_i^\rho S_j^\rho + \frac{\Delta^4\beta^2}{N^2} \sum_{\rho\sigma} S_i^\rho S_j^\rho S_i^\sigma S_j^\sigma)] \\ &\times \prod_{i<j} \exp(\frac{\Delta^2\beta^2}{2N} \sum_{\rho\sigma} S_i^\rho S_j^\rho S_i^\sigma S_j^\sigma). \end{aligned} \tag{4.20}$$

Efetuando o cálculo da integral Gaussiana sobre o conjunto $\{\delta J_{ij}\}$, que resulta ser igual à unidade, ficamos com um termo bastante simples para $I_{SK}$:

$$I_{SK} = \exp(\frac{\Delta^2\beta^2}{2N} \sum_{i<j} \sum_{\rho,\sigma} S_i^\rho S_j^\rho S_i^\sigma S_j^\sigma), \tag{4.21}$$

que podemos reescrever de maneira mais conveniente, removendo a restrição sobre a soma nos sítios e somando sobre as réplicas com a condição $\rho < \sigma$

$$I_{SK} = \exp(\frac{\Delta^2\beta^2(N-1)n}{4}) \exp(\frac{\Delta^2\beta^2}{2N} \sum_{\rho<\sigma} [\sum_i S_i^\rho S_i^\sigma]^2) \exp(-\frac{\Delta^2\beta^2 n(n-1)}{4}). \qquad (4.22)$$

Combinando a primeira e a última exponencial e tendo em mente que consideraremos o limite do número de réplicas tendendo a zero, obtemos uma expressão para $I_{SK}$ que depende apenas de um sítio

$$I_{SK} = \exp(\frac{\Delta^2\beta^2 Nn}{4}) \exp(\frac{\Delta^2\beta^2}{2N} \sum_{\rho<\sigma} [\sum_i S_i^\rho S_i^\sigma]^2). \qquad (4.23)$$

Desse modo a equação (4.19) pode ser escrita como

$$\begin{aligned} \langle Z^n \rangle_c &= e^{-\frac{1}{2}\beta nps} e^{\frac{1}{4}\Delta^2\beta^2 Nn} Tr_{S^\rho} \langle \{ \underbrace{\prod_{\mu\nu\rho} \exp[\frac{\beta}{2N} \sum_i (\xi_i^{\mu\nu} S_i^\rho)^2] \prod_\rho \exp(-\beta h^1 \sum_{i\rho} \xi_i^\mu S_i^\rho)}_{rede\ completamente\ conexa} \} \\ &\times \underbrace{\{\prod_{\rho<\sigma} \exp \frac{\Delta^2\beta^2}{2N} [\sum_i S_i^\rho S_i^\sigma]^2\}}_{rede\ diluída} \rangle. \end{aligned} \qquad (4.24)$$

Nesse ponto, podemos juntar a contribuição devido à diluição ao termo de Hopfield completamente conexo, e proceder aos cálculos da forma usual como desenvolvemos no capítulo anterior.

Introduz-se os parâmetros $q_{\rho\sigma} = \frac{1}{N}\sum_i S_i^\rho S_i^\sigma$ e $r_{\rho\sigma}$ através da delta de Dirac

$$1 = \int dq_{\rho\sigma}\ \delta(q_{\rho\sigma} - \frac{1}{N} \sum_i S_i^\rho S_i^\sigma), \qquad (4.25)$$

e sua representação integral

$$\delta(q_{\rho\sigma} - \frac{1}{N} \sum_i S_i^\rho S_i^\sigma) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} dr_{\rho\sigma}\ \ \exp[-r_{\rho\sigma}(q_{\rho\sigma} - \frac{1}{N} \sum_i S_i^\rho S_i^\sigma)]. \qquad (4.26)$$

Após utilizar a propriedade da automediação e desconsiderar fatores multiplicativos que não contribuem para a energia livre nos limites considerados, obtemos para a função de partição

$$\langle Z^n \rangle = \int \prod_{\rho\nu} dm_\rho^{1\nu} \int \prod_{\rho<\sigma} dq_{\rho\sigma} dr_{\rho\sigma} \exp\{-N\beta H(\{m_\rho^{\mu\nu}, q_{\rho\sigma}, r_{\rho\sigma}\}), \qquad (4.27)$$

onde

$$\begin{aligned} H(\{m_\rho^{\mu\nu}, q_{\rho\sigma}, r_{\rho\sigma}\}) &= \frac{1}{2}\sum_{\rho\nu}(m_\rho^{1\nu})^2 + \frac{\alpha c\beta}{2}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} - \frac{\beta\Delta^2}{4}\sum_{\rho\neq\sigma} q_{\rho\sigma}^2 \\ &- \frac{\alpha c}{\beta}\ln G(q_{\rho\sigma}) - \frac{1}{\beta}\langle \ln Tr_{S^\rho}\exp(\beta H_\xi)\rangle, \end{aligned} \tag{4.28}$$

sendo $H_\xi$ e $G(q_{\rho\sigma})$ dados pelas equações (3.92) e (3.89), respectivamente.

Realizando a integração através do método do ponto de sela no limite termodinâmico, obtemos a seguinte densidade de energia livre

$$f = \lim_{n\to 0}\left[\frac{1}{2n}\sum_{\rho\nu}(m_\rho^{1\nu})^2 + \frac{\alpha c\beta}{2n}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} - \frac{\beta\Delta^2}{4n}\sum_{\rho\neq\sigma} q_{\rho\sigma}^2 - \frac{\alpha c}{n\beta}\ln G(q_{\rho\sigma}) - \frac{1}{n\beta}\langle \ln Tr_{S^\rho} e^{\beta H_\xi}\rangle\right]. \tag{4.29}$$

A interpretação física dos parâmetros de ordem $m_\rho^{\mu\nu}$, $q_{\rho\sigma}$ e $r_{\rho\sigma}$ é a mesma obtida para a rede completamente conexa, dada pelas equações (3.95), (3.96) e (3.97).

Antes de efetuarmos o limite em que o número de réplicas tende a zero ($n \to 0$), assumimos que todas as réplicas são equivalentes, implicando a *simetria de réplicas*, que consiste, do ponto de vista estritamente matemático, na independência dos parâmetros de ordem dos índices de réplicas

$$m_\rho^{\mu\nu} = m^{\mu\nu}, \tag{4.30}$$

$$q_{\rho\sigma} = q, \quad \rho \neq \sigma, \tag{4.31}$$

$$r_{\rho\sigma} = r, \quad \rho \neq \sigma. \tag{4.32}$$

Desse modo, a densidade de energia livre para réplicas simétricas toma a forma

$$\begin{aligned} f &= \frac{1}{2}\sum_{\nu}(m^{1\nu})^2 + \frac{C\alpha cr}{2} + \frac{\beta\Delta^2}{4}q^2 - \frac{\alpha c}{\beta}\ln G(q) \\ &- \frac{1}{\beta}\int \mathcal{D}z\langle \ln\{2\cosh[\beta(\sqrt{\alpha rc}z + \sum_{\nu} m^{1\nu}\xi^{1\nu} - h^1\xi^1)]\}\rangle, \end{aligned} \tag{4.33}$$

sendo $\ln G(q)$ a mesma função obtida no capítulo anterior (equação (3.102)).

A partir das equações de ponto de sela, que minimizam a densidade de energia livre,

$$\frac{\partial f}{\partial m} = 0, \quad \frac{\partial f}{\partial q} = 0, \quad \frac{\partial f}{\partial r} = 0 \tag{4.34}$$

obtemos os parâmetros de ordem que descrevem o comportamento de equilíbrio do sistema (tomando $h^\mu = 0$)

$$m^{1\nu} = \langle \int \mathcal{D}z\, \xi^{1\nu} \tanh[\beta(\sqrt{\alpha r c}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle \tag{4.35}$$

$$q = \langle \int \mathcal{D}z\, \tanh^2[\beta(\sqrt{\alpha r c}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle \tag{4.36}$$

$$r = sq\frac{[1-C(1-b^2)(1-b^2+sb^2)]^2 + (s-1)b^4}{[1-C(1-b^2)]^2[1-C(1-b^2+sb^2)]^2} + q\frac{\Delta^2}{\alpha c} \tag{4.37}$$

e, a partir da equação (4.15), obtemos

$$m^1 = \langle \int \mathcal{D}z\, \xi^1\ \tanh[\beta(\sqrt{\alpha r c}z + \sum_\nu m^{1\nu}\xi^{1\nu})]\rangle. \tag{4.38}$$

Nota-se, das equações para os parâmetros de ordem, que a diluição das sinapses tem uma influência direta sobre o ruído estocástico da rede através da dispersão $\Delta^2$. Note, também, que $\alpha = \alpha_H/c$, sendo $\alpha_H$ o parâmetro de armazenamento do modelo de Hopfield completamente conexo.

No limite em que $s = 1 = b$, recuperamos as equações para a densidade de energia livre e para os parâmetros de ordem do modelo de Hopfield simetricamente diluído [56], e, no limite de diluição extrema $c \to 0$, recuperamos as equações de Watkin e Sherrington [57]. No regime de plena conectividade ($c = 1$), recuperamos as equações do capítulo anterior.

### 4.3.1 Soluções de categorização

Como nosso objetivo é analisar o efeito da diluição sobre a habilidade de categorização do modelo de Hopfield, vamos nos ater à solução que leva em consideração justamente esse tipo de comportamento. Já vimos no capítulo anterior que os estados simétricos desempenham o papel fundamental no problema de categorização. Consideraremos, então, soluções simétricas com $s$ exemplos, $m^{\mu\nu} = m_s\delta_{\mu 1}$, $\nu = 1, \ldots, s$, o que nos permite escrever

$$\sum_\nu m^{1\nu} = m_s x_s, \tag{4.39}$$

em termos da soma simétrica dos $s$ exemplos, $x_s = \sum_\nu \xi^{1\nu}$, que é uma variável aleatória que satisfaz a distribuição binomial (3.112).

Efetuando-se o cálculo da média sobre os exemplos e os conceitos e mantendo a média binomial explícita, as equações para os parâmetros de ordem são escritas da seguinte forma:

$$m_s = \frac{1}{2s}\int \mathcal{D}z \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} (2k-s) P_+(k) \tanh[\beta(\sqrt{\alpha r c}z + m_s(2k-s))], \tag{4.40}$$

$$q = \frac{1}{2}\int \mathcal{D}z \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_+(k) \tanh^2[\beta(\sqrt{\alpha r c}z + m_s(2k-s))], \tag{4.41}$$

sendo $P_\pm(k)$ dada pela equação (3.63), e a superposição com os conceitos passa a ser escrita como

$$m^1 = \frac{1}{2}\int \mathcal{D}z \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_-(k) \tanh[\beta(\sqrt{\alpha r c}z + m_s(2k-s))]. \tag{4.42}$$

Para essa classe de soluções, a densidade de energia livre é

$$f = \frac{1}{2}s\, m_s^2 + \frac{C\alpha c r}{2} - \frac{\alpha}{\beta}\ln G(q) - \frac{1}{2\beta}\int_{-\infty}^{\infty} Dz \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_+(k) \ln 2\cosh[\beta(\sqrt{\alpha r c}z + m_s(2k-s))]. \tag{4.43}$$

**Limite de ruído nulo ($T = 0$)**

Na ausência de ruído sináptico, as equações para os parâmetros de ordem são dadas por

$$m_s = \frac{1}{2s}\sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} (2k-s) P_+(k) \mathrm{erf}\left(\frac{m_s(2k-s)}{\sqrt{2\alpha r c}}\right), \tag{4.44}$$

onde a função erro, $\mathrm{erf}(x)$, é definida pela equação (2.80). Nesse limite, onde $\beta \to \infty$, temos que o parâmetro $q \to 1$ e $C = \beta(1-q)$ permanece finito, sendo, então, o parâmetro de ordem relevante

$$C = \frac{1}{\sqrt{2\pi\alpha r}} \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} P_+(k) \exp\left[-\frac{m_s^2(2k-s)^2}{2\alpha r c}\right], \tag{4.45}$$

e o parâmetro $r$ passa a ser

$$r = s\frac{[1-C(1-b^2)(1-b^2+sb^2)]^2+(s-1)b^4}{[1-C(1-b^2)]^2[1-C(1-b^2+sb^2)]^2}+\frac{\Delta^2}{\alpha c}. \tag{4.46}$$

A medida da categorização, dada pela superposição com os conceitos, passa a ser escrita como

$$m^1 = \frac{1}{2}\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right) P_-(k)\mathrm{erf}(\frac{m_s(2k-s)}{\sqrt{2\alpha rc}}), \tag{4.47}$$

e densidade de energia livre, por

$$\begin{aligned} f &= \frac{1}{2}sm_s^2+\frac{\alpha rC}{2}-\frac{\alpha sc[1-C(1-b^2)(1-b^2+sb^2)]}{2[1-C(1-b^2)][1-C(1-b^2+sb^2)]} \\ &- \frac{1}{2}\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right) P_+(k)\left\{\sqrt{\frac{2\alpha rc}{\pi}}\exp[-\frac{m_s^2(2k-s)^2}{2\alpha rc}]\right. \\ &+ \left. m_s(2k-s)\,\mathrm{erf}(\frac{m_s(2k-s)}{\sqrt{2\alpha rc}})\right\}. \end{aligned} \tag{4.48}$$

**Linha de Almeida-Thouless**

As equações de campo médio, na teoria de simetria de réplicas para o modelo de Hopfield simetricamente diluído no problema de recuperação de memórias [56] [57], e para o problema de vidros de spin de Sherrington e Kirkpatrick [23], não são válidas abaixo da linha de Almeida-Thouless (AT) [58]. É, portanto, importante que se determine a linha AT e, em particular, o ponto onde essa encontra o contorno de fase entre a fase de categorização e a fase de vidro de spin. Os cálculos para a obtenção da linha AT no problema de categorização com exemplos ponderados foram desenvolvidos por A. Theumann [59] e posteriormente aplicados ao problema de categorização com diluição simétrica por W. K. Theumann. A equação que determina a linha AT para a solução de categorização é dada por

$$\int \mathcal{D}z\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right) P_+(k)\mathrm{sech}^4[\beta(\sqrt{\alpha rc}z+m_s(2k-1))] = \frac{q}{\alpha rc}, \tag{4.49}$$

que deve ser resolvida simultaneamente com as equações para os parâmetros $m_s$ e $q$.

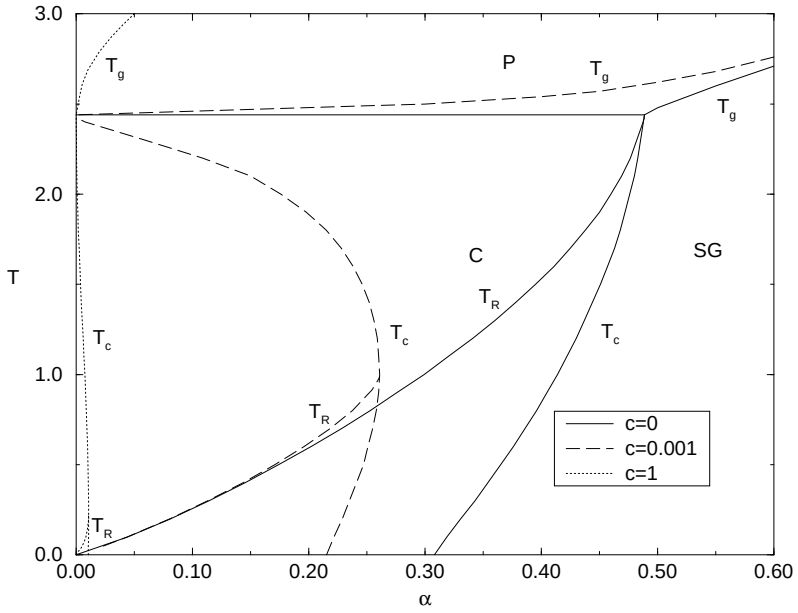



**Fig. 4.1:** *Diagrama de fases $\alpha \times T$ para correlação $b = 0.4$, $s = 10$ exemplos e conectividades $c = 1,\ 0.001$ e $0\,(c \to 0)$.*

## 4.4 Resultados Numéricos

A solução numérica das equações de ponto de sela, juntamente com a equação para a linha AT, permite analisar o comportamento da rede em função da correlação $b$, do número de exemplos $s$, da conectividade $c$, do ruído sináptico $T$ e do parâmetro de armazenamento $\alpha$.

Inicialmente, obtivemos o diagrama de fases para a temperatura $T$ versus o parâmetro de armazenamento $\alpha$, apresentado na figura 4.1, para vários valores de conectividade $c$, quando a correlação entre os exemplos e o conceito correspondente é $b = 0.4$, para $s = 10$ exemplos. A fase de categorização (C), onde $m_s \neq 0$ e $q \neq 0$, aparece à esquerda do contorno de fases $T_c$, enquanto que a fase vidro de spin (SG), onde $m_s = 0$ e $q \neq 0$, existe em toda a região abaixo do contorno de fases delimitado por $T_g$, exceto para $c = 0$. Em particular, os estados de vidro de spin tornam-se instáveis na fase de categorização, no limite de diluição extrema. Na região limitada pelo contorno $T_c$, onde coexistem estados estáveis de categorização e de

vidro de spin para valores finitos de diluição, os estados de categorização são mais estáveis para pequenos valores de $\alpha$, enquanto os estados de vidro de spin são mais estáveis para grandes valores de $\alpha$. Esse comportamento é semelhante à competição existente entre os estados de recuperação e de vidro de spin observado no modelo de Hopfield simetricamente diluído, para o problema de memorização [57]. Estados de recuperação devem aparecer para baixos valores de $\alpha$ [60].

Observam-se um aumento considerável da fase de categorização com o aumento da diluição da rede e, simultaneamente, uma redução dos estados de vidro de spin globalmente estáveis, principalmente no limite $c \to 0$. É instrutivo comparar o nosso diagrama de fases (figura 4.1) com o diagrama de fases correspondente obtido no problema de categorização com diluição *assimétrica* extrema [61] [62]. Enquanto no caso *simétrico* a região de categorização *aumenta* com a temperatura, até encontrar a fase para magnética (P) onde $m_s = 0 = q$, no caso *assimétrico* a região de categorização *diminui* com o aumento da temperatura. Entretanto, à $T = 0$, o valor de $\alpha_c$ é o mesmo para os dois modelos.

A linha de transição de fases $T_c$ é de primeira ordem para valores finitos de conectividade, caracterizando uma transição descontínua C-SG, tornando-se uma linha de segunda ordem estritamente no limite de conectividade extrema (c=0), onde o parâmetro $m_s$ vai continuamente a zero na transição C-SG. A linha de transição $T_g$ é sempre de segunda ordem, caracterizando uma transição contínua, tanto na transição categorização-paramagnética, quanto na transição vidro de spin-paramagnética.

A linha de Almeida-Thouless $T_R$, abaixo da qual as soluções com simetria de réplicas para os parâmetros de ordem tornam-se instáveis às perturbações da quebra de simetria de réplicas, também são indicadas na figura 4.1. Dentro da precisão dos cálculos efetuados, essas linhas $T_R$ encontram as linhas $T_c$ justamente nos pontos em que essas sofrem a inflexão que marca o início da reentrância à baixas temperaturas.

Como observamos na rede completamente conexa, a diminuição da correlação entre conceitos e exemplos ($b$) tem como efeito reduzir a região de categorização como pode ser verificado na figura 4.2*(a)* ao compará-la com a figura 4.1. O efeito do ruído sináptico ($T$)

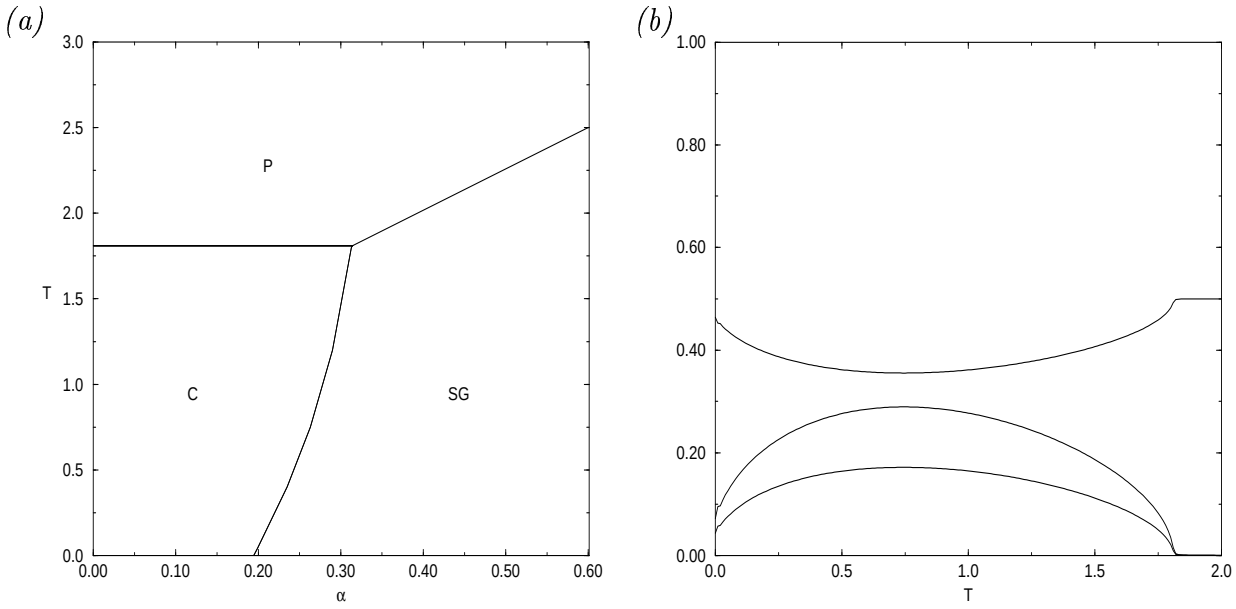



**Fig. 4.2:** *(a) Diagrama de fases $\alpha \times T$ para $b = 0.3$, $s = 10$ e $c = 0$, (b) corte do diagrama em $\alpha = 0.192$ mostrando o comportamento de $m_s$, $m^1$ e $\epsilon$ em função da temperatura (ruído sináptico), de baixo para cima respectivamente.*

sobre a habilidade de categorização pode ser visualizado na figura 4.2*(b)*, que representa um corte da figura 4.2*(a)* em $\alpha = 0.192$. Observa-se que, para valores moderados de temperatura, há um favorecimento da categorização, visto que o erro $\epsilon(T)$ diminui até um valor mínimo para um valor determinado de temperatura, a partir do qual o erro passa a aumentar.

Para verificarmos em que medida a diluição afeta a habilidade de categorização, analisamos a performance da rede. Observamos uma melhor performance da rede na fase de categorização à medida que aumentamos a diluição, como podemos verificar na figura 4.3, para $T = 0.1$, $b = 0.4$, $s = 10$ e conectividade $c = 0$ e $c = 0.001$. No limite de diluição extrema, as superposições simétricas com os exemplos $m_s$ e com o conceito $m^1$ desaparecem continuamente ao se aproximarem do contorno de fase $T_c$ para valores crescentes de $\alpha$, enquanto que, para conectividade finita, observamos uma descontinuidade para $m_s$ e $m^1$, no contorno de fase, como observado no rede completamente conexa. Na medida em que

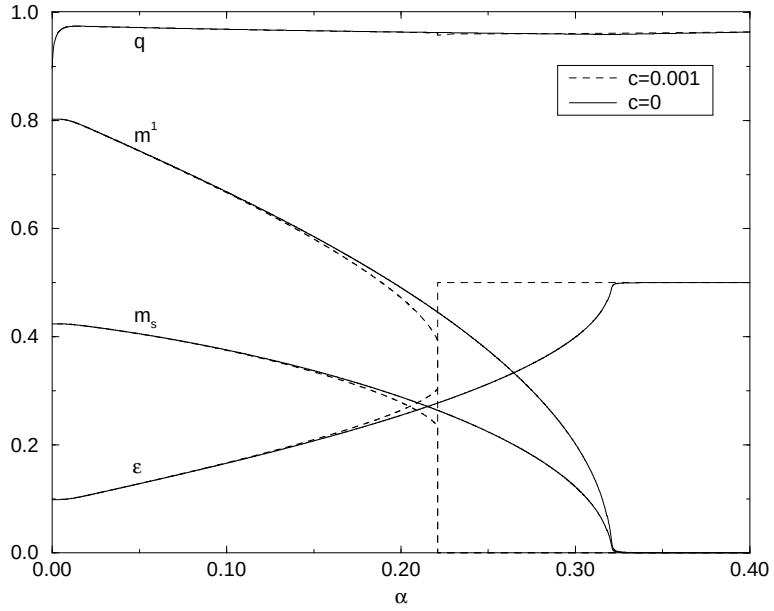



**Fig. 4.3:** *Superposições para os conceitos* $m^1$ *e exemplos* $m_s$*, parâmetro de ordem de vidros de spins* $q$*, erro de categorização* $\epsilon$ *para* $T = 0.1$*,* $b = 0.4$*,* $s = 10$ *e para conectividades* $c = 0$ *e* $0.001$*.*

aumentamos a diluição da rede, a descontinuidade torna-se cada vez menor e, no limite de extrema diluição, passa a ser contínua. Como é de se esperar, as superposições com o conceito e com os exemplos diminuem gradualmente com o aumento de $\alpha$, e, conseqüentemente, o erro de categorização aumenta. Como o valor crítico $\alpha_c$ aumenta com o contorno de fases C-SG na medida que a conectividade decresce, a superposição com o conceito e a habilidade de categorização da rede diminuiem na criticalidade.

A melhoria na habilidade de categorização com o aumento da diluição fica evidenciada na figura 4.4, onde mostramos as curvas para o erro de categorização $\epsilon(s)$ em função do número de exemplos, para conectividades $c = 0$, $c = 0.1$, $c = 0.5$ e $c = 1$ com $\alpha = 0.03125$ e $b = 0.5$, tanto para $T = 0$, quanto para $T = 0.8$. Para um número de exemplos $s$ menor que o valor crítico $s_c$, observamos um erro de 0.5 devido à superposição com o conceito $m^1$ ser nula e à existência de estados de vidro de spin. Quando o número de exemplos usados para treinar a rede encontra o valor crítico $s_c$, observamos uma queda abrupta no erro de categorização,

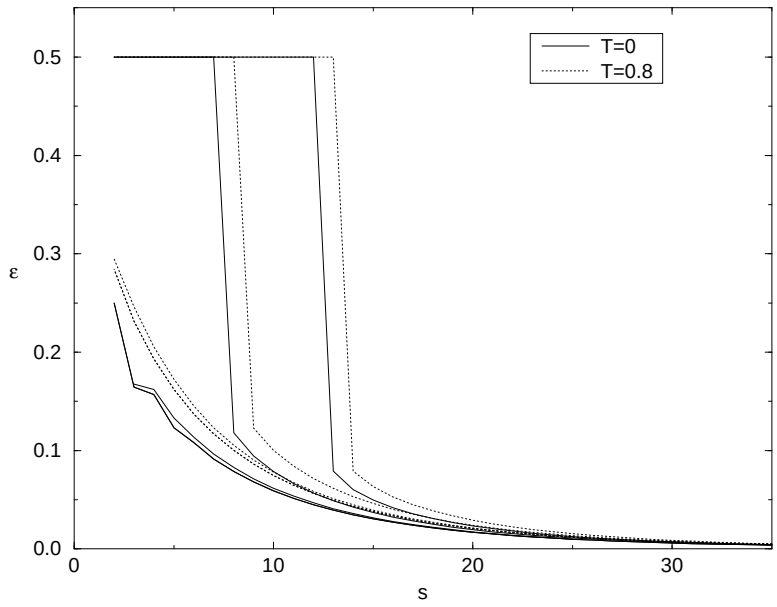



**Fig. 4.4:** *Curvas de categorização* $\epsilon(s)$ *para* $c = 0,\ 0.1,\ 0.5$ *e* $1$*, da esquerda para a direita, com* $\alpha = 0.03125$*,* $b = 0.5$ *a* $T = 0$ *e* $T = 0.8$*.*

sinalizando a entrada da rede no regime de categorização. Esse mesmo comportamento foi observado na rede completamente conexa. Quanto menor a conectividade da rede, tanto menor é o número crítico de exemplos $s_c$ com os quais devemos treinar a rede para atingirmos o regime de categorização. Desse modo, observamos que a rede diluída categoriza com um número menor de exemplos que a rede completamente conexa.

Na figura 4.5, podemos observar o efeito do ruído estocástico na performance da rede na fase de categorização devido à presença de um número macroscópico de conceitos. As curvas de categorização foram obtidas para vários valores de $\alpha/\alpha_0$ ($\alpha_0 = 2/\pi$), com $b = 0.2$, $c = 0$ e $T = 0$ (sem ruído sináptico). O ponto de partida $s = 1$ corresponde à recuperação de um exemplo. O erro de categorização inicialmente decresce monotonamente com o aumento do número de exemplos $s$ para baixos valores de $\alpha/\alpha_0$, enquanto que aumenta, para valores maiores de $\alpha/\alpha_0$, até um determinado número de exemplos, diminuindo posteriormente. A razão para esse comportamento é a competição entre estados simétricos que favorecem a categorização e estados de vidro de spin que tendem a destruí-la. Para valores de $\alpha/\alpha_0$

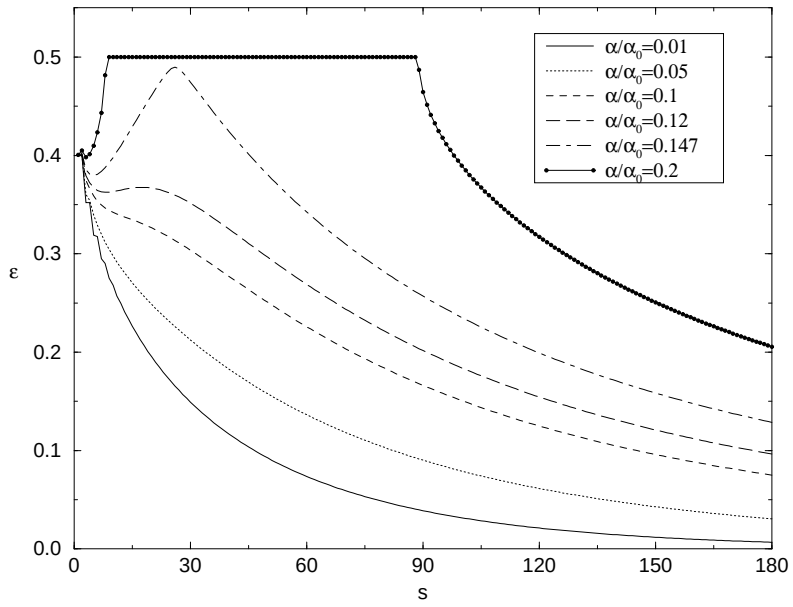



**Fig. 4.5:** *Curvas de categorização $\epsilon(s)$ a $T = 0$, $b = 0.2$ e $c = 0$, para vários valores de $\alpha/\alpha_0$, como indicado, onde $\alpha_0 = 2/\pi$ é a capacidade de armazenamento do modelo de Hopfield extremamente diluído.*

maiores de aproximadamente 0.15, encontram-se, de maneira contínua, a fase de vidro de spin onde o erro de categorização é 0.5, indicando a incapacidade da rede em categorizar. Com o aumento do número de exemplos, a rede passa continuamente para o regime de categorização, reconhecendo os conceitos. Resultados similares são obtidos para outros valores do parâmetro de correlação $b$. Para valores maiores de $b$, é necessário um número menor de exemplos, para que a rede opere como um dispositivo de categorização.

A figura 4.6 apresenta as curvas de categorização para $c = 0.001$, $b = 0.2$, $T = 0$ e vários valores de $\alpha/\alpha_0$. Observamos que, quando o valor crítico $\alpha/\alpha_0 = 0.0936$ é atingido, a curva de erro de categorização passa a apresentar uma descontinuidade no contorno de fase, tendo um platô em 0.5 para $s < s_c$, correspondendo à fase de vidro de spin, em concordância com a natureza da transição discutida no contexto da figura 4.1. Os resultados, em ambos os casos, indicam que um aumento no nível de ruído estocástico, devido ao número macroscópico de conceitos, sempre prejudica a performance da rede.

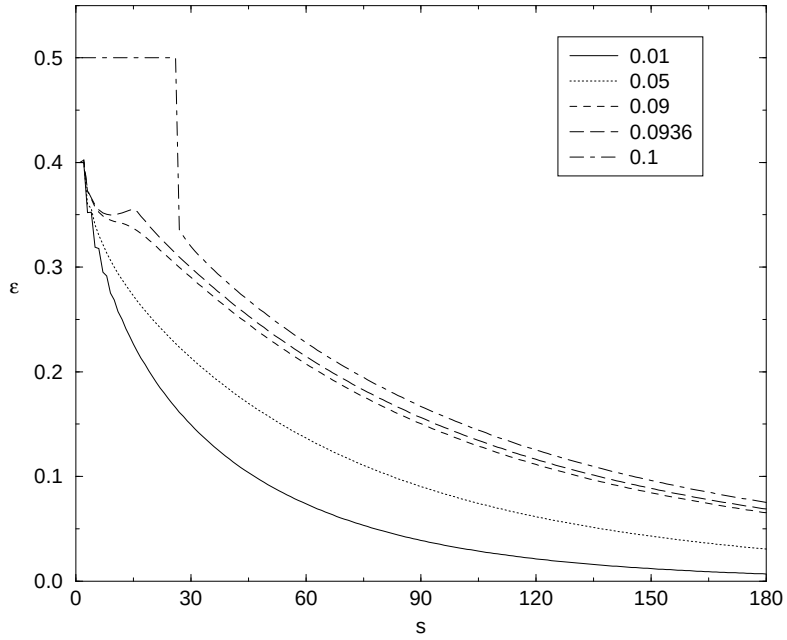



**Fig. 4.6:** *Curvas de categorização* $\epsilon(s)$ *a* $T = 0$, $b = 0.2$ *e* $c = 0.001$, *para vários valores de* $\alpha/\alpha_0$, *como indicado.*

Para valores relativamente altos de ruído estocástico $\alpha$, podemos observar o efeito do ruído sináptico $T$ sobre a habilidade de categorização da rede. As figuras 4.7 e 4.8 apresentam as curvas de categorização com $\alpha/\alpha_0 = 0.3$, $b = 0.3$, para $c = 0$ e $c = 0.001$, respectivamente. Podemos observar que, para $c = 0$ e valores de ruído inferiores a aproximadamente $T = 0.4$, o erro de categorização aumenta com o número de exemplos até aproximadamente $s = 12$ exemplos, acima do qual o erro de categorização diminui progressivamente. O aumento do ruído desde $T = 0$ até aproximadamente $T = 1.2$, favorece a categorização. Para valores acima de $T = 1.2$ é necessário um número de exemplos crescente para que a rede atinja a fase de categorização. Observa-se que a rede é robusta ao ruído, pois para valores relativamente altos deste, é ainda possível categorizar, bastando que se exponha a rede a um número suficientemente alto de exemplos. Para $c = 0.001$, observamos que o aumento do ruído sináptico favorece a categorização, porém a transição é descontínua.

A dependência da performance da categorização para ambos tipos de ruído (estocástico e sináptico) pode ser melhor observada na evolução do erro de categorização com o aumento de $\alpha$ ou $T$, apresentada na figura 4.9 para $c = 0$, $b = 0.3$ e $s = 10$ exemplos. Em contraste

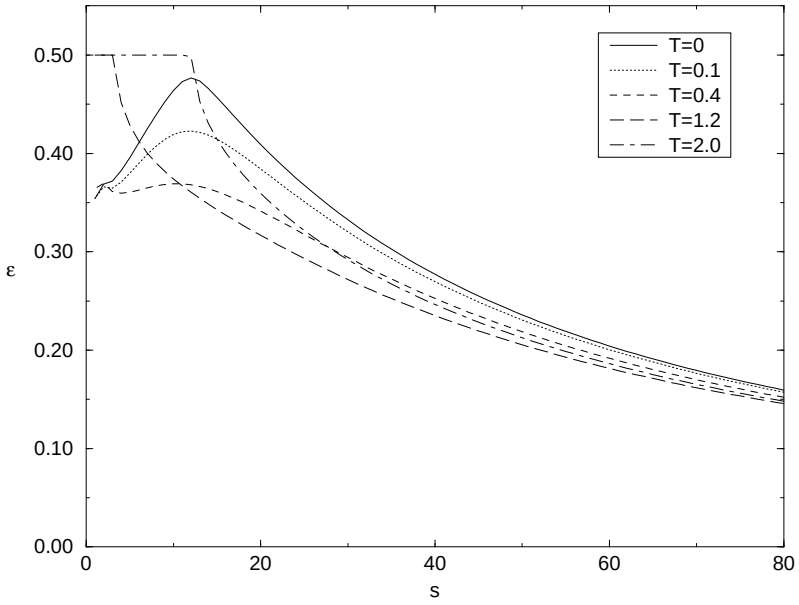



**Fig. 4.7:** *Curvas de categorização $\epsilon(s)$ com $\alpha/\alpha_0 = 0.3$, $b = 0.3$, e vários valores de temperatura $T$ para $c = 0$.*

com a dependência em $\alpha$, a dependência em $T$ é claramente não monótona para valores pequenos de $\alpha$. Nesse caso, o aumento do ruído sináptico provoca um aumento no erro de categorização. Porém, com o aumento do ruído estocástico, observamos uma inversão no comportamento do erro de categorização resultando na diminuição desse, à medida que aumentamos o ruído sináptico. Como resultado, o efeito do aumento no nível de ruído estocástico pode ser compensado, até certo ponto, pelo aumento do ruído sináptico $T$, permitindo que a rede represente um maior número de conceitos.

A figura 4.10 mostra os valores críticos das superposições dos estados simétricos $m_s$ e do conceito $m^1$, do erro crítico de categorização e da capacidade crítica de armazenamento $\alpha_c$ em função da conectividade $c$, para $T = 0$. A linha crítica $\alpha_c(c)$ separa as fases de categorização e vidro de spin. Para valores de $\alpha \leq \alpha_c(c)$, a rede encontra-se na fase de categorização, enquanto que, para valores de $\alpha > \alpha_c(c)$, a rede encontra-se na fase de vidro de spin. Observa-se que a capacidade crítica de armazenamento $\alpha_c$ aumenta com o aumento da diluição (diminuição da conectividade).

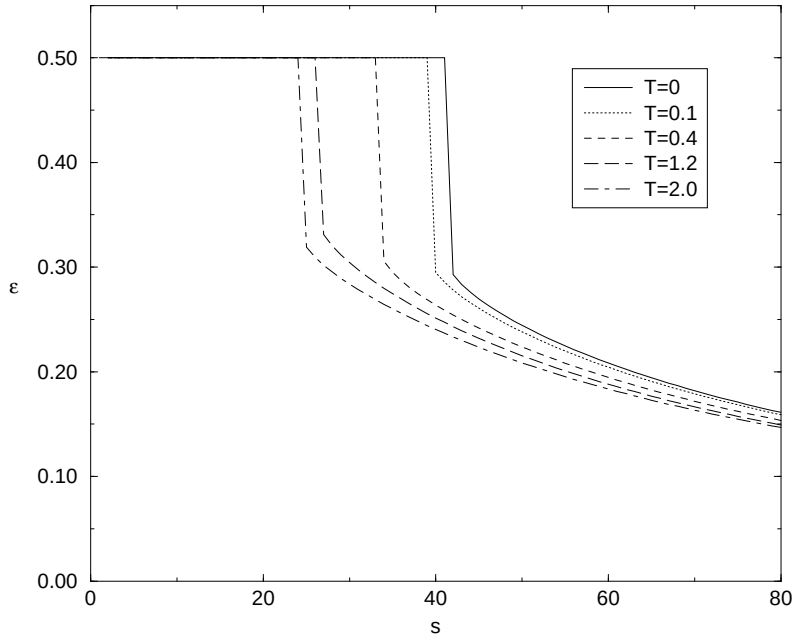



**Fig. 4.8:** *Curvas de categorização* $\epsilon(s)$ *com* $\alpha/\alpha_0 = 0.3$, $b = 0.3$, *e vários valores de temperatura* $T$ *para* $c = 0.001$.

Nos limites em que a correlação entre exemplos e conceitos tende a um ($b = 1$), o número de exemplos $s$ tende a um ($s = 1$) e $c \to 0$, o valor crítico de $\alpha_c \to 2/\pi$, que é o resultado conhecido no problema de memorização na rede extremamente diluída [57] [56].

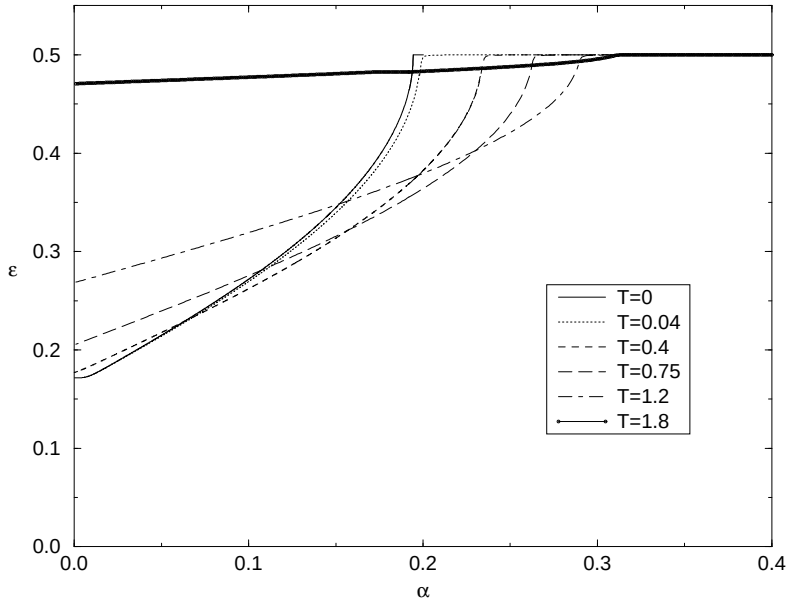



**Fig. 4.9:** *Dependência do erro de categorização $\epsilon$ com $\alpha$ para vários valores de temperatura $T$, com $b = 0.3$, $s = 10$ e $c = 0$.*

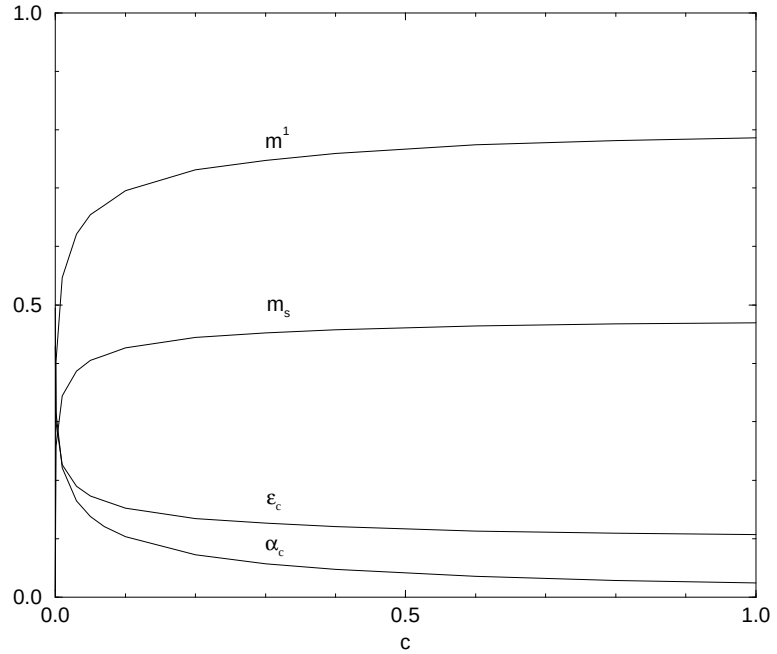



**Fig. 4.10:** *Valores críticos do erro de categorização $\epsilon_c$, da capacidade de armazenamento $\alpha_c$, das superposições $m_s$ e $m^1$ em função da conectividade $c$, para $T = 0$, $b = 0.5$ e $s = 10$.*

# Conclusões

O objetivo central desta tese é analisar os efeitos da introdução de ingredientes inspirados na biologia, tais como o ruído sináptico e a diluição sináptica, na habilidade de categorização do modelo de Hopfield. Para realizar esta análise, utilizando-se a mecânica estatística de equilíbrio, é necessário impor certas simplificações que, por vezes, não são biologicamente plausíveis, como a simetria das conexões sinápticas. Entretanto, essas simplificações mantêm as características essenciais que reproduzem comportamentos biológicos de interesse.

O modelo de Hopfield foi intensamente estudado como um dispositivo de memória endereçada por conteúdo (memória associativa). As variações na regra de aprendizagem buscavam sempre um aumento na capacidade de armazenamento de padrões. O aparecimento de estados simétricos, cuja superposição é igual para um conjunto de padrões, sempre foi visto como um efeito negativo do modelo. Algumas regras de aprendizagem foram capazes de suprimir completamente esses estados. Porém, ao se pretender que o modelo de Hopfield não apenas operasse como um dispositivo de memória associativa, mas também realizasse tarefas computacionais mais elaboradas como generalização e categorização, verificou-se que os estados simétricos desempenham um papel fundamental [42]-[45],[47]-[50], [54, 55],[60]-[73]. Esses estados capturam informações comuns a conjuntos de padrões de maneira espontânea, representando, portanto, manifestações autônomas da própria rede. É de importância fundamental compreender as relações entre a estrutura estatística dos padrões utilizados no processo de treinamento da rede e a estrutura das representações criadas por ela (mínimos de energia), que extraem as informações comuns aos dados de treinamento (exemplos) e que governam o processo dinâmico de recuperação destas informações (conceitos).

A introdução do ruído sináptico se justifica pela observação biológica de neurônios que emitem potenciais de ação sem serem estimulados. Esse comportamento atua como um ruído sináptico e pode ser modelado formalmente através da temperatura. Para compreendermos em que medida o ruído sináptico (temperatura) modifica a habilidade de categorização do modelo de Hopfield, analisamos separadamente o caso em que o número de conceitos é finito ($\alpha = 0$) e o caso em que o número de conceitos é proporcional ao número de neurônios ($p = \alpha N$) [44].

Para um número finito de conceitos, $\alpha = 0$, observamos a existência de estados de recuperação de exemplos, correspondendo às soluções assimétricas ($m^{11} \neq m_{s-1} \neq 0$), para um número de exemplos $s$ menor que o valor crítico $s_c$, a partir do qual a rede entra descontinuamente no regime de categorização, onde apenas a solução simétrica ($m_s$) existe (ver figura 3.2), além da solução trivial. Com o aumento do ruído sináptico $T$, o número crítico de exemplos $s_c$ que devem ser apresentados à rede, para que se inicie a categorização, diminui gradativamente até $T_P = 1$, evidenciando o favorecimento da categorização. Entretanto, como podemos verificar na figura 3.4, o erro de categorização aumenta com o ruído para um número fixo de exemplos. Para valores de $T > T_P$, identificamos uma fase paramagnética, onde $m^{11} = m_{s-1} = m_s = 0$, desfavorecendo a categorização. Porém, aumentando o número de exemplos, é sempre possível alcançar a fase de categorização. A transição da fase paramagnética para a fase de categorização é de segunda ordem. Com a diminuição da correlação entre exemplos e conceitos ($b$), a fase de recuperação de exemplos aumenta em detrimento da fase de categorização, porém o ruído sináptico continua a favorecer a categorização. Observamos, também, que o controle do nível de ruído sináptico na rede permite escolher entre a recuperação de um exemplo ou o reconhecimento de um conceito, quando $s < s_c$. As simulações de Monte Carlo que realizamos corroboram os resultados obtidos através da teoria de campo médio com uma concordância notável.

Para um número de conceitos proporcional ao número de neurônios ($\alpha \neq 0$), construímos o diagrama de fases ($\alpha \times T$) onde identificamos as fases de categorização, vidro de spin e paramagnética (ver figura 3.5). Tanto na fase vidro de spin, quanto na paramagnética, a

rede é incapaz de categorizar e/ou recuperar exemplos. A transição de fase categorização-vidro de spin é de primeira ordem, enquanto que a transição vidro de spin-paramagnética é de segunda ordem. Nesse caso, o ruído desfavorece a categorização devido à competição entre os estados simétricos e de vidro de spin, como pode ser verificado nas figuras 3.7 e 3.8, não sendo possível controlar a recuperação de exemplos ou conceitos, como observado para $\alpha = 0$. A recuperação de exemplos acontece para pequenos valores de correlação $b$ e exemplos $s$. A fase de categorização aumenta com o número de exemplos (ver figura 3.6) e com a correlação entre entre exemplos e conceitos. Um aspecto interessante que observamos no diagrama $(\alpha \times T)$ é a forma reentrante da linha $T_c$ para baixos valores de ruído sináptico, que é uma conseqüência da aproximação de réplicas simétricas. Os resultados obtidos através de simulações de Monte Carlo confirmam as previsões da teoria de campo médio com simetria de réplicas e evidenciam um forte efeito do tamanho finito do sistema.

Os sistemas biológicos apresentam estruturas neurais complexas que conectam grupos de neurônios. Tipicamente, um neurônio conecta-se em média a outros $10^4$ neurônios. Como uma primeira aproximação dessa característica biológica, mantendo ainda uma conectividade de ordem $\mathcal{O}(N)$ no limite termodinâmico, estudamos o efeito da diluição simétrica das conexões sinápticas no modelo de Hopfield [50]. Do ponto de vista tecnológico, um dispositivo neural eletrônico seria beneficiado com uma arquitetura de conexões diluída, consumindo menor quantidade de energia e facilitando sua dissipação.

Tendo em mente essas considerações, investigamos os efeitos da diluição simétrica através da teoria de campo médio com a aproximação de simetria de réplicas. A figura 4.1 exibe o efeito da diluição sobre a rede, evidenciando o favorecimento da categorização com a diminuição da conectividade $c$, o que pode também ser verificado nas figuras 4.3, 4.4 e 4.10, esta última mostrando o aumento do valor crítico de armazenamento $\alpha_c$ para redes diluídas. Identificamos as fases de categorização, vidro de spin e paramagnética, nas quais observamos o aumento significativo da fase de categorização, à medida em que a rede é diluída. Em particular, verifica-se que o ruído sináptico desempenha um papel interessante, pois, para diluição finita, seu efeito é desfavorecer a categorização, enquanto que, no limite

de diluição extrema $c = 0$, produz um favorecimento da categorização. Um resultado importante é a desestabilização dos estados de vidro de spin em toda a fase de categorização no limite de diluição extrema. Para valores finitos de diluição, essa desestabilização é gradual. A transição de fase paramagnética-vidro de spin é sempre de segunda ordem, enquanto que a transição categorização-vidro de spin é de segunda ordem no limite de diluição extrema e de primeira ordem para valores intermediários de diluição. A comparação das figuras 4.1 e 4.2 indica que o aumento da correlação entre exemplos e conceitos $b$ favorece a categorização. O ruído estocástico, que representa o efeito do parâmetro $\alpha$, atua no sentido de desfavorecer a categorização, exigindo um número maior de exemplos à medida que seu valor aumenta (ver figuras 4.5 e 4.6). O ruído sináptico desempenha um papel de favorecimento da categorização até valores moderados, a partir dos quais passa a desfavorecê-la (figuras 4.7 e 4.8), dependendo do número de exemplos.

Como possíveis extensões do estudo desenvolvido nesta tese, podemos citar a análise dos efeitos do ruído sináptico e da diluição simétrica para conceitos correlacionados e a realização de simulações de Monte Carlo para esse caso, bem como considerar uma estrutura hierárquica de padrões com mais de dois níveis. Outra extensão interessante seria investigar o papel da relação entre a estrutura da rede e função, no problema de categorização no modelo de Hopfield, para topologias de redes regulares, redes aleatórias, redes sem escala de Barabási-Albert e redes "small-world" de Watts-Strogatz que combinam regularidade e aleatoriedade. Essas topologias poderiam contemplar, de maneira mais realista, a diluição observada em sistemas biológicos.

Todos os resultados obtidos nesta tese estão em consonância com as idéias de que a habilidade de categorização está associada às limitações da memória associativa. Essas idéias podem ser resumidas na seguinte frase de M. A. Virasoro: *"we categorize not because we want to but because we cannot do otherwise"* [49].

## Apêndice A

# Geração de Padrões Hierárquicos

No terceiro nível, geramos $p_3$ padrões $\{\xi_i^{\mu\nu\lambda}\}$, com $\lambda = 1, ..., p_3$, para cada padrão $\{\xi_i^{\mu\nu}\}$, a partir da probabilidade

$$P(\xi_i^{\mu\nu\lambda}) = \frac{1}{2}(1 + c\xi_i^{\mu\nu})\delta(\xi_i^{\mu\nu\lambda} - 1) + \frac{1}{2}(1 - c\xi_i^{\mu\nu})\delta(\xi_i^{\mu\nu\lambda} + 1). \tag{A.1}$$

Para esses padrões, a correlação e a atividade podem ser escritas em termos das correlações e atividades dos níveis anteriores

$$\langle \xi_i^{\mu\nu\lambda}\xi_i^{\mu'\nu'\lambda'}\rangle = [c^2 + (1-c^2)\delta_{\mu\mu'}\delta_{\nu\nu'}\delta_{\lambda\lambda'}]\underbrace{[b^2 + (1-b^2)\delta_{\mu\mu'}\delta_{\nu\nu'}][a^2 + (1-a^2)\delta_{\mu\mu'}]}_{\langle \xi_i^{\mu\nu}\xi_i^{\mu'\nu'}\rangle}, \tag{A.2}$$

$$\langle \xi_i^{\mu\nu\lambda}\rangle = \langle \xi_i^{\mu\nu}\rangle c. \tag{A.3}$$

As correlações com os padrões dos níveis precedentes são

$$\langle \xi_i^{\mu\nu\lambda}\xi_i^{\mu'\nu'}\rangle = \langle \xi_i^{\mu\nu}\xi_i^{\mu'\nu'}\rangle c, \tag{A.4}$$

$$\langle \xi_i^{\mu\nu\lambda}\xi_i^{\mu'}\rangle = \langle \xi_i^{\mu\nu}\xi_i^{\mu'}\rangle c. \tag{A.5}$$

Esse procedimento pode ser generalizado para uma estrutura hierárquica de $k$ níveis, utilizando-se os $p_{k-1}$ padrões $\{\xi_i^{\alpha_1,...,\alpha_{k-1}}\}$ para gerar os $p_k$ descendentes $\{\xi_i^{\alpha_1,...,\alpha_k}\}$ a partir da distribuição de probabilidade

$$P(\xi_i^{\alpha_1,...,\alpha_k}) = \frac{1}{2}(1 + a_k\xi_i^{\alpha_1,...,\alpha_{k-1}})\delta(\xi_i^{\alpha_1,...,\alpha_k} - 1) + \frac{1}{2}(1 - a_k\xi_i^{\alpha_1,...,\alpha_{k-1}})\delta(\xi_i^{\alpha_1,...,\alpha_k} + 1), \tag{A.6}$$

originando correlações e atividades que podem ser obtidas a partir das seguintes relações de recorrência

$$\langle \xi_i^{\alpha_1,...,\alpha_k} \xi_i^{\alpha_1',...,\alpha_k'} \rangle = [a_k^2 + (1 - a_k^2) \prod_{j=1}^{k} \delta_{\alpha_j \alpha_j'}] \langle \xi_i^{\alpha_1,...,\alpha_{k-1}} \xi_i^{\alpha_1',...,\alpha_{k-1}'} \rangle, \tag{A.7}$$

$$\langle \xi_i^{\{\alpha\}} \xi_i^{\{\beta\}} \rangle = \langle \xi_i^{\{\beta\}} \xi_i^{\{\gamma\}} \rangle a_k, \tag{A.8}$$

$$\langle \xi_i^{\{\alpha\}} \rangle = \prod_{j=1}^{k} a_j, \tag{A.9}$$

onde $\{\alpha\} = \alpha_1, ..., \alpha_k$, $\{\beta\} = \beta_1, ..., \beta_l$ e $\{\gamma\} = \gamma_1, ..., \gamma_{k-1}$ com $l < k$. O valor do parâmetro de atividade $a$ determina se os padrões terão mais neurônios ativos ($\xi_i^\mu = +1$) que inativos ($\xi_i^\mu = -1$) e, conseqüentemente, a correlação entre padrões pertencentes a classes diferentes.

## Apêndice B

# Cálculo das Médias para $\alpha = 0$

**Obtenção das equações (3.47), (3.48), (3.51) e (3.53):**

Partindo da equação (3.43), procedemos ao cálculo da média na ordem indicada. Calculando-se a média sobre o exemplo $\xi^{11}$, com a distribuição (3.4), temos

$$\begin{aligned} m^{11} &= \left\langle \left(\frac{1+b\xi^1}{2}\right) \left\langle \tanh[\beta(m^{11}+m_{s-1}x_{s-1})]\right\rangle_{\{x_{s-1}\}} \right\rangle_{\{\xi^1\}} \\ &+ \left\langle \left(\frac{1-b\xi^1}{2}\right) \left\langle \tanh[\beta(m^{11}-m_{s-1}x_{s-1})]\right\rangle_{\{x_{s-1}\}} \right\rangle_{\{\xi^1\}}, \end{aligned} \tag{B.1}$$

onde utilizamos a relação $\tanh(x) = -\tanh(-x)$. Explicitamos, agora, a média sobre $x_{s-1}$

$$\begin{aligned} m^{11} &= \left\langle \left(\frac{1+b\xi^1}{2}\right) \sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} \left(\frac{1+b\xi^1}{2}\right)^k \left(\frac{1-b\xi^1}{2}\right)^{s-1-k} \right. \\ &\times \quad \tanh[\beta(m^{11}+m_{s-1}(2k-s+1))] \\ &+ \quad \left(\frac{1-b\xi^1}{2}\right) \sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} \left(\frac{1+b\xi^1}{2}\right)^k \left(\frac{1-b\xi^1}{2}\right)^{s-1-k} \\ &\times \quad \left. \tanh[\beta(m^{11}-m_{s-1}(2k-s+1))] \right\rangle_{\{\xi^1\}}. \end{aligned} \tag{B.2}$$

Calculando a média sobre o conceito $\xi^1$, com a distribuição (3.14), obtemos

$$\begin{aligned} m^{11} &= \frac{1}{2}\sum_{k=0}^{s-1} \begin{pmatrix} s-1 \\ k \end{pmatrix} \left\{ \left[ \left(\frac{1+b}{2}\right)^{k+1} \left(\frac{1-b}{2}\right)^{s-1-k} + \left(\frac{1-b}{2}\right)^{k+1} \left(\frac{1+b}{2}\right)^{s-1-k} \right] \right. \\ &\times \quad \tanh[\beta(m^{11}+m_{s-1}(2k-s+1))] \\ &+ \quad \left. \left[ \left(\frac{1+b}{2}\right)^{k} \left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1-b}{2}\right)^{k} \left(\frac{1+b}{2}\right)^{s-k} \right] \right. \end{aligned}$$

$$\times \quad \tanh[\beta(m^{11} - m_{s-1}(2k-s+1))]\Bigg\}. \tag{B.3}$$

que é a equação (3.47). Esse mesmo procedimento é utilizado para obtermos as equações (3.48) e (3.51).

No caso da densidade de energia livre, equação (3.53), partimos da equação (3.37), na qual inserimos a solução de recuperação dada pela equação (3.41), tornando-se

$$f_r = \frac{1}{2}[(m^{11})^2 + (s-1)m_{s-1}^2] - \frac{1}{\beta}\underbrace{\langle \ln 2\cosh[\beta(m^{11}\xi^{11} + m_{s-1}x_{s-1})]\rangle_{\{\xi^{11}\}\{x_{s-1}\}\{\xi^1\}}}_{f_{\ln}}. \tag{B.4}$$

A média é calculada seguindo o procedimento descrito acima, utilizado no cálculo das superposições. Inicialmente calculamos a média sobre os exemplos

$$\begin{aligned} f_{\ln} &= \left\langle \left(\frac{1+b\xi^1}{2}\right) \ln 2\cosh[\beta(m^{11} + m_{s-1}x_{s-1})] \right\rangle_{\{x_{s-1}\}\{\xi^1\}} \\ &+ \left\langle \left(\frac{1-b\xi^1}{2}\right) \ln 2\cosh[\beta(-m^{11} + m_{s-1}x_{s-1})] \right\rangle_{\{x_{s-1}\}\{\xi^1\}}. \end{aligned} \tag{B.5}$$

Explicitando a média sobre $x_{s-1}$, temos

$$\begin{aligned} f_{\ln} &= \sum_{k=0}^{s-1} \binom{s-1}{k} \Bigg\langle \left(\frac{1+b\xi^1}{2}\right)^{k+1} \left(\frac{1-b\xi^1}{2}\right)^{s-1-k} \ln 2\cosh[\beta(m^{11} + m_{s-1}(2k-s+1))] \\ &+ \left(\frac{1+b\xi^1}{2}\right)^{k} \left(\frac{1-b\xi^1}{2}\right)^{s-k} \ln 2\cosh[\beta(m^{11} - m_{s-1}(2k-s+1))] \Bigg\rangle_{\{\xi^1\}}, \end{aligned} \tag{B.6}$$

onde utilizamos a relação $\cosh(x) = \cosh(-x)$.

Finalmente, calculamos a média sobre o conceito $\xi^1$, resultando

$$\begin{aligned} f_{\ln} &= \frac{1}{2}\sum_{k=0}^{s-1} \binom{s-1}{k} \Bigg\{ \left[ \left(\frac{1+b}{2}\right)^{k+1} \left(\frac{1-b}{2}\right)^{s-1-k} + \left(\frac{1+b}{2}\right)^{s-1-k} \left(\frac{1-b}{2}\right)^{k+1} \right] \\ &\times \quad \ln 2\cosh[\beta(m^{11} + m_{s-1}(2k-s+1))] \\ &+ \left[ \left(\frac{1+b}{2}\right)^{k} \left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1+b}{2}\right)^{s-k} \left(\frac{1-b}{2}\right)^{k} \right] \\ &\times \quad \ln 2\cosh[\beta(m^{11} - m_{s-1}(2k-s+1))] \Bigg\}. \end{aligned} \tag{B.7}$$

Desse modo, a densidade de energia livre resulta em

$$f_r = \frac{1}{2}[(m^{11})^2 + (s-1)m_{s-1}^2] - \frac{1}{\beta} f_{\ln}, \tag{B.8}$$

que é a equação (3.53).

**Obtenção das equações (3.61), (3.62) e (3.64):**

A equação (3.61) é obtida introduzindo-se a solução de categorização $m^{\mu\nu} = m_s \delta_{\mu 1}$ na equação (3.39) resultando na equação (3.59). Inicialmente realiza-se o cálculo da média sobre os exemplos, resultando em

$$m_s = \frac{1}{s} \langle \langle x_s \tanh(\beta m_s x_s) \rangle_{\{x_s\}} \rangle_{\{\xi^1\}}. \tag{B.9}$$

Explicitando a média sobre $x_s$, temos

$$m_s = \frac{1}{s} \left\langle \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} \left(\frac{1+b\xi^1}{2}\right)^k \left(\frac{1-b\xi^1}{2}\right)^{s-k} (2k-s) \tanh[\beta m_s (2k-s)] \right\rangle_{\{\xi^1\}}. \tag{B.10}$$

Calculando a média sobre o conceito $\xi^1$, obtemos

$$m_s = \frac{1}{2s} \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} \left[ \left(\frac{1+b}{2}\right)^k \left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1+b}{2}\right)^{s-k} \left(\frac{1-b}{2}\right)^k \right] (2k-s) \tanh[\beta m_s (2k-s)], \tag{B.11}$$

que é a equação (3.61). Procedendo da mesma forma, obtemos a equação (3.62).

A densidade de energia livre (3.64) é obtida a partir da equação (3.37) de modo similar à obtenção da equação (3.55), porém com a solução de categorização, resultando, após a média sobre os exemplos, em

$$f_c = \frac{1}{2} s \; m_s^2 - \frac{1}{\beta} \underbrace{\langle \ln 2 \cosh[\beta(m_s x_s)] \rangle_{\{x_s\}\{\xi^1\}}}_{f_{\ln}}. \tag{B.12}$$

Calculando a média sobre $x_s$, temos

$$f_{\ln} = \left\langle \sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} \left(\frac{1+b\xi^1}{2}\right)^k \left(\frac{1-b\xi^1}{2}\right)^{s-k} \ln 2 \cosh[\beta m_s (2k-s)] \right\rangle_{\{\xi^1\}}, \tag{B.13}$$

e, realizando a média sobre o conceito, temos

$$f_{\ln} = \frac{1}{2}\sum_{k=0}^{s}\begin{pmatrix} s \\ k \end{pmatrix}\left[\left(\frac{1+b}{2}\right)^{k}\left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1+b}{2}\right)^{s-k}\left(\frac{1-b}{2}\right)^{k}\right]\ln 2\cosh[\beta m_s(2k-s)]. \tag{B.14}$$

Obtemos, assim, a equação (3.64)

$$f_c = \frac{1}{2}s\ m_s^2 - \frac{1}{\beta}f_{\ln}. \tag{B.15}$$

**Obtenção das equações (3.57) e (3.67):**

Para obtermos as energias livres no limite de $T = 0$ ($\beta \to \infty$), partimos da equação (3.37) escrita em notação vetorial

$$f = \frac{1}{2}\vec{m}\cdot\vec{m} - \frac{1}{\beta}\langle \ln 2\cosh[\beta(\vec{m}\cdot\vec{\xi})]\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}. \tag{B.16}$$

No limite $\beta \to \infty$, temos

$$\lim_{\beta\to\infty}\frac{1}{\beta}\langle \ln 2\cosh[\beta(\vec{m}\cdot\vec{\xi})]\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}} = \langle|\vec{m}\cdot\vec{\xi}|\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}} \tag{B.17}$$

e a energia livre passa a ser escrita como

$$f = \frac{1}{2}\vec{m}\cdot\vec{m} - \langle|\vec{m}\cdot\vec{\xi}|\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}. \tag{B.18}$$

A superposição dos exemplos dada pela equação (3.39) é

$$\vec{m} = \langle\vec{\xi}\tanh[\beta(\vec{m}\cdot\vec{\xi})]\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}} \tag{B.19}$$

e, no limite $\beta \to \infty$, usamos a relação $\lim_{\beta\to\infty}\tanh[\beta(\vec{m}\cdot\vec{\xi})] = \mathrm{sgn}(\vec{m}\cdot\vec{\xi})$ para escrevê-la como

$$\vec{m} = \langle\vec{\xi}\,\mathrm{sgn}(\vec{m}\cdot\vec{\xi})\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}. \tag{B.20}$$

Porém, multiplicando escalarmente pela esquerda a equação acima por $\vec{m}$, obtemos

$$\vec{m}\cdot\vec{m} = \langle(\vec{m}\cdot\vec{\xi})\,\mathrm{sgn}(\vec{m}\cdot\vec{\xi})\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}, \tag{B.21}$$

e, notando que $x\,\mathrm{sgn}(x) = |x|$, resulta

$$\vec{m}\cdot\vec{m} = \langle|\vec{m}\cdot\vec{\xi}|\rangle_{\{\xi^{\mu\nu}\}\{x_s\}\{\xi^{\mu}\}}. \tag{B.22}$$

Desse modo, a densidade de energia livre pode ser escrita como

$$f = \frac{1}{2}(\vec{m} \cdot \vec{m}) - (\vec{m} \cdot \vec{m}) = -\frac{1}{2}(\vec{m} \cdot \vec{m}). \tag{B.23}$$

Substituindo nessa expressão geral a solução de recuperação e de categorização, obtemos, respectivamente, a equação (3.57)

$$f_r = -\frac{1}{2}[(m^{11})^2 + (s-1)m_{s-1}^2], \tag{B.24}$$

e a equação (3.67)

$$f_c = -\frac{1}{2}m_s^2. \tag{B.25}$$

## Apêndice C

# Obtenção das Densidades de Energia Livre para $\alpha \neq 0$

**Obtenção da equação (3.93):**

A partir da equação (3.90), utilizamos o método do ponto de sela para calcular a integral. No limite termodinâmico, o integrando é dominado pelos seus pontos de sela de tal forma que a função de partição é dada por

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} = \exp\{-N\beta[\frac{1}{2}\sum_{\rho\nu}(m_{\rho}^{1\nu})^2 + \frac{\alpha\beta}{2}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} - \frac{\alpha}{\beta}\ln G(q_{\rho\sigma}) - \frac{1}{\beta}\langle \ln Tr_{S^{\rho}} e^{\beta H_{\xi}}\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}]\}. \tag{C.1}$$

Como a densidade de energia livre é dada por

$$f = \lim_{N\to\infty, n\to 0} \frac{1}{N\beta}\frac{\langle Z^n\rangle - 1}{n}, \tag{C.2}$$

expandimos a exponencial, resultando

$$\langle Z^n \rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} = 1 - \frac{\beta N}{2}\sum_{\rho\nu}(m_{\rho}^{1\nu})^2 - \frac{\alpha\beta^2 N}{2}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} + \frac{\alpha\beta N}{\beta}\ln G(q_{\rho\sigma}) - \beta N[\frac{1}{\beta}\langle \ln Tr_{S^{\rho}} e^{\beta H_{\xi}}\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}]. \tag{C.3}$$

Substituindo a expansão da exponencial na equação para a densidade de energia livre acima, obtemos a equação (3.93)

$$f = \lim_{n\to 0}\left[\frac{1}{2n}\sum_{\rho\nu}(m_{\rho}^{1\nu})^2 + \frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} - \frac{\alpha}{n\beta}\ln G(q_{\rho\sigma}) - \frac{1}{n\beta}\langle \ln Tr_{S^{\rho}} e^{\beta H_{\xi}}\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}}\right]. \tag{C.4}$$

**Obtenção da equação (3.101):**

Assumindo a simetria de réplicas, expressa pelas equações (3.98), (3.99) e (3.100), reescrevemos a equação (C.4) termo a termo, no limite $n \to 0$

$$\frac{1}{2n}\sum_{\rho\nu}(m_\rho^{1\nu})^2 = \frac{1}{2}\sum_{\nu}(m^{1\nu})^2, \tag{C.5}$$

$$\frac{\alpha\beta}{2n}\sum_{\rho\neq\sigma} q_{\rho\sigma}r_{\rho\sigma} = -\frac{\alpha\beta}{2}qr, \tag{C.6}$$

$$\frac{\alpha}{n\beta}\ln G(q_{\rho\sigma}) = \frac{\alpha}{\beta}\ln G(q), \tag{C.7}$$

pois o argumento da exponencial da equação (3.89) pode ser expresso como

$$-\frac{1}{2}\sum_{\rho\sigma\nu\lambda} y_{\rho\nu}y_{\sigma\lambda}\underbrace{[\delta_{\rho\sigma}\delta_{\nu\lambda} - \beta B_{\nu\lambda}Q_{\rho\sigma}]}_{M_{\rho\nu\sigma\lambda}}, \tag{C.8}$$

permitindo a utilização da identidade

$$\int \frac{d\vec{y}}{\sqrt{2\pi}}\exp(-\frac{1}{2}\vec{y}\cdot\tilde{M}\cdot\vec{y}) = |\tilde{M}|^{-\frac{1}{2}} \tag{C.9}$$

e $|\tilde{M}| = \prod_i \Lambda_i$, sendo $\Lambda_i$ os auto-valores da matriz $\tilde{M}$ dados por

$$\Lambda_1 = 1 - \beta[1 + (s-1)b^2][1 + (n-1)q] \qquad \text{não degenerado,} \tag{C.10}$$

$$\Lambda_2 = 1 - \beta[1 + (s-1)b^2][1 - q] \qquad (n-1) \text{ vezes degenerado,} \tag{C.11}$$

$$\Lambda_3 = 1 - \beta[1 + (n-1)q][1 - b^2] \qquad (s-1) \text{ vezes degenerado,} \tag{C.12}$$

e

$$\Lambda_4 = 1 - \beta[1 - q][1 - b^2] \qquad (n-1)(s-1) \text{ vezes degenerado.} \tag{C.13}$$

Desse modo, $G(q) = |\tilde{M}|^{-\frac{1}{2}}$ e, portanto,

$$\frac{1}{n}\ln G(q) = -\frac{1}{2n}\ln[\Lambda_1\Lambda_2^{(n-1)}\Lambda_3^{(s-1)}\Lambda_4^{(n-1)(s-1)}], \tag{C.14}$$

que resultará na equação (3.102). O último termo resulta

$$\frac{1}{n\beta}\langle \ln Tr_{S^\rho}\exp[\beta(\sum_{\rho\nu} m_\rho^{1\nu}\xi^{1\nu}S^\rho + \frac{\alpha\beta}{2}\sum_{\rho\neq\sigma} r_{\rho\sigma}S^\rho S^\sigma - h^1\xi^1\sum_\rho S^\rho)]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}} =$$
$$\frac{\alpha\beta r}{2} - \frac{1}{\beta}\int_{-\infty}^{\infty} Dz\langle \ln[2\cosh(\beta\Delta)]\rangle_{\{\xi^{\mu\nu}\}\{\xi^\mu\}}. \tag{C.15}$$

Rearranjando os termos, obtém-se a equação (3.101).

**Obtenção da equação (3.117):**

Substituindo a solução de categorização na equação (3.101), observamos que apenas o primeiro e o último termo mudam, resultando em

$$\frac{1}{2}\sum_{\nu}(m^{1\nu})^2 = \frac{1}{2}s\,m_s^2 \tag{C.16}$$

e, definindo

$$\% = \frac{1}{\beta}\int_{-\infty}^{\infty} Dz\langle \ln[2\cosh[\beta(z\sqrt{\alpha r} + \sum_{\nu} m^{1\nu}\xi^{1\nu})]]\rangle_{\{\xi^{\mu\nu}\}\{\xi^{\mu}\}} \tag{C.17}$$

obtemos, após realizar a média sobre os exemplos e sobre $x_s$

$$\begin{aligned} \% \;\; = \;\; & \langle\frac{1}{\beta}\int_{-\infty}^{\infty} Dz\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right)\left(\frac{1+b\xi^1}{2}\right)^k\left(\frac{1-b\xi^1}{2}\right)^{s-k} \\ & \ln[2\cosh[\beta(z\sqrt{\alpha r} + m_s(2k-s))]]\rangle_{\{\xi^1\}}. \end{aligned} \tag{C.18}$$

Efetuando-se a média sobre o conceito

$$\begin{aligned} \% \;\; = \;\; & \frac{1}{2\beta}\int_{-\infty}^{\infty} Dz\sum_{k=0}^{s}\left(\begin{array}{c} s \\ k \end{array}\right)\left[\left(\frac{1+b}{2}\right)^k\left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1+b}{2}\right)^{s-k}\left(\frac{1-b}{2}\right)^k\right] \\ & \ln[2\cosh[\beta(z\sqrt{\alpha r} + m_s(2k-s))]]. \end{aligned} \tag{C.19}$$

Substituindo as equações (C.16) e (C.19) na equação (3.101), obtemos a equação (3.117) para a densidade de energia livre de categorização.

**Obtenção da equação (3.121):**

No limite em que $\beta \to \infty$, apenas o terceiro e o quarto termo da equação (3.117) dependem da temperatura e precisam ser recalculados. Para o terceiro termo, temos

$$\begin{aligned} \lim_{\beta\to\infty}\frac{\alpha}{\beta}\ln G(q) \;\; = \;\; & -\lim_{\beta\to\infty}\frac{1}{2\beta}[(s-1)\ln(1-C(1-b^2)) + \ln(1-C(1-b^2+sb^2)) \\ & -\frac{\beta q s(1-C(1-b^2)(1-b^2+sb^2))}{(1-C(1-b^2))(1-C(1-b^2+sb^2))}] \end{aligned} \tag{C.20}$$

resultando

$$\lim_{\beta\to\infty}\frac{\alpha}{\beta}\ln G(q) = \frac{\alpha s(1-C(1-b^2)(1-b^2+sb^2))}{2(1-C(1-b^2))(1-C(1-b^2+sb^2))}. \tag{C.21}$$

Levando em conta que $\lim_{\beta\to\infty} \frac{1}{\beta}\ln[2\cosh(\beta\Lambda)] = |\Lambda|$, podemos escrever o quarto termo como

$$\% = \lim_{\beta\to\infty} \frac{1}{\beta} \int_{-\infty}^{\infty} Dz \langle |\beta(z\sqrt{\alpha r} + m_s x_s)| \rangle_{\{x_s\}\{\xi^1\}}. \tag{C.22}$$

Efetuando as médias sobre $x_s$ e $\xi^1$, obtemos

$$\% = \frac{1}{2}\sum_{k=0}^{s} \begin{pmatrix} s \\ k \end{pmatrix} \left[ \left(\frac{1+b}{2}\right)^k \left(\frac{1-b}{2}\right)^{s-k} + \left(\frac{1+b}{2}\right)^{s-k} \left(\frac{1-b}{2}\right)^k \right] \underbrace{\int_{-\infty}^{\infty} Dz |z\sqrt{\alpha r} + m_s x_s|}_{I_m}. \tag{C.23}$$

Resolvendo a integral, obtemos

$$I_m = \sqrt{\frac{2\alpha r}{\pi}} \exp[-\frac{m_s^2(2k-s)^2}{2\alpha r}] + m_s(2k-s)\mathrm{erf}\left(\frac{m_s(2k-s)}{\sqrt{2\alpha r}}\right). \tag{C.24}$$

Substituindo as equações (C.24), (C.23) e (C.21) na equação (3.117), obtemos a equação (3.121).

# Referências


[1] PARISI, G. Attractor neural networks. Disponível em: <http://arXiv.org/abs/cond-mat/9412030>. Acesso em: 10 dez. 1994.

[2] AMIT, D. J. *Modeling brain function.* Cambridge: Cambridge University Press, 1989.

[3] MULLER, B.; REINHARDT, J.; STRICKLAND, M. T. *Neural networks:* an introduction. Berlin: Springer-Verlag, 1995.

[4] HERTZ, J.; KROGH, A.; PALMER, R. *Introduction to the theory of neural computation.* Reading: Addison-Wesley, 1991.

[5] PERETTO, P. *An introduction to the modeling of neural networks.* Cambridge: Cambridge University Press, 1992.

[6] McCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, Chicago, v. 5, p. 115-133, Dec. 1943.

[7] HEBB, D. O. *The organization of behaviour:* a neuropsychological theory. New York: John Wiley, 1949.

[8] ROSENBLATT, F. *Principles of neurodynamics:* perceptron and the theory of brain machanisms. Washington: Spartan, 1962.

[9] MINSKY, M.; PAPERT, S. *Perceptron:* an introduction to computational geometry. Cambridge: MIT Press, 1969.

[10] LITTLE, W. A. The existence of percistent states in the brain. *Math. Biosci.*, New York, v. 19, p. 101-120, 1974.

[11] HOPFIELD, J. J. Neural networks and physical systems with emergent colletive computational abilities. *Proc. Natl. Acad. Sci. USA*, Washington, v. 79, p. 2554-2558, Apr. 1982.

[12] AMIT, D. J.; GUTFREUND, H.; SOMPOLINSKY, H. Spin-glass models of neural networks. *Phys. Rev. A*, Woodbury, v. 32, n. 2, p. 1007-1018, Aug. 1985.

[13] AMIT, D. J.; GUTFREUND, H.; SOMPOLINSKY, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.*, Woodbury, v. 55, n. 14, p. 1530-1533, Sept. 1985.

[14] AMIT, D. J.; GUTFREUND, H.; SOMPOLINSKY, H. Statistical mechanics of neural networks near saturation. *Ann. Phys.*, New York, v. 173, n. 1, p. 30-67, Jan. 1987.

[15] AMIT, D. J.; GUTFREUND, H.; SOMPOLINSKY, H. Information storage in neural networks with low levels of activity. *Phys. Rev. A*, Woodbury, v. 35, n. 5, p. 2293-2303, Mar. 1987.

[16] DERRIDA, B.; GARDNER, E.; ZIPPELIUS, A. An exactly solvable asymmetric neural network model. *Europhys. Lett.*, Les Ulis, v. 4, n. 2, p. 167-173, July 1987.

[17] GARDNER, E. The space of interactions in neural network models. *J. Phys A*, Bristol, v. 21, n. 1, p. 257-270, Jan. 1988.

[18] GARDNER E.; DERRIDA, B. Optimal storage properties of neural network models. *J. Phys A*, Bristol, v. 21, n. 1, p. 271-284, Jan. 1988.

[19] SANTOS, A. F.; MATTOS, J. G. Z.; KREBS, P. R. Estudo preliminar da utilização de redes neurais na previsão de temperatura média diária para a cidade de Pelotas-RS. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 12., 2002, Foz de Iguaçu. *Anais*. 2002. p. 3885-3889. 1 CD-R0M.

[20] HUANG, K. *Statistical mechanics.* 2. ed. New York: John Wiley, 1987.

[21] LITTLE, W. A.; SHAW, G. L. Analytic study of the memory storage capacity of a neural network. *Math. Biosci.*, New York, v. 39, n. 3/4, p. 281-290, June 1978.

[22] PERETTO, P. Collective properties of neural networks: a statistical physics approach. *Biol. Cybern.*, Berlin, v. 50, n. 1, p. 51-62, 1984.

[23] KIRKPATRICK, S.; SHERRINGTON, D. Infinite-ranged models of spin-glasses. *Phys. Rev B*, New York, v. 17, n. 11, p. 4384-4403, June 1978.

[24] MÉZARD, M.; PARISI, G.; VIRASORO, M. A. *Spin glass theory and beyond.* Singapore: World Scientific, 1987.

[25] FONTANARI, J. F.; KÖBERLE, R. Information storage and retrieval in synchronous neural networks. *Phys. Rev A*, Woodbury, v. 36, n. 5, p. 2475-2477, Sept. 1987.

[26] FONTANARI, J. F.; KÖBERLE, R. Information processing in synchronous neural networks. *J. Phys.*, Les Ulis, v. 49, n. 1, p. 13-23, Jan. 1988.

[27] GLAUBER, R. J. Time-dependent statistical of the ising model. *J. Math. Phys.*, New York, v. 4, n. 2, p. 294-307, Feb. 1963.

[28] COOLEN, A. C. C. *Statistical mechanics of neural networks.* Disponível em: <http://www.mth.kcl.ac.uk/~tcoolen/>. Acesso em: 5 abr. 2003.

[29] HEERMANN, D. W. *Computer simulation methods in theoretical physics.* Berlin: Springer-Verlag, 1986.

[30] KOHRING, G. A. A high-precision study of the Hopfield model in the phase of broken replica symmetry. *J. Stat. Phys.*, New York, v. 59, n. 3/4, p. 1077-1086, May 1990.

[31] STRIEFVATER, TH.; MULLER, K-R.; KÜHN, R. Averaging and finite-size analysis for disorder: the Hopfield model. *Physica A*, Amsterdam, v. 232, n. 1/2, p. 61-73, Oct. 1996.

[32] VOLK, D. On the phase transition of Hopfield networks: another Monte Carlo study. *Int. J. Mod. Phys C*, Singapore, v. 9, n. 5, p. 693-700, July 1998.

[33] CRISANTI, A.; AMIT, D. J.; GUTFREUND, H. Saturation level of the Hopfield model for neural network. *Europhys. Lett.*, Les Ulis, v. 2, n. 4, p. 337-341, Aug. 1986.

[34] PARISI, G. The order parameter for spin glasses: a function on the interval 0-1. *J. Phys. A*, Bristol, v. 13, n. 3, p. 1101-1112, Mar. 1980.

[35] STEFFAN, H.; KÜHN, R. Replica symmetry breaking in attractor neural network models. *Z. Phys. B*, Berlin, v. 95, n. 2, p. 249-260, 1994.

[36] GYÖRGYI, G. Techniques of replica symmetry breaking and the storage problem of the McCulloch-Pitts neurons. *Phys. Rep.*, Amsterdam, v. 342, n. 4/5, p. 263-392, Feb. 2001.

[37] GARDNER, H. *A Nova ciência da mente:* uma história da revolução cognitiva. São Paulo: Edusp, 1995.

[38] PARGA, N.; VIRASSORO, M. A. The ultrametric organization of memories in a neural network. *J. Phys.*, Les Ulis, v. 47, n. 11, p. 1857-1864, Nov. 1986.

[39] DOTSENKO, V. Hierarchical model of memory. *Physica A*, Amsterdam, v. 140, n. 1/2, p. 410-415, Dec. 1986.

[40] FEIGELMAN, M. V.; IOFFE, L. B. The augmented models of associative memory asymmetric interaction and hierarchy of patterns. *Int. J. Mod. Phys. B*, Singapore, v. 1, n. 1, p. 51-68, Apr. 1987.

[41] GUTFREUND, H. Neural networks with hierarchically correlated patterns. *Phys. Rev. A*, Woodbury, v. 37, n. 2, p. 570-577, Jan. 1988.

[42] FONTANARI, J. F.; MEIR, R. Learning noisy patterns in a Hopfield network. *Phys. Rev. A*, Woodbury, v. 40, n. 5, p. 2806-2809, Sept. 1989.

[43] FONTANARI, J. F. Generalization in a Hopfield network. *J. Phys.*, Les Ulis, v. 51, n. 21, p. 2421-2430, Nov. 1990.

[44] KREBS, P. R.; THEUMANN, W. K. Generalization in a Hopfield network with noise. *J. Phys. A*, Bristol, v. 26, n. 16, p. 3983-3993, Aug. 1993.

[45] STARIOLO, D. A.; TAMARIT, F. A. Generalization in an analog neural network. *Phys. Rev. A*, Woodbury, v. 46, n. 8, p. 5249-5252, Oct. 1992.

[46] NAEF, J-P.; CANNING, A. Reentrant spin glass behaviour in the replica symmetric solution of the Hopfield neural network model. *J. Phys. I*, Les Ulis, v. 2, n. 3, p. 247-250, Mar. 1992.

[47] MIRANDA, E. N. Generalization in the Hopfield model: numerical results. *J. Phys. I*, Les Ulis, v. 1, n. 7, p. 999-1004, July 1991.

[48] BRANCHTEIN, M. C.; ARENZON, J. J. Categorization and generalization in the Hopfield model. *J. Phys. I*, Les Ulis, v. 2, n. 11, p. 2019-2024, Nov. 1992.

[49] VIRASORO, M. A. Categorization in neural networks and prosopagnosia. *Phys. Rep.*, Amsterdam, v. 184, n. 2/4, p. 301-306, Dec. 1989.

[50] KREBS, P. R.; THEUMANN, W. K. Categorization in the symmetrically dilute Hopfield network. *Phys. Rev. E*, Melville, v. 60, n. 4, p. 4580-4587, Oct. 1999.

[51] SOMPOLINSKY, H. Neural networks with nonlinear synapses and a static noise. *Phys. Rev. A*, Woodbury, v. 34, n. 3, p. 2571-2574, Sept. 1986.

[52] SOMPOLINSKY, H. The theory of neural networks: the hebb rule and beyond. In: VAN HEMMEN, J. L.; MORGENSTERN, I. (Eds.). *Heidelberg colloquium on glassy dynamics*. Berlin: Springer-Verlag, 1987. p. 485-527. (Lecture notes in physics, 275).

[53] VIANA, L.; BRAY, A. J. Phase diagrams for dilute spin glasses. *J. Phys. C*, Bristol, v. 18, n. 15, p. 3037-3051, May 1985.

[54] THEUMANN, W. K. Mean-field dynamics of sequence processing neural networks with finite connectivity. *Physica A*, Amsterdam, v. 328, n. 1/2, p. 1-12, Oct. 2003.

[55] THEUMANN, W. K.; ERICHSEN JR., R. Retrieval behavior and thermodynamic properties of symmetrically diluted Q-Ising neural networks. *Phys. Rev. E*, Melville, v. 64, n. 6, 061902 11p., Nov. 2001.

[56] CANNING, A.; NAEF, J-P. Phase diagrams and the instability of the spin glass states for the diluted Hopfield neural network model. *J. Phys. I*, Les Ulis, v. 2, n. 9, p. 1791-1801, Sept. 1992.

[57] WATKIN, T. L. H.; SHERRINGTON, D. A neural network with symmetric connectivity. *Europhys. Lett.*, Les Ulis, v. 14, n. 8, p. 791-796, Apr. 1991.

[58] ALMEIDA, J. R. L. de; THOULESS, D. J. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *J. Phys. A*, Bristol, v. 11, n. 5, p. 983-390, May 1978.

[59] THEUMANN, A. Porto Alegre: Instituto de Física - UFRGS, 1998. Comunicação pessoal.

[60] COSTA, R. L.; THEUMANN, A. Categorization in a Hopfield network trained with weighted examples: extensive number of concepts. *Phys. Rev. E*, Melville, v. 61, n. 5, p. 4860-4865, May 2000.

[61] SILVA, C. R. da; TAMARIT, F. A.; LEMKE, N.; ARENZON, J. J.; CURADO, E. M. F. Generalization in a diluted neural network. *J. Phys. A*, Bristol, v. 28, n. 6, p. 1593-1602, Mar. 1995.

[62] SILVA, C. R. da. *Propriedades dinâmicas de redes de neurônios com períodos refratários e diliução assimétrica*. 1997. 131 f. Tese (Doutorado em Ciências) - Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, 1997.

[63] DOMINGUEZ, D. R. C.; THEUMANN, W. K. Generalization in a multi-state neural network. *J. Phys. A*, Bristol, v. 29, n. 4, p. 749-761, Feb. 1996.

[64] DOMINGUEZ, D. R. C. Inference and chaos by a network of nonmonotonic neurons. *Phys. Rev. E*, Woodbury, v. 54, n. 4, p. 4066-4070, Feb. 1996.

[65] DOMINGUEZ, D. R. C.; THEUMANN, W. K. Generalization and chaos in a layered neural network. *J. Phys. A*, Bristol, v. 30, n. 5, p. 1403-1414, Mar. 1997.

[66] DOMINGUEZ, D. R. C.; BOLLÉ, D. Categorization by a three-state attractor neural network. *Phys. Rev. E*, Woodbury, v. 56, n. 6, p. 7306-7309, Dec. 1997.

[67] DOMINGUEZ, D. R. C. Information capacity of a hierarchical neural network. *Phys. Rev. E*, Woodbury, v. 58, n. 4, p. 4811-4815, Oct. 1998.

[68] RODRIGUEZ NETO, C.; FONTANARI, J. F. Categorization in the pseudo-inverse neural network. *J. Phys. A*, Bristol, v. 31, n. 2, p. 531-540, Jan. 1998.

[69] MARTINS, J. A.; THEUMANN, W. K. Categorization in a layered neural network. *Physica A*, Amsterdam, v. 253, n. 1/4, p. 38-56, May 1998.

[70] COSTA, R. L.; THEUMANN, A. Categorization in a Hopfield network trained with weighted examples. I. Finite number of concepts. *Physica A*, Amsterdam, v. 268, n. 3/4, p. 499-512, June 1999.

[71] ERICHSEN JR., R.; THEUMANN, W. K. , DOMINGUEZ, D. R. C. Categorization in fully connected multistate neural network models. *Phys. Rev. E*, Melville, v. 60, n. 6, p. 7321-7331, Dec. 1999.

[72] KATAYAMA, K.; HORIGUCHI, T. Generalization ability of Hopfield neural network with spin-s ising neurons. *J. Phys. Soc. Jpn*, Tokyo, v. 69, n. 9, p. 2816-2824, Sept. 2000.

[73] SILVA, C. R. da Categorization ability in a biologically motivated neural network. *Physica A*, Amsterdam, v. 301, n. 1/4, p. 362-274, Dec. 2001.