

**Centro Brasileiro de Pesquisas Físicas**

Coordenação de Física Teórica

Tese de Doutorado

**Efeito do Estímulo Externo  
no Reconhecimento de  
Padrões em uma Rede  
Neuronal**

**Nilo Barrantes Melgar**

**Orientador: Evaldo Mendonça Fleury Curado**

# Agradecimentos

Eu gostaria expressar meus mais sinceros agradecimentos:

- Ao prof. Dr. Evaldo Mendonça Fleury Curado pela proposição do problema, incentivo, paciência, empenho, compreensão e pela sua orientação neste trabalho.
- Ao prof. Dr. Fernando Dantas Nobre pelo incentivo, críticas e sugestões extremamente orientadoras.
- Aos meus filhos João Nilo e Joaquim Nilo pois eles são o motivo pelo qual todos os dias decido seguir em frente.
- A Fabiana Santana Rodrigues, uma mãe maravilhosa, que me deu dois grandes presentes, meus filhos.
- Aos Doutores Arthur Constantino Scardua, Leonardo José Lessa Cirto, Maike Antonio Faustino dos Santos e Carlos Eduardo Cedeño Montaña pelo apoio, compreensão e amizade nestes anos de doutorando.
- À minha mãe, Rosa Melgar Palomino, que sempre estará no meu coração.
- A Gustavo dos Santos Inez Borges, por sua importante ajuda nas correções do português.
- A Deus, por me conceder saúde e disposição de vencer mais um desafio em minha vida.

# Resumo

Neste estudo, abordamos a capacidade de armazenamento de um novo modelo de rede neuronal, com o qual se tenta modelar o comportamento do cérebro no processo de recuperação de memórias armazenadas. Este modelo, chamado de rede neuronal dependente do estímulo, considera uma maneira do sistema neuronal atenuar o ruído presente no processo de reconhecimento de uma memória armazenada. O objetivo deste modelo é propor um mecanismo básico de funcionamento que pode ser utilizado em um grande número de modelos de redes neuronais, e que visa aumentar a capacidade e qualidade de armazenamento destas redes.

Verificamos como a aplicação do modelo dependente do estímulo na rede de Hopfield pode aumentar significativamente a sua capacidade de armazenamento máxima,  $\alpha_C = 0.138$ . Desta forma, apresentamos uma expressão analítica das equações de ponto sela para este modelo, assim como mediante um estudo sinal-ruído fornecemos um valor aproximado da intensidade que o padrão externo deve ter para reduzir o ruído presente no processo de reconhecimento de uma memória armazenada. Uma análise numérica, por meio de simulações de Monte Carlo, corrobora os resultados analíticos.

A ideia do modelo dependente do estímulo é também aplicada a uma rede de Hopfield diluída e a uma rede de Hopfield com padrões armazenados correlacionados. Portanto estudamos numericamente como se pode atenuar o efeito do ruído no processo de reconhecimento de padrões em diversas situações diferentes. Em todas estas situações o nosso modelo leva a um grande aumento na capacidade de reconhecimento das memórias armazenadas.

*Palavras Chave*— armazenamento de padrões, rede totalmente conectada, rede neuronal diluída, padrões correlacionados, modelo Hopfield, erro de recuperação.

# Abstract

In this study, we address the storage capacity of a new model of neural network, with which we attempt to model the behavior of the brain in the process of recovering stored memories. This model, called the stimulus-depend neural network, considers a way for the neuronal system to attenuate the noise present in the process of recognizing a stored memory. The purpose of this model is to propose a basic functioning mechanism that can be used in a large number of neural network models, and which aims to increase the capacity and storage quality of these networks.

Thus, we verified how the application of the dependent stimulus model in the Hopfield network can significantly increase its maximum storage capacity,  $\alpha_c = 0.138$ . In this way, we present an analytical expression of the saddle point equations for this model, as well as through a signal-to-noise study we provide an approximate value of the intensity that the external standard must have to reduce the noise present in the process of recognizing a stored memory. A numerical analysis, using Monte Carlo simulations, corroborates the analytical results.

The idea of the dependent stimulus model is also applied to a diluted Hopfield network and to a Hopfield network with correlated stored patterns. Therefore, we study numerically how to mitigate the effect of noise in the pattern recognition process in several different situations. In all investigated cases, our model leads to a great increase in the capacity to recognize stored memories.

**Keywords**— pattern storage, fully connected network, diluted neuronal network, correlated patterns, Hopfield model, retrieval error.

# Sumário

<b>Agradecimentos</b>	<b>ii</b>
<b>1 Um breve Histórico</b>	<b>1</b>
1.1 Por que redes neuronais artificiais? . . . . .	2
1.2 Os primórdios das Redes Neuronais . . . . .	4
1.2.1 Arquiteturas de rede . . . . .	7
1.2.2 Aprendizagem . . . . .	8
<b>2 Modelagem de uma rede de neurônios</b>	<b>11</b>
2.1 O neurônio Biológico . . . . .	12
2.1.1 Dinâmica do neurônio . . . . .	12
2.1.2 Rede Neuronal biológica . . . . .	14
2.2 O neurônio de McCulloch e Pitts . . . . .	15
2.3 Hipótese de Hebb . . . . .	17
2.4 Modelo de Hopfield de uma rede neuronal . . . . .	18
2.4.1 Potencial Pós-sináptico . . . . .	18
2.4.2 A regra de atualização de um neurônio . . . . .	19
2.4.3 Relação com modelos magnéticos . . . . .	20
2.4.4 Regra de Hebb . . . . .	21
2.4.5 Dinâmica do sistema . . . . .	22
2.4.6 Hamiltoniano do sistema . . . . .	23
2.4.7 Parâmetro de ordem de Recuperação . . . . .	25
2.5 Análise de ruído Gaussiano . . . . .	27
2.5.1 Verificação da estabilidade de padrões armazenados . . . . .	27

2.5.2	Capacidade de armazenamento para $\alpha = p/N$ finito . . . . .	29
2.5.3	Recuperação de padrões sem erros . . . . .	31
<b>3</b>	<b>Teoria de Campo Médio no modelo de Hopfield</b>	<b>34</b>
3.1	TCM - Limite de $p$ finito . . . . .	35
3.2	TCM - Limite de $\alpha = p/N$ finito . . . . .	38
3.2.1	Soluções de simetria de Réplicas . . . . .	43
3.2.2	Equações de campo médio . . . . .	44
3.3	Soluções em $T \rightarrow 0$ . . . . .	45
3.3.1	Campo acoplado com um único padrão . . . . .	45
3.3.2	Campo Intrínseco Nulo ( $\kappa = 0$ ) . . . . .	48
3.3.3	Campo Intrínseco Ortogonal aos padrões armazenados . . . . .	49
<b>4</b>	<b>Rede Neuronal Dependente do Estímulo</b>	<b>52</b>
4.1	RNA's são adequadas para descrever o cérebro? . . . . .	54
4.2	A importância do estímulo externo . . . . .	57
4.3	Rede neuronal Dependente do Estímulo (RNDE) . . . . .	59
4.4	Estimativa do $\kappa$ ótimo . . . . .	61
4.5	Teoria de Campo Médio na RNDE . . . . .	64
4.5.1	Parâmetros de Ordem . . . . .	65
4.5.2	Teoria de simetria de réplicas . . . . .	66
4.5.3	Equações de campo médio . . . . .	67
4.6	Soluções no limite $T \rightarrow 0$ . . . . .	68
4.7	Simulação Computacional . . . . .	71
4.7.1	Cálculo numérico de $\kappa_C$ . . . . .	72
4.7.2	Comparação dos resultados analíticos com simulações Numéricas . . . . .	79
4.7.3	Limite de capacidade de armazenamento . . . . .	82
4.8	RNDE sob mudanças no padrão externo . . . . .	84
<b>5</b>	<b>Modificações na RNDE: correlação entre os padrões e diluição das sinapses</b>	<b>88</b>
5.1	Reconhecimento de padrões correlacionados . . . . .	89
5.1.1	Simulação computacional . . . . .	90

5.2	Reconhecimento em uma rede neuronal diluída . . . . .	95
5.2.1	RNDE diluída . . . . .	95
5.2.2	Estimativa de $\kappa$ ótimo . . . . .	96
5.2.3	Simulação computacional . . . . .	98
<b>6</b>	<b>Conclusões</b>	<b>101</b>
<b>A</b>	<b>Estudo analítico do modelo RNDE</b>	<b>105</b>
A.1	Teoria do Campo médio . . . . .	105
A.2	Soluções da simetria de réplicas . . . . .	112
A.3	Equações de campo médio . . . . .	115
A.3.1	Soluções em $T = 0$ . . . . .	116
<b>B</b>	<b>Algoritmo da RNDE</b>	<b>122</b>
<b>C</b>	<b>Auto média</b>	<b>123</b>
C.1	Auto-média da energia livre . . . . .	123
<b>D</b>	<b>O ruído no modelo de Hopfield</b>	<b>125</b>
<b>E</b>	<b>Autovalores e autovetores de K</b>	<b>128</b>
E.1	Autovalores . . . . .	128
E.2	Autovetores . . . . .	129
E.2.1	Para $\lambda_1 = 1 - \beta - (n - 1) \beta q$ . . . . .	130
E.2.2	Para $\lambda_2 = 1 - \beta + \beta q$ . . . . .	131

# Lista de Figuras

1.1	Taxonomia da arquitetura de redes diretas e recorrentes. . . . .	8
2.1	O neurônio biológico . . . . .	13
2.2	O córtex cerebral . . . . .	14
2.3	Diagrama Esquemático do neurônio formal. . . . .	15
2.4	Tipos de funções de ativação. . . . .	16
2.5	A hipótese de Hebb . . . . .	17
2.6	Superfície de quase energia do modelo de Hopfield. . . . .	26
2.7	Gráfico da probabilidade de erro $P(x)$ . . . . .	30
3.1	Curvas $m(\alpha)$ , para campos acoplados com um único padrão. . . . .	47
3.2	Curvas $m(\kappa)$ , para campos acoplados com um único padrão. . . . .	47
3.3	Esquema da solução da equação (3.47) . . . . .	48
3.4	Curvas $m(\alpha)$ , para campos ortogonais. . . . .	51
3.5	Curvas $m(\kappa)$ , para campos ortogonais. . . . .	51
4.1	Dois grandes personagens da historia contemporânea; (a) Albert Einstein, (b) Winston Churchill. Frente a um estímulo (foto a), a maioria de pessoas, poderá reconhecer rapidamente Eintein (memória armazenada), mas dificilmente reconhecerá Churchill (foto b). . . . .	58
4.2	Paisagem de quase energia usual de uma RNA de Hopfield com $\alpha < \alpha_c$ . . . . .	63
4.3	Paisagem de quase energia usual de uma RNDE na ausência (a) e presença (b) de um estímulo externo, respetivamente. . . . .	64
4.4	Gráficos da aproximação de campo médio em $T = 0$ , para $m_\rho$ , $m_\perp$ e $\Delta m$ , como função de $\kappa$ e $\gamma$ . . . . .	70

4.5	Resultados numéricos $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ . . . . .	74
4.6	Resultados numéricos $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 0.9$ . . . . .	77
4.7	Resultados numéricos $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 0.74$ . . . . .	78
4.8	Comparação teórica e numérica de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ . . . . .	81
4.9	Resultados numéricos $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ e $\alpha \gg 1$ . . . . .	83
4.10	Resultados numéricos da evolução temporal de $m_\rho(t)$ , $m_\perp(t)$ e $\Delta m(t)$ . . . . .	85
5.1	Resultados numéricos de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para dois padrões correlacionados. . . . .	91
5.2	Gráfico de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ e valores de $\alpha \leq 1.0$ , para o caso de três padrões correlacionados com o padrão $\{\xi_i^\rho\}$ . . . . .	93
5.3	Gráfico de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ e valores de $\alpha > 1.0$ , para o caso de três padrões correlacionados com o padrão $\{\xi_i^\rho\}$ . . . . .	94
5.4	Gráfico de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ e $\alpha = 0.5$ e valores típicos do parâmetro de diluição $d$ . . . . .	99
5.5	Gráfico de $m_\rho$ , $m_\perp$ e $\Delta m$ vs. $\kappa$ , para $\gamma = 1.0$ e $\alpha = 1.0$ e valores típicos do parâmetro de diluição $d$ . . . . .	100

# Lista de Tabelas

1.1	Computador de Von Neumann versus sistema neuronal biológico. . . . .	3
-----	--	---

# Capítulo 1

## Um breve Histórico

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.”

– *Isaac Newton.*

O cérebro humano é uma máquina altamente poderosa e complexa capaz de processar uma grande quantidade de informações em tempo mínimo. As unidades principais do cérebro são os neurônios e é por meio deles que as informações são transmitidas e processadas. As tarefas realizadas pelo cérebro intrigam os pesquisadores, por exemplo, a capacidade do cérebro de reconhecer um rosto familiar dentre uma multidão em apenas milésimos de segundo. As respostas sobre alguns enigmas do funcionamento do cérebro ainda não foram respondidas e se perpetuam até os dias de hoje. O que é conhecido sobre o funcionamento do cérebro é que o mesmo desenvolve suas regras através da experiência adquirida em situações vividas anteriormente.

O desenvolvimento do cérebro humano ocorre principalmente nos dois primeiros anos de vida, mas se arrasta por toda a vida. Inspirando-se neste modelo, diversos pesquisadores, das mais variadas disciplinas científicas, tentam modelar o funcionamento do cérebro, principalmente o processo de aprendizagem por experiência, a fim de criar sistemas inteligentes capazes de realizar tarefas como classificação, reconhecimento de padrões, processamento de imagens, entre outras atividades.

Abordagens convencionais foram propostas para resolver esses problemas. Embora aplicações bem-sucedidas possam ser encontradas em determinados ambientes bem restritos, nenhuma é flexível o suficiente para ter um bom desempenho fora de seu domínio. As redes

neurônais artificiais (RNArt) fornecem alternativas emocionantes, e muitas aplicações podem se beneficiar do seu uso.

Este capítulo é dedicado àqueles leitores com pouco ou nenhum conhecimento de RNArt, e tem por objetivo ajudá-los a entender os outros capítulos desta tese. A seguir, discutiremos as motivações por trás do desenvolvimento de RNArt, vemos uma breve história de como se desenvolveram estas redes, delinaremos as arquiteturas da rede e os processos de aprendizagem e apresentaremos alguns dos modelos de RNArt mais usados.

## 1.1 Por que redes neurônais artificiais?

O longo curso da evolução deu ao cérebro humano muitas características desejáveis que não estão presentes na proposta de Von Neumann ou em computadores paralelos modernos. Esses incluem

- paralelismo massivo,
- representação distribuída e computação,
- capacidade de aprendizado,
- capacidade de generalização,
- adaptabilidade,
- processamento de informações contextuais inerentes,
- tolerância a falhas e
- baixo consumo de energia.

Espera-se que dispositivos baseados em redes neurônais biológicas possuam algumas dessas características desejáveis.

Os computadores digitais modernos superam os humanos no domínio da computação numérica e da manipulação de símbolos relacionados. No entanto, os humanos podem facilmente resolver problemas de percepção complexos (como reconhecer um homem na multidão por um simples vislumbre de seu rosto). Por que há uma diferença tão notável em seu desempenho?

A arquitetura do sistema neuronal biológico é completamente diferente da arquitetura de Von Neumann (vide Tabela 1.1). Essa diferença afeta significativamente os tipos de funções que cada modelo computacional pode executar melhor.

	<b>Computador de Von Neumann</b>	<b>Sistema Neuronal Biológico</b>
Processador	Complexo Alta velocidade Um ou alguns	Simples Baixa velocidade Um grande número
Memória	Separado do processador Localizado Sem Conteúdo endereçável	Integrado no processador Distribuído Conteúdo endereçável
Computação	Centralizado Sequencial Programas armazenados	Distribuído Paralelo Autoaprendizagem
Confiabilidade	Muito vulnerável	Robusto
Expertise	Manipulações numéricas e simbólicas	Problemas perceptivos
Ambiente operacional	Bem definido Bem restrito	Mal definido Sem restrições

Tabela 1.1: Computador de Von Neumann versus sistema neuronal biológico.

Inúmeros esforços para desenvolver programas “inteligentes” baseados na arquitetura centralizada de Von Neumann não resultaram em programas inteligentes de uso geral. Inspiradas em redes neuronais biológicas, as RNArt são sistemas de computação massivamente paralelos que consistem em um número extremamente grande de processadores simples com muitas interconexões. Os modelos de RNArt tentam usar alguns princípios “organizacionais” que se acredita serem usados no cérebro humano. Modelar um sistema nervoso biológico usando RNArt também pode aumentar nossa compreensão das funções biológicas. A tecnologia de hardware de computador de última geração (como GLSI, ULSI, VLSI e óptica) tornou essa modelagem viável.

Um estudo aprofundado das RNArt requer conhecimentos de neurofisiologia, ciência cognitiva / psicologia, física (mecânica estatística), teoria de controle, ciência da computação, inteligência artificial, estatística / matemática, reconhecimento de padrões, visão computacional, processamento paralelo e hardware (digital / analógico / VLSI / óptico). Novos desenvolvimentos nessas disciplinas nutrem continuamente o campo. Por outro lado, as RNArt também incentivam essas disciplinas na forma de novas ferramentas e representações. Essa simbiose é necessária para a vitalidade da pesquisa em redes neuronais. As comunicações entre essas disciplinas devem ser incentivadas.

## 1.2 Os primórdios das Redes Neuronais

Um dos primeiros registros do surgimento das RNArt se deu com a publicação de um artigo de Warren McCulloch e Walter Pitts em 1943, no qual se definiu o neurônio formal, mostrando que são unidades capazes de executar operações lógicas. Eles criaram um modelo computacional para redes neuronais baseadas em matemática e algoritmos denominados lógica de limiar (threshold logic). Este modelo abriu o caminho para a pesquisa da rede neuronal dividida em duas abordagens: uma abordagem focada em processos biológicos no cérebro, enquanto a outra focada na aplicação de redes neuronais à inteligência artificial.

Já em 1949, Donald Hebb apresentou o primeiro método de treinamento das RNArt, no seu livro “The Organization of Behavior”, uma obra que apontou o fato de que os caminhos neuronais são fortalecidos cada vez que são usados, um conceito fundamentalmente essencial para a maneira como os humanos aprendem. Se dois nervos dispararem ao mesmo tempo, argumentou, a conexão entre eles é melhorada.

À medida que os computadores se tornaram mais avançados na década de 1950, finalmente foi possível simular uma hipotética rede neuronal. O primeiro passo para isso foi feito por Nathaniel Rochester dos laboratórios de pesquisa da IBM. Infelizmente para ele, a primeira tentativa de fazê-lo falhou.

No entanto, ao longo deste tempo, os defensores das “máquinas pensantes” continuaram a desenvolver suas pesquisas. Em 1956, o Projeto de Pesquisa de Verão de Dartmouth sobre Inteligência Artificial proporcionou um impulso tanto à Inteligência Artificial como às Redes Neuronais. Um dos resultados deste processo foi estimular a pesquisa em IA na parte de pro-

cessamento neuronal.

Nos anos seguintes ao Projeto Dartmouth, John von Neumann sugeriu imitar funções simples de neurônios usando relés telegráficos ou tubos de vácuo. Além disso, entre 1957 e 1958, Frank Rosenblatt, um neurobiologista, começou a trabalhar no Perceptron. Ele estava intrigado com o funcionamento do olho de uma mosca. Grande parte do processamento feito por uma mosca ao decidir fugir, é feito em seus olhos. Essa pesquisa teve como resultado o desenvolvimento do primeiro neurocomputador construído em hardware e, no período de 1958 a 1962, Rosenblatt criou uma grande classe de RNArt denominada Perceptrons. Um Perceptron de camada única foi útil para classificar um conjunto de entradas de valor contínuo em uma de duas classes. O Perceptron calcula uma soma ponderada das entradas, subtrai um limite e passa um dos dois valores possíveis como resultado.

Por volta do mesmo período (1962), Bernard Widrow e Marcian Hoff, de Stanford, desenvolveram modelos de RNArt denominados “ADALINE” e “MADALINE”. Em uma exibição típica do amor de Stanford por siglas, os nomes provêm do uso de múltiplos elementos ADaptive LINEar. ADALINE foi desenvolvido para reconhecer padrões binários de modo que, se ele estivesse lendo bits de transmissão de uma linha telefônica, poderia prever o próximo bit. MADALINE foi a primeira rede neuronal aplicada a um problema do mundo real, usando um filtro adaptativo que elimina ecos nas linhas telefônicas. Embora o sistema seja tão antigo como os sistemas de controle de tráfego aéreo, ele ainda está em uso comercial.

Os resultados obtidos nessas pesquisas motivaram vários pesquisadores a estudarem as RNArt, até que, em 1969, Minsky e Papert, demonstraram matematicamente as limitações das redes constituídas de uma única camada, como o Perceptron e o Adaline. No livro *Perceptrons — An Introduction to Computational Geometry*, eles usam um simples problema de lógica denominado XOR (OU exclusivo) para mostrarem que essas redes eram incapazes de resolvê-lo (Fausett, 1994). Além disso, esses autores não acreditavam que uma arquitetura adequada, juntamente com um algoritmo de ajuste de pesos, pudessem ser desenvolvidos de forma a superar esta limitação. A publicação deste trabalho causou um grande impacto entre os pesquisadores da neurocomputação, fazendo com que o interesse pela área ficasse bastante reduzido. Isso, de certa forma, ocasionou a ausência de novos resultados para a área por um longo período (Da Silva et al.). Nesse período de relativa turbulência das pesquisas envolvendo RNArt, foi implementada a rede ART (Adaptive Resonance Theory). Num trabalho de Grossberg, em

1980, foi realizada a formulação de mapas auto-organizáveis de Kohonen.

Em 1982, vários eventos provocaram um renovado interesse. John Hopfield apresentou um documento à Academia Nacional de Ciências. A abordagem de Hopfield não era simplesmente modelar cérebros, mas criar dispositivos úteis. Com clareza e análise matemática, ele mostrou como essas redes poderiam funcionar e o que poderiam fazer. Hopfield, juntamente com D. Tank, pesquisador da AT&T, começaram a desenvolver diversos modelos de redes neuronais baseadas em pesos fixos e ativações adaptativas.

Estas redes podem ser usadas como memórias associativas e para resolver problemas de otimização restrita como o caso do “Caixeiro Viajante”. A rede de Hopfield pode ser considerada como um sistema dinâmico com um número finito de estados de equilíbrio, de forma que, o sistema invariavelmente irá evoluir para um destes estados ou para uma sequência periódica de estados a partir de uma condição inicial. É também natural que a localização destes estados de equilíbrio possa ser controlada pela intensidade das conexões (pesos) da rede neuronal.

A conclusão interessante adotada por Hopfield foi que tais estados de equilíbrio podem ser utilizados como dispositivos de memória. De forma distinta daquela utilizada pelos computadores convencionais, em que o acesso à informação armazenada se dá por meio de um endereço, o acesso ao conteúdo da memória de uma rede de Hopfield se dá permitindo que a rede evolua com o tempo para um de seus estados de equilíbrio. Tais modelos de memória são denominados memórias endereçáveis por conteúdo.

A partir do trabalho de Hopfield, a neurocomputação voltou a receber a atenção dos pesquisadores, principalmente, matemáticos, físicos e engenheiros.

Não obstante, a teoria das RNArt conseguiu realmente se reestabelecer a partir da publicação do livro *Parallel Distributed Processing* de Rumelhart, Hinton e Williams, publicado em 1986. Nele, mostra-se um algoritmo, denominado retropropagação (*backpropagation*), capaz de treinar redes com múltiplas camadas e que, por sua vez, eram capazes de resolver as limitações apresentadas por Minsk e Papert. Assim, Rumelhart conseguiu estender a regra Widrow-Hoff para várias camadas apresentando uma ideia semelhante, com as denominadas redes Backpropagation, pois estas distribuem os erros de reconhecimento de padrões em toda a rede. As redes híbridas utilizavam apenas duas camadas, essas redes de Backpropagation utilizam muitas. O resultado é que as redes de Backpropagation “aprendem” de forma mais lenta, pois necessitam, possivelmente, de milhares de iterações para aprender, mas geram um resultado muito preciso.

Coincidentemente esse acontecimento deu-se no momento em que estavam sendo desenvolvidos computadores com maior capacidade de processamento (Da Silva et al.), os quais permitiram melhores simulações das RNArt. Após a publicação, pesquisadores das mais diferentes áreas passaram a visualizar interessantes aplicações para redes neurais artificiais.

Agora, as redes neurais são usadas em várias aplicações. A ideia fundamental por trás da natureza das redes neurais é que, se ela funcionar na natureza, deve ser capaz de funcionar em computadores. O futuro das redes neurais, no entanto, reside no desenvolvimento de hardware. As redes neurais rápidas e eficientes dependem do hardware especificado para seu eventual uso.

### **1.2.1 Arquiteturas de rede**

É possível visualizar as RNArt como grafos direcionados ponderados nos quais os neurônios artificiais são nós e as arestas dirigidas (com pesos) são conexões entre saídas neuronais e entradas neuronais. Com base no padrão de conexão (arquitetura), as RNArt podem ser agrupadas em duas categorias (veja a figura 1.1):

- redes diretas (ou feedforward), nas quais os grafos não possuem ciclos (ou loops), e
- redes recorrentes (ou feedback), nas quais ocorrem ciclos devido a conexões de realimentação.

Na família mais comum de redes diretas, chamada perceptron multicamada, os neurônios são organizados em camadas que possuem conexões unidirecionais entre eles. Na figura 1.1 mostram-se redes típicas para cada categoria.

Os diferentes comportamentos da rede são produto das diferentes conectividades. As redes diretas são estáticas, ou seja, elas produzem, para uma determinada entrada, apenas um conjunto de valores de saída. As redes diretas não têm memória no sentido de que sua resposta a uma entrada é independente do estado anterior da rede.

As redes recorrentes, por outro lado, são sistemas dinâmicos. Quando um novo padrão de entrada é apresentado, as saídas dos neurônios são calculadas. Por causa dos caminhos de recuperação, as entradas para cada neurônio são modificadas, o que leva a rede a entrar em um novo estado. Elas podem ter uma dinâmica complicada e isso pode torná-las muito difíceis de treinar. Entretanto, estes tipos são biologicamente realistas.

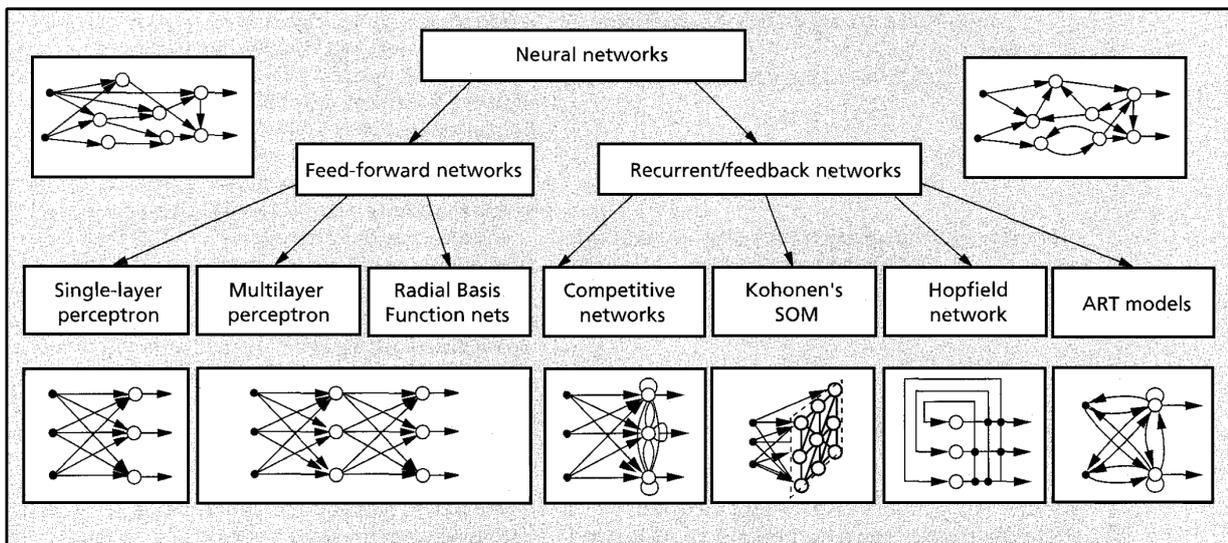


Figura 1.1: Uma taxonomia da arquitetura de redes diretas/feedforward e recorrentes/feedback. (Soares e Fagundes, 2012).

Uma variante de redes recorrentes são aquelas cujas conexões entre unidades são simétricas (elas têm o mesmo peso em ambas as direções) e são chamadas de “redes conectadas simetricamente”. Estas redes são mais fáceis de analisar do que as redes recorrentes. Elas também são mais restritas no que podem fazer porque obedecem a uma função de energia. As redes conectadas simetricamente sem unidades ocultas são chamadas de “Redes Hopfield”. Já, as redes conectadas simetricamente com unidades ocultas chamam-se de “Máquinas de Boltzmann”.

Arquiteturas de rede diferentes requerem algoritmos de aprendizado apropriados. A próxima subseção fornece uma visão geral dos processos de aprendizagem.

## 1.2.2 Aprendizagem

A capacidade de aprender é uma característica fundamental da inteligência. Embora seja difícil formular uma definição precisa de aprendizado, um processo de aprendizagem no contexto das RNArt pode ser visto como o problema de atualizar a arquitetura de rede e os pesos de conexão para que uma rede possa executar com eficiência uma tarefa específica. A rede geralmente deve aprender os pesos de conexão dos padrões de treinamento disponíveis. O desempenho é aprimorado com o tempo, atualizando iterativamente os pesos na rede. A capacidade de aprender automaticamente os exemplos torna as RNArt atraentes e empolgantes. Em vez de seguir um conjunto de regras especificado por humanos, as RNArt aprendem regras subjacentes

(como relações de entrada-saída) da coleção de exemplos representativos. Essa é uma das principais vantagens das redes neurais sobre os sistemas tradicionais.

Para entender ou projetar um processo de aprendizado, primeiro é necessário ter um modelo do ambiente no qual a rede neuronal irá operar, ou seja, deve-se saber quais informações estarão disponíveis para a rede. Nós nos referimos a esse modelo como um paradigma de aprendizado. Segundo, você deve entender como os pesos da rede são atualizados, ou seja, quais regras de aprendizado regem o processo de atualização. Um algoritmo de aprendizado refere-se a um procedimento no qual as regras de aprendizado são usadas para ajustar os pesos.

Existem três principais paradigmas de aprendizado: supervisionado, não supervisionado e híbrido. Na aprendizagem supervisionada, ou com um “professor”, a rede recebe uma resposta (saída) correta para cada padrão de entrada. Os pesos são determinados para permitir que a rede produza respostas o mais próximo possível das respostas corretas conhecidas. Por outro lado, o aprendizado não supervisionado não requer uma resposta correta associada a cada padrão de entrada no conjunto de dados de treinamento. Ele explora as correlações entre padrões nos dados e organiza os padrões em categorias a partir dessas correlações.

O aprendizado híbrido combina aprendizado supervisionado e não supervisionado. Parte dos pesos geralmente é determinada através de aprendizado supervisionado, enquanto os outros são obtidos através de aprendizado não supervisionado.

É possível dividir as regras ou algoritmos de aprendizado em cinco tipos básicos: aprendizagem por correção de erro, aprendizagem hebbiana, aprendizagem de Boltzmann, aprendizado competitivo e aprendizagem baseada em memória. Estas regras permitem ir muito além do alcançável por filtros adaptativos lineares (válido para redes diretas), tanto em termos de capacidade como em universalidade.

Este trabalho está dividido da seguinte forma: no capítulo 2, abordamos brevemente uma das redes neurais atratoras (RNA) mais conhecidas, a rede de Hopfield, e alguns problemas relacionados a sua capacidade de armazenamento. No capítulo 3, com ajuda do método de réplicas, calculamos a capacidade máxima de armazenamento de uma rede neuronal de Hopfield. No capítulo 4, apresentamos nosso modelo e discutimos os novos conceitos envolvendo essa nova abordagem, incluindo o número de padrões armazenados e como esse número pode ser muito maior do que nos modelos de RNA. Essa abordagem, ou estrutura, pode ser usada em muitos modelos de redes neurais, mas a ilustraremos aqui com o modelo de Hopfield, de-

vido à sua simplicidade e à possibilidade de obter resultados analíticos. Apresentamos também, alguns resultados numéricos que fundamentaram nossos cálculos analíticos. Já na seção 5, aplicamos a abordagem dependente do estímulo à rede de Hopfield diluída e à rede de Hopfield com padrões correlacionados, mostrando alguns resultados numéricos, focando principalmente no número de padrões armazenados e na qualidade do reconhecimento de padrões. Na seção 6, apresentamos nossos comentários finais.

## Capítulo 2

# Modelagem de uma rede de neurônios

“Nenhum cientista pensa com fórmulas.”

– *Albert Einstein.*

A forma como a informação memorizada é codificada, armazenada e recuperada é uma das tarefas que o cérebro corriqueiramente realiza no dia a dia e entender este processo é um dos maiores desafios da ciência. A melhor forma de abordar o estudo de um problema tão complexo, é propondo um modelo simples cujas características consigam imitar o máximo possível algumas das funções do cérebro humano. Para atingir esse objetivo, os componentes básicos do modelo devem estar inspirados nos elementos neurofisiológicos, neurônios e sinapses, e um conjunto altamente interconectado destes elementos básicos deve ter propriedades computacionais coletivas úteis.

É evidente que estamos mais interessados na modelização das maiores redes neuronais naturais; o córtex, é de particular interesse pois é um dos mais complexos e é o local do processamento neuronal mais sofisticado e distinto. Este é também o sistema para o qual o uso de técnicas de mecânica estatística poderia ser importante.

A memória associativa<sup>1</sup> é um modelo plausível para a memória biológica, onde um grande número de neurônios conectados age individualmente de uma maneira aparentemente aleatória, mas, coletivamente, constitui um órgão que executa uma tarefa complicada específica de maneira robusta. A rede neuronal de Hopfield (1982) é um modelo de memória associativa pois é capaz de recuperar a partir de pistas parciais ou ruidosas, o estado correto.

---

<sup>1</sup>A memória associativa define-se como a capacidade de recuperar informações armazenadas usando como dicas conjuntos parciais ou conjuntos corrompidos dessas informações.

Apesar das simplificações do modelo, a essência dos sistemas biológicos, representada pela capacidade computacional que emerge do comportamento coletivo de um grande número de neurônios altamente estilizados, é resguardada, como veremos neste capítulo.

## 2.1 O neurônio Biológico

O anatomista italiano Camillo Golgi, inventou a técnica de manchar um tecido nervoso com objetivo de ter maior contraste nas suas imagens de microscópio (Schiller, 2001). Afortunadamente a coloração mostrou que somente algumas partes bem delimitadas do tecido estavam manchadas, demonstrando que essas partes estavam separadas por membranas do resto. Assim, pela primeira vez, se conseguia enxergar que o tecido consiste em células nervosas separadas, chamadas neurônios.

Felizmente a pesquisa dos neurônios individuais permitiu compreender melhor as relações entre morfologia e função. O neurônio é um processador elementar de informação. Seu corpo ou soma cuida do sustento do neurônio e faz parte do processamento; as extensões curtas no soma, chamadas dendritos são seus dispositivos de entrada. A extensão longa ou axônio como todos os seus ramos é o dispositivo de saída. Finalmente, o pé pequeno e grosso na extremidade de cada ramo do axônio, tocando outro neurônio (geralmente em um de seus dendritos), é o mais importante para nós: uma sinapse, um dispositivo adaptável e ensinável para transferir informações de um neurônio para outro. Acredita-se que a adaptação das sinapses seja o passo básico do armazenamento da memória no sistema nervoso.

### 2.1.1 Dinâmica do neurônio

Na ausência de estímulos, o interior do neurônio está a um potencial negativo de  $-70\text{ mV}$  em relação ao exterior (estado passivo), sustentada pela ação de bombas moleculares que transportam íons sódio e potássio através da membrana celular. Quando sinais químicos estimulam o neurônio observam-se mudanças na diferença de potencial (DDP) através de sua membrana. No caso de hiperpolarização, DDP mais negativa, o neurônio permanece passivo e no caso de despolarização, DDP menos negativa, somente no caso dela atingir um certo potencial limiar (cerca de  $-55\text{ mV}$ ), gera-se um pulso elétrico como consequência de uma rápida variação de potencial, conhecida como potencial de ação. O potencial de ação propaga-se pelo axônio ra-

pidamente, dividindo-se em pulsos completos em cada ramificação, e ao chegar aos terminais sinápticos estimula a liberação de substâncias químicas, chamadas neurotransmissores, capazes de induzir uma despolarização ou hiperpolarização (da ordem de  $0,1\text{ mV}$ ) na membrana do neurônio receptor.

O potencial produzido pela alteração da permeabilidade iônica da membrana neuronal é chamado de potencial pós-sináptico (PPS) e será inibitório caso o neurotransmissor tenha produzido uma hiperpolarização ou excitatório no caso de despolarização.

O resultado da adição dos potenciais pós-sinápticos, que chegam quase simultaneamente ao soma do neurônio receptor, deve exceder o limiar de  $-55\text{ mV}$  para tornar o neurônio ativo. Caso contrário, o neurônio permanecerá passivo e a diferença de potencial através de sua membrana decairá ao seu valor de equilíbrio ( $-70\text{ mV}$ ) quase que exponencialmente.

A partir de estudos biológicos, sabe-se que cada neurônio pode ter apenas um tipo de sinapse: sejam todas excitatórias ou todas inibitórias, enquanto pode receber potenciais de ação de sinapses de qualquer tipo. Portanto, os próprios neurônios podem ser classificados como excitatórios ou inibitórios (Gesztí, 1990).

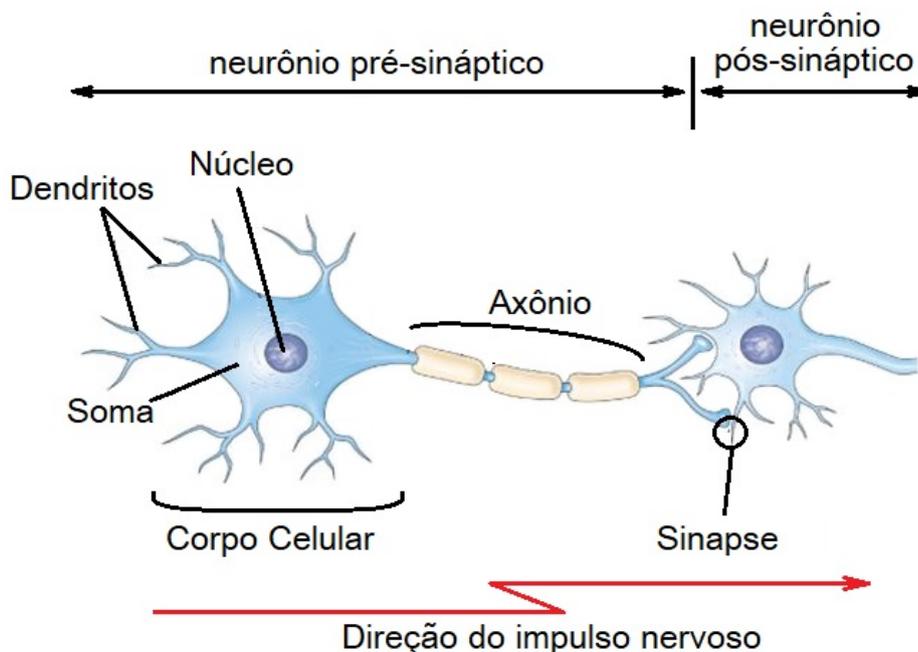


Figura 2.1: Representação de um neurônio biológico (Soares e Fagundes, 2012).

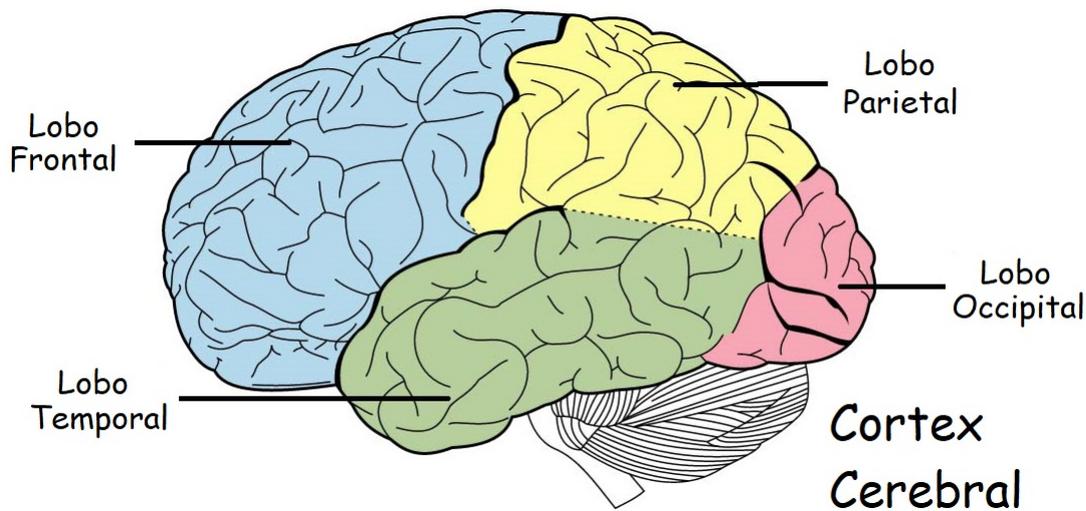


Figura 2.2: Vista lateral do cérebro, mostrando as principais áreas do córtex cerebral (Queensland Brain Institute, 2005).

### 2.1.2 Rede Neuronal biológica

O cérebro humano contém cerca de  $10^{11}$  neurônios, que é aproximadamente o número de estrelas na Via Láctea. Os neurônios são maciçamente conectados, muito mais complexos e densos que as redes telefônicas. Cada neurônio está conectado com aproximadamente  $10^4$  outros neurônios. No total, o cérebro humano contém um aproximado de  $10^{15}$  interconexões ou sinapses (Herculano, 2014). O córtex cerebral, encarregado de processar e armazenar a informação, é uma fina camada de neurônios com cerca de 2 a 3 milímetros de espessura e com uma área de cerca de  $2200 \text{ cm}^2$  (Aghoghovwia, 2020), aproximadamente o dobro da área de um teclado de computador comum.

Os neurônios se comunicam através de um trem muito curto de pulsos, normalmente de milissegundos de duração. A mensagem é modulada na frequência de transmissão de pulsos. Essa frequência pode variar de alguns a várias centenas de hertz, que é milhões de vezes mais lento que a velocidade de comutação mais rápida nos circuitos eletrônicos. No entanto, decisões perceptivas complexas, como reconhecimento facial, são tipicamente tomadas por seres humanos dentro de algumas centenas de milissegundos. Essas decisões são tomadas por uma rede de neurônios cuja velocidade operacional é de apenas alguns milissegundos. Isso implica que os cálculos não podem levar mais do que cerca de 100 estágios seriais.

Em outras palavras, o cérebro executa programas paralelos com cerca de 100 passos para

essas tarefas perceptivas. Isso é conhecido como *regra dos cem passos* (Feldman et al., 1988). As mesmas considerações de tempo mostram que a quantidade de informações enviadas de um neurônio para outro deve ser muito pequena (alguns bits). Isso implica que informações críticas não são transmitidas diretamente, mas capturadas e distribuídas nas interconexões - daí o nome, modelo conexionista, usado para descrever as RNArt.

## 2.2 O neurônio de McCulloch e Pitts

O primeiro modelo computacional de um neurônio foi proposto por McCulloch e Pitts (1943), os quais, ignorando completamente a dinâmica interna do neurônio biológico, introduziram a noção do “neurônio formal” como um elemento de dois estados; ativo ( $x_j = 1$ ) ou passivo ( $x_j = 0$ ), dependendo de se o neurônio disparou ou não um sinal eletroquímico. Para isso calcula-se somas ponderadas das entradas com os pesos preestabelecidos, podendo ser negativos no caso de uma sinapse inibitória, ou positivos no caso excitatório. Se essa soma for maior que um dado limiar, então o neurônio emite um pulso. Com esse mecanismo simples é possível implementar uma rede de neurônios que realize o cálculo de qualquer função booleana.

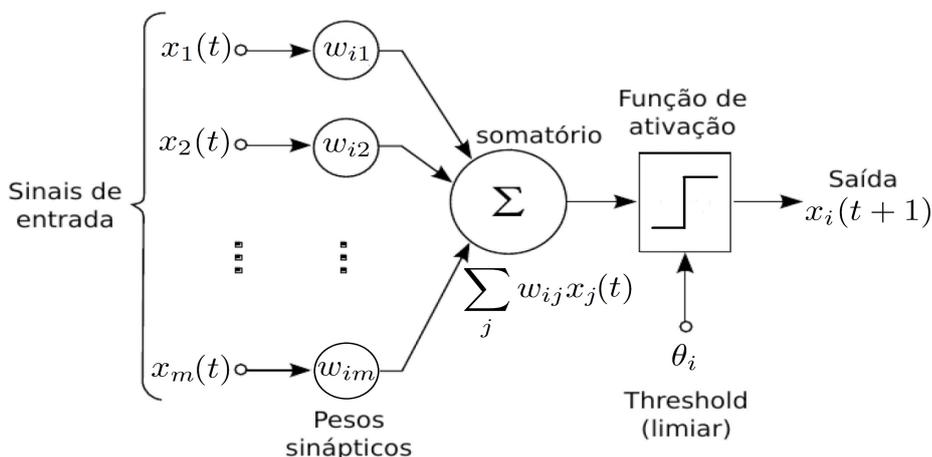


Figura 2.3: Diagrama Esquemático do neurônio formal (Yegnanarayana, 2009).

O neurônio matemático, mostrado na figura 2.3, calcula uma soma ponderada de seus sinais de entrada  $x_j$ ,  $j = 1, 2, \dots, m$  e gera uma saída de 1 se essa soma estiver acima de um determinado limite  $\theta_j$ . Caso contrário, resulta uma saída de 0. Matematicamente,

$$x_i(t+1) = U\left(\sum_j w_{ij}x_j(t) - \theta_i\right)$$

onde  $U(x)$  é uma função que vale 1 se  $x > 0$  e 0 se  $x < 0$  e  $w_{ij}$  é o peso da sinapse associado à  $j$ -ésima entrada. Para simplificar a notação, geralmente considera-se o limiar  $\theta_i$  como outro peso  $w_0 = -\theta_i$  anexado ao neurônio com uma entrada constante  $x_0 = 1$ .

McCulloch e Pitts provaram que, em princípio, pesos adequadamente escolhidos permitem que um arranjo síncrono de tais neurônios realize cálculos universais. Há uma analogia grosseira aqui para um neurônio biológico; fios e interconexões modelam axônios e dendritos, pesos de conexão representam sinapses e a função limiar se aproxima à atividade em um soma. O modelo de McCulloch e Pitts, no entanto, contém uma série de suposições simplificadoras que não refletem o verdadeiro comportamento dos neurônios biológicos.

O neurônio de McCulloch-Pitts foi generalizado de várias maneiras. Uma óbvia é o uso de funções de ativação que não sejam uma função limiar, como linear por partes, sigmoide ou gaussiana, como mostra a figura 2.4.

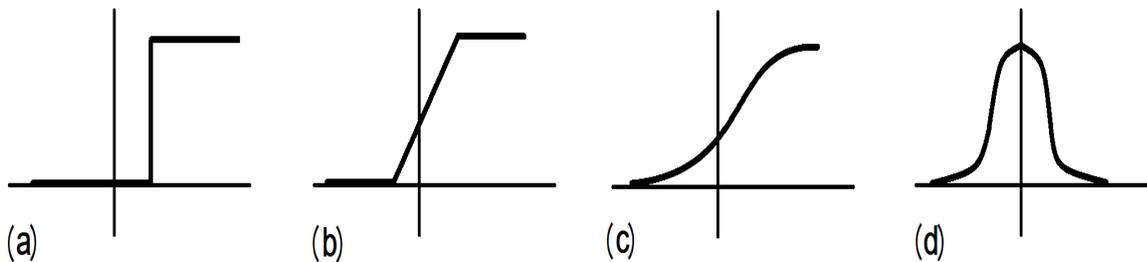


Figura 2.4: Diferentes tipos de funções de ativação: (a) limiar, (b) linear por partes, (c) sigmoide e (d) gaussiana.

A função sigmoide é de longe a mais frequentemente usada em RNArt. É uma função estritamente crescente que exibe suavidade e tem as propriedades assintóticas desejadas. A função sigmoide padrão é definida por

$$f(x) = \frac{1}{1 + \exp(-\beta x)}$$

onde  $\beta$  é o parâmetro de inclinação.

## 2.3 Hipótese de Hebb

É fato que uma função muito importante do sistema nervoso é o aprendizado/memória, pois muitas outras funções tais como a capacidade de adaptação a variações ambientais e processamento visual, que devido a sua complexidade são adquiridas por aprendizado na primeira infância, dependem dele.

Mesmo com diversos candidatos, a ferramenta mais robusta e suficientemente flexível para lhe confiar o papel da memória, é de longe, a força de conexão sináptica  $J_{ij}$ . Com base nesta ideia, Donald Hebb (1949), descreveu um mecanismo básico da plasticidade sináptica chamada de “Hipótese de Hebb”, a qual afirma que;

“Quando um axônio do neurônio A é próximo o suficiente de excitar um neurônio B e repetidamente ou persistentemente toma parte em dispará-la, algum processo de crescimento ou mudança metabólica acontece em uma ou ambas as células, tal que a eficiência de A, em contribuir para o disparo de B, é aumentada”.

Essa hipótese inspirou muitos pesquisadores e o primeiro mecanismo de apoio a ela, a “potenciação de longo prazo (LTP)”, foi descoberto no início dos anos 70. Ainda hoje, a Hipótese de Hebb, continua sendo um dos principais fatores para prever quais sinapses serão fortalecidas em uma rede neuronal.

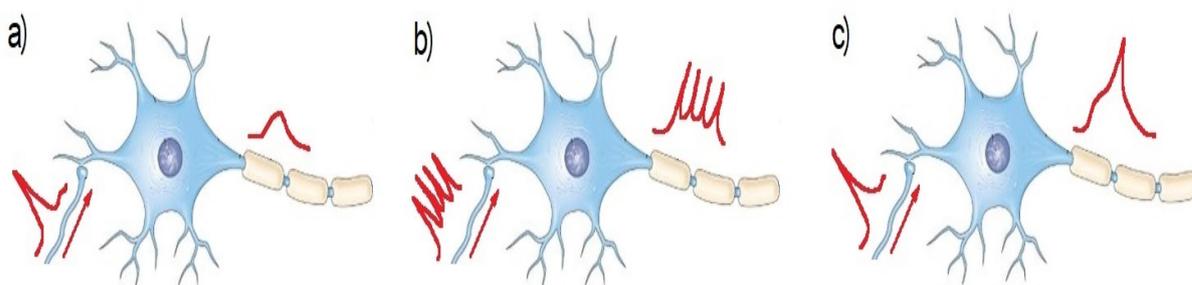


Figura 2.5: Hipótese de Hebb. (a) A sinapse do neurônio não é eficiente o suficiente para desencadear um potencial de ação. (b) Atividade simultânea intensa ocorre entre ambos os neurônios. (c) A sinapse do neurônio, fortalecida por essa atividade simultânea, desencadeia um potencial de ação (Bruno, 2002).

## 2.4 Modelo de Hopfield de uma rede neuronal

As redes Neurais artificiais (RNArt) fundamentam-se nos estudos sobre a estrutura do cérebro humano para tentar modelar sua forma inteligente de processar informação. Alguns estudos da neurofisiologia consideram que a riqueza computacional do cérebro humano está associada ao grande número de neurônios, interconectados por uma rede complexa de sinapses (Carvalho, 1988).

Circuitos eletrônicos complicados usando arquiteturas semelhantes a neurônios podem ser feitos na tentativa de reproduzir aspectos da memória biológica. No entanto, esses circuitos são bastante complexos e altamente ordenados. Parece altamente improvável que tais mecanismos surjam naturalmente e sejam usados como blocos básicos de construção da memória biológica. Em vez disso, se um grande número de neurônios tivesse propriedades coletivas úteis computacionalmente, surgindo simplesmente devido ao seu número, o acaso favoreceria o uso de blocos de construção que são mais simples e menos ordenados. Hopfield (1982) mostrou que um grande número de neurônios altamente “estilizados” possui propriedades coletivas. Ele descobriu que um conjunto de neurônios formais não-lineares de operação assíncrona consegue armazenar informação com estabilidade e eficiência, e é capaz de recuperá-las com alguma capacidade de correção de erros. Além disso, seu modelo é bastante robusto e deve funcionar mesmo se mais detalhes neurológicos forem adicionados. O modelo de Hopfield, baseia-se nas seguintes suposições:

### 2.4.1 Potencial Pós-sináptico

Assume-se que o potencial de membrana  $V_i$  em cada neurônio, é dado pela soma linear de todos os potenciais pós-sinápticos que chegam ao neurônio  $i$ , quase simultaneamente, isto é, entre o tempo  $t$  e  $t + \Delta t$ , ponderado pelo fator  $J_{ij}$ . Ou seja

$$V_i = \sum_{j \neq i}^N J_{ij} \left( \frac{S_j + 1}{2} \right), \quad (2.1)$$

onde  $S_j = +1$  ou  $S_j = -1$  de acordo a se o neurônio  $j$  disparou ou não um sinal eletroquímico. Aqui se faz a suposição de que a soma dos potenciais de ação é idêntica a soma dos potenciais pós-sinápticos  $V_i$ . A eficácia sináptica  $J_{ij}$ , mede a contribuição da atividade do  $j$ -ésimo neurônio pré-sináptico para o potencial pós-sináptico atuante sobre o  $i$ -ésimo neurônio pós-sináptico.

O  $J_{ij}$  será positivo se a sinapse é excitatória e negativo caso a sinapse seja inibitória.

### 2.4.2 A regra de atualização de um neurônio

Se o sistema não sofre algum tipo de perturbação externa (ausência de ruído), cada neurônio poderá disparar um “potencial de ação” ou pulso eletroquímico ao longo de seu axônio, se a soma dos potenciais pós-sinápticos (“potencial de membrana”)  $V_i$  exceder um potencial limiar  $U_i$ . Se o potencial de membrana está abaixo do limiar, os neurônios estão em um estado quiescente. A atividade de um neurônio é representada pela variável  $S_j$ , que toma o valor  $-1$  no estado quiescente e  $+1$  no estado de taxa máxima de disparo. Este processo dinâmico pode ser descrito pela seguinte equação:

$$S_i(t+1) = \operatorname{sgn}[V_i(t) - U_i(t)] = \begin{cases} +1 & V_i > U_i \\ -1 & V_i < U_i \end{cases}, \quad (2.2)$$

onde  $\operatorname{sgn}(x)$  é a função sinal de  $x$  e  $S_i(t+1)$  indica o estado atualizado. A equação anterior é equivalente à condição  $S_i \cdot h_i > 0$ , que o estado interno  $S_i$ , do neurônio  $i$ , deve obedecer, com objetivo de cumprir o requisito de alinhamento com seu campo local (“estímulo”)  $h_i$ , o qual define-se como  $h_i = V_i - U_i$ .

#### Inclusão do ruído na atualização

O mecanismo de transmissão dos pulsos eletroquímicos está sujeito a uma série de fontes de ruído (pelo fato dos neurônios dispararem com atrasos nas sinapses, entre outros), de modo que o valor do potencial pós-sináptico total  $V_i$ , dado na equação (2.1) é realmente um valor médio. Para ter em conta a estocasticidade do potencial pós-sináptico, se introduz uma “temperatura” efetiva  $1/\beta$ , que indica o nível de ruído no sistema (não tem nada a ver com a verdadeira temperatura física do cérebro.).

Sob um campo magnético,  $h_i$  o spin  $i$  tem uma energia potencial  $\epsilon_i = -h_i S_i$ ; por conseguinte, num ambiente ruidoso a probabilidade do neurônio  $i$  ter um dado valor  $S_i$ , será proporcional ao fator de Boltzmann  $e^{-\beta \epsilon_i} = e^{+\beta h_i S_i}$ , que deve ser adequadamente normalizado para os dois possíveis valores de  $S_i$ . Após um arranjo algébrico, a probabilidade  $P(S_i)$  para o novo

valor do spin  $i$  no tempo  $t + 1$ , será (Amit, 1992):

$$P(S_i) = \frac{\exp(\beta h_i S_i)}{\exp(\beta h_i) + \exp(-\beta h_i)} = \frac{1}{2} [1 + \tanh(\beta h_i S_i)], \quad (2.3)$$

onde  $\beta$  representa o inverso da temperatura  $T$ ,  $h_i$  é o campo molecular em  $i$  e  $P(S_i)$  assume uma forma sigmoidal. Quando  $\beta \rightarrow \infty$ , recupera-se a descrição determinista dada em (2.2).

### 2.4.3 Relação com modelos magnéticos

Com objetivo de reforçar a analogia com os modelos magnéticos, resulta conveniente reescrever o campo local  $h_i$  da seguinte forma

$$\begin{aligned} h_i &= \sum_{j \neq i}^N J_{ij} (S_j + 1) - U_i \\ &= \sum_{j \neq i}^N J_{ij} S_j + \left( \sum_{j \neq i}^N J_{ij} - U_i \right), \end{aligned} \quad (2.4)$$

onde, redimensionou-se as forças de acoplamento, escrevendo  $\frac{1}{2} J_{ij}$  com  $J_{ij}$ . O lado direito da expressão anterior, pode interpretar-se como a soma de um “campo efetivo” agindo no spin  $i$ , originado por suas interações com outros spins, com um “campo magnético intrínseco”,  $h_i^{int} = \sum_{j \neq i}^N J_{ij} - U_i$ , que não depende dos outros spins  $S_j$  do sistema. Assim, o campo local que age no neurônio  $i$ , toma a forma

$$h_i = \sum_{j \neq i}^N J_{ij} S_j + h_i^{int}. \quad (2.5)$$

Acredita-se que, no caso do potencial médio de membrana  $\bar{V}_i = \sum_{j \neq i}^N J_{ij}$  ser aproximadamente igual ao limiar  $U_i$ , a resposta de um neurônio será mais sensível a mudanças em suas entradas. É possível mostrar que a operação na vizinhança do limiar neuronal aumenta a capacidade de armazenamento da rede neuronal (Amit et al., 1985a). Pode-se, portanto, simplificar o modelo assumindo que

$$h_i^{int} = 0. \quad (2.6)$$

Essa suposição, muito usada na literatura, implica que os limiares reais dos neurônios sejam balanceados pela atividade média na rede. A eliminação dos limiares na descrição do modelo, permite focalizar a atenção no papel das eficácias sinápticas,  $J_{ij}$ , no controle do desempenho coletivo das redes neuronais artificiais.

Outra consequência importante é que, uma expressão idêntica à obtida em (2.3) (com mudanças apropriadas no significado das variáveis), descreve a dinâmica de spin único de um sistema de Ising interagindo com um banho térmico a temperatura  $T = \beta^{-1}$ , conhecida também como a dinâmica de Glauber (1963).

#### 2.4.4 Regra de Hebb

Como desejamos que a rede funcione como uma memória associativa, suas configurações estáveis devem estar correlacionadas com certas configurações  $\xi^\mu$ , que são fixadas pelo processo de aprendizagem (Peretto, 1992). Para conseguir isso, constrói-se as junções sinápticas de força  $J_{ij}$  a partir das configurações  $\xi^\mu$  que devem ser memorizadas. Assim, baseados no mecanismo sugerido por Hebb, obtemos a expressão chamada de “regra de Hebb” para os  $J_{ij}$ :

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu (1 - \delta_{ij}) \quad (\xi_i^\mu = \pm 1), \quad (2.7)$$

onde os  $p$  conjuntos binários  $\{\xi_i^\mu\}$ , com  $\mu (= 1, \dots, p)$  enumerando as diferentes memórias e  $i (= 1, \dots, N)$  enumerando os neurônios, assumem valores  $+1$  e  $-1$  com probabilidades prescritas. A escolha de (2.7) permite que os  $\{\xi^\mu\}$  sejam as configurações estáveis da rede em condições apropriadas, dando assim à rede uma capacidade de aprendizagem e memória.

Pode-se observar, de (2.7), que: o fator  $N^{-1}$ , é necessário para a obtenção do limite termodinâmico, se considerarmos que o alcance dos neurônios é praticamente infinito; as sinapses são simétricas ( $J_{ij} = J_{ji}$ ), cujo equivalente biológico é um número igual de sinapses excitatórias e inibitórias; os neurônios não estão conectados a si mesmos ( $J_{ii} = 0$ ), o qual é vantajoso no tratamento mecânico estatístico das propriedades de equilíbrio.

Cabe ressaltar que, a Regra de Hebb funciona muito bem se todos os padrões de entrada são ortogonais ou não correlacionados. O requisito de ortogonalidade impõe sérias limitações à Regra de Aprendizagem Hebbiana. Uma regra de aprendizado mais poderosa é a regra delta, que utiliza a discrepância entre a saída desejada e a real de cada unidade de saída para alterar os pesos que alimentam a mesma.

### 2.4.5 Dinâmica do sistema

O estado de uma rede de  $N$  neurônios, no tempo  $t$ , é definido como a configuração instantânea de todas as variáveis de spin  $S_i$  no tempo  $t$ :

$$\mathbf{S} = \{S_i\} = (S_1, S_2, \dots, S_N) \quad (S_i = \pm 1). \quad (2.8)$$

A evolução dinâmica desse estado, no espaço de fase de  $2^N$  estados é determinada pelas interações entre pares de neurônios, dadas através dos acoplamentos  $J_{ij}$  chamados sinapses.

Devido a que o modelo de Hopfield não envolve nenhum mecanismo dinâmico específico é necessário especificar como se vai implementar a dinâmica do sistema. Obviamente, capturar a complicada dinâmica de uma rede neuronal não é nada fácil e, por esse motivo, duas versões simplificadas da dinâmica de rede, que são tratáveis, se tornaram populares.

- O primeiro, a dinâmica síncrona ou paralela, pressupõe que todos os neurônios atualizam seus estados de atividade simultaneamente em intervalos de tempo discretos  $n\tau$ , onde  $n = 1, 2, \dots$ , como se governados por um relógio.
- A segunda é a dinâmica assíncrona ou sequencial, na qual os neurônios são atualizados um a um, em alguma sequência prescrita ou em ordem aleatória.

#### Dinâmica Assíncrona

Mesmo sendo uma idealização da realidade, a dinâmica assíncrona, está mais perto de capturar a natureza assíncrona e estocástica da evolução temporal de uma rede neuronal. A ideia subjacente é que, dentro de um intervalo de tempo unitário  $\tau$ , cada um dos  $N$  neurônios é atualizado em média uma vez. Como cada um dos neurônios pode ser atualizado em qualquer tempo dentro do intervalo  $\tau$ , o potencial pós-sináptico de um neurônio pode ser parcialmente determinado pelas atividades neuronais no intervalo de tempo anterior e parcialmente pelo novo estado de atividade dos neurônios no presente. Este efeito é capturado dividindo o intervalo de tempo elementar em  $N$  subintervalos iguais,  $t = \tau/N$  (Amit, 1992). Assim, dois estados sucessivos da rede diferenciam-se no máximo pelo estado de um neurônio.

Nestes processos dinâmicos assíncronos, a matriz de probabilidades de transição  $W(I | J)$ , que é da ordem  $2^N \times 2^N$ , é escassamente preenchida com entradas diferentes de zero. Se a sequência de atualização for aleatória, a matriz terá apenas  $N$  entradas por linha, além do

elemento diagonal, cada um correspondendo à inversão no estado de um neurônio distinto. No espaço de configurações, a dinâmica assíncrona implica uma trajetória conectando apenas vértices adjacentes.

Uma diferença interessante, ao considerar entre uma ou outra versão da dinâmica da rede, consiste no fato que no modelo de Hopfield (dinâmica assíncrona) a probabilidade de salto é igualmente compartilhada entre os sub estados degenerados, já no modelo de Little (dinâmica síncrona) os loops entre os estados degenerados podem subsistir. Na verdade, isso acontece apenas para o estado fundamental e os caminhos, quando  $1/\beta = 0$ , terminam em loops de comprimento um ou em loops de comprimento dois “piscando” entre um estado  $\{S_i\}$  e seu anti estado correspondente  $\{-S_i\}$ . Assim que o nível de ruído for diferente de zero, a distribuição limite leva a probabilidades iguais de encontrar o sistema em um dos sub estados degenerados, exatamente como faz a distribuição limite no modelo de Hopfield [vide Peretto (1984)].

### 2.4.6 Hamiltoniano do sistema

Em geral, a dinâmica neuronal pode ser muito complexa, mas é consideravelmente mais simples em redes simétricas, nas quais os coeficientes sinápticos  $J_{ij}$  e  $J_{ji}$  são iguais para cada par de neurônios. Uma quantidade extensiva chamada hamiltoniano pode ser definida, aproveitando a analogia com sistemas magnéticos. Para dinâmica assíncrona, a função de energia de uma rede de neurônios terá a mesma forma que o Hamiltoniano de um sistema de spins de Ising:

$$\begin{aligned} \mathcal{H} &= - \sum_i h_i S_i \\ &= - \frac{1}{2} \sum_{i,j} J_{ij} S_i S_j - \sum_{i=1}^N h_i^{int} S_i. \end{aligned} \quad (2.9)$$

Se o sistema fizer uma transição da configuração  $J$ , na qual o neurônio  $k$  assume o estado  $-S_k$ , para a configuração  $I$ , na qual o neurônio  $k$  se encontra no estado  $S_k$ , a variação de energia é dada por

$$\begin{aligned} \Delta \mathcal{H} &= \mathcal{H}(I) - \mathcal{H}(J) \\ &= -2 \sum_{i \neq k} J_{ik} S_i(J) S_k(J) - 2h_k^{int} S_k(J) \\ &= -2 \left( \sum_{i \neq k} J_{ik} S_i(I) + h_k^{int} \right) S_k(I) \end{aligned}$$

Para  $T = 0$ , a regra de atualização em (2.2) implica que na transição  $J \rightarrow I$ , se obtenha

$$\Delta \mathcal{H} = 2h_k S_k < 0. \quad (2.10)$$

Isso mostra que, em redes com conexões sinápticas simétricas a dinâmica assíncrona conduz o sistema sempre a um mínimo local de  $\mathcal{H}$ . Esses mínimos locais são configurações de spin nos quais cada spin é alinhado com seu campo local.

Na presença de ruído sináptico ( $T \neq 0$ ), a noção de mínimos no espaço de configurações é mais sutil. De acordo com (2.3) existe para cada neurônio uma probabilidade não nula de alinhar-se contrariamente ao campo local. Portanto, a probabilidade do sistema passar de uma configuração  $J$ , na qual o spin  $i$  está no estado  $S_i$ , para a configuração  $I$ , na qual esse spin está no estado  $-S_i$ , é  $P(-S_i)$  e de um estado em que o mesmo spin vai de  $-S_i$  para  $S_i$  é  $P(S_i)$ . A divisão expressa a probabilidade a priori de escolher o spin em particular. Consequentemente:

$$\frac{W(I | J)}{W(J | I)} = \frac{P(-S_i)}{P(S_i)} = \frac{\exp(-\beta h_i S_i)}{\exp(\beta h_i S_i)} = \exp(-2\beta h_i S_i).$$

Observe que era essencial ter  $h_i$  independente de  $S_i$  para garantir o balanço detalhado. Como apontado na expressão (2.9), o campo local total no spin  $i$  é apenas o coeficiente de  $S_i$  na energia total  $\mathcal{H}$ . Portanto, a expressão no expoente do lado direito da última equação é apenas a diferença de energia entre as duas configurações  $I$  e  $J$ , com os mesmos spins em todos os sítios, exceto em  $i$ . Assim, a equação anterior pode ser reescrita na forma

$$W(I | J) \exp[-\beta \mathcal{H}(J)] = W(J | I) \exp[-\beta \mathcal{H}(I)], \quad (2.11)$$

que representa a condição de balanço detalhado, satisfeita pela dinâmica de Glauber. Em consequência, no equilíbrio termodinâmico, as flutuações térmicas farão o sistema transitar ergódica e erráticamente entre “vales” que satisfazem a distribuição de Gibbs, isto é, a probabilidade  $P(I)$  de encontrar o sistema na configuração de equilíbrio  $I$  é dada por

$$P(I) = \frac{1}{\mathcal{Z}} \exp[-\beta \mathcal{H}(I)], \quad (2.12)$$

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

$$\mathcal{Z} = \sum_I \exp[-\beta \mathcal{H}(I)]. \quad (2.13)$$

Se as barreiras de energia que cercam um vale crescem com o tamanho do sistema, no entanto, a probabilidade de escapar do vale pode desaparecer no limite termodinâmico,  $N \rightarrow \infty$ , a

baixas temperaturas. Nesse caso, os vales de energia tornam-se disjuntos ou desconectados em escalas de tempo finitas, e se diz que a ergodicidade foi quebrada. Cada um desses vales de temperatura finita representa um estado termodinâmico distinto, ou fase. Assim, chama-se de “atratores” do processo dinâmico aos mínimos locais e globais da energia livre definida pela distribuição (2.12), a saber, os mínimos de

$$\mathcal{F}(\beta) = -\frac{1}{\beta} \ln \mathcal{Z}, \quad (2.14)$$

mas apenas aqueles mínimos que seriam cercados por barreiras de energia infinitamente altas no limite termodinâmico.

É esse fato que permite a aplicação das ferramentas da mecânica estatística no estudo das propriedades dinâmicas assintóticas das redes neuronais. No entanto, estados metastáveis são importantes, já que o último diz apenas respeito aos mínimos globais. Aqui os mínimos locais desempenham um papel importante, às vezes positivo, outros negativo.

### 2.4.7 Parâmetro de ordem de Recuperação

Como já foi mencionado, o sistema servirá como memória apenas se os estados associados aos atratores tiverem uma superposição significativa com alguns dos padrões memorizados. Um conjunto de quantidades que são medidas naturais dessas superposições seria

$$m^\mu \equiv \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i \rangle_T, \quad \mu = 1, \dots, p. \quad (2.15)$$

Para  $N$  grande, essas são  $p$  variáveis contínuas, cada uma no intervalo  $(-1, 1)$ . A média,  $\langle \dots \rangle_T$ , é sobre as configurações entre as quais o sistema vagará em períodos assintoticamente longos. Em  $T = 0$ , esses serão, geralmente, estados únicos, mas serão fases termodinâmicas para  $T$  finito. Na literatura, a dinâmica da rede neuronal de Hopfield à temperatura finita ( $T \neq 0$ ) é estudada amplamente, por Amit et al. (1987b).

Enquanto  $N$  for finito, as flutuações da soma em (2.15) atrapalharão qualquer discernimento entre  $m^\mu$ 's grandes e pequenos. No limite  $N \rightarrow \infty$ , o limite termodinâmico, as flutuações aleatórias não contribuirão para  $m^\mu$ . A ergodicidade será caracterizada por todos os  $m^\mu = 0$ , enquanto a quebra de ergodicidade por  $m^\mu \neq 0$  (para alguns  $\mu$ 's). As últimas serão fases com superposições macroscópicas (condensadas), que são os análogos diretos de fases, ou estados,

com magnetização macroscópica. O aparecimento de  $m^\mu$ 's finitos indica uma transição de fase com uma quebra de simetria, que é comumente detectada no limite termodinâmico.

Em um atrator no qual apenas um único  $m^\mu$  é diferente de zero, é possível obter uma recuperação inequívoca dessa memória específica. Esses estados serão chamados de “estados de recuperação”. Mas, geralmente, também existem atratores com vários  $m^\mu$ 's diferentes de zero, que serão chamados de “estados mistura”, ou “estados espúrios”.

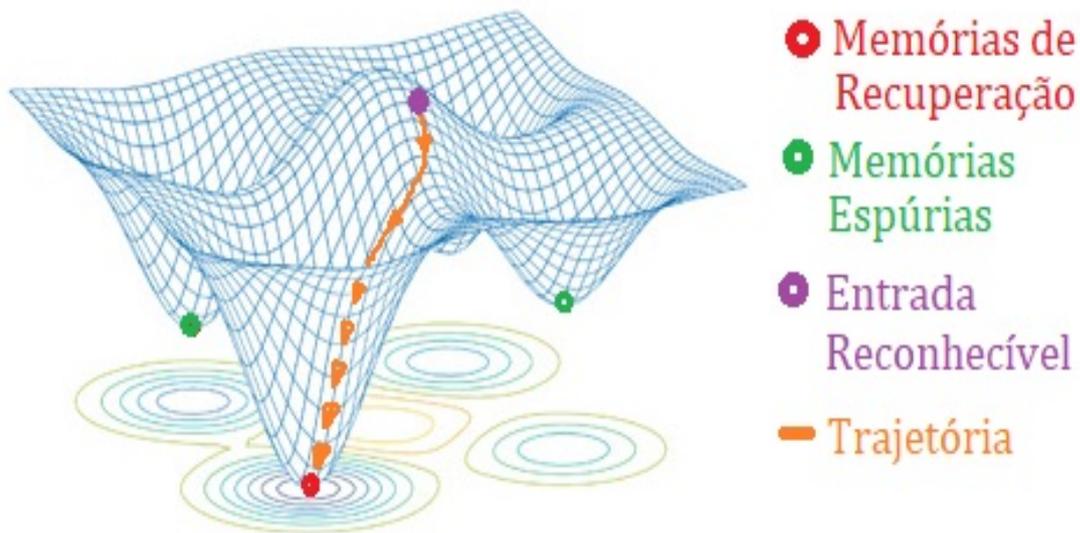


Figura 2.6: Superfície de quase energia do modelo de Hopfield: Os poços mais profundos são atratores que correspondem aos padrões armazenados, os demais correspondem as memórias espúrias (Geszi, 1990).

## 2.5 Análise de ruído Gaussiano

Voltando nossa atenção para o desempenho de uma rede de memória associativa no modo de recuperação, assumimos que o ciclo de aprendizado foi completado e os  $J_{ij}$ 's atingiram um determinado limite estático.

O desempenho é caracterizado por vários parâmetros. Um dos mais importantes é a capacidade, isto é, o número máximo de memórias que podem ser simultaneamente estabilizadas na rede. A natureza associativa da recordação é caracterizada pelo tamanho das bacias de atração dos vales correspondentes aos estados de memória. Estas bacias são limitadas em tamanho pela presença de outros atratores espúrios, que no caso de redes simétricas são mínimos locais de  $\mathcal{H}$  além dos estados de memória. Os tempos de convergência dentro das bacias determinam a velocidade da recordação. Outra questão importante é a robustez da rede à presença de ruído ou falhas de neurônios e sinapses.

A análise teórica da maioria dessas questões é praticamente impossível, a menos que se considere o limite termodinâmico,  $N \rightarrow \infty$ , onde as respostas podem ser derivadas de um estudo mecânico estatístico.

### 2.5.1 Verificação da estabilidade de padrões armazenados

Considerando como regra de aprendizado a prescrição de Hebb, dada por  $J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu$ , o campo local agindo no neurônio  $i$ , na ausência de campo intrínseco ( $h^{int} = 0$ ), terá a forma

$$h_i(\mathbf{S}) = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu S_j. \quad (2.16)$$

Para analisar a estabilidade dinâmica dos padrões armazenados, fazemos a substituição  $\mathbf{S} = \xi^\nu$  na equação anterior, esta escolha do padrão é arbitrária, uma vez que os padrões entram simetricamente na matriz sináptica e o que vale para um padrão será válido para qualquer outro.

Separamos  $h_i(\xi^\nu)$  em dois termos, correspondentes ao padrão  $\nu$  e todos os outros  $\mu \neq \nu$ :

$$\begin{aligned} h_i(\xi^\nu) &= \frac{1}{N} \sum_{j \neq i}^N \xi_i^\nu \xi_j^\nu \xi_j^\nu + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \\ &= \frac{N-1}{N} \xi_i^\nu + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \\ &= S + R. \end{aligned} \quad (2.17)$$

O primeiro termo, chamado de sinal ( $S$ ), corresponde ao padrão cuja estabilidade está sendo investigada. Já o segundo termo, denominado de ruído ( $R$ ) inclui a contribuição de todos os padrões armazenados restantes ao potencial pós-sináptico. A ideia crucial é que se  $|R|$  é menor que  $|S| \approx 1$ , o sinal de  $h_i(\xi^\nu)$  não será afetado e, portanto, podemos concluir que  $\xi^\nu$  é um ponto fixo da dinâmica da rede. Acontece que  $|R|$  é menor que  $|S|$  se  $p$  for pequeno e  $N$  for muito grande. Para entender por que esse é o caso, observe que a soma do termo de ruído contém uma soma de  $(N-1)(p-1) \approx Np$  variáveis aleatórias com a mesma probabilidade de ser  $+1$  ou  $-1$  (ver Orhan, 2014, pág. 2). Considerando que essas variáveis aleatórias binárias são decorrelacionadas por pares, a soma das  $Np$  variáveis aleatórias no termo de ruído é equivalente ao caminho aleatório unidimensional de  $Np$  passos de tamanho unidade. Em tal caminhada, são feitos passos para a frente e para trás. A soma - o ponto final da caminhada - irá flutuar em torno de zero. Seu deslocamento quadrático médio da origem é  $Np$ . Considerando que a soma é dividida por  $N$ , pode-se estimar o termo de ruído por  $\sqrt{p/N}$ . Ou seja o tamanho do ruído, no limite termodinâmico ( $N \rightarrow \infty$ ), será aproximadamente

$$|R| \approx \sqrt{\frac{p}{N}}. \quad (2.18)$$

Portanto, conclui-se que, se  $p$  é mantido fixo, quando o número  $N$  é muito grande, o ruído torna-se desprezível em relação ao sinal. Este argumento demonstra que os padrões  $\xi^\mu$  são pontos fixos da dinâmica da rede, desde que  $p$  seja pequeno comparado com  $N$ . Isso significa que se a rede for inicializada com um padrão  $\mathbf{S} = (S_1, \dots, S_N)$  que está próximo a um dos padrões armazenados  $\xi^\mu$ , a soma presente no campo local  $h_i \left( \equiv \sum_{j \neq i}^N J_{ij} S_j \right)$  será dominada pelos bits idênticos em  $\mathbf{S}$  e  $\xi^\mu$ , e conseqüentemente, a rede será rapidamente puxada em direção ao padrão  $\xi^\mu$ . Logo, os padrões armazenados  $\xi^\mu$  não são apenas pontos fixos da dinâmica da rede, mas pontos fixos estáveis, conforme desejado.

A seguir, mostra-se como uma análise probabilista fornece a capacidade máxima de memória da rede sob uma determinada regra de aprendizado. Baseia-se no seguinte fato observado nas simulações numéricas: quando o estado inicial da rede é um estado aprendido, a rede permanece nesse estado após o relaxamento (a recuperação é então perfeita) ou se afasta, e isso sugere que uma análise baseada em pequenas perturbações deve ser capaz de prever a capacidade de memória de uma rede com uma determinada regra de aprendizado.

### 2.5.2 Capacidade de armazenamento para $\alpha = p/N$ finito

Na seção anterior, vimos que o campo local experimentado pelo neurônio  $i$ , quando o sistema está no estado  $\xi^\nu$  é apenas

$$h_i(\xi^\nu) = S + R, \quad (2.19)$$

onde  $S$  é o termo de sinal, de tamanho  $(N - 1)/N$ , e  $R$  é o termo de ruído, o qual, pelo teorema do limite central, pode ser aproximado por uma variável aleatória gaussiana com distribuição de probabilidade

$$\mathcal{P}(x) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left(-\frac{x^2}{2\sigma_R^2}\right), \quad (2.20)$$

média zero e variância  $\sigma_R^2 = (N - 1)(p - 1)/N^2$ .

A probabilidade de termos um erro, ou seja, de que o campo local  $h_i$  seja negativo, é igual à probabilidade do termo de ruído  $R$  ultrapassar o termo de sinal  $S$ , ou seja

$$P_{erro} = \mathcal{P}[R \geq S] = \int_{-\infty}^{-S} \mathcal{P}(x) dx = \int_S^{\infty} \mathcal{P}(x) dx = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-t^2/2} dt, \quad (2.21)$$

onde  $S = (N - 1)/N$  e  $z (\equiv S/\sigma_R)$  é o termo  $S$ , normalizado pelo desvio padrão. É possível reescrever esta equação, usando a função erro,  $\text{erf}(x)$ , dada por

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (2.22)$$

Reorganizando os limites de integração e usando a mudança de variável  $t = \sqrt{2}s$ , obtemos

$$\begin{aligned} P_{erro} &= \frac{1}{\sqrt{2\pi}} \left( \int_0^{\infty} e^{-t^2/2} dt - \int_0^z e^{-t^2/2} dt \right) \\ &= \frac{1}{\sqrt{\pi}} \left( \int_0^{\infty} e^{-s^2} ds - \int_0^{z/\sqrt{2}} e^{-s^2} ds \right). \end{aligned}$$

No caso de  $N$  e  $p$  grandes,  $z \equiv S/\sigma_R \approx \sqrt{N/p}$ , assim

$$P_{erro} = \frac{1}{2} \left[ 1 - \text{erf} \left( \sqrt{\frac{N}{2p}} \right) \right]. \quad (2.23)$$

Considerando a variância de  $R$  ( $\sigma_R^2 \approx p/N$ ), como variável na equação (2.23), construímos uma nova função

$$P(x) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^{1/\sqrt{x}} e^{-t^2/2} dt, \quad (2.24)$$

onde  $x = p/N$ . Para pequenas variações, a função  $P(x)$ , é linear em  $x$  na vizinhança de  $x_1 = 1/3$ , o ponto de inflexão. Ele pode ser aproximado (vide fig. 2.7) por uma linha reta passando

pelo ponto  $(x_1, P(x_1))$ , com  $P(x_1) = 0.042$  e inclinação  $\frac{dP}{dx} \Big|_{x_1} = \frac{1}{\sqrt{\pi}} \left(\frac{3}{2e}\right)^{3/2} \cong 0.2313$ , que cruza o eixo  $x$  em  $x_0 = 0.153$ . Para  $x < x_0$ ,  $P(x) \cong 0$  (Gordon, 1987). Depois do ponto de cruzamento em 0.153, erros na recuperação são esperados. O número máximo de padrões que podem ser aprendidos antes dos erros na recuperação tornarem-se importantes é, portanto,

$$p = 0.153N. \quad (2.25)$$

Este cálculo representa a estabilidade inicial dos padrões. Se escolhermos  $p < 0.185N$ , por exemplo, ele nos diz que não mais que 1% dos bits do padrão serão instáveis inicialmente. Mas se iniciarmos o sistema em um padrão específico  $\xi_i^v$  e cerca de 1% dos bits mudarem, o que acontece a seguir? Pode ser que os primeiros façam com que mais bits mudem. No pior dos casos, poderia haver um fenômeno de avalanche no qual mais e mais bits virem até o estado final ter pouca ou nenhuma semelhança com o padrão original. Portanto, nossas estimativas para  $p_C$  são limites superiores; valores menores podem ser necessários para manter os atratores finais próximos aos padrões desejados. O cálculo mais sofisticado apresentado na Subseção 3.3.2 trata desse problema e mostra que uma avalanche ocorre se  $p = 0.138N$ , tornando toda a memória inútil. Assim,  $p_C = 0.138N$ , se estivermos dispostos a aceitar os erros que ocorrem até esse ponto.

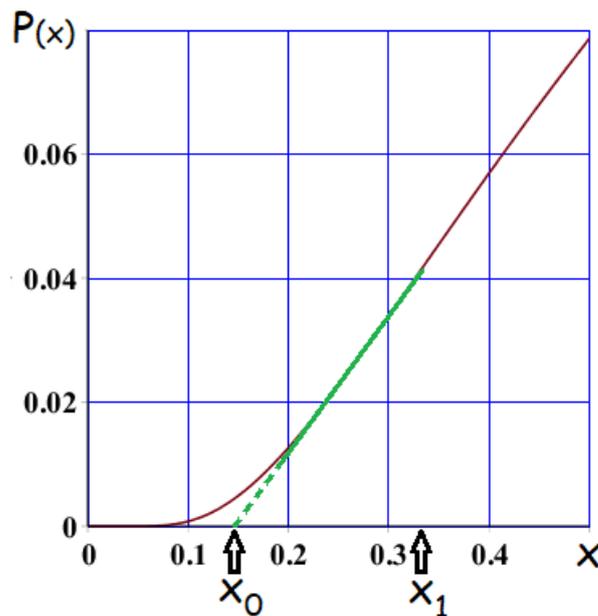


Figura 2.7: Gráfico da probabilidade  $P(x)$  (eq. 2.24) como função de  $x = p/N$ .

### 2.5.3 Recuperação de padrões sem erros

Considera-se que os campos locais  $h_i(\xi^\nu)$  e  $h_k(\xi^\nu)$ , são variáveis aleatórias independentes para  $i \neq k$ . De fato, elas são somas de termos independentes (ver eq. (2.17)), além do termo  $J_{ij}\xi_j^\nu$  que é comum para os dois. No entanto, para um modelo com conectividade macroscópica, em que cada spin (ou neurônio) está conectado a infinitamente muitos outros quando  $N \rightarrow \infty$ , a contribuição desse único termo comum é desprezível em relação aos muitos outros termos independentes (ver Geszti, 1990, pág. 28). Usando essa independência, a probabilidade  $P_0$ , de que todos os estados armazenados sejam estáveis (nenhum erro) é dada por

$$P_0 = (1 - P_{erro})^N. \quad (2.26)$$

Se desejamos que esta probabilidade  $P_0$  seja próxima de 1,  $P[R \geq S]$  será da ordem de 0 e a função  $\text{erf}(x)$ , em consequência, da ordem de 1. O argumento  $\sqrt{N/2p}$  da função  $\text{erf}$  deve, portanto, ser grande e a seguinte expansão assintótica pode ser usada:

$$\text{erf}(x) \simeq 1 - \frac{1}{x\sqrt{\pi}}e^{-x^2}. \quad (2.27)$$

Assim, da eq. 2.26, temos

$$P_0 = \left[ \frac{1}{2} \left( 1 + \text{erf} \left( \sqrt{\frac{N}{2p}} \right) \right) \right]^N \simeq \left[ 1 - \sqrt{\frac{p}{2N\pi}} \exp \left( -\frac{N}{2p} \right) \right]^N. \quad (2.28)$$

Tomando o logaritmo dos dois lados desta equação e usando a expansão  $\ln(1-x) \simeq -x$  para  $x$ 's pequenos, encontra-se uma equação implícita para  $\alpha$

$$\sqrt{\frac{\alpha}{2\pi}} \exp \left( -\frac{1}{2\alpha} \right) = \frac{-1}{N} \ln P_0. \quad (2.29)$$

Toma-se novamente o logaritmo da equação anterior, depois de elevar ao quadrado ambos os membros, obtemos

$$2 \ln N - \frac{N}{p} - \ln \frac{N}{p} = \ln [2\pi (\ln P_0)^2]. \quad (2.30)$$

Observa-se das simulações numéricas que o valor de  $\alpha$  é praticamente invariável quando  $P_0$  está entre 0.1 e 0.9, ou seja, o lado direito permanece finito. Quando  $N \rightarrow \infty$ , o último termo do lado esquerdo da equação é insignificante, por tanto  $1/\alpha$  deverá variar como  $2 \ln N$ , para satisfazer o requisito de finitude (Weisbuch e Fogelman-Soulié, 1985). Obtendo-se assim o limite superior, abaixo do qual o requisito é cumprido

$$\alpha \leq \frac{1}{2 \ln N}. \quad (2.31)$$

Além disso, é necessário ter em conta uma degeneração fundamental da rede, a saber:

Se um estado  $\mathbf{S}$  é estável, então o estado  $-\mathbf{S}$  é também estável.

Esta é uma propriedade das redes neuronais conectadas através de interações quadráticas. Claramente, isto permanece válido mesmo quando as interações  $J_{ij}$  não são simétricas.

Finalmente, o número máximo  $p_C$  de padrões aleatórios independentes que uma rede Hebiana pode armazenar e recuperar sem nenhum erro é

$$p_C = \frac{N}{4 \ln N}. \quad (2.32)$$

É possível remover estas rígidas limitações no número de padrões armazenados, se permitirmos uma pequena fração finita de bits desalinhados. Na verdade, não é necessário nos preocuparmos com a estabilidade exata dos padrões memorizados. Para fins de armazenamento e recuperação, é suficiente que existam pontos estacionários do processo dinâmico que estejam próximos o suficiente dos padrões armazenados, de modo que o mecanismo de leitura possa tirar as conclusões corretas.

Derivações mais detalhadas dos resultados de  $N/\ln N$  estão disponíveis em Weisbuch e Fogelman-Soulié (1985) e McEliece et al. (1987). Em resumo, o número máximo de padrões armazenados  $p_C$ , é proporcional a  $N$  (não superior a  $0.138 N$ ) se estivermos dispostos a aceitar uma pequena porcentagem de erros em cada padrão, sendo proporcional a  $N/\ln N$  se insistirmos que a maioria ou todos os padrões sejam recuperados perfeitamente.

Os padrões reais geralmente não serão aleatórios, embora algumas pré-codificações possam torná-los mais. O modelo de Hopfield é geralmente estudado com padrões aleatórios por conveniência matemática, embora o efeito de padrões correlacionados também tenha sido examinado (ver Hertz, 2018). No outro extremo, se os diferentes padrões armazenados forem estritamente ortogonais, ou seja,

$$m^{\mu\nu} = \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu = 0, \quad \forall \mu \neq \nu. \quad (2.33)$$

Neste caso ortogonal, o número máximo de padrões armazenados  $p_C$  é aparentemente  $N$ , porque um máximo de  $N$  cadeias de bits mutuamente ortogonais de comprimento  $N$  podem ser construídas. Mas a capacidade útil é menor. Tentar incorporar  $N$  padrões ortogonais à regra

de Hebb na verdade torna todos os estados instáveis; o sistema permanece onde quer que seja iniciado e é inútil como memória. Isso ocorre porque as condições de ortogonalidade (2.33) levam necessariamente a<sup>2</sup>

$$J_{ij} = \begin{cases} 1 & i = j; \\ 0 & \text{outro caso.} \end{cases} \quad (2.34)$$

ou seja, cada unidade está conectada apenas a si mesma. Para definir uma medida útil de capacidade para esse caso, é claramente necessário insistir em uma bacia finita de atração em torno de cada padrão desejado. Isso leva a uma capacidade máxima de memória menor que  $N$ .

Nesta seção mostrou-se que a análise sinal-ruído é uma ferramenta útil para estimar o local da falha nas variantes do modelo. Esta ferramenta permitiu estimar que o erro de recuperação, no modelo de Hopfield, alcançaria o valor crítico  $\alpha_C = 0.153$ . Esta estimativa apoia-se no cálculo de tipo campo médio feito na seção 3.2.

A visão de que é o nível geral de ruído que determina a evolução na bacia de atração de um único padrão, foi retomada em um cálculo dinâmico mais quantitativo por Krauth et al. (1988). No caso das amplitudes de aprendizagem dos diferentes padrões forem desiguais, um limitante é que, um padrão suficientemente forte pode desestabilizar o mais fraco muito antes que o ruído chegue a um nível perigoso (PazmAndi e Geszti, 1989).

---

<sup>2</sup>Considere a matriz  $\chi$  com componentes  $\chi_{\mu i} = \xi_i^\mu$ . A equação (2.22) implica que  $\chi\chi^T = N\mathbb{I}$ , onde  $\mathbb{I}$  é a matriz unidade, enquanto a regra de Hebb (2.9) pode ser escrita como  $J = N^{-1}\chi^T\chi$ . Usando  $(AB)^T = B^T A^T$  leva imediatamente a  $J = 1$ .

## Capítulo 3

# Teoria de Campo Médio no modelo de Hopfield

“Em algum lugar, alguma coisa incrível está esperando para ser descoberta.”

– *Carl Sagan.*

**D**EVIDO à complexidade matemática no estudo analítico de grandes modelos estocásticos, é comum mesmo para sistemas em equilíbrio se utilizar métodos aproximados para a determinação de algumas propriedades críticas de modelos desse tipo.

A aproximação mais simples utilizada é a aproximação de campo médio. A ideia consiste em isolar um determinado spin (ou neurônio) do sistema e assumir que a soma de todos os outros spins atua como um campo magnético médio que interage com o spin fixado, transformando assim um problema mecânico estatístico de muitos corpos em um problema de um só corpo. A ideia da teoria de campo médio (TCM), apareceu por primeira vez na física, no trabalho de Pierre Weiss (1907) para descrever transições de fase.

Obviamente, estes tipos de procedimentos não são exatos, mas sim muito úteis no cálculo dos diagramas de fase e outras propriedades termodinâmicas. Os expoentes críticos, determinados por esta teoria, são constantes e se tornam exatos acima de uma dimensão crítica.

Resumindo, o objetivo da TCM é substituir todas as interações por um corpo com uma interação média ou efetiva, às vezes chamada “campo local ou molecular”.

### 3.1 TCM - Limite de $p$ finito

Em 1985, Amit, Gutfreund e Sompolinsky estudaram o modelo proposto por Hopfield, com objetivo de analisar o comportamento coletivo de grandes redes neuronais. A análise desenvolvida nesta seção limita-se ao caso em que o número de padrões armazenados,  $p$ , é pequeno em comparação ao número de neurônios,  $N$ , do sistema.

Considerando-se que as “memórias” armazenadas  $\xi_i^\mu$ , são variáveis aleatórias independentes temperadas, e que obedecem a seguinte distribuição de probabilidade

$$\begin{aligned}\mathcal{P}(\{\xi_i^\mu\}) &= \prod_{i,\mu} \mathcal{P}(\xi_i^\mu), \\ \mathcal{P}(\xi_i^\mu) &= \frac{1}{2} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)].\end{aligned}\quad (3.1)$$

ou seja, os  $\xi_i^\mu$  assumem valores  $+1$  e  $-1$  com igual probabilidade.

Para determinar qualquer propriedade física do sistema deve-se tomar uma média dessa propriedade sobre a distribuição de probabilidades (3.1). Como estamos interessados no cálculo da densidade de energia livre do sistema, a média desta propriedade será dada por

$$f(\beta) = - \lim_{N \rightarrow \infty} \frac{1}{\beta N} \langle \ln \mathcal{Z} \rangle_\xi, \quad (3.2)$$

onde  $\langle \dots \rangle_\xi$  representa a média sobre a distribuição temperada  $\{\xi_i^\mu\}$ .

Para a análise termodinâmica da rede de spins, precisamos definir o hamiltoniano do sistema. Introduzindo  $h_i^{int} = \sum_{\mu=1}^p \kappa^\mu \xi_i^\mu$  no hamiltoniano da equação (2.9) e considerando que  $J_{ij} = J_{ji}$ , obtêm-se

$$\mathcal{H}(I) = -\frac{1}{2} \sum_{i,j \neq i}^N J_{ij} S_i S_j - \sum_{\mu=1}^p \kappa^\mu \sum_{i=1}^N \xi_i^\mu S_i, \quad (3.3)$$

onde, por conveniência, os  $\kappa^\mu$  são campos externos uniformes, conjugados a  $\sum_i \xi_i^\mu S_i$ . Substituindo a equação anterior, que representa o hamiltoniano do sistema, na definição da função de partição (vide (2.13)), temos

$$\mathcal{Z}(\beta, \kappa) = \text{Tr}_s \exp \left[ \beta \left( \frac{1}{2} \sum_{i,j \neq i}^N J_{ij} S_i S_j + \sum_{\mu=1}^p \kappa^\mu \sum_{i=1}^N \xi_i^\mu S_i \right) \right]. \quad (3.4)$$

Como as forças de conexão sinápticas obedecem a regra de Hebb (2.7), é possível chegar a seguinte identidade,

$$\sum_{i,j \neq i}^N J_{ij} S_i S_j = \frac{1}{N} \sum_{\mu=1}^p \left( \sum_{i=1}^N \xi_i^\mu S_i \right)^2 - p. \quad (3.5)$$

Introduzindo (3.5) em (3.4), temos

$$\mathcal{Z}(\beta, \kappa^\mu) = e^{-\beta p/2} \text{Tr}_s \exp \beta \left[ \frac{1}{2N} \sum_{\mu=1}^p \left( \sum_{i=1}^N \xi_i^\mu S_i \right)^2 + \sum_{\mu=1}^p \kappa^\mu \sum_{i=1}^N S_i^\mu \right].$$

Fazendo uso da variável  $x^\mu = N^{-1} \sum_i \xi_i^\mu S_i$ , e definindo-se os vetores  $p$  dimensionais  $\mathbf{x} = \sum_{\mu=1}^p x^\mu \hat{e}_\mu$  e  $\boldsymbol{\kappa} = \sum_{\mu=1}^p \kappa^\mu \hat{e}_\mu$ , temos

$$\mathcal{Z}(\beta, \kappa^\mu) = e^{-\beta p/2} \text{Tr}_s \exp \left[ \frac{\beta N}{2} \mathbf{x}^2 + \beta N \boldsymbol{\kappa} \cdot \mathbf{x} \right]. \quad (3.6)$$

Para desacoplarmos os termos quadráticos na equação anterior, se usa um truque chamado de identidade de Hubbard-Stratonovich (Hubbard, 1959), que é uma observação do resultado de uma integral gaussiana,

$$\exp(\mathbf{a}^2) = \int_{-\infty}^{+\infty} \prod_{\mu=1}^p \frac{dy^\mu}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \mathbf{y}^2 + \sqrt{2} \mathbf{a} \cdot \mathbf{y} \right). \quad (3.7)$$

Assim, usando (3.7), é possível rescrever o termo quadrático em (3.6), na forma

$$\begin{aligned} \mathcal{Z}(\beta, \kappa^\mu) &= e^{-\beta p/2} \text{Tr}_s \int_{-\infty}^{+\infty} \prod_{\mu=1}^p \frac{dy^\mu}{\sqrt{2\pi}} \exp \left[ \left( \beta N \boldsymbol{\kappa} + \sqrt{\beta N} \mathbf{y} \right) \cdot \mathbf{x} \right] \cdot \exp \left( -\frac{\mathbf{y}^2}{2} \right) \\ &= e^{-\beta p/2} \int_{-\infty}^{+\infty} \prod_{\mu=1}^p \frac{dy^\mu}{\sqrt{2\pi}} \text{Tr}_s \exp \left[ \left( \beta N \boldsymbol{\kappa} + \sqrt{\beta N} \mathbf{y} \right) \cdot \mathbf{x} \right] \cdot \exp \left( -\frac{\mathbf{y}^2}{2} \right). \end{aligned}$$

Após inverter a ordem do traço com a integral, se usa a definição de  $x^\mu$ , e se calcula o traço sobre  $\{S_i\}$ , obtendo

$$\mathcal{Z}(\beta, \kappa^\mu) = e^{-\beta p/2} \int_{-\infty}^{+\infty} \prod_{\mu=1}^p \frac{dy^\mu}{\sqrt{2\pi}} \prod_{i=1}^N \left[ 2 \cosh \left( \left( \beta N \boldsymbol{\kappa} + \sqrt{\beta N} \mathbf{y} \right) \cdot \boldsymbol{\xi}_i / N \right) \right] \cdot \exp \left( -\frac{\mathbf{y}^2}{2} \right).$$

Nessa última expressão, é possível realizar facilmente o traço sobre os estados, devido a que este pode-se escrever como um produto de  $N$  termos, cada um dependendo de um único  $S_i$  com diferente  $i$ .

Fazendo o reescalamento  $y^\mu \rightarrow \sqrt{\beta N} y^\mu$ , na expressão anterior, com o objetivo de garantir a extensividade da energia livre, obtêm-se

$$\mathcal{Z}(\beta, \kappa^\mu) = (\beta N)^{p/2} e^{-\beta p/2} \int_{-\infty}^{+\infty} \prod_{\mu=1}^p \frac{dy^\mu}{\sqrt{2\pi}} \exp \left\{ \sum_{i=1}^N \ln [2 \cosh \beta (\mathbf{y} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}_i] - \frac{\beta N}{2} \mathbf{y}^2 \right\}.$$

No limite termodinâmico, quando  $N \rightarrow \infty$ , é possível resolver a integral em  $y^\mu$  pelo método de ponto de sela, o qual usa a seguinte aproximação

$$\int_{-\infty}^{+\infty} dx e^{-N\beta f(x)} \simeq e^{-N\beta f(x_0)} \sqrt{\frac{2\pi}{N\beta f''(x_0)}}, \quad \text{onde } f'(x_0) = 0. \quad (3.8)$$

Enquanto  $p$  permanece finita, a integral sobre  $y$  é dominada por seu valor de ponto sela,

$$f(\beta, \kappa^\mu) = -\frac{1}{N\beta} \ln \mathcal{Z} = \frac{1}{2} \mathbf{y}^2 - \frac{1}{N\beta} \sum_{i=1}^N \ln [2 \cosh \beta (\mathbf{y} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}_i]. \quad (3.9)$$

Das equações do método de ponto de sela,  $\partial f / \partial y^\mu = 0$ , obtemos os valores de  $y^\mu$  que minimizam o lado direito de (3.9),

$$y^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \tanh [\beta (\mathbf{y} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}_i]. \quad (3.10)$$

No limite termodinâmico, no caso de  $p$  finito, os lados das equações (3.9), (3.10), apresentarão a propriedade de auto média. Ou seja, as somas  $N^{-1} \sum_i$ , produzirão uma média sobre a distribuição de padrões aleatórios  $\xi$ , matematicamente;

$$\frac{1}{N} \sum_{i=1}^N (\dots) = \langle \dots \rangle_\xi. \quad (3.11)$$

Obtemos assim, as correspondentes equações de campo médio

$$\begin{aligned} f(\beta, \kappa^\mu) &= \frac{1}{2} \sum_{\mu=1}^p (y^\mu)^2 - \frac{1}{\beta} \langle \ln [2 \cosh \beta (\mathbf{y} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}] \rangle_\xi \\ \mathbf{y} &= \langle \xi \tanh [\beta (\mathbf{y} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}] \rangle_\xi. \end{aligned} \quad (3.12)$$

Com o objetivo de interpretar fisicamente o valor que o parâmetro de ordem  $x^\mu$  assume no ponto de sela, calcula-se a média de ensemble das superposições a partir da sua definição:

$$m^\mu = \left\langle \frac{1}{N} \sum_{i=1}^N \langle S_i \rangle_T \xi_i^\mu \right\rangle_\xi.$$

Como  $\langle S_i \rangle_T = \mathcal{Z}^{-1} \sum_{\{S_i\}} S_i \exp(-\beta \mathcal{H})$  é a média térmica de spin no sitio  $i$ . Conclui-se que

$$m^\mu = \left\langle \frac{1}{\mathcal{Z}} \sum_{\{S_i\}} \left( \frac{1}{N} \sum_{i=1}^N S_i \xi_i^\mu \right) \exp(-\beta \mathcal{H}) \right\rangle_\xi = -\frac{\partial f(\beta, \kappa^\mu)}{\partial \kappa^\mu}.$$

Para achar a expressão de  $m^\mu$ , derivamos (3.12) respeito de  $\kappa^\mu$ , obtendo-se

$$\mathbf{m} = \langle \xi \tanh [\beta (\mathbf{m} + \boldsymbol{\kappa}) \cdot \boldsymbol{\xi}] \rangle_\xi. \quad (3.13)$$

Da comparação da equação anterior com (3.12), verifica-se que

$$\mathbf{y} = \mathbf{m}. \quad (3.14)$$

Ou seja, o valor que a variável  $\mathbf{y}$ , assume no ponto de sela é exatamente a da “superposição”  $\mathbf{m}$ . Onde o vetor  $\mathbf{m}$ , que representa o conjunto de “superposições”, é dado por

$$\mathbf{m} = (m^1, m^2, \dots, m^p) \quad (3.15)$$

No modelo estudado se pressupõe interações de alcance infinito entre os neurônios do sistema. O vetor  $\mathbf{m}$ , que caracteriza macroscopicamente os estados de equilíbrio, é dado pela solução das equações de campo médio (3.13), que corresponde à solução exata do modelo.

### 3.2 TCM - Limite de $\alpha = p/N$ finito

Com objetivo de analisar a capacidade de armazenamento do sistema, apresentamos as características gerais da teoria de campo médio do modelo de Hopfield no limite em que o número de padrões armazenados  $p$  aumenta com o tamanho  $N$  da rede, como  $p = \alpha N$ . Nós, nos concentramos aqui no caso de padrões aleatórios não correlacionados e seguimos os passos de Amit Gutfreund Sompolinsky (AGS) em (Amit et al., 1985b).

Sabe-se que o hamiltoniano para o modelo de Hopfield é dado por,

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j=1}^N J_{ij} S_i S_j - \sum_{\nu=1}^s \kappa^\nu \sum_{i=1}^N \xi_i^\nu S_i \quad (S_i = \pm 1), \quad (3.16)$$

onde  $p$  é o número de padrões aleatórios aprendidos e  $\kappa^\nu$  é um campo conjugado a um dos  $s$  “padrões condensados”  $\{\xi_i^\nu\}$  finitos ( $s \ll p$ ). A força de interação  $J_{ij}$ , construída a partir dos  $p$  padrões aleatórios memorizados ( $\xi_i^\mu = \pm 1$ ,  $\mu = 1, \dots, p$ ) obedece a regra de Hebb (2.7).

Nós tratamos os  $J_{ij}$ 's como congelados ou temperados, pelo que a influência do processo de recuperação da memória é ignorada. Ou seja, assumimos que as conexões sinápticas mudam muito mais lentamente que os estados dos neurônios. Isto é, o modelo de Hopfield trata apenas os aspectos dinâmicos dos neurônios, e não de todo o processo de aprendizagem, ou seja, trata-se de um sistema em que o aprendizado foi completado.

Trataremos então com estados iniciais (estímulos) que possuam correlações finitas apenas com um número finito  $s$  de padrões. A grandeza de interesse a ser calculada neste sistema, é a energia livre média por neurônio

$$f(\beta) = \lim_{N \rightarrow \infty} \left( \frac{-1}{\beta N} \right) \langle \ln \mathcal{Z} \rangle_\xi, \quad (3.17)$$

onde  $\langle \dots \rangle_\xi$  indica a média sobre o desordem temperado  $\{\xi_i^\mu\}$ . Para o cálculo da energia livre, é necessário determinar a média

$$\langle \ln \mathcal{Z} \rangle_\xi \equiv \int dJ_{ij} P(\{J_{ij}\}) \ln [\text{Tr}_S \exp(-\beta \mathcal{H})], \quad (3.18)$$

onde o símbolo  $\text{Tr}_S$  refere-se ao traço tomado sobre os  $N$  spins de Ising e  $P(\{J_{ij}\})$  é a distribuição de todos os padrões binários aleatórios  $\{\xi_i^\mu\}$ . Estimando a média  $\langle \ln \mathcal{Z} \rangle_\xi$  corretamente, pode-se obter as quantidades médias das grandezas que queremos conhecer. Infelizmente, é muito difícil calcular essa média diretamente, e ela não é idêntica a  $\ln \langle \mathcal{Z} \rangle_\xi$  (Kirkpatrick e Sherrington, 1978).

Afortunadamente existe uma técnica, chamada de “método de réplica”, que nos permite calcular a média de  $\ln \mathcal{Z}$ . Como a média de uma potência de  $\mathcal{Z}$  é mais facilmente obtida, usa-se o limite (Mézard et al., 1987):

$$\ln \mathcal{Z} = \lim_{n \rightarrow 0} \frac{\mathcal{Z}^n - 1}{n}, \quad (3.19)$$

que dá a energia livre por neurônio calculada sobre os padrões temperados na forma,

$$f(\beta) = \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \left( \frac{-1}{\beta N} \right) \frac{\langle \mathcal{Z}^n \rangle_\xi - 1}{n}. \quad (3.20)$$

Portanto, uma discussão central é a estimativa da média da função de partição  $n$ -replicada,  $\mathcal{Z}^n$ , e esta é a ideia básica do método de réplica. A função de partição  $n$ -replicada, é dada pelo produto de  $n$  funções de partição, correspondentes a  $n$  sistemas, ou réplicas, idênticas ao sistema original, assim

$$\mathcal{Z}^n = \prod_{a=1}^n \mathcal{Z}(\{S_i^a\}) = \text{Tr}_{\{S_i^a\}} \exp \left( -\beta \sum_{a=1}^n \mathcal{H}^a \right). \quad (3.21)$$

A somatória na exponencial da equação anterior, representa a energia de um sistema de  $n \cdot N$  spins. Note-se que o traço é feito sobre o conjunto total de variáveis de spin.

Usando (3.16), o hamiltoniano da  $a$ -ésima réplica do sistema, terá a seguinte forma:

$$\mathcal{H}^a = -\frac{1}{2N} \sum_{i,j \neq i}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu S_i^a S_j^a - \sum_{\nu=1}^s \kappa^\nu \sum_{i=1}^N \xi_i^\nu S_i^a, \quad (3.22)$$

assim, a expressão de  $\langle \mathcal{Z}^n \rangle_\xi$  será dada por

$$\langle \mathcal{Z}^n \rangle_\xi = \left\langle \text{Tr}_S \left[ \prod_{a=1}^n \exp \left( \frac{\beta N}{2} \sum_{\mu=1}^p \left( \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i^a \right)^2 - \frac{\beta p}{2} + \beta \sum_{\nu=1}^s \kappa^\nu \sum_{i=1}^N \xi_i^\nu S_i^a \right) \right] \right\rangle_\xi. \quad (3.23)$$

Para o cálculo da média envolvida na equação anterior, desacoplamos os termos quadráticos usando a identidade de Hubbard-Stratonovich, pelo que a equação de  $\langle \mathcal{Z}^n \rangle_\xi$ , toma a forma

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_\xi &= e^{-\beta p n / 2} \left\langle \text{Tr}_s \left[ \int \prod_{a=1}^n \prod_{\mu=1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( - \sum_{\mu=s+1}^p \sum_{a=1}^n \frac{(y_a^\mu)^2}{2} + \sqrt{\frac{\beta}{N}} \sum_{\mu=s+1}^p \sum_{a=1}^n y_a^\mu \sum_{i=1}^N \xi_i^\mu S_i^a \right) \right. \right. \\ &\quad \left. \left. \times \exp \left( - \sum_{\nu=1}^s \sum_{a=1}^n \frac{(y_a^\nu)^2}{2} + \sum_{\nu=1}^s \sum_{a=1}^n \left( \sqrt{N\beta} y_a^\nu + \beta N \kappa^\nu \right) \frac{1}{N} \sum_{i=1}^N \xi_i^\nu S_i^a \right) \right] \right\rangle_\xi. \end{aligned} \quad (3.24)$$

Tendo em conta que faremos uma média sobre  $s$  padrões  $\xi$ 's discretos e uma média sobre um número de padrões,  $p - s$ , eventualmente, gaussiano. Seguimos os seguintes passos

1. Calcula-se primeiramente o termo ( $\equiv L_1$ ), que corresponde à média aleatória da exponencial para “padrões não condensados”  $\{\xi_i^\mu\}$  ( $\mu = s + 1, \dots, p$ ), obtendo

$$L_1 = \exp \left[ - \sum_{\mu=s+1}^p \sum_{a=1}^n \frac{(y_a^\mu)^2}{2} + \sum_{\mu=s+1}^p \sum_{i=1}^N \ln \left( \cosh \left( \sqrt{\frac{\beta}{N}} \sum_{a=1}^n y_a^\mu S_i^a \right) \right) \right], \quad (3.25)$$

que, no limite  $N \rightarrow \infty$ , toma a forma

$$L_1 \approx \exp \left[ - \sum_{\mu=s+1}^p \sum_{a=1}^n \frac{(y_a^\mu)^2}{2} + \frac{\beta}{2N} \sum_{\mu=s+1}^p \sum_{i=1}^N \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right]. \quad (3.26)$$

2. Integrando os  $y_a^\mu$ 's presentes em  $L_1$ , e após o reescalonamento  $y_a^\mu \rightarrow \sqrt{N\beta} y_a^\mu$ , temos

$$\begin{aligned} \int \prod_{a=1}^n \prod_{\mu=s+1}^p L_1 &= \int \prod_{a \neq b} dr_{ab} \int \prod_{a \neq b} dq_{ab} \exp \left[ - \frac{p}{2} \text{Tr} [\ln ((1 - \beta) \mathbf{I} - \beta \mathbf{Q})] \right] \\ &\quad \times \exp \left[ - \frac{N\alpha\beta^2}{2} \sum_{a \neq b} r_{ab} \left( q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b \right) \right], \end{aligned} \quad (3.27)$$

com  $a, b = 1, \dots, n$ . Para finalizarmos os cálculos nos dois primeiros passos, foi preciso usar a equação de Hubbard-Stratonovich duas vezes e uma das propriedades da Delta de Dirac, os quais permitiram a introdução de três novas variáveis  $y_a^\mu$ ,  $q_{ab}$  e  $r_{ab}$ , chamadas “parâmetros de ordem”.

3. Ao introduzir o resultado obtido em (3.27) na equação (3.24) e tendo em conta o fato de que a média temperada sobre o número finito de  $\xi^\nu$ 's ( $\nu = 1, \dots, s$ ) pode ser efetuada pela automedia, devido a que  $N \rightarrow \infty$ , encontra-se

$$\langle \mathcal{Z}^n \rangle_\xi \propto \int \prod_{\nu, a} dy_a^\nu \int \prod_{a \neq b} dq_{ab} dr_{ab} \exp [-N\beta\mathcal{F}(y_a^\nu, q_{ab}, r_{ab})] \quad (3.28)$$

onde

$$\begin{aligned} \mathcal{F} = & \frac{\alpha n}{2} + \frac{1}{2} \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu)^2 + \frac{\alpha\beta}{2} \sum_{a \neq b}^n r_{ab} q_{ab} + \frac{\alpha}{2\beta} \text{Tr} [\ln ((1 - \beta) \mathbf{I} - \beta \mathbf{Q})] \\ & - \frac{1}{\beta} \left\langle \ln \text{Tr}_{s^a} \left[ \exp \left( \frac{\alpha\beta^2}{2} \sum_{a \neq b}^n r_{ab} S^a S^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu + \kappa^\nu) \xi^\nu S^a \right) \right] \right\rangle_{\xi^\nu} \end{aligned} \quad (3.29)$$

Aqui nós notamos que a média  $\langle O \{ \xi^\nu \} \rangle_{\xi^\nu} \equiv 2^{-s} \sum_{\nu=1}^s \sum_{\xi^\nu = \pm 1} O \{ \xi^\nu \}$  denota a média aleatória para  $s (\ll p)$  “padrões condensados”.  $\mathbf{I}$  é uma matriz unitária com  $n \times n$  elementos, onde  $n$  é um número total de réplicas, enquanto  $\mathbf{Q}$  é uma matriz de réplica que fornece o parâmetro de ordem  $q_{ab}$ , e  $a$  (ou  $b$ ) denota um índice para a réplica  $a$ -ésima. Com  $\text{Tr}_{s^a}$ , denota-se explicitamente que o traço corresponde agora a uma soma sobre os valores de spin para cada réplica em um único “sitio”.

Note-se que termos de ordem  $n$ , no argumento da exponencial em (3.28), foram depreciados pois no limite termodinâmico, os termos persistentes são os termos de ordem  $nN$ . Podemos concluir que o método das réplicas nos permite por tanto, desacoplar a soma sobre os sítios, ao preço de um acoplamento sobre as réplicas.

As integrais em (3.28), podem ser resolvidas assintoticamente usando o método de Laplace (de Bruijn, 1981), devido a que o argumento da função exponencial, é proporcional ao número de neurônios  $N$ . Segundo o método de Laplace, o valor da integral será dado pelo integrando em seu valor máximo, a menos de termos da ordem de  $N^{-1}$ . Obtendo-se assim,

$$\langle \mathcal{Z}^n \rangle = \exp [-N\beta\mathcal{F}(y_a^\nu, q_{ab}, r_{ab})] + \mathcal{O}(N) \quad (3.30)$$

onde  $\mathcal{F}(y_a^\nu, q_{ab}, r_{ab})$ , dada pela equação (3.29), deverá ser mínima. A aplicação do método de Laplace pressupõe que a inversão na ordem dos limites em  $N$  e  $n$  é lícita, pois a princípio deve-se calcular o limite  $n \rightarrow 0$  antes de avaliar o limite no número de partículas  $N$ . Dita proposição, usada nos trabalhos de Sherrington e Kirkpatrick (SK), foi analisada por Van Hemmen e Palmer (1979), que concluíram que, ao menos no modelo SK, não existe problema em trocar a ordem dos limites.

A densidade de energia livre,  $f$ , é dada então por

$$f = \frac{\alpha}{2} + \frac{1}{2n} \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu)^2 + \frac{\alpha\beta}{2n} \sum_{a \neq b}^n r_{ab} q_{ab} + \frac{\alpha}{2\beta n} \text{Tr} [\ln ((1 - \beta) \mathbf{I} - \beta \mathbf{Q})] - \frac{1}{\beta n} \langle \ln \mathcal{Z}_0 \rangle_{\xi^\nu} \quad (3.31)$$

onde o operador  $\mathcal{Z}_0$ , tem a forma

$$\mathcal{Z}_0 = \text{Tr}_{S^a} \left[ \exp \left( \frac{\alpha\beta^2}{2} \sum_{a \neq b}^n r_{ab} S^a S^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu + \kappa^\nu) \xi^\nu S^a \right) \right], \quad (3.32)$$

e as variáveis  $m_a^\nu$ ,  $q_{ab}$  e  $r_{ab}$  obedecem às equações de ponto sela

$$\frac{\partial \mathcal{F}}{\partial m_a^\nu} = 0, \quad \frac{\partial \mathcal{F}}{\partial q_{ab}} = 0, \quad \frac{\partial \mathcal{F}}{\partial r_{ab}} = 0. \quad (3.33)$$

As equações (3.31) e (3.32) descrevem completamente os estados estáveis do sistema analisado, no limite  $N \rightarrow \infty$ . Esta teoria de campo médio é exata porque a rede está totalmente conectada, como no modelo Sherrington-Kirkpatrick (Amit et al., 1987b). Aqui, enfatizamos que o parâmetro  $\alpha \equiv p/N$  se refere a um nível de armazenamento ou taxa de carregamento de memória, que altera a fase do sistema.

A definição dos parâmetros de ordem  $m_a^\nu$ ,  $r_{ab}$  e  $q_{ab}$  em (3.31), podem ser obtidos tomando os valores de ponto sela em todas as integrais de (3.23). Ao fazer  $\partial \langle \mathcal{Z}^n \rangle / \partial m_a^\nu = 0$  para os padrões recuperados (condensados)  $\nu \leq s$  obtemos

$$-m_a^\nu \langle \mathcal{Z}^n \rangle + \dots \text{Tr}_{S^a} \int \left\langle \dots \frac{1}{N} \sum_i \xi_i^\nu S_i^a \right\rangle_\xi = 0,$$

ou seja,

$$m_a^\nu = \left\langle \frac{1}{N} \sum_{i=1}^N \xi_i^\nu \langle S_i^a \rangle_T \right\rangle_\xi, \quad (3.34)$$

que é o parâmetro de ordem chamado de “magnetização”, que indica a qualidade de recuperação do  $\nu$ -ésimo padrão memorizado para o equilíbrio térmico temperado da réplica  $a$ . Em (3.34),  $\langle \dots \rangle_T$ , denota uma média térmica, isto é, uma média sobre aquela parte do espaço de estados que a dinâmica da rede permite. A média pode ser considerada como uma média temporal ou como uma média de ensemble (conjunto).

De forma semelhante, a partir de  $\partial \langle \mathcal{Z}^n \rangle / \partial r_{ab} = 0$ , é possível obter o parâmetro de ordem de Edwards-Anderson

$$q_{ab} = \left\langle \frac{1}{N} \sum_{i=1}^N \langle S_i^a \rangle_T \langle S_i^b \rangle_T \right\rangle_\xi, \quad (3.35)$$

que na imagem da réplica caracteriza as correlações causadas pela igualdade de padrões temperados entre duas réplicas termicamente independentes. Ao fazermos  $\partial \langle \mathcal{Z}^n \rangle / \partial q_{ab} = 0$ , identifica-se o multiplicador de Lagrange  $r_{ab}$ , dado por

$$r_{ab} = \frac{1}{\alpha} \sum_{\mu=s+1}^p \langle m_a^\mu m_b^\mu \rangle_\xi, \quad (3.36)$$

que é o parâmetro de ordem que descreve o ruído devido aos padrões não condensados (não recuperados). Em outras palavras, se a rede estiver em um estado com grandes superposições (macroscópicas) com alguns dos padrões memorizados (condensados), o acúmulo das superposições aleatórias com todos os outros padrões cria uma quantidade significativa de ruído.

Para realizar o cálculo do limite  $n \rightarrow 0$  na equação (3.31), que representa a densidade de energia livre, é necessário considerarmos algum tipo de condição inicial sobre a estrutura de  $m_a^\nu$ ,  $r_{ab}$  e  $q_{ab}$ . Além disso, a introdução destas variáveis, apenas como uma abstrata manipulação algébrica, torna possível o cálculo da energia livre, sem permitir, entretanto, a análise termodinâmica do modelo, que só é possível se estas variáveis possuírem algum sentido físico (Geszi, 1990).

### 3.2.1 Soluções de simetria de Réplicas

A maior parte da discussão de AGS foi realizada dentro da teoria de “simétrica de réplicas”, condição inicialmente adotada por Sherrington e Kirkpatrick (1975). Na “simetria de réplicas” se assume, como uma primeira aproximação, só um pequeno subconjunto dos parâmetros variacionais ( $y_a^\nu$ ,  $q_{ab}$ ,  $r_{ab}$ ) no qual, a todas as réplicas se lhes atribui os mesmos valores dos parâmetros. Em outras palavras, se assume que cada parâmetro de ordem é simétrico sob uma permutação de índices de réplica como

$$\begin{aligned} m_a^\nu &= m^\nu, \\ q_{ab} &= q, \quad a \neq b, \\ r_{ab} &= r. \quad a \neq b, \end{aligned} \tag{3.37}$$

pelo que é possível minimizar a densidade de energia livre,  $f$ , usando apenas estes três parâmetros de ordem. Aliás agora é mais fácil calcular  $\text{Tr}[\ln((1 - \beta)\mathbf{I} - \beta\mathbf{Q})]$ , pois a matriz  $(1 - \beta)\mathbf{I} - \beta\mathbf{Q}$ , sob as condições de simetria de réplicas, possuirá, um valor próprio  $1 - \beta - (n - 1)\beta q$  não degenerado, e outro valor próprio  $1 - \beta(1 - q)$  que é  $(n - 1)$  vezes degenerado (vide apêndice E). A densidade de energia livre pode ser representada, usando os parâmetros de ordem da simetria de réplicas dados em (3.37), após fazer o correspondente processo limite  $n \rightarrow 0$ , como

$$f = \frac{\alpha}{2} + \frac{1}{2} \sum_{\nu=1}^s (m^\nu)^2 + \frac{\alpha\beta}{2} r (1 - q) + \frac{\alpha}{2\beta} \left[ \ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q} \right] - \frac{1}{\beta} \left\langle \ln \left[ 2 \cosh \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s (m^\nu + \kappa^\nu) \xi^\nu \right) \right] \right\rangle_z, \quad (3.38)$$

onde  $\langle \dots \rangle_z \equiv \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \langle \dots \rangle_{\xi^\nu}$  representa uma média dupla: sobre os padrões condensados (recuperados)  $\xi^\nu$ , e sobre o ruído multiplicativo gaussiano  $z$ . A essência da derivação do último termo é a aplicação da identidade de Hubbard-Stratonovitch (3.7) ao termo quadrático  $(\sum_a S^a)^2$  que aparece em (3.32).

### 3.2.2 Equações de campo médio

Todas as soluções de  $f$ , incluindo o mínimo local, são estados estacionários do processo dinâmico, com barreiras de  $\mathcal{O}(N)$ . Fazendo a variação da energia livre  $f$  (sob condições de simetria de réplicas) com respeito a  $y^\nu$ ,  $q$  e  $r$ , obtém-se às equações para estado estacionário. Assim, depois de fazer uma integração por partes, temos:

$$m^\nu = \left\langle \xi^\nu \tanh \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s (m^\nu + \kappa^\nu) \xi^\nu \right) \right\rangle_z, \quad (3.39)$$

$$r = \frac{q}{(1 - \beta + \beta q)^2}, \quad (3.40)$$

$$q = \left\langle \tanh^2 \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s (m^\nu + \kappa^\nu) \xi^\nu \right) \right\rangle_z, \quad (3.41)$$

onde a média  $\langle \dots \rangle_z$  refere-se à média combinada sobre os  $\xi^\nu$ 's e sobre o ruído gaussiano dado pelo operador integral  $\int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ . Aqui notamos que, no ponto de sela, o valor dos parâmetros dados pelas eqs. (3.39)-(3.41) tem apenas os significados físicos definidos nas eqs. (3.34)-(3.36), respetivamente.

Note-se que o campo local (vide o argumento das funções) consiste de três partes; uma parte “ferromagnética”,  $\mathbf{m}$ , que resulta das superposições condensadas, uma parte “vidro de spin”,  $z\sqrt{\alpha r}$ , gerada pela soma das superposições com o resto de padrões e uma parte “campo externo”, originada pelo agente externo ao sistema.

As equações (3.39) - (3.41), quando  $\kappa = 0$  tem dois tipos de soluções que são localmente estáveis para variações em  $\mathbf{m}$ ,  $q$  e  $r$ .

1. Uma solução com  $\mathbf{m} = 0$ ,  $q$  e  $r \neq 0$ . Ela representa um estado vidro de spin que não tem uma superposição macroscópica com qualquer um dos padrões.
2. Soluções ferromagnéticas com  $\mathbf{m} \neq 0$ , além de  $q$  e  $r$ . Essas soluções, que existem para  $\alpha$  suficientemente pequeno, tornam o sistema útil para a memória associativa.

As soluções ferromagnéticas mais importantes são caracterizadas por superposições macroscópicas com um único padrão  $m^\nu = m \delta^{\nu\rho}$ , os chamados estados Mattis com  $n = 1$ . Existem em total  $2N\alpha$  soluções degeneradas de este tipo. Quando  $\alpha \rightarrow 0$  eles se aproximam dos  $p$  finitos estados Mattis (Amit et al., 1985b), que é o caso de duas soluções.

### 3.3 Soluções em $T \rightarrow 0$

Com o objetivo de estudar a capacidade de armazenamento de memória do modelo de Hopfield, analisa-se as equações de AGS no limite de ruído zero ( $T \rightarrow 0$ ) nos casos de o campo externo estar acoplado com um único padrão armazenado e quando o campo externo é ortogonal com todos os padrões armazenados.

#### 3.3.1 Campo acoplado com um único padrão

A continuação, lidamos com a solução para a chamada “fase de recuperação ferromagnética” (FM) correspondente ao caso  $m^\nu = m \delta^{\nu\rho}$ , ou seja, soluções com uma única superposição finita. Com objetivo de estudar este regime analiticamente, observa-se que no limite  $T \rightarrow 0$  (ou  $\beta \rightarrow \infty$ );

$$\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh \beta (Az + x) \rightarrow \operatorname{erf} (x/\sqrt{2}A) \quad (3.42)$$

e que

$$\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2 \beta (Az + x) \rightarrow 1 - \frac{1}{\beta A} \sqrt{\frac{2}{\pi}} e^{-x^2/2A^2}. \quad (3.43)$$

Observe-se que a equação (3.42) foi reescrita usando a função erro,  $\operatorname{erf}(x)$ , definida por

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dz e^{-z^2}. \quad (3.44)$$

Introduzindo estas aproximações, quando  $T \rightarrow 0$ , nas equações (3.39) e (3.41), obtemos

$$m \simeq \operatorname{erf}\left(\frac{m+h}{\sqrt{2\alpha r}}\right), \quad (3.45)$$

$$q \simeq 1 - \frac{1}{\beta} \sqrt{\frac{2}{\pi\alpha r}} e^{-(m+\kappa)^2/2\alpha r}. \quad (3.46)$$

Observa-se no denominador da equação (3.40), uma ambiguidade aparente pois o produto  $\beta(1-q)$  é indeterminado para  $\beta \rightarrow \infty$ . Porém, o limite pode-se expressar através de  $m$  e  $r$ . Com esse objetivo, substitui-se (3.46) na equação (3.40) e tendo em conta que a temperatura zero, o parâmetro de Edwards-Anderson tende a unidade ( $q \rightarrow 1$ ), obtemos

$$y = \frac{\operatorname{erf}(y+x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-(y+x)^2}}, \quad (3.47)$$

onde  $m = y\sqrt{2\alpha r}$  e  $\kappa = x\sqrt{2\alpha r}$ . Esta equação proporciona a dependência da qualidade de recuperação  $m$  com o nível de armazenamento  $\alpha$  e  $\kappa$ , quando  $T \rightarrow 0$ . A capacidade de armazenamento é simplesmente o valor de  $\alpha$  acima do qual a equação não tem solução, exceto  $y = 0$ .

Na figura 3.1 apresentam-se curvas de magnetização  $m(\alpha)$ , para  $\kappa$  fixo. As linhas correspondem à magnetização de um padrão marcado. O gráfico interno à direita, mostra a magnetização corrigida com objetivo de assegurar a inviolabilidade da concavidade da energia livre do sistema respeito da magnetização<sup>1</sup>. Os saltos presentes nas curvas corrigidas, para  $\kappa = 0.1$  e  $\kappa = 0.3$ , caracterizam uma transição fase de primeira ordem.

A grandes valores de  $\alpha$ , apenas o estado Vidro de Spin é estável. Na presença do campo externo há uma superposição não nula,  $m$ , com o padrão marcado, o qual se desenvolve continuamente a partir de zero, e apresenta um comportamento inicial linear com o campo  $\kappa$  (vide fig. 3.2). Quando  $\alpha$  diminui, um estado metaestável aparece em  $\alpha_C(\kappa)$ . Este é um estado de recuperação FM, com um alto valor de  $m$ . À medida que  $\alpha$  é reduzido ainda mais, esse estado se torna globalmente estável. Finalmente, em  $\alpha$  muito pequeno, o estado SG, isto é, o estado com  $m$  baixo, desaparece e apenas o estado FM persiste.

A figura 3.2 mostra a dependência da magnetização  $m$ , com o campo externo  $\kappa$ , para diferentes valores de  $\alpha$ . Observa-se que para  $\kappa$  muito grande a magnetização sempre tende a unidade. Já, para  $\kappa$  e  $\alpha$  pequeno, obtém-se três soluções para a magnetização, duas das quais

<sup>1</sup>A energia livre, deve ser uma função concava de  $m$  por uma questão de estabilidade dinâmica.

serão estáveis e a terceira solução, além de possuir maior energia livre, será também instável (convexa) do ponto de vista termodinâmico.

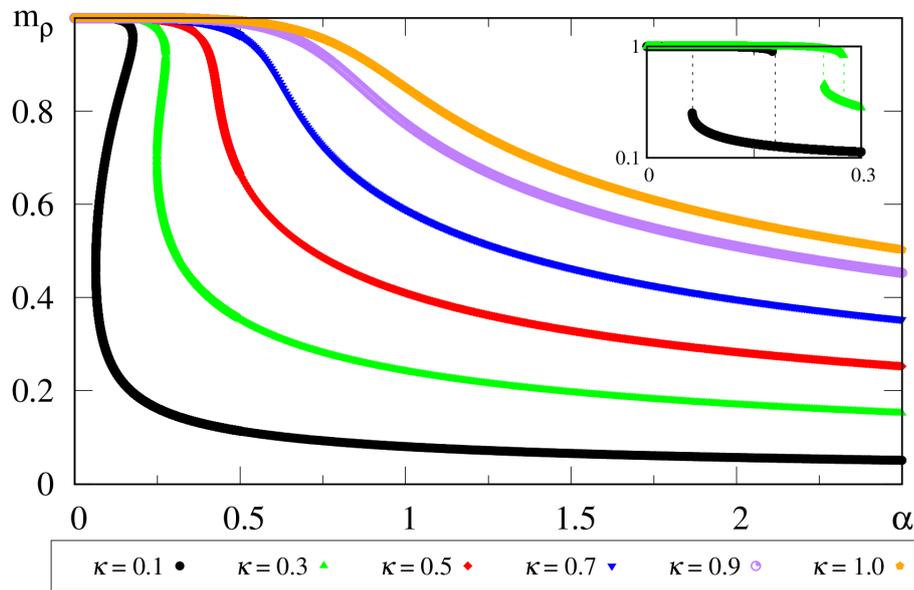


Figura 3.1: Curvas de magnetização  $m$  vs  $\alpha$ , para diversos valores de  $\kappa$ . As linhas correspondem à magnetização de um padrão “marcado”. O gráfico interno à direita, mostra a magnetização corrigida como função de  $\alpha$ , para  $\kappa = 0.1$  e  $\kappa = 0.3$ .

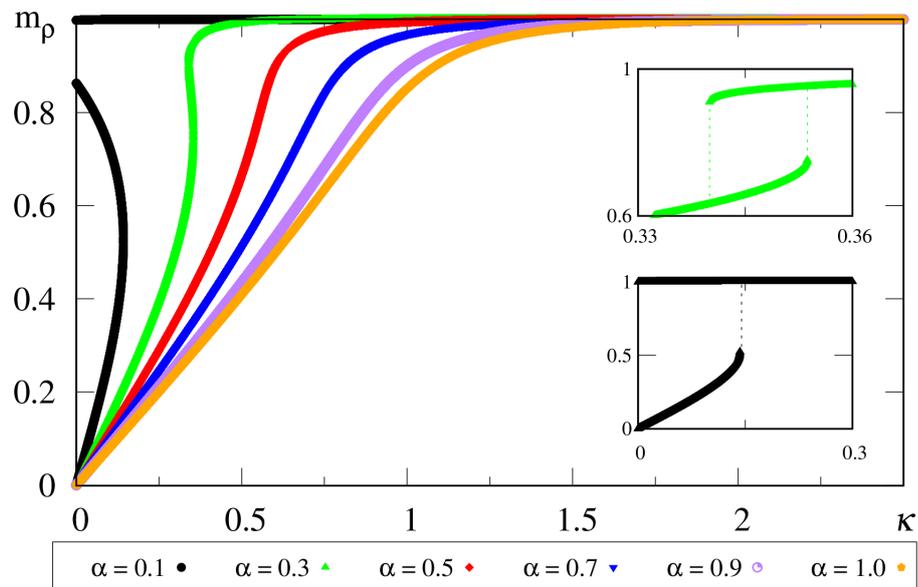


Figura 3.2:  $m(h)$ , para diversos valores de  $\alpha$ , no caso de reconhecimento de um padrão armazenado. Na figura inferior direita se mostra a curva corrigida para  $\alpha = 0.1$  e  $\alpha = 0.3$ .

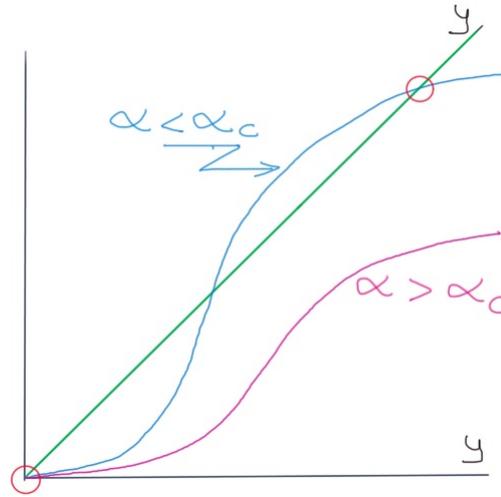


Figura 3.3: Solução da equação (3.47) quando  $x \rightarrow 0$ . (esquemático)

### 3.3.2 Campo Intrínseco Nulo ( $\kappa = 0$ )

No caso de campo intrínseco nulo, a equação (3.47) se reduz a uma simples equação para a variável  $y (= m/\sqrt{2\alpha r})$ , ou seja

$$y = \frac{\text{erf}(y)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}}e^{-y^2}}. \quad (3.48)$$

A solução gráfica de (3.48), mostra-se na figura 3.3. Graficamente, uma solução da equação  $y = \phi(y)$  é a abscissa do ponto de intersecção da reta  $z = y$  e da curva  $z = \phi(y)$ . Na figura 3.3 se mostra o gráfico da linha reta do lado esquerdo e da curva do lado direito, plotada para dois valores de  $\alpha$ , um acima e outro abaixo do valor crítico

$$\alpha_C = 0.138, \quad (3.49)$$

obtido numericamente. Esse valor de  $\alpha_C$  é atualmente o adotado na literatura. Para  $\alpha < \alpha_C$ , temos três interseções, o vidro de spin com  $m = 0$  e os estados de recuperação ferromagnética com  $m$  alto, são dinamicamente estáveis e também mínimos de energia. Quando  $\alpha > \alpha_C$ , somente a interseção  $m = 0$  persiste. Esta figura torna intuitivamente claro que o desaparecimento da solução de recuperação ocorre abruptamente.

Para uma análise detalhada do sistema em todo o espaço  $T - \alpha$  (temperatura-memória de carga), consultar os dois artigos de Amit, Gutfreund e Sompolinsky (1985a; 1985b).

### 3.3.3 Campo Intrínseco Ortogonal aos padrões armazenados

A aplicação de um campo intrínseco ortogonal a qualquer configuração da rede, induzirá uma superposição diferente de zero com esse estado. Por exemplo, se

$$h_i = \kappa \eta_i, \quad (3.50)$$

onde  $\{\eta_i\}$  é uma configuração de rede, não correlacionada com qualquer um dos  $\{\xi_i\}$  padrões que estão embutidos nos acoplamentos  $J_{ij}$ , então para valores finitos de  $p$  ( $\alpha = 0$ ) há um estado metaestável em  $T = 0$ , com  $\{S_i = \eta_i\}$ , para  $h$  arbitrário. No entanto, assim que  $\alpha$  se torna finito, a superposição deste estado metaestável com  $\{\eta_i\}$  diminui muito acentuadamente. A potencial utilidade de um campo acoplado a um padrão memorizado, está no fato de que a superposição do estado FM, condensado nesse padrão, permanece alto mesmo quando  $\alpha$  não é muito pequena.

As equações para a magnetização são derivadas, neste caso, pelo procedimento descrito nas Seções (3.2), exceto que agora todos os  $\xi$ 's são calculados na transição da Eq. (3.25) a (3.27). O resultado é,

$$m = \operatorname{erf} \left( \frac{\kappa}{w} \right) \quad (3.51)$$

e  $w = \sqrt{2\alpha r}$  é determinada por

$$w = \sqrt{2\alpha} + (2/\sqrt{\pi}) e^{-\kappa^2/w^2}. \quad (3.52)$$

Para  $\alpha$  pequeno, a eq. (3.52) tem duas soluções estáveis, uma das quais tem  $w$  muito pequeno e, portanto,  $m \simeq 1$ . Na figura 3.4 mostram-se curvas de magnetização  $m(\alpha)$ , para  $\kappa$  fixo. As linhas correspondem à magnetização de um padrão aleatório, que não foi aprendido. O gráfico interno à direita, mostra a magnetização corrigida. Deve-se perceber que o aumento no valor de  $\alpha_C$ , descrito acima, se aplica apenas ao padrão marcado: o acoplado com  $\kappa$ . De fato, a presença de  $\kappa$  em um padrão produz ruído aleatório para os outros padrões. Isso resulta em uma redução no valor de  $\alpha_C$ , para os outros padrões (Amit et al., 1987b).

Consequentemente, a melhoria na recuperação de um padrão marcado pode ocorrer às custas da capacidade de recuperar o restante dos padrões. Em outras palavras, se a rede estiver um pouco abaixo da saturação, a marcação de um padrão reduzirá o erro em sua recuperação. Contudo, pode ao mesmo tempo impossibilitar a recuperação de todo o resto dos padrões memorizados, reduzindo o  $\alpha$  crítico abaixo do nível de armazenamento real. Por outro lado, se a

rede está inicialmente acima da saturação, um padrão marcado pode ser recuperável, enquanto o dano ao resto dos padrões memorizados é irrelevante, já que eles estavam com magnetização praticamente igual a zero.

Nas figuras 3.4 e 3.5 (de forma semelhante a 3.1 e 3.2) é preciso assegurar a inviolabilidade da concavidade da energia livre do sistema. Por tal motivo, foi tirada uma parte das curvas de magnetização. Observe-se que, somente foi desprezada uma parte da curva onde tem-se três soluções para  $m$ . Duas dessas soluções sempre serão estáveis para um intervalo de valores de  $\kappa$ , mas a solução correta será aquela de menor energia livre. A terceira solução, além de possuir uma energia livre maior, é também instável (convexa) do ponto de vista termodinâmico.

Todas estas figuras foram obtidas, ao resolver numericamente as equações (3.47), (3.48) e (3.52) com ajuda do programa MAPLE XV. A estratégia usada para a obtenção das curvas é descrita brevemente na subseção 4.7.2.

No próximo capítulo apresentaremos o modelo de rede neuronal dependente do estímulo, no qual empregaremos as diversas ferramentas matemáticas usadas pelos pioneiros no estudo das redes neuronais atratoras; Amit, Gutfreund e Sompolinsky.

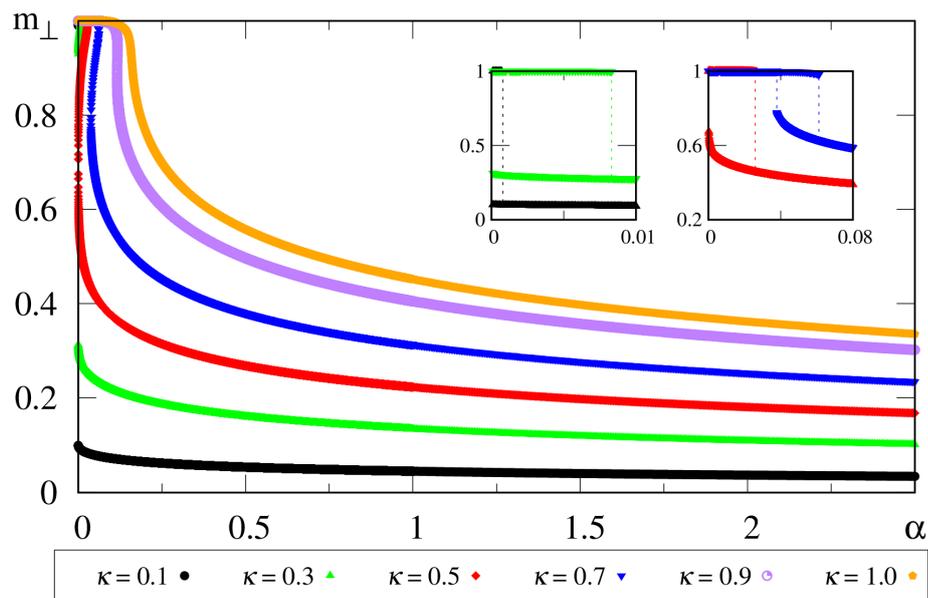


Figura 3.4: Curvas de magnetização  $m$  vs  $\alpha$ , para  $\kappa$  fixo. As linhas correspondem à magnetização de um padrão ortogonal. O gráfico interno à direita, mostra a magnetização corrigida como função de  $\alpha$ , para  $\kappa = 0.1$ ,  $\kappa = 0.3$  e  $\kappa = 0.9$ .

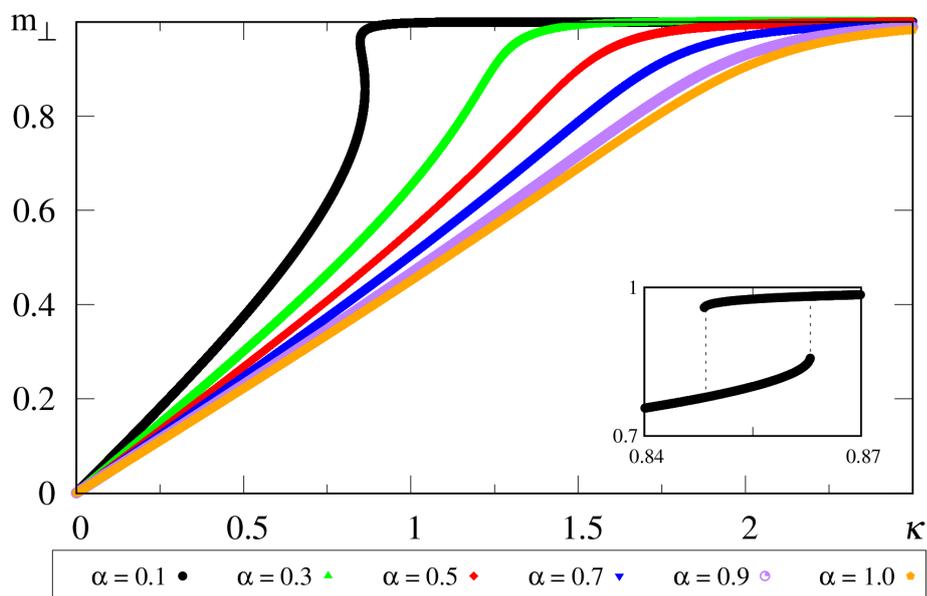


Figura 3.5:  $m(\kappa)$ , para diversos valores de  $\alpha$ , no caso de reconhecimento de um padrão ortogonal. Na figura inferior direita se mostra a curva corrigida para  $\alpha = 0.1$ .

## Capítulo 4

# Rede Neuronal Dependente do Estímulo

“O sucesso é aprender a ir de fracasso em fracasso sem perder o entusiasmo.”

– *Winston Churchill.*

**E**MBORA a área de redes neuronais tenha experimentado avanços impressionantes nas últimas décadas (Amit, 1992; Peretto, 1992; Hertz, 2018), características essenciais e reações do cérebro ainda estão longe de serem alcançadas de forma satisfatória por esses modelos. Na verdade, eles foram introduzidos para reproduzir algumas funções primárias do cérebro (McCulloch e Pitts, 1943), e sua primeira aplicação prática principal foi o reconhecimento de padrões (Pitts e McCulloch, 1947). Até agora, as redes neuronais têm sido mais bem-sucedidas em aplicações de inteligência artificial, como processamento de informações e imagens, do que emular reações de seres vivos na natureza.

Uma das pedras angulares da ciência moderna, é a teoria da evolução, cujo primeiro anúncio público foi feito em um trabalho desenvolvido por Darwin e Wallace (1858), que impulsionou a publicação do famoso livro “On the origin of species by means of natural selection” apresentado por Darwin (1859). Segundo eles, os seres vivos mais bem adaptados ao meio, têm maiores probabilidades de sobrevivência do que os menos adaptados, deixando um número maior de descendentes. Os seres vivos mais bem adaptados são, portanto, selecionados para aquele ambiente. É possível concluir partindo desta premissa, que todos os órgãos das criaturas vivas também foram moldados pela evolução. O cérebro, como os outros órgãos, também cresceu e se desenvolveu, obviamente, pela ação da seleção natural. O seu funcionamento deve então, respeitar sua história evolutiva. Mas, o que significa sua história evolutiva? Significa

---

o desenvolvimento, de geração em geração, de espécie para espécie, de um sistema complexo de respostas rápidas a estímulos externos, uma condição fundamental para sobreviver. De fato, um ser vivo tem que reagir, em muitos casos instantaneamente, a escapar de um predador, a pegar uma presa, a evitar um acidente ou por muitas outras razões. Isso revela que os sistemas nervosos dos animais foram desenvolvidos como as ferramentas fundamentais desses organismos para reagir rapidamente às mudanças do ambiente. Após milhões de anos de desenvolvimento do sistema nervoso, geralmente aumentando seus tamanhos e complexidade, apareceu uma região concentrada de células nervosas, o cérebro primitivo, que rapidamente aumentou seu tamanho. Foi um avanço no desenvolvimento dos seres vivos. Mas a tarefa era a mesma, permitindo que estes indivíduos reagissem rápida e eficientemente às mudanças no ambiente. Essa forma de trabalhar do cérebro, dependente de estímulos, deve estar por trás de todas as tarefas que o cérebro realiza, desde as antigas (reações essencialmente instintivas) e as biológicas mais recentes (relacionadas à cognição).

Neste capítulo discutiremos um modelo de rede neuronal dependente de estímulos para reconhecimento de padrões, mas certamente um cenário semelhante pode ser construído para outras atividades cerebrais. A seguir listamos resumidamente os principais passos no processo de reconhecimento de padrões feito pelos seres vivos.

- (a) Os seres vivos, através de sua história, acumularam memórias (alguns padrões de entrada no passado), que são armazenados de alguma forma.
- (b) Sem estímulos externos, eles não deveriam reconhecer nenhum padrão, permanecendo em um estado “ruidoso”.
- (c) Na presença de um estímulo externo, associado a algum padrão armazenado, eles devem reconhecer esse padrão; se o estímulo externo não tem relação com nenhum padrão armazenado, nada deve ser reconhecido.
- (d) Se o estímulo externo associado a alguma memória desaparecer, a efetividade do padrão reconhecido deve diminuir, tornando-se praticamente zero após algum tempo, induzindo o retorno a seu estado “ruidoso”.
- (e) Se o estímulo externo mudar abruptamente, como geralmente acontece na natureza, eles devem mudar rapidamente o padrão reconhecido associado ao estímulo anterior, para um

novo, associado ao estímulo posterior.

- (f) Esses passos são repetidos seguidamente, na presença de cada novo estímulo, ou sequência de estímulos.

Um modelo de rede neuronal de reconhecimento de padrões será biologicamente mais realista se for capaz de reproduzir as etapas esboçadas linhas acima (pelo menos, a maioria delas). Essas etapas têm muitos recursos bons, indo na direção dos novos conceitos de sistemas complexos, que vivem na fronteira da ordem do caos ou limites desordenados e ordenados. Sistemas caóticos, ordenados ou desordenados nunca poderão ser bons modelos para redes neuronais ou organismos vivos. Assim, um modelo de rede neuronal realmente eficaz deve funcionar na fronteira da ordem do caos.

Seguindo essas linhas, apresentamos aqui um modelo de rede neuronal dependente de estímulos, e sua eficácia será mostrada essencialmente para o reconhecimento de padrões; certamente, cenários semelhantes podem ser reconstruídos para outras atividades cerebrais.

## 4.1 RNA's são adequadas para descrever o cérebro?

Um dos modelos paradigmáticos de redes neuronais até hoje é a classe de modelos neuronais denominada “rede neuronal atratora” (RNA), que foi inicialmente proposta por Hopfield (1982). Inspirado nos modelos magnéticos aleatórios, particularmente aqueles do modelo de vidro de spin de Ising, considerou-se as analogias, *neurônios*  $\leftrightarrow$  *spins* e *intensidades sinápticas*  $\leftrightarrow$  *constantes de acoplamento*. Desta forma, o modelo de Hopfield é definido por meio de um Hamiltoniano,

$$\mathcal{H}(t) = - \sum_{(i,j)} J_{ij} S_i(t) S_j(t), \quad (4.1)$$

com  $S_i(t) = \pm 1$  ( $i = 1, 2, \dots, N$ ) representando o estado do  $i$ -ésimo neurônio no tempo  $t$  (McCulloch e Pitts, 1943), ativado ou em repouso. Além disso,  $J_{ij}$  representa a intensidade das sinapses entre os neurônios  $i$  e  $j$  (considerados simétricos), enquanto  $\sum_{(i,j)}$  denota uma soma sobre todos os pares distintos de neurônios, correspondendo a uma rede neuronal totalmente conectada. A intensidade das sinapses  $J_{ij}$  é expressa em termos das  $p$  memórias armazenadas  $\{\xi_i^\mu\}$ , de acordo com a regra de Hebb (1949),

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad (i \neq j). \quad (4.2)$$

Essas memórias devem permanecer fixas ao longo de toda a evolução temporal (ou seja, variáveis temperadas) e são assumidas como ortogonais em média [para  $N$  finito podem ocorrer superposições entre as memórias de  $O(1/\sqrt{N})$ ] e seguem

$$P(\xi_i^\mu) = \frac{1}{2} \delta(\xi_i^\mu - 1) + \frac{1}{2} \delta(\xi_i^\mu + 1). \quad (4.3)$$

A definição do hamiltoniano acima despertou o interesse de muitos físicos, particularmente considerando que técnicas de mecânica estatística poderiam ser aplicadas (Amit et al., 1985a; Peretto, 1984; Mézard et al., 1987). No entanto, um questionamento imediato diz respeito à suposição de acoplamentos simétricos  $\{J_{ij}\}$ , tornando-se uma séria desvantagem para uma emulação adequada do cérebro, uma vez que é bem sabido que as sinapses reais violam essa simetria, ou seja,  $J_{ij} \neq J_{ji}$ , caso em que nenhum hamiltoniano pode ser definido.

O modelo de Hopfield deu início à classe das RNAs, sendo sistemas onde os estados fundamentais tornam-se as memórias armazenadas, definindo um espaço de fase não trivial, onde cada memória possui sua própria bacia de atração. Nas RNAs um padrão inicial apresentando uma superposição significativa com uma memória armazenada, e então pertencente a sua bacia de atração, evolui por meio de uma dinâmica apropriada (estocástica ou não) para o estado de menor energia, ou seja, para a memória armazenada, levando ao reconhecimento desse padrão (Amit, 1992; Peretto, 1992; Hertz, 2018; McEliece et al., 1987). No entanto, uma vez que o padrão é reconhecido, o sistema permanece nesse estado para sempre, mesmo que o padrão de entrada tenha agido apenas no início do processo. Devido a esse aspecto, é fácil perceber que, entre os itens listados na seção anterior, a maioria nem sempre é cumprida.

Observe que, geralmente, as RNAs possuem três limitações básicas, as quais são listadas e discutidas abaixo.

- (i) Não operam na fronteira da ordem do caos, como qualquer sistema complexo deveria, mas sim no lado ordenado;
- (ii) Apresentam limite máximo na capacidade de armazenamento;
- (iii) São ineficientes em reagir às mudanças no ambiente externo.

O lado ordenado, onde uma multiplicidade de estados de baixa energia aparece, representa o ingrediente principal de uma RNA; Como propostas para modelar um dos sistemas mais complexos da natureza, a restrição (i) faz com que as RNAs estejam muito longe de atingir abordagens adequadas para o cérebro. A restrição (ii) vem do fato de que as memórias correspondem a estados fundamentais, cada um apresentando sua própria bacia de atração, de forma que o tamanho das bacias diminui à medida que aumenta o número de memórias armazenadas, limitando severamente seu número, mesmo que essas memórias não sejam correlacionadas. Em geral, a capacidade de armazenamento das RNAs, com  $N$  unidades neuronais, pode ser expressa como  $p = \alpha N$ , com  $\alpha$  caracterizada por valores limites ( $\alpha_c$ ), e na maioria dos casos  $\alpha_c$  é muito menor que a unidade. No modelo padrão de Hopfield, é bem conhecido que  $\alpha_c \simeq 0,14$  (Amit et al., 1985b), de modo que o sistema se torna incapaz de reconhecer qualquer outro padrão para  $\alpha > 0,14$ . A limitação (iii) está relacionada ao fato acima mencionado de que, uma vez reconhecido um padrão, o sistema permanecerá no estado correspondente; essa característica evita reações rápidas a mudanças externas. A fim de considerar tais reações no modelo de Hopfield padrão, para  $\alpha < \alpha_c$ , deve-se alterar o estado inicial  $\{S_i(0)\}$ , o que pode levar a um pulo no espaço de fase para uma bacia de atração diferente. Como será mostrado mais adiante, as reações após mudanças no ambiente externo ocorrem naturalmente, de forma suave, para o modelo a ser definido a seguir.

Com objetivo de adicionar algumas características biológicas ausentes no modelo padrão de Hopfield, foram propostas modificações tais como; levar em consideração a assimetria das constantes de acoplamento (neste caso não há hamiltoniano) (Stariolo et al., 1996; Folli et al., 2017), correlacionar ou acoplar pares de memórias (Hertz, 2018; Sompolinsky, 1987) diluição da rede (Treves e Amit, 1988), modificação do campo local (Engel et al., 1989), etc. Estas modificações geram pequenas mudanças, mas geralmente falham em satisfazer alguns dos requisitos para uma descrição apropriada do cérebro [passos (b) - (e) da seção anterior], e não superam as desvantagens descritas nos itens (i) - (iii) acima. As constantes de acoplamento assimétrico, correlações entre memórias e diluição podem levar a pontos fixos da dinâmica, ou ciclos, que aparecem no lado ordenado. Esses ingredientes também podem levar a um estado caótico (Sompolinsky et al., 1988; Wainrib and Touboul, 2013); em ambas as situações (estados ordenados ou desordenados) não estamos na fronteira da ordem do caos, mas de um lado ou do outro, e ambos os lados são inadequados para lidar com um sistema complexo como o cérebro.

## 4.2 A importância do estímulo externo

As rápidas reações frente às mudanças na vizinhança são tão importantes para a vida que muitas delas são escritas em nosso DNA e são as mais primitivas. Eles são comumente chamados de reações instintivas. O mamífero recém-nascido que suga tudo o que entra em sua boca ilustra uma reação instintiva, fundamental nas primeiras semanas de vida de qualquer mamífero. Movimentos rápidos e involuntários para escapar dos predadores são outros, dentre muitos, exemplos de reações instintivas dos animais. Claramente, o cérebro foi forjado pela evolução na tarefa de analisar rapidamente qualquer estímulo externo e desencadear uma reação muscular, se necessário. Com a evolução, outros sistemas de estímulos externos, que não são instintivos, foram desenvolvidos pelo sistema nervoso de muitos animais, geralmente levando mais tempo para analisar e reconhecer o estímulo externo e, posteriormente, para reagir. Esse cenário de reação imediata a estímulos externos deve ser onipresente para qualquer atividade cerebral, incluindo a importante tarefa de reconhecimento de padrões.

No processo de reconhecimento de padrões, em suma, o cérebro, com um certo número de memórias (padrões) armazenadas, provenientes de experiências anteriores, é submetido a um estímulo externo. Se um estímulo externo relacionado a um desses padrões for apresentado ao sistema, este poderá reconhecê-lo como um dos padrões armazenados. Já se o estímulo externo não está relacionado com nenhum padrão armazenado, o sistema não o reconhece, não é associado a nenhuma das “memórias”. Esse comportamento acontece mesmo com o cérebro “cheio” de informação, produto de experiências passadas. Ou seja, o cérebro faz uso de algum mecanismo para diminuir o “ruído presente no sistema”, produto de memórias já armazenadas, e recupera um dado padrão com muita eficiência. Na Fig.4.1 ilustramos o comportamento usual do cérebro frente a um estímulo externo, como descrito nas linhas acima.

É conhecido que o cenário típico das RNAs para reconhecimento de padrões gera uma topografia, em um espaço matemático apropriado, onde as memórias estão no fundo dos vales e o estímulo externo pode estar (inicialmente) localizado nas encostas das montanhas, dando origem ao processo de reconhecimento. Uma dinâmica apropriada leva o estímulo externo localizado nas encostas das montanhas ao fundo do vale, associando-o à sua memória correspondente. Nestes tipos de modelos de RNA, a influência de um estímulo externo em um neurônio específico pertencente à rede neuronal, pode ser essencialmente dividida em duas contribuições,

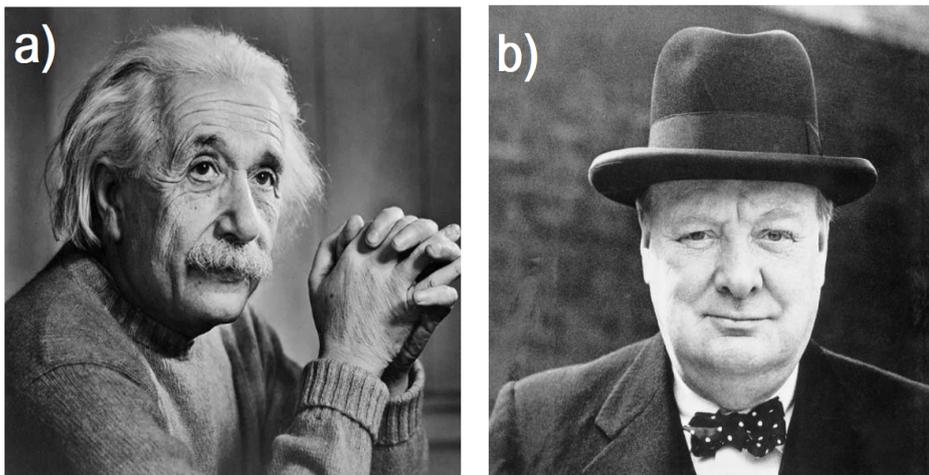


Figura 4.1: Dois grandes personagens da história contemporânea; (a) Albert Einstein, (b) Winston Churchill. Frente a um estímulo (foto a), a maioria de pessoas, poderá reconhecer rapidamente Einstein (memória armazenada), mas dificilmente reconhecerá Churchill (foto b).

um sinal, conectado ao estímulo externo e correlacionado com um padrão armazenado (memória), bem como um ruído produzido por todas as outras memórias armazenadas que não estão correlacionadas ao estímulo externo. Esse estímulo externo produz o estado inicial de atividade neuronal que, evoluindo por algum processo interno, leva o sistema a reconhecer a memória associada ao estímulo inicial. Conforme o número de padrões armazenados (memórias) aumenta, a largura da distribuição do ruído também aumenta, de tal forma que quando o sistema atinge um número suficientemente grande de memórias armazenadas, esta largura passa a ser da ordem do sinal, de forma que a rede neuronal não é mais capaz de reconhecer nenhum outro padrão armazenado adicional. A capacidade da RNA de reconhecer memórias armazenadas, falha porque a largura da distribuição do ruído cancela (aproximadamente) o sinal do estímulo inicial externo.

Inspirados no comportamento do cérebro frente a estímulos externos, assumimos como hipótese principal que, a evolução de longo prazo permite que os seres vivos calibrem a influência do estímulo externo de forma a cancelar (minimizar), tanto quanto possível, o ruído causado pelas memórias não correlacionadas com o estímulo externo. Em consequência, é possível expressar a influência ( $h_i$ ) exercida sobre um determinado neurônio  $i$  pelos outros neurônios, na presença de um estímulo externo correlacionado com alguma memória armazenada, nas seguin-

tes contribuições (vide Curado, Melgar, and Nobre (2021)),

$$h_i = \text{sinal} + \text{ruído} + \text{estímulo externo} . \quad (4.4)$$

Se o estímulo externo conseguir cancelar a contribuição do ruído, causado pelos padrões não correlacionados com a memória associada ao estímulo externo, o sinal predominará, permitindo o reconhecimento do estímulo, se associado a algum padrão armazenado. O estímulo externo, que não se corresponde com o estado inicial da rede neuronal, dá início ao processo de reconhecimento e estará presente durante todo este processo. Nesta estrutura, o estado inicial  $\{S_i(0)\}$  torna-se irrelevante; o ponto fundamental diz respeito ao ajuste da intensidade do estímulo externo para cancelar precisamente (ou aproximadamente) o termo de ruído. A partir de agora, nos referimos a este esquema como uma rede neuronal dependente de estímulos (RNDE).

### 4.3 Rede neuronal Dependente do Estímulo (RNDE)

Para ilustrar a construção descrita na seção anterior, devemos escolher um modelo de rede neuronal simples, com o qual seja possível obter resultados analíticos, ajudando-nos a entender melhor a rede neuronal dependente do estímulo. Usaremos como modelo de memória biológica a rede neuronal de Hopfield, por ser um dos modelos mais conhecidos e estudados na literatura e porque permite obter muitos resultados analíticos, mas, de fato, essa abordagem pode ser usada em um grande número de modelos de rede neuronal.

Ao aplicarmos a construção dependente do estímulo (RNDE) a uma rede de Hopfield padrão, o campo local  $h_i$  no tempo  $t$ , atuando no  $i$ -ésimo neurônio do sistema, torna-se:

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t) + \kappa \eta_i , \quad (4.5)$$

onde  $S_i = \pm 1$  com  $i = 1, \dots, N$ . Se percebe no primeiro termo do lado direito da equação anterior a contribuição usual devido aos  $(N - 1)$  neurônios, enquanto o segundo termo corresponde ao estímulo externo que age sobre o neurônio  $i$ . Esta última contribuição deve permanecer fixa durante todo o tempo de evolução, com  $\eta_i = \pm 1$  e  $\kappa$  ( $\kappa \geq 0$ ) representando sua intensidade.

A ideia principal é manter a influência do padrão externo durante todo o processo, e não apenas como um estado inicial, como acontece nas RNAs típicas. A seguir, vamos considerar

duas situações usuais, relacionadas a este estímulo externo  $\eta_i$ :

- (i) Deve apresentar uma superposição com uma memória específica armazenada (por exemplo, memória  $\rho$ ,  $1 \leq \rho \leq p$ ), sendo ortogonal a todas as outras memórias e levando a um padrão externo correlacionado com uma memória específica;
- (ii) Deve ser ortogonal a todas as memórias armazenadas, não possuindo superposição com nenhuma memória, correspondendo a um padrão externo não correlacionado com nenhuma memória armazenada.

O caso (i) pode ser expresso assumindo que o conjunto  $\{\eta_i\}$  obedece à distribuição de probabilidade

$$P(\eta_i) = \gamma\delta(\eta_i - \xi_i^\mu) + (1 - \gamma)\delta(\eta_i + \xi_i^\mu), \quad (4.6)$$

onde  $1/2 < \gamma \leq 1$  para  $\mu = \rho$  e  $\gamma = 1/2$  para  $\mu \neq \rho$ . O caso particular  $\gamma = 1$  significa que o conjunto de sinais de estímulo externo  $\{\eta_i\}$  produz o mesmo padrão que a memória  $\rho$ , cobrindo situações reais frequentes, onde se deve reconhecer um padrão externo que coincide precisamente com um padrão armazenado. Por outro lado, o caso (ii) é caracterizado por

$$P(\eta_i) = \frac{1}{2}\delta(\eta_i - \xi_i^\mu) + \frac{1}{2}\delta(\eta_i + \xi_i^\mu), \quad (1 \leq \mu \leq p). \quad (4.7)$$

Vamos investigar o campo local da Eq. (4.5), separando a contribuição da memória  $\rho$ , daqueles de outras memórias ( $\mu \neq \rho$ ) na Eq. (4.2); obtemos,

$$h_i(t) = \frac{1}{N} \sum_{j \neq i}^N \xi_i^\rho \xi_j^\rho S_j(t) + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \xi_i^\mu \xi_j^\mu S_j(t) + \kappa \eta_i. \quad (4.8)$$

Comparando a equação anterior com a Eq. (4.4), identifica-se imediatamente cada uma das contribuições no seu lado direito; a primeira é o sinal, quando a configuração neuronal  $\{S_i\}$  é praticamente igual à memória armazenada  $\{\xi_i^\rho\}$ ; o segundo representa o ruído, induzido por outras memórias distintas de  $\rho$ ; a terceira contribuição corresponde ao estímulo externo. Considerando que no instante  $t$ , o estado da rede neuronal  $\{S_i(t)\}$ , bem como o estímulo  $\{\eta_i\}$ , são os mesmos que o padrão  $\{\xi_i^\rho\}$ , ou seja  $\gamma = 1$  e  $S_i(t) = \eta_i = \xi_i^\rho$ , então

$$h_i(t) = \frac{N-1}{N} \xi_i^\rho + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \xi_i^\mu \xi_j^\mu \xi_j^\rho + \kappa \xi_i^\rho, \quad (4.9)$$

mostrando que se pode calibrar  $\kappa$  apropriadamente, permitindo que o estímulo externo cancele aproximadamente o termo de ruído, a fim de favorecer a contribuição do sinal, permitindo que o sistema reconheça o padrão externo submetido. Observe que,  $\kappa$  não pode ser tão pequeno, já que não poderá cancelar o termo de ruído, nem muito grande, pois nesse caso dominará campo local, forçando um alinhamento com o estímulo externo, sendo correlacionado com uma memória ou não. Encontrar o valor ideal para  $\kappa$  representa um ponto fundamental desta estrutura.

De acordo com Penna (2009), é possível fazer uma analogia do algoritmo conhecido por “simulated annealing” e o reconhecimento de padrões numa RNDE. Desse modo, ao invés de termos um campo externo “fixo” que aponta na direção do padrão  $\eta_i$ , podemos imaginar que no processo de reconhecimento o estímulo externo é dependente do tempo (ou seja  $\eta_i = \eta_i(t)$ ) e cumpre a relação  $\eta_i(t+1) = S_i(t)$ . Seria interessante fazer simulações futuras, analisando possíveis consequências deste tipo de dinâmica.

## 4.4 Estimativa do $\kappa$ ótimo

Nosso objetivo é escolher um valor do parâmetro  $\kappa$ , de tal forma que o termo de estímulo externo seja capaz de cancelar aproximadamente o termo de ruído, permitindo ao sistema o reconhecimento do padrão externo enviado. Com esse intuito, vamos analisar o termo de ruído presente no lado direito da equação (4.9):

$$h_i^{\text{ruído}}(t) = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \xi_i^\mu \xi_j^\mu \xi_j^\rho. \quad (4.10)$$

Considerando que nenhum padrão  $\mu$  está correlacionado com o padrão  $\rho$  e os sítios  $i$  e  $j$  são independentes, podemos considerar que o termo de ruído contém uma soma de  $(N-1)(p-1)$  termos aleatórios independentes, cada um dando uma contribuição  $+1$  ou  $-1$  para a soma. Esse é o famoso problema de passeio aleatório e o resultado, para valores grandes de  $N$ , é que o ruído do sistema tende a uma variável aleatória Gaussiana com um desvio padrão  $\frac{1}{N} \sqrt{(N-1)(p-1)} \approx \sqrt{\alpha}$ . Para verificar este resultado, basta calcular os dois primeiros

momentos do termo de ruído:

$$\begin{aligned}\langle h_i^{\text{ruído}}(t) \rangle &= \left\langle \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \right\rangle_\xi = 0, \\ \langle (h_i^{\text{ruído}}(t))^2 \rangle &= \left\langle \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{k \neq i}^N \sum_{S \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \xi_i^S \xi_k^S \xi_k^\nu \right\rangle_\xi = \frac{(N-1)(p-1)}{N^2}.\end{aligned}$$

Assim o desvio padrão, ou largura da distribuição, para valores suficientemente grandes de  $N$  e  $p$ , é:

$$\sigma_{\text{ruído}} = \sqrt{\langle (h_i^{\text{ruído}})^2 \rangle - \langle h_i^{\text{ruído}} \rangle^2} = \sqrt{\frac{(N-1)(p-1)}{N^2}} \simeq \sqrt{\frac{p}{N}} = \sqrt{\alpha}. \quad (4.11)$$

Por tanto, para um sistema com padrões armazenados independentes, uma boa escolha para o parâmetro  $\kappa$  deve ser aproximadamente  $\kappa \simeq \sqrt{\alpha}$ ; as simulações numéricas a serem apresentadas posteriormente mostram uma boa concordância com esta escolha. Este procedimento pode resultar no reconhecimento de padrões para valores de  $p$  muito maiores do que aqueles restritos pelo limite superior do modelo de Hopfield. No entanto, nos modelos em que os cálculos analíticos para o valor ótimo de  $\kappa$  não são viáveis, ainda deve ser possível estimar esse valor numericamente, na verdade, essa é uma tarefa fácil, como será mostrado mais adiante.

Da Eq. (4.8), lembrando que os acoplamentos  $\{J_{ij}\}$  são simétricos no modelo de Hopfield, pode-se definir um hamiltoniano no tempo  $t$ ,

$$\begin{aligned}\mathcal{H}(t) &= - \sum_{i=1}^N h_i(t) S_i(t) \\ &= - \frac{1}{N} \sum_{(i,j)} \xi_i^p \xi_j^p S_i(t) S_j(t) - \frac{1}{N} \sum_{(i,j)} \sum_{\mu \neq \rho}^p \xi_i^\mu \xi_j^\mu S_i(t) S_j(t) - \kappa \sum_{i=1}^N \eta_i S_i(t),\end{aligned} \quad (4.12)$$

onde  $\sum_{(i,j)}$  denotam somas sobre todos os pares distintos de neurônios, correspondendo a uma rede neuronal totalmente conectada.

Nas simulações a serem apresentadas a seguir, nós focamos essencialmente em dois parâmetros, a saber, a superposição macroscópica (ou superposição) de um estado de neurônio  $\{S_i(t)\}$  com um determinado padrão  $\{\xi_i^\mu\}$  no momento  $t$ ,

$$m_\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i(t) \quad (\mu = 1, 2, \dots, p), \quad (4.13)$$

bem como na superposição deste estado de neurônio com o caso especial onde o sinal de estímulo externo  $\{\eta_i\}$  são ortogonais a todas as memórias armazenadas [isto é, caso (ii) na Eq. (4.7) com  $\gamma = 1/2$  para todas as memórias], definido como

$$m_{\perp}(t) = \frac{1}{N} \sum_{i=1}^N \eta_i S_i(t). \quad (4.14)$$

A quantidade acima representa uma superposição macroscópica com um padrão externo, não armazenado e ortogonal a todos os padrões armazenados, para os quais,  $\eta_i = \pm 1$  com iguais probabilidades.

Frequentemente, o conjunto  $\{m_{\mu}(t)\}$  é considerado como componentes de um vetor  $p$ -dimensional,  $\mathbf{m}(t) = (m_1(t), \dots, m_{\rho}(t), \dots, m_p(t))$ , e daremos uma ênfase especial a  $m_{\rho}(t)$ , associado a um único padrão condensado e identificado no primeiro termo do hamiltoniano na Eq. (4.12). Todos os outros componentes  $m_{\mu}(t)$  ( $\mu \neq \rho$ ) aparecem na contribuição de ruído do hamiltoniano.

Perceba que o modelo RNDE, descrito linhas acima, leva um novo cenário de uma rede neuronal conceitualmente diferente. Como descrito na seção 2.4.7; em modelos RNA, as memórias estão associadas a vales em uma paisagem, ocupando volume no espaço de fases e limitando a capacidade de recuperação de memórias. Por exemplo, para a rede de Hopfield, com  $\alpha < \alpha_c = 0,14$ , a paisagem terá um formato semelhante ao da figura 4.2. Dessa forma, uma vez que o sistema alcança um vale, associado a uma memória armazenada, ele permanecerá nesse vale em definitivo.

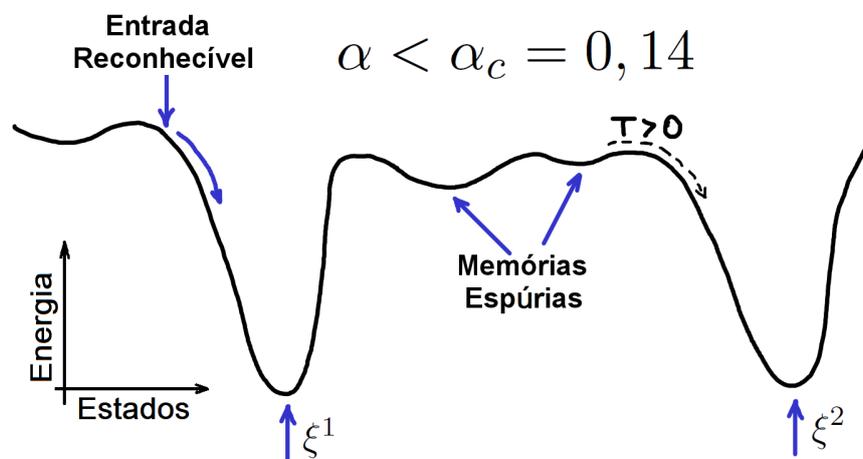


Figura 4.2: Paisagem de quase energia usual de uma RNA de Hopfield com  $\alpha < \alpha_c$ .

Em contraste com a RNA, numa RNDE, onde  $\alpha \gg \alpha_c$ , as memórias não são vales e não ocupam volume no espaço de fases na ausência de um estímulo externo (vide figura 4.3a). Na presença de um estímulo externo, com a intensidade apropriada, um vale é criado e está fortemente correlacionado à memória associada ao estímulo externo (ver figura 4.3b). Se o estímulo externo não está associado a nenhuma memória, nenhum vale é criado.

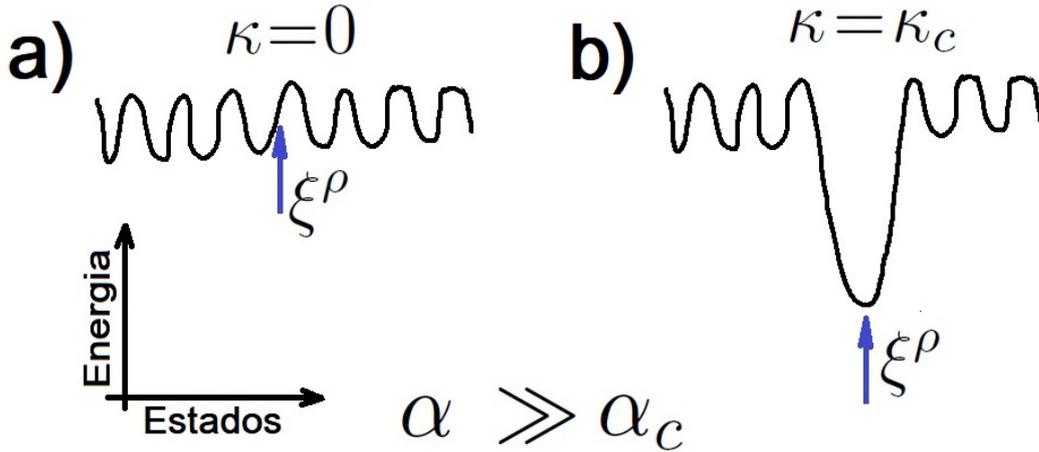


Figura 4.3: Paisagem de quase energia usual de uma RNDE na ausência (a) e presença (b) de um estímulo externo, respetivamente.

## 4.5 Teoria de Campo Médio na RNDE

Nesta seção, realizamos cálculos analíticos, que correspondem essencialmente a uma abordagem de campo médio, seguindo o caminho de Amit et al. (1985a,b). Assumimos que o sistema definido pelo hamiltoniano na Eq. (4.12) atinge, após um tempo suficientemente longo, estados de equilíbrio térmico bem definidos (para temperaturas finitas,  $T > 0$ ), junto com seu limite de temperatura zero ( $T \rightarrow 0$ ). Para uma dada realização da desordem temperada ( $\{\xi_i^\mu\}, \{\eta_i\}$ ), pode-se definir uma função de partição  $\mathcal{Z} \equiv \mathcal{Z}(\{\xi_i^\mu\}, \{\eta_i\})$ , de modo que a energia livre por neurônio se torna

$$-\beta f = \lim_{N \rightarrow \infty} \frac{1}{N} \langle \ln \mathcal{Z}(\{\xi_i^\mu\}, \{\eta_i\}) \rangle_{\eta, \xi}, \quad (4.15)$$

onde  $\beta = 1/T$  (trabalhamos em unidades  $k_B = 1$ ). Acima,  $\langle \dots \rangle_{\eta, \xi}$  indica médias temperadas em  $\{\eta_i\}$  e  $\{\xi_i^\mu\}$ , que, de acordo com as Eqs. (4.3) e (4.6), deve ser realizado tomando a média

em  $\{\eta_i\}$ 's antes da média em  $\{\xi_i^\mu\}$ 's. Seguindo o procedimento padrão, aplicamos o método de réplicas para calcular a energia livre (Amit, 1992; Peretto, 1992),

$$-\beta f = \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \frac{1}{Nn} \left( \langle \mathcal{Z}^n(\{\xi_i^\mu\}, \{\eta_i\}) \rangle_{\eta, \xi} - 1 \right), \quad (4.16)$$

onde  $\mathcal{Z}^n(\{\xi_i^\mu\}, \{\eta_i\})$  corresponde à função de partição de  $n$  réplicas idênticas e independentes, para uma dada realização de desordem.

No equilíbrio mencionado acima, pode-se realizar as médias usuais, de modo que a quantidade da Eq. (4.13) leva a

$$m_\mu = \left\langle \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle S_i \rangle \right\rangle_{\eta, \xi} \quad (\mu = 1, \dots, p), \quad (4.17)$$

enquanto que, para o caso em que o estímulo externo é ortogonal a todas as memórias armazenadas, a superposição na Eq. (4.14) resulta

$$m_\perp = \left\langle \frac{1}{N} \sum_{i=1}^N \eta_i \langle S_i \rangle \right\rangle_{\eta, \xi}, \quad (4.18)$$

com  $\langle S_i \rangle$  correspondendo às magnetizações locais.

### 4.5.1 Parâmetros de Ordem

Seguindo os passos de Amit et al. (1987b) e tendo em conta que quando  $N \rightarrow \infty$ , a aproximação de ponto sela pode ser usada, a energia livre temperada, pode-se expressar em termos dos seguintes parâmetros de ordem:

(a) a superposição macroscópica com  $s$  padrões condensados ou recuperados

$$m_{\nu, a} = \left\langle \frac{1}{N} \sum_{i=1}^N \xi_i^\nu \langle S_i^a \rangle_T \right\rangle_{\eta, \xi}, \quad (4.19)$$

(b) o parâmetro de ordem de Edwards-Anderson

$$q_{ab} = \left\langle \frac{1}{N} \sum_{i=1}^N \langle S_i^a \rangle_T \langle S_i^b \rangle_T \right\rangle_{\eta, \xi}, \quad (4.20)$$

(c) o ruído devido aos  $p - s$  padrões não condensados

$$r_{ab} = \frac{1}{\alpha} \sum_{\mu=s+1}^p \langle m_a^\mu m_b^\mu \rangle_{\eta, \xi}. \quad (4.21)$$

Aqui  $\langle \dots \rangle_T$  representa a média térmica e  $\nu = 1, \dots, s$ . Cálculos extensos (mas bem estabelecidos) para a energia livre da Eq. (4.16) são descritos no apêndice A, onde dividimos os padrões  $\{\xi_i^\mu\}$  em dois conjuntos, a saber, os  $s$  padrões condensados, e  $(p - s)$  não condensados, obtendo-se:

$$f = \frac{\alpha}{2} + \frac{1}{2n} \sum_{a=1}^n \sum_{\nu=1}^s (m_a^\nu)^2 + \frac{\alpha\beta}{2n} \sum_{a \neq b}^n r_{ab} q_{ab} + \frac{\alpha}{2\beta n} \text{Tr} [\ln ((1 - \beta) \mathbf{I} - \beta \mathbf{Q})] - \frac{1}{n\beta} \left\langle \langle \ln \mathcal{Z}_0 \rangle_\eta \right\rangle_{\xi^\nu}, \quad (4.22)$$

onde o operador  $\mathcal{Z}_0$ , é dado por

$$\mathcal{Z}_0 = \text{Tr}_{S^a} \left[ \exp \left( \frac{\alpha\beta^2}{2} \sum_{a \neq b}^n r_{ab} S^a S^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s m_{\nu,a} \xi^\nu S^a + \beta\kappa \sum_{a=1}^n \eta S^a \right) \right]. \quad (4.23)$$

A média  $\langle \dots \rangle_\eta \equiv \int (\dots) P(\eta) d\eta$  refere-se à média sobre os  $\eta$  e a média  $\langle \dots \rangle_{\xi^\nu}$  indica uma média sobre os  $s$  padrões condensados  $\xi^\nu$ . Lembre-se que é necessário tomar primeiramente uma média em  $\eta$ , devido a que sua distribuição de probabilidades  $P(\eta)$  depende do padrão  $\xi^\rho$ .

Observe que, na estrutura RNDE, temos um padrão externo associado a uma memória e que todas as outras memórias armazenadas, são ortogonais. Portanto, teremos um único padrão condensado  $\nu = \rho$  (ou seja,  $s = 1$ ), o que significa que o estímulo externo apresenta uma superposição com a memória  $\rho$ , correspondente ao caso (i) da Eq. (4.6), para o qual temos a superposição

$$m_\rho = \left\langle \frac{1}{N} \sum_{i=1}^N \xi_i^\rho \langle S_i \rangle \right\rangle_{\eta, \xi}. \quad (4.24)$$

Mesmo com a situação descrita acima, continuaremos trabalhando com o formalismo geral, ou seja, considerando que temos  $s$  padrões condensados.

## 4.5.2 Teoria de simetria de réplicas

Para calcular a energia livre, faremos uso do método de réplicas, o qual consiste em considerar réplicas idênticas, não interagentes entre si, do sistema real. Ao assumirmos o *ansatz* de simetria de réplicas, considera-se que cada parâmetro de ordem é simétrico sob permutação de seus índices de réplica, ou seja:

$$m_{\nu,a} = m_\nu \quad q_{ab} = q \quad (a \neq b) \quad r_{ab} = r \quad (a \neq b). \quad (4.25)$$

Como pode ser visto no apêndice (novamente os cálculos são análogos ao caso Hopfield), ao permitirmos  $n$  tender a zero na expressão da energia livre (4.22), obtemos finalmente o seguinte resultado:

$$f = \frac{\alpha}{2} + \frac{1}{2} m_\rho^2 + \frac{\alpha\beta}{2} r (1 - q) + \frac{\alpha}{2\beta} \left[ \ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q} \right] - \frac{1}{\beta} \int \mathcal{D}z \left\langle \ln \left[ 2 \cosh \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s m_\nu \xi^\nu + \kappa \eta \right) \right] \right\rangle_{\eta, \xi^\nu} \quad (4.26)$$

onde

$$\int Dz \dots \equiv \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \dots, \quad (4.27)$$

e as médias  $\langle \dots \rangle_{\eta, \xi^\nu}$  devem ser consideradas apenas sobre o único padrão condensado  $\xi^\rho$ .

### 4.5.3 Equações de campo médio

A solução de equilíbrio provém do cálculo das derivadas parciais da energia livre  $f$  com respeito aos parâmetros de ordem  $m_\nu$ ,  $q$  e  $r$ , levando às equações de ponto sela:

$$m_\nu = \int \mathcal{D}z \left\langle \xi^\nu \tanh \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s m_\nu \xi^\nu + \kappa \eta \right) \right\rangle_{\eta, \xi}, \quad (4.28)$$

$$q = \int \mathcal{D}z \left\langle \tanh^2 \beta \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s m_\nu \xi^\nu + \kappa \eta \right) \right\rangle_{\eta, \xi}, \quad (4.29)$$

$$r = \frac{q}{(1 - \beta + \beta q)^2}. \quad (4.30)$$

Para o caso (ii) na Eq. (4.7), onde o estímulo externo é ortogonal a todas as memórias, não há padrão condensado e o parâmetro  $m_\perp$  da Eq. (4.18) pode ser calculado a partir de

$$m_\perp = \frac{1}{\beta N} \left\langle \frac{\partial}{\partial \kappa} \ln \mathcal{Z}(\{\xi_i^\mu\}, \{\eta_i\}) \right\rangle_{\eta, \xi} = \frac{1}{\beta N} \frac{\partial}{\partial \kappa} \langle \ln \mathcal{Z}(\{\xi_i^\mu\}, \{\eta_i\}) \rangle_{\eta, \xi} = -\frac{\partial f}{\partial \kappa}, \quad (4.31)$$

onde assumimos, como de costume, que a derivada comuta com as operações médias, uma vez que estas são essencialmente expressas por somas e integrais. É importante ressaltar que as médias acima devem ser consideradas sobre os padrões ortogonais, para os quais  $m_\rho = 0$  e  $\eta_i = \pm 1$  com probabilidades iguais. Nesse caminho,

$$m_\perp = \int \mathcal{D}z \langle \eta \tanh[\beta (z\sqrt{\alpha r} + \kappa \eta)] \rangle_\eta. \quad (4.32)$$

## 4.6 Soluções no limite $T \rightarrow 0$

Agora vamos analisar o caso limite de temperatura zero (ou  $\beta \rightarrow \infty$ ). Trataremos com a solução da chamada “fase de recuperação ferromagnética (FMR)”, correspondente ao caso em que  $m_\nu = \delta_{\nu\rho}$ . Neste caso,  $q = 1 + O(T)$ , de modo que a energia interna por neurônio torna-se

$$u = \frac{1}{2} \alpha(1 - r) - \frac{1}{2} m_\rho^2 - \kappa m_\perp. \quad (4.33)$$

As médias envolvidas no cálculo de  $m_\rho$ ,  $q$  e  $r$  nas equações Eqs. (4.28)–(4.32), sob estas considerações, podem ser escritas da seguinte forma:

$$m_\rho = \gamma \operatorname{erf} \left( \frac{m_\rho + \kappa}{\sqrt{2\alpha r}} \right) + (1 - \gamma) \operatorname{erf} \left( \frac{m_\rho - \kappa}{\sqrt{2\alpha r}} \right), \quad (4.34)$$

$$r = \frac{q}{(1 - \beta + \beta q)^2}, \quad (4.35)$$

$$q = 1 - \frac{1}{\beta} \sqrt{\frac{2}{\pi \alpha r}} \left[ \gamma \exp \left( -\frac{(m_\rho + \kappa)^2}{2\alpha r} \right) + (1 - \gamma) \exp \left( -\frac{(m_\rho - \kappa)^2}{2\alpha r} \right) \right], \quad (4.36)$$

Já no caso em que o estímulo externo é ortogonal a todas as memórias armazenadas obtemos

$$m_\perp = \operatorname{erf} \left( \frac{\kappa}{\sqrt{2\alpha r}} \right), \quad (4.37)$$

onde  $\operatorname{erf}(x)$  é a função de erro, definida por:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.38)$$

Observe que no limite  $T \rightarrow 0$  (ou  $\beta \rightarrow \infty$ ), ao substituir a Eq. (4.36) na Eq. (4.35) e usando a Eq. (4.34), podemos obter as seguintes equações para o modelo RNDE com padrões armazenados descorrelacionados, no caso do reconhecimento de um padrão  $\gamma\%$  diferente do padrão armazenado

$$\begin{aligned} y &= \frac{\gamma \operatorname{erf}(y + x) + (1 - \gamma) \operatorname{erf}(y - x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} [\gamma e^{-(y+x)^2} + (1 - \gamma) e^{-(y-x)^2}]}, \\ m_\rho &= \gamma \operatorname{erf}(y + x) + (1 - \gamma) \operatorname{erf}(y - x), \end{aligned} \quad (4.39)$$

onde  $m_\rho = y\sqrt{2\alpha r}$  e  $\kappa = x\sqrt{2\alpha r}$ .

Já no caso particular de  $\gamma = 1$ , é possível recuperar a Eq. (3.47), que corresponde ao caso do reconhecimento de um estímulo externo exatamente igual ao padrão armazenado  $\xi^\rho$ ,

$$\begin{aligned} y &= \frac{\operatorname{erf}(y + x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-(y+x)^2}}, \\ m_\rho &= \operatorname{erf}(y + x), \end{aligned} \quad (4.40)$$

e se assumimos que  $y + x \rightarrow x$ , na Eq. (4.39), obtemos as equações que correspondem a um estímulo externo que não está relacionado a nenhum padrão armazenado, (vide Eqs. (3.51) e (4.37)),

$$\begin{aligned} y &= \frac{\operatorname{erf}(x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}}e^{-x^2}}, \\ m_\rho &= \operatorname{erf}(x). \end{aligned} \quad (4.41)$$

As equações obtidas a temperatura zero, podem ser resolvidas numericamente para determinados valores de  $\gamma$ ,  $\alpha$  e  $\kappa$ ; na Fig. 4.4 ilustramos o caso particular  $\gamma = 1$ , com curvas para as superposições  $m_\rho$  e  $m_\perp$  resultantes da solução das Eqs. (4.37) e (4.40). A Fig. 4.4 condensa os resultados mostrados nas figuras 3.4,3.5,3.1 e 3.2, do capítulo anterior.

As superposições  $m_\rho$  and  $m_\perp$  versus  $\alpha$  (valores típicos de  $\kappa$ ), ou versus  $\kappa$  (valores típicos de  $\alpha$ ), são apresentadas nas Figs. 4.4(a) e (b), ou Figs. 4.4(c) e (d), respectivamente. Aqui novamente, enfatizamos que duas situações diferentes para o estímulo externo estão sendo consideradas, ou seja, um estímulo externo totalmente correlacionado (gráficos para  $m_\rho$ ) e um estímulo externo ortogonal a todas as memórias (gráficos para  $m_\perp$ ).

Percebe-se certos intervalos de  $\kappa$  e  $\alpha$  onde várias soluções para essas quantidades aparecem, caracterizando as transições de fase de primeira ordem. Alguns casos típicos são ilustrados nas inserções correspondentes, mostrando a coexistência de mais de uma solução; nestes casos, as localizações precisas das descontinuidades podem ser calculadas através de construções de Maxwell, onde se equaciona a energia interna da Eq. (4.33) para ambas as soluções. Para fins de reconhecimento de padrões, espera-se que as quantidades  $m_\rho$  e  $m_\perp$  variem suavemente, de modo que as descontinuidades, onde pode haver grandes variações em qualquer uma dessas superposições, para mudanças infinitesimais de determinados parâmetros, não correspondem a situações comuns em sistemas naturais.

Portanto, para o resto deste trabalho nos concentraremos em soluções onde os parâmetros  $m_\rho$  e  $m_\perp$ , associados a dois tipos diferentes de estímulos externos, variam suavemente em qualquer um dos parâmetros  $\kappa$  ou  $\alpha$ ; mais particularmente, vamos nos concentrar no papel desempenhado por  $\kappa$  na região de não recuperação do modelo Hopfield padrão, ou seja,  $\alpha > 0.14$ , ao longo do qual se tem  $m_\rho = m_\perp = 0$  para  $\kappa = 0$ . Chamamos a atenção para aqueles gráficos para valores mais altos de  $\kappa$  (ou  $\alpha$ ) na Fig. 4.4, focando nas regiões relevantes para o reconhecimento de padrões ( $\kappa \simeq \sqrt{\alpha}$ ), em torno do qual se nota valores significativos para os

$m_\rho$  superpostos (normalmente,  $m_\rho > 0.8$ ).

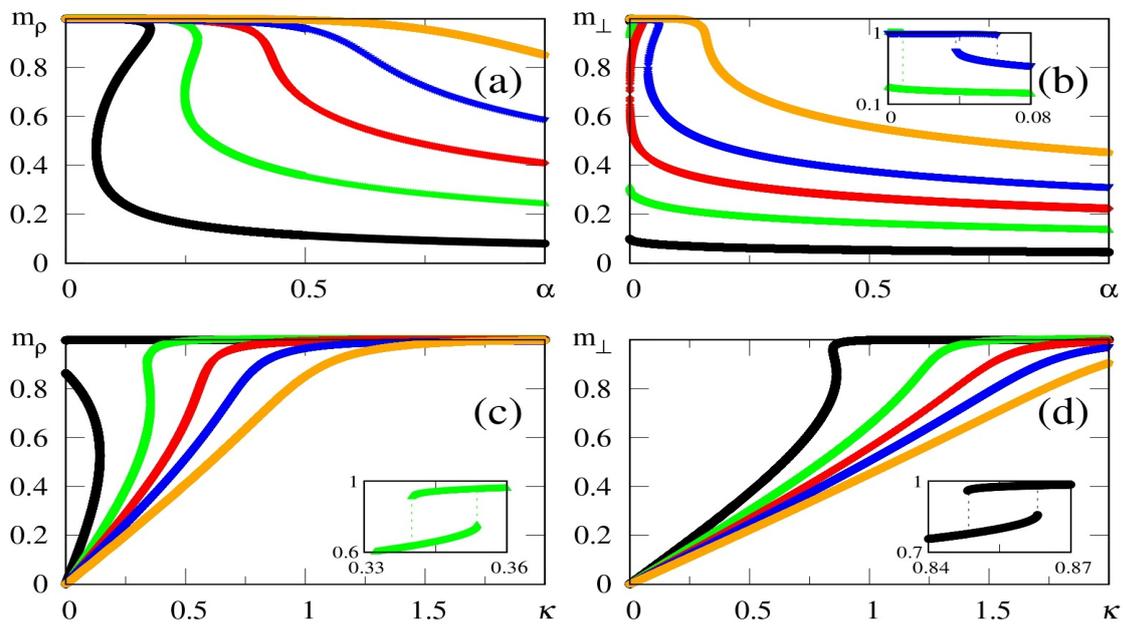


Figura 4.4: Resultados da aproximação de campo médio no limite de temperatura zero, para as superposições  $m_\rho$  [Eq. (4.40)] e  $m_\perp$  [Eq. (4.37)], são exibidos para o caso  $\gamma = 1$ :  $m_\rho$  e  $m_\perp$  versus  $\alpha$  são mostrados (valores típicos de  $\kappa$ ) nos painéis (a) e (b), e representados versus  $\kappa$  (valores típicos de  $\alpha$ ) nos painéis (c) e (d), respectivamente. Em (a) e (b) temos  $\kappa = 0.1, 0.3, 0.5, 0.7$  e  $1.0$  (de baixo para cima), enquanto em (c) e (d) temos  $\alpha = 0.1, 0.3, 0.5, 0.7$  e  $1.0$  (de cima para baixo). Para valores menores de  $\kappa$  e  $\alpha$ , pode-se observar várias soluções para ambos  $m_\rho$  e  $m_\perp$ , típicas de transições de fase de primeira ordem. As inserções ilustram alguns casos em que duas soluções coexistem para certos intervalos desses parâmetros: (b)  $\kappa = 0.3$  e  $0.7$ ; (c)  $\alpha = 0.3$ ; (d)  $\alpha = 0.1$ .

## 4.7 Simulação Computacional

Para simular numericamente uma RNDE a temperatura zero, cujo campo local  $h_i(t)$  é definido pelas Eqs. (4.5)-(4.6), consideramos uma rede de  $N = 10^4$  neurônios (vide. Oliveira (1991)) e  $p = \alpha N$  memórias armazenadas, com diversos valores dos parâmetros  $\alpha$  e  $\gamma$ . O conjunto de sinapses  $J_{ij}$  e o estímulo externo  $\eta$ , definidos pelas Eqs. (4.2), (4.3), e (4.6), são consideradas variáveis temperadas, ou seja, permanecem fixas ao longo de cada evolução temporal. Uma versão resumida do algoritmo usado na simulação computacional é mostrada no apêndice B. Por praticidade, consideramos a configuração inicial da rede  $\{S_i(0)\}$ , como aleatória, ou de forma equivalente, o estado de cada neurônio será  $S_i = \pm 1$  com igual probabilidade. Estas variáveis dinâmicas serão atualizadas de forma sequencial, obedecendo à dinâmica de Glauber a “temperatura zero” ( $T \rightarrow 0$ ), ou seja, de acordo com

$$S_i(t + \Delta t) = \text{sign}(h_i), \quad (i = 1, \dots, N), \quad (4.42)$$

onde cada unidade de tempo  $\Delta t$ , equivale à atualização de um único neurônio. Seguindo este procedimento, o mínimo local de energia mais próximo é atingido, após um tempo suficientemente longo  $t_e$  ( $t_e \simeq 10^2 N$ , ou seja, cada neurônio será visitado  $10^2$  vezes), atingimos tipicamente um estado de equilíbrio, de modo que as quantidades macroscópicas apresentando pequenas variações para  $t > t_e$ .

Em um dado mínimo local de energia, ou seja, para  $t > t_e$ , enfatizamos nossa análise para as superposições do estado  $\{S_i(t)\}$  com um dado padrão  $\xi_i^\rho$

$$m_\rho = \left\langle \frac{1}{N} \sum_{i=1}^N \xi_i^\rho \langle S_i \rangle_T \right\rangle_{\eta, \xi}, \quad (4.43)$$

e a superposição do estado  $\{S_i(t)\}$  com o estímulo externo  $\eta_i$ ,

$$m_\perp = \left\langle \frac{1}{N} \sum_{i=1}^N \eta_i \langle S_i \rangle_T \right\rangle_{\eta, \xi}. \quad (4.44)$$

e de acordo com as Eqs. (4.3) e (4.6), a média em  $\{\eta_i\}$ 's deve ser tomada antes da média em  $\{\xi_i\}$ 's. A média das superposições macroscópicas  $m_\rho$  e  $m_\perp$  é feita sobre  $10^3$  configurações diferentes das variáveis temperadas  $\eta$  e  $\xi$ .

### 4.7.1 Cálculo numérico de $\kappa_C$

Em um sistema RNDE, devido a rede estar sobrecarregada, as superposições  $m_\rho$  e  $m_\perp$  existirão, somente no caso de termos um estímulo externo, por esse motivo, a intensidade do estímulo  $\kappa$ , desempenha um papel importante no modelo. Caso  $\kappa = 0$  temos que  $m_\rho = m_\perp = 0$ . A intensidade do estímulo externo  $\kappa > 0$  produz duas situações distintas em relação ao padrão externo, a saber,  $1/2 < \gamma \leq 1$  para  $m_\rho(\kappa)$  e  $\gamma = 1/2$  para  $m_\perp(\kappa)$ ; de acordo com isso, valores suficientemente grandes de  $\kappa$  levam as seguintes tendências  $m_\rho \rightarrow 2\gamma - 1$  e  $m_\perp \rightarrow 1$ . Esses dois casos limites ( $\kappa$  muito pequeno ou grande) não são interessantes para o reconhecimento de padrões numa RNDE; por outro lado, um valor bem calibrado de  $\kappa$ , que denominaremos de  $\kappa_C$ , escolhido de forma a diminuir o máximo possível os efeitos do termo de ruído na Eq. (4.8), conforme discutido na seção 3, representa o ponto crucial da estrutura RNDE. Se  $\kappa \ll \kappa_C$ , o campo externo não será suficiente para cancelar o ruído, enquanto que para  $\kappa \gg \kappa_C$ , o estímulo externo dominará o campo local total e não haverá reconhecimento de padrão nenhum, neste último caso, o campo local simplesmente reproduzirá os estímulos externos, associados ou não a uma memória. Para identificar o valor ótimo de  $\kappa = \kappa_C$ , propõe-se monitorar o valor absoluto da diferença entre as superposições das Eqs. (4.18) e (4.24), que correspondem a dois tipos diferentes de estímulos externos,

$$\Delta m = |m_\rho - m_\perp| = \left\langle \frac{1}{N} \sum_{i=1}^N |\xi_i^\rho - \eta_i| \langle S_i \rangle_T \right\rangle_{\eta, \xi} \quad (4.45)$$

Note que,  $m_\rho$  é calculado em uma RNDE, na qual o estímulo externo  $\{\eta_i\}$  está correlacionado  $\gamma\%$  com um padrão armazenado  $\xi^\rho$ . Já  $m_\perp$  se calcula quando  $\{\eta_i\}$  é um padrão ortogonal a todos os padrões armazenados. Das definições de  $m_\rho$  e  $m_\perp$ , concluímos que, apenas as memórias ortogonais(ou seja, não armazenadas) deverão contribuir para  $\Delta m$ , de modo que a escolha apropriada para  $\kappa_c$  deve corresponder ao seu valor máximo,  $\max(\Delta m)$ .

Os resultados das simulações numéricas são mostrados na Fig. 4.5, onde as quantidades definidas nas Eqs. (4.18), (4.24) e (4.45) são plotadas versus  $\kappa$ , para  $\gamma = 1.0$ , e valores típicos de  $\alpha$ . Em todos os casos,  $\kappa_c$  é claramente identificado por meio de  $\max(\Delta m)$  e uma vez que a capacidade de armazenamento  $p$  cresce com  $\alpha$ , a variância da contribuição do ruído também aumenta, com o resultado analítico aproximado  $\kappa_c \simeq \sqrt{\alpha}$  [compare com Eq. (4.11)] mostrando boa concordância com as estimativas numéricas da Fig. 4.5.

Como outra consequência deste aumento, verifica-se uma tendência de diminuição da diferença  $\Delta m$ , bem como da magnitude dos valores  $\max(\Delta m)$ . Em geral, calcular estimativas extremamente precisas para  $\kappa_c$  não representa o objetivo central na presente estrutura, uma vez que pequenas variações em torno dos valores de  $\max(\Delta m)$  fornecem resultados igualmente bons, isto está diretamente relacionado às flutuações na contribuição do ruído, mostrado na Seção 4.2. Deve ser enfatizado que todos os valores de  $\alpha$  considerados na Fig. 4.5 são maiores (ou muito maiores) do que o valor crítico  $\alpha_c \simeq 0.14$ , acima do qual a rede neuronal padrão de Hopfield não reconhece nenhum padrão. Além disso, observam-se os limites  $m_\rho = m_\perp = 0$  ( $\kappa = 0$ ), bem como  $m_\rho \rightarrow 2\gamma - 1$  e  $m_\perp \rightarrow 1$  ( $\kappa \gg \kappa_c$ ).

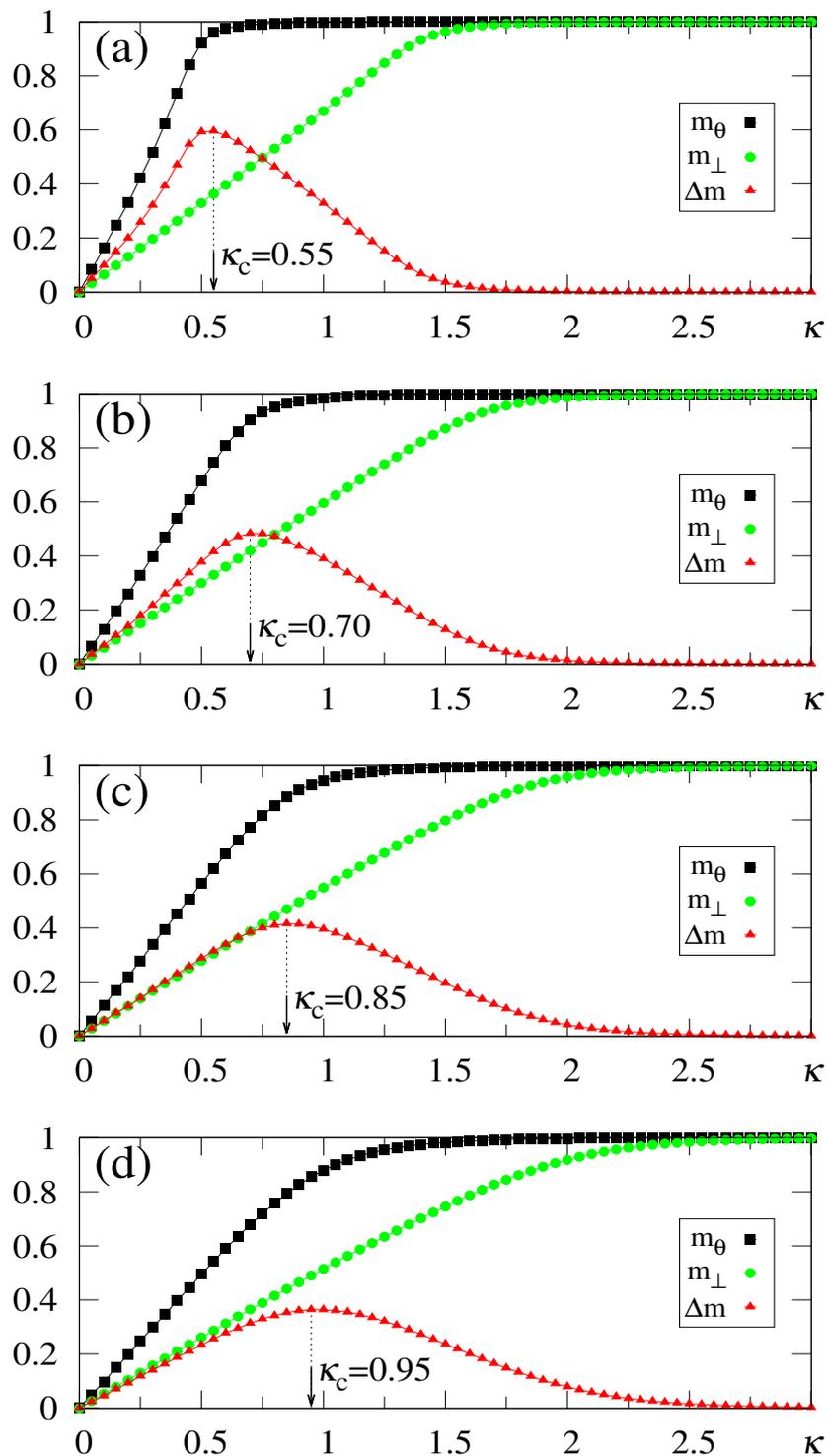


Figura 4.5: Resultados das simulações numéricas para as superposições  $m_\rho$  [Eq.(4.43)],  $m_\perp$  [Eq.(4.44)], e o módulo de sua diferença,  $\Delta m$  [Eq.(4.45)], são plotados versus  $\kappa$ , para  $\gamma = 1.0$ , e valores típicos de  $\alpha$ : (a)  $\alpha = 0.4$ ; (b)  $\alpha = 0.6$ ; (c)  $\alpha = 0.8$ ; (d)  $\alpha = 1.0$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ . As linhas que interpolam os símbolos são guias para os olhos.

Um efeito interessante aparece na Fig. 4.6, onde apresentamos resultados de simulações numéricas para as grandezas definidas nas Eqs. (4.18), (4.24) e (4.45) para  $\gamma = 0.9$ , e os mesmos valores de  $\alpha$  usados na Fig. 4.5. Para  $\kappa < \kappa_c$ , todos os gráficos são qualitativamente semelhantes aos da Fig. 4.5 e, particularmente, estimativas muito próximas para  $\kappa_c$  (ou mesmo iguais em alguns casos) foram obtidos; deve-se notar que embora o resultado analítico  $\kappa_c \simeq \sqrt{\alpha}$  da Eq. (4.11) foi derivado para o caso  $\gamma = 1$ , é válido para  $\gamma = 0.9$  como uma boa aproximação. No entanto, os limites  $m_\rho \rightarrow 2\gamma - 1 = 0.8$  e  $m_\perp \rightarrow 1$  são cumpridos para  $\kappa \gg \kappa_c$ , de modo que seus gráficos se cruzem ( $m_\rho = m_\perp$ ) em algum valor de  $\kappa$ , produzindo  $\Delta m = 0$ . À direita deste ponto de cruzamento tem-se que  $m_\perp > m_\rho$ , sinalizando que o estímulo externo passa a dominar o campo local total, conforme se aproxima de um limite não relevante para o RNDE. Nessa região, o reconhecimento de padrões começa a se deteriorar e, para  $\kappa \gg \kappa_c$ , o campo local simplesmente reproduzirá os estímulos externos, associados ou não a uma memória. Associado à diferença máxima entre as superposições  $m_\rho$  e  $m_\perp$ ,  $\kappa = \kappa_c$ , representa o valor apropriado, que permite o reconhecimento de um padrão armazenado, para  $\alpha > 0.14$ , não pode ser muito grande, de forma a evitar que o estímulo externo domine os outros termos do campo local na Eq. (4.8). Esse equilíbrio sutil se torna o coração desta abordagem.

Uma das principais virtudes da RNDE, ilustrada nas Figs. 4.5 e 4.6, diz respeito à faixa de valores de  $\alpha$ , acima do valor crítico do modelo padrão de Hopfield, e ainda permitindo o reconhecimento do padrão de forma significativa. Considerando como exemplo ilustrativo o caso  $\alpha = 1$  [cf. Figs. 4.5(d) e 4.6(d)], encontra-se  $\kappa_c \simeq 0.95$  em ambos os casos  $\gamma = 1.0$  e  $\gamma = 0.9$ , e a superposição macroscópica com um dado padrão  $\rho$  apresenta valores expressivos,  $m_\rho \simeq 0.9$  [Fig. 4.5(d)] e  $m_\rho \simeq 0.7$  [Fig. 4.6(d)], deve-se lembrar que as superposições com memórias  $\mu \neq \rho$  ainda estão presentes, sendo da ordem de  $1/\sqrt{N}$ . Este aumento na capacidade de reconhecimento de padrões está diretamente relacionado ao fato de que a RNDE não apresenta bacias de atração, embora crie uma enorme e única bacia de atração quando  $\kappa \simeq \kappa_c$ . Porém, ao definir  $\kappa = 0$ , esta bacia de atração desaparece e as memórias não ocupam nenhum volume no espaço de fase, para qualquer  $\alpha > \alpha_c$ . Como será mostrado mais adiante, dentro dessa estrutura o reconhecimento funciona bem, mesmo para valores de  $\alpha \gg 1$ .

Conforme definido na Eq. (4.6), o parâmetro  $\gamma$  mede a perturbação associada ao reconhecimento de um dado padrão armazenado  $\rho$ . Portanto, à medida que  $\gamma$  diminui da unidade, deve-se encontrar diferenças entre o estímulo externo e o padrão armazenado, de modo que o

reconhecimento do padrão deve falhar para algum  $\gamma < 1$ . Esta característica do modelo pode ser observada na Fig. 4.7. Claramente, não é possível observar uma diferença entre as duas curvas  $m_\rho$  e  $m_\perp$ , bem como não existe mais o máximo na diferença  $\max(\Delta m)$ . Portanto, não é possível escolher um  $\kappa_C$  adequado para  $\gamma \lesssim 0.74$ , de modo que não se pode distinguir um padrão armazenado de um novo padrão, não armazenado e ortogonal a outros padrões armazenados. Ou seja, podemos dizer que o limite de perturbação para o reconhecimento de um padrão armazenado é  $\gamma \simeq 0.74$ .

Portanto, nesta abordagem podemos deformar o padrão armazenado perto de 25% e a RNDE ainda irá reconhecê-lo, mesmo para valores de  $\alpha$  maiores que o limite crítico do modelo de Hopfield. Essa peculiaridade representa um desempenho impressionante, quando comparado com outros modelos de redes neuronais conhecidos na literatura.

Das três Figs., 4.5, 4.6 e 4.7, é possível distinguir uma dependência interessante entre os parâmetros do modelo RNDE. Conforme  $\alpha$  cresce,  $\kappa_C$  aumenta e se  $\gamma$  cresce,  $\kappa_c$  diminui. Esse é um comportamento esperado dado que, quando o número de padrões armazenados  $\alpha$  cresce, o ruído no sistema aumenta e em consequência  $\kappa_c$  deve crescer. Caso  $\gamma$  aumente, o ruído do campo local diminui, pois o padrão que se tenta recuperar é mais próximo do padrão armazenado  $\{\xi_i^p\}$ , conseqüentemente  $\kappa_c$  diminui.

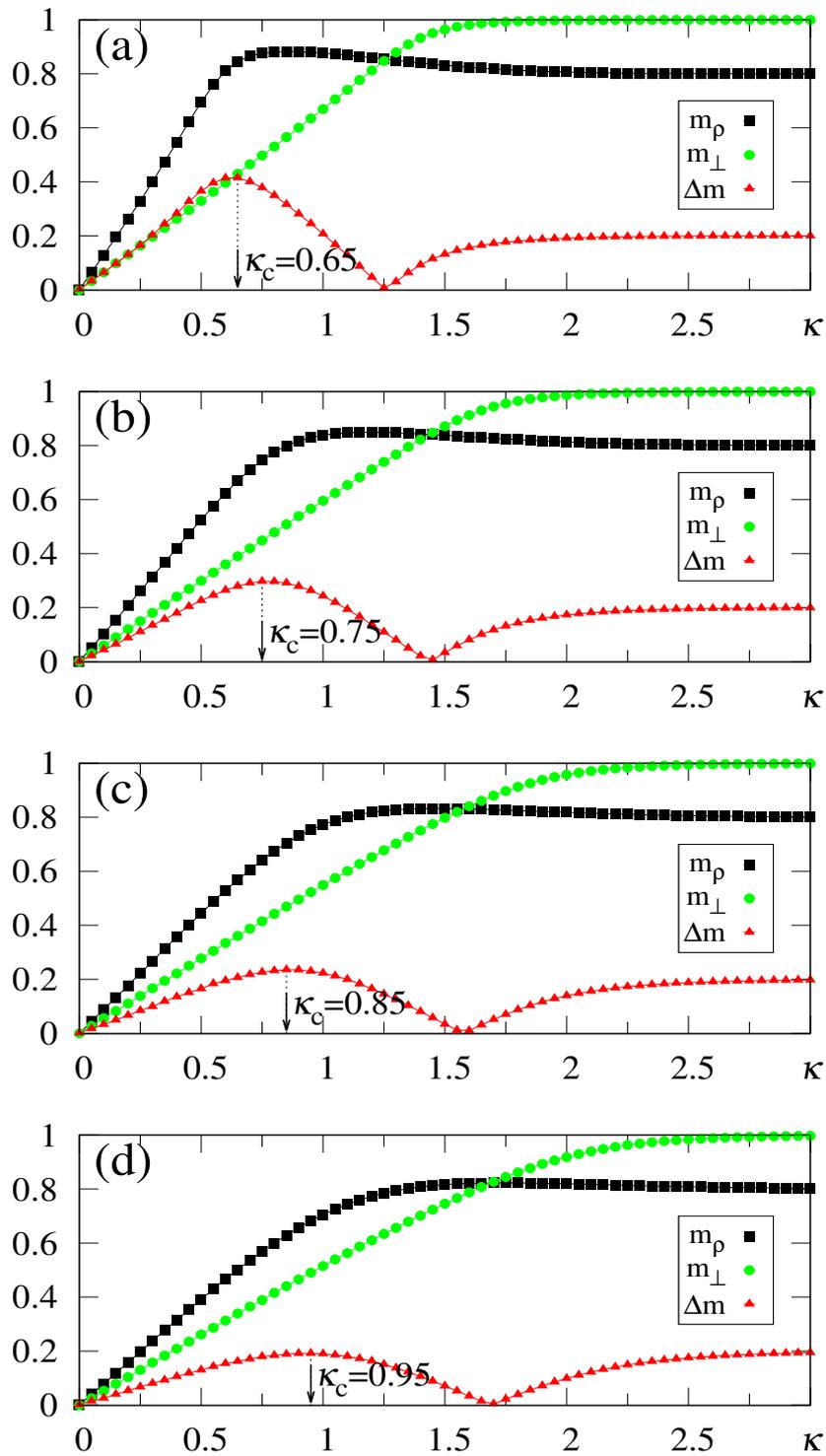


Figura 4.6: Resultados das simulações numéricas para as superposições  $m_\rho$  [Eq.(4.43)],  $m_\perp$  [Eq.(4.44)], e o módulo de sua diferença,  $\Delta m$  [Eq.(4.45)], são plotados versus  $\kappa$ , para  $\gamma = 0.9$ , e valores típicos de  $\alpha$ : (a)  $\alpha = 0.4$ ; (b)  $\alpha = 0.6$ ; (c)  $\alpha = 0.8$ ; (d)  $\alpha = 1.0$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ . As linhas que interpolam os símbolos são guias para os olhos.

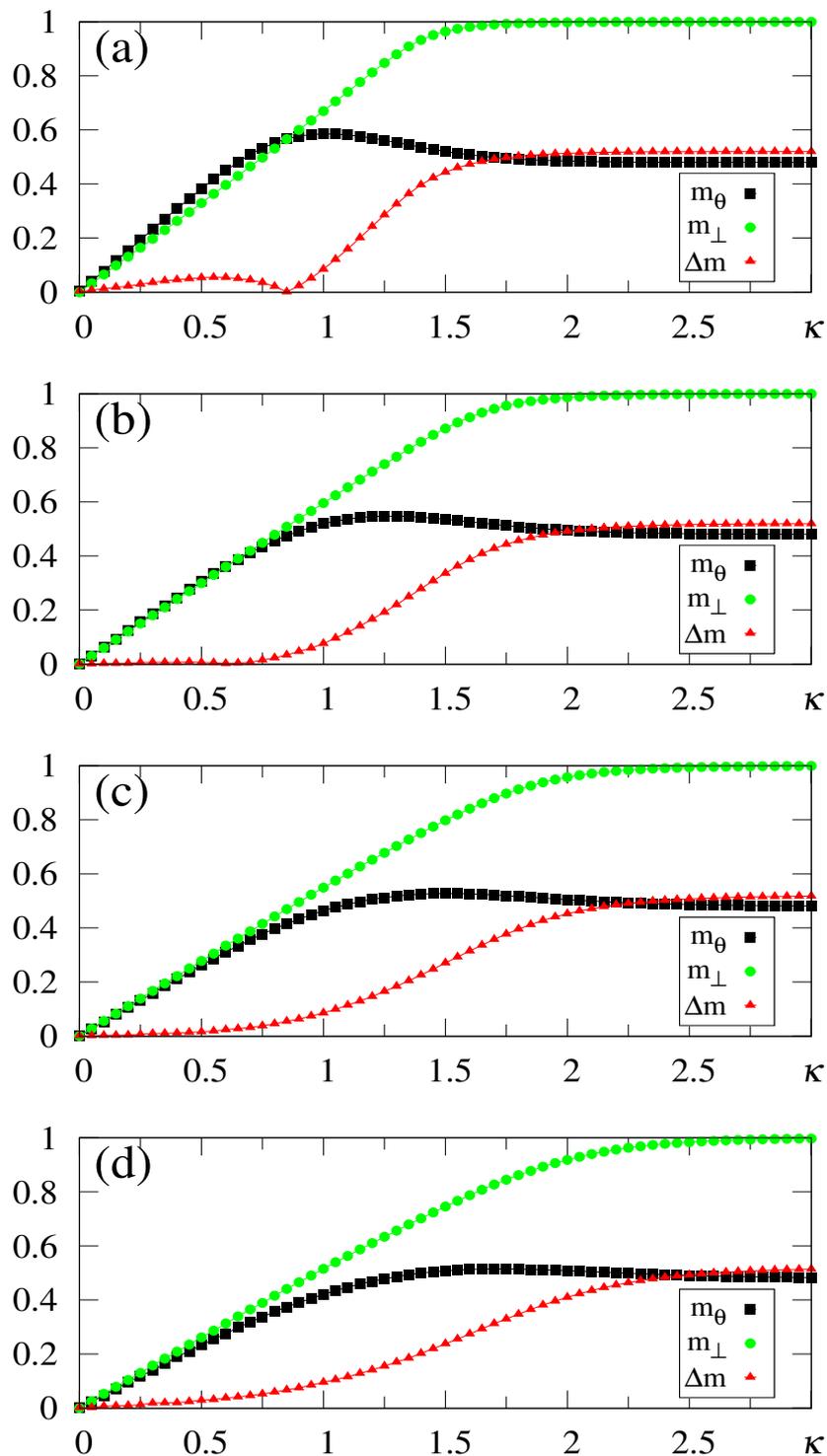


Figura 4.7: Resultados das simulações numéricas para as superposições  $m_\rho$  [Eq.(4.43)],  $m_\perp$  [Eq.(4.44)], e o módulo de sua diferença,  $\Delta m$  [Eq.(4.45)], são plotados versus  $\kappa$ , para  $\gamma = 0.74$ , e valores típicos de  $\alpha$ : (a)  $\alpha = 0.4$ ; (b)  $\alpha = 0.6$ ; (c)  $\alpha = 0.8$ ; (d)  $\alpha = 1.0$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ . As linhas que interpolam os símbolos são guias para os olhos.

### 4.7.2 Comparação dos resultados analíticos com simulações Numéricas

Os resultados teóricos e analíticos do modelo desenvolvido (RNDE), precisam ser contrastados com a simulação computacional do sistema. Esta simulação, além de verificar as hipóteses teóricas permitirá prever o comportamento futuro, isto é, os efeitos produzidos por alterações no sistema. Para simular numericamente uma rede neuronal dependente do estímulo, foi necessário o desenvolvimento de um algoritmo construído sob medida, com base no método de Monte Carlo.

Na Fig. 4.8 comparamos dados de simulações numéricas de  $N = 10^4$  neurônios (símbolos) com resultados de cálculos analíticos no limite de temperatura zero (linhas inteiras), considerando escolhas típicas para  $\alpha$  e  $\gamma$ . Os resultados analíticos para  $m_{\perp}$  seguem das Eqs. (4.37) e (4.35), enquanto aqueles para  $m_{\rho}$  foram obtidos das Eqs. (4.34) e (4.35); as estimativas numéricas correspondentes foram calculadas a partir das Eqs. (4.18) e (4.24), respectivamente, com  $\Delta m = |m_{\rho} - m_{\perp}|$  dado pela Eq. (4.45). Em todos os casos percebe-se uma boa concordância entre as duas abordagens, com pequenas discrepâncias entre elas (no máximo da ordem de 0.05), que se tornam mais significativas para pequenos valores de  $\kappa$ , bem como em torno da melhor escolha  $\kappa_c$ , como esperado, sinalizando a região mais relevante para o reconhecimento de padrões.

No entanto, os resultados das duas abordagens coincidem essencialmente quando o estímulo externo se torna muito grande, uma vez que neste limite as flutuações no reconhecimento de padrões desaparecem. Na Fig. 4.8(a), apresentamos resultados para  $\gamma = 0.74$ , ou seja, muito próximos do limite de perturbação para o reconhecimento de um padrão armazenado, conforme ilustrado na Fig. 4.7, de modo que não há um valor bem definido para  $\kappa_c$ . No entanto,  $\kappa_c$  pode ser obtido claramente nas Figs. 4.8(b)–(d), e nota-se nesses casos que as estimativas analíticas (dadas pelo máximo de  $\Delta m$ ) são sempre superestimadas em relação às numéricas [mesmo nos casos  $\gamma = 0.9$  dos painéis (b) e (c)], de acordo com as discussões anteriores referentes às aproximações que deram origem à Eq. (4.11). As discrepâncias acima mencionadas são essencialmente devidas a efeitos de tamanho finito, uma vez que os resultados analíticos e numéricos devem coincidir para  $N \rightarrow \infty$ , em cujo limite a aproximação de campo médio torna-se exata para uma rede totalmente conectada. (Amit, 1992; Peretto, 1992).

Um detalhe a ter em conta é que, a resolução das equações (4.39) - (4.41), para valores conhecidos de  $\alpha$  e  $\gamma$ , não é algo trivial, mesmo com o uso de software especializado como

MAPLE XV. Uma forma de facilitar a resolução dessas equações com o MAPLE, consiste em introduzir a variável  $w = \sqrt{2\alpha r}$  e reescreve-las da seguinte forma:

- No modelo RNDE, caso do reconhecimento de um padrão  $\gamma\%$  diferente do armazenado:

$$\begin{aligned} w &= \sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} \left[ \gamma e^{-\left(\frac{m_\rho + \kappa}{w}\right)^2} + (1 - \gamma) e^{-\left(\frac{m_\rho - \kappa}{w}\right)^2} \right], \\ m_\rho &= \gamma \operatorname{erf}\left(\frac{m_\rho + \kappa}{w}\right) + (1 - \gamma) \operatorname{erf}\left(\frac{m_\rho - \kappa}{w}\right). \end{aligned} \quad (4.46)$$

- No caso do reconhecimento de um padrão armazenado ( $\gamma = 1$ )

$$\begin{aligned} w &= \sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{m_\rho + \kappa}{w}\right)^2}, \\ m_\rho &= \operatorname{erf}\left(\frac{m_\rho + \kappa}{w}\right). \end{aligned}$$

- Para o caso do reconhecimento de um padrão ortogonal

$$\begin{aligned} w &= \sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{\kappa}{w}\right)^2}, \\ m_\rho &= \operatorname{erf}\left(\frac{\kappa}{w}\right). \end{aligned}$$

As equações de superposição, dadas no formato indicado linhas acima, permitem que softwares como MAPLE, possam lidar com as múltiplas soluções, no caso do sistema estar numa transição de fase.

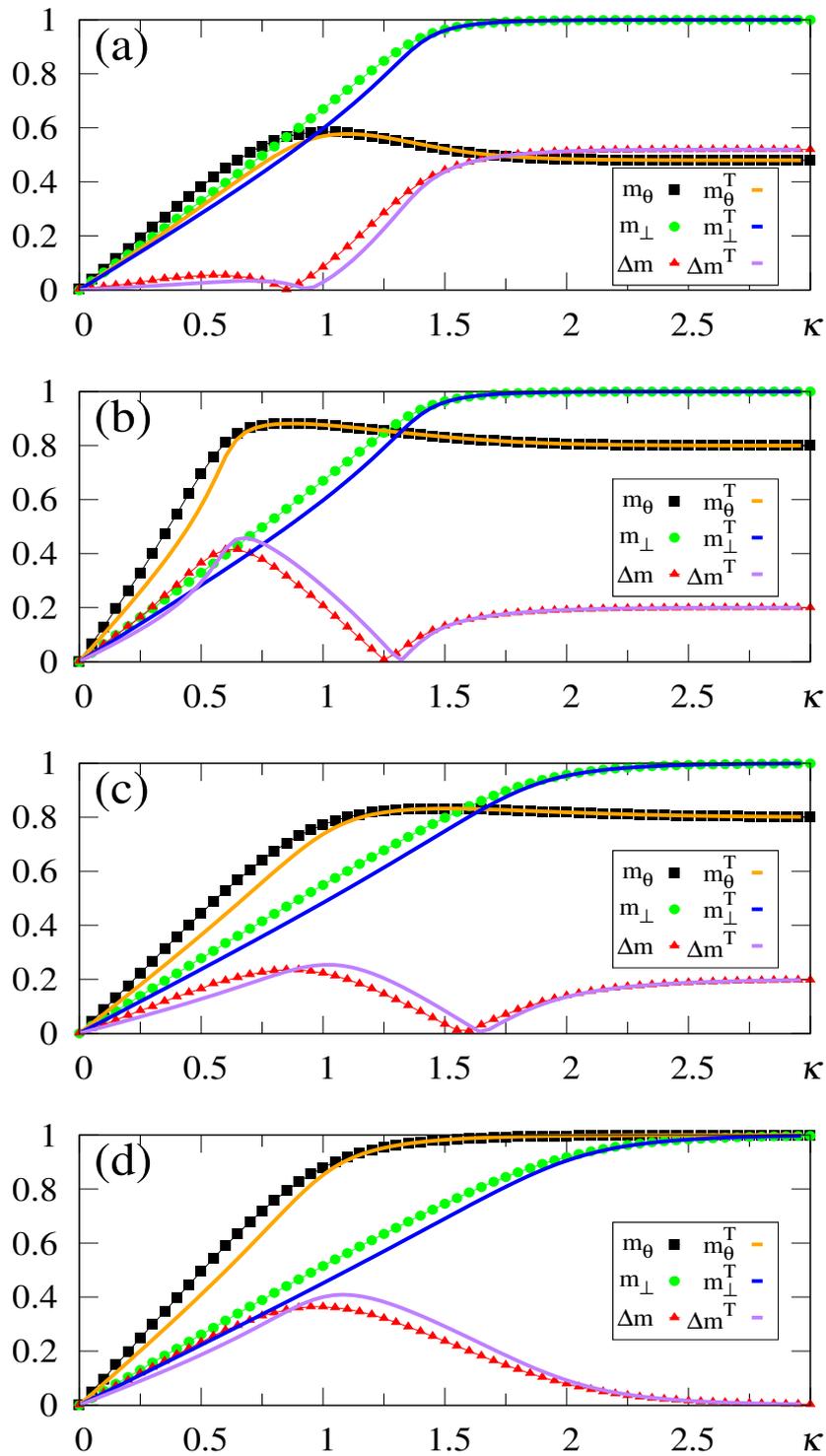


Figura 4.8: Resultados dos cálculos analíticos a temperatura zero de  $m_\rho$  e  $m_\perp$  [cf. Eqs. (4.37), (4.34)], são plotados versus  $\kappa$  e comparados com os de simulações numéricas [cf. Eqs. (4.18), (4.24)], respetivamente, enquanto o módulo teórico da diferença entre essas quantidades é comparado com o numérico  $\Delta m$  [cf. Eq. (4.45)], para valores típicos de  $\alpha$  e  $\gamma$ : (a)  $\alpha = 0.4$  e  $\gamma = 0.74$ ; (b)  $\alpha = 0.4$  e  $\gamma = 0.9$ ; (c)  $\alpha = 0.8$  e  $\gamma = 0.9$ ; (d)  $\alpha = 1.0$  e  $\gamma = 1.0$ . Para distinguir os dados teóricos dos numéricos, é usado o índice superior T.

### 4.7.3 Limite de capacidade de armazenamento

Simulações considerando valores de  $\alpha$  muito mais altos do que  $\alpha = 1$ , também foram realizadas. Nestas simulações, se considerou que o padrão externo se correspondia com  $\rho$ , ou seja,  $\gamma = 1$ . Os resultados, são mostrados na Fig. 4.9. Vê-se que não há limite bem definido para a capacidade de armazenamento, mostrando que o procedimento usado para calcular  $\kappa_c$  como o correspondente ao  $\max(\Delta m)$ , se pode aplicar também a valores consideravelmente grandes de  $\alpha$ . Particularmente, conforme apresentado na Fig. 4.9 (c), essa técnica foi facilmente implementada para  $\alpha = 16$ ; neste caso, obtemos  $\kappa_c \simeq 3.3$ , junto com  $m_\rho \simeq 0.7$ , enquanto a diferença  $\Delta m$  permanece bastante significativa (revelando um valor máximo  $\Delta m \simeq 0.1$ ), e ainda permitindo o reconhecimento claramente. Deve-se notar que o valor de  $\alpha$  usado na Fig. 4.9(c) é mais de  $10^2$  vezes maior do que o limite do modelo Hopfield padrão ( $\alpha_c \simeq 0.14$ ) Amit et al. (1985b). Analisando em conjunto os resultados das Figs. 4.4 e 4.9 (todas elas para  $\gamma = 1$ ), duas características merecem discussão, conforme apresentado a seguir. (i) O resultado analítico aproximado (derivado para  $\gamma = 1$  e  $N \rightarrow \infty$ ),  $\kappa_c \simeq \sqrt{\alpha}$  [cf. Eq. (4.11)], mostra uma boa concordância com a maioria das estimativas numéricas (normalmente dentro de erros de 10%), embora as discrepâncias aumentem para valores de  $\alpha$  maiores. Percebe-se que o cálculo analítico produz superestimativas para  $\kappa_c$ , no que diz respeito aos resultados numéricos.

Essas discrepâncias são principalmente devido às aproximações que levam à Eq. (4.11), bem como os tamanhos finitos usados nas simulações. Lembre-se que nosso sistema possui  $10^4$  neurônios<sup>1</sup>, com os quais se intenta extrapolar um sistema de tamanho muito maior como é o córtex cerebral (aprox.  $10^{10}$  neurônios). (ii) Considerando valores ainda maiores para  $\alpha$ , ocorre um efeito de saturação, no sentido de que  $m_\perp \rightarrow m_\rho$  vindo de baixo, resultando em  $\Delta m \rightarrow 0$ . Esta característica é ilustrada na Fig. 4.9(d) onde consideramos  $\alpha = 32$ , mostrando que as curvas para  $m_\rho$  e  $m_\perp$  tornam-se muito próximas, levando a um comportamento plano em  $\Delta m$ . Neste regime, identificar claramente  $\kappa_c$  torna-se difícil, explicando o aumento das discrepâncias entre os resultados analíticos e numéricos. Verificamos que essa saturação ocorre tipicamente para  $\alpha \gtrsim 40$ . No entanto, podemos afirmar que, para valores relativamente altos de  $\alpha$ , como ilustrado nas Figs. 4.9(b) e 5(c), a RNDE é capaz de produzir um reconhecimento de padrões sem dificuldades.

<sup>1</sup>Para reduzir os efeitos de tamanho da rede é possível usar um algoritmo de otimização, como o descrito por Penna e Oliveira (1989; 1990), para aumentar o número de neurônios presentes em nosso sistema.

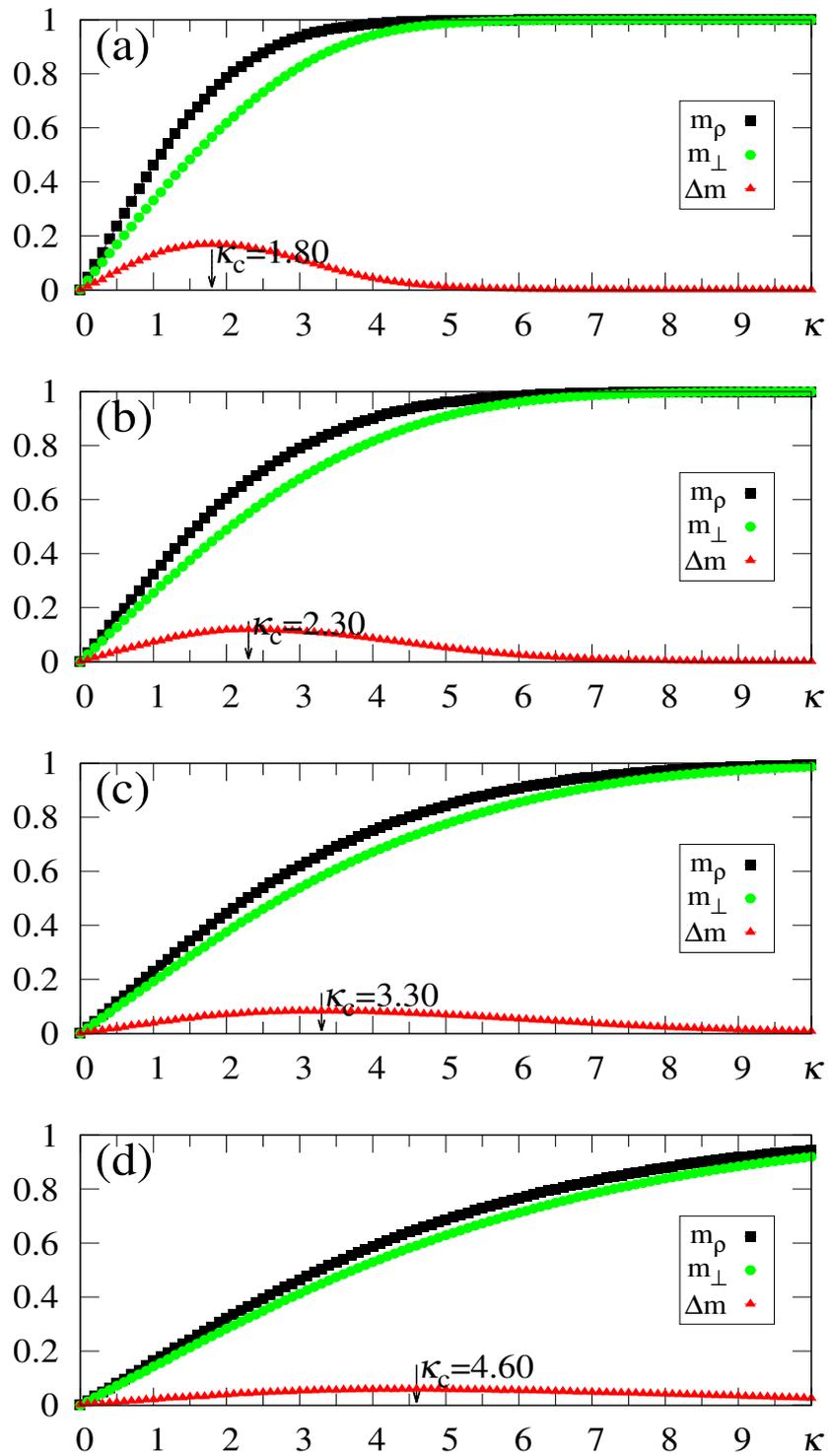


Figura 4.9: Resultados das simulações numéricas para as superposições  $m_\rho$  [cf. Eq. (4.24)],  $m_\perp$  [cf. Eq. (4.18)], e o módulo de sua diferença,  $\Delta m$  [cf. Eq. (4.45)], são plotados versus  $\kappa$ , para  $\gamma = 1.0$ , e valores típicos de  $\alpha$ : (a)  $\alpha = 4$ ; (b)  $\alpha = 8$ ; (c)  $\alpha = 16$ ; (d)  $\alpha = 32$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ .

## 4.8 RNDE sob mudanças no padrão externo

Uma característica bem conhecida dos seres vivos, fundamental para sua sobrevivência, consiste em sua capacidade de reagir às mudanças do ambiente. Nesta seção revelamos como a RNDE se ajusta, sendo capaz de substituir um padrão previamente reconhecido por um novo, de acordo com as mudanças no estímulo externo. Consideremos então uma situação real típica onde em um tempo  $t_0$  apresentamos um padrão associado a uma memória armazenada, digamos a memória  $\rho$ , e então, em um momento posterior  $t_1$  o retiramos, apresentando outro padrão, associado à memória  $\nu$ . A fim de ilustrar a eficácia da RNDE em responder a essas mudanças, agora permitimos que o conjunto de sinais de estímulo externo mude com o tempo, ou seja,  $\{\eta_i\} \equiv \{\eta_i(t)\}$ . Além disso, assumimos mudanças abruptas nessas variáveis em dois momentos dados,  $t = t_0$  e  $t = t_1$ , de modo que a distribuição de probabilidade na Eq. (4.6) é substituído pelo conjunto dependente do tempo

$$P(\eta_i(t)) = \gamma_1 \delta(\eta_i(t) - \xi_i^\rho) + (1 - \gamma_1) \delta(\eta_i(t) + \xi_i^\rho); \quad (t_0 \leq t \leq t_1), \quad (4.47)$$

e

$$P(\eta_i(t)) = \gamma_2 \delta(\eta_i(t) - \xi_i^\nu) + (1 - \gamma_2) \delta(\eta_i(t) + \xi_i^\nu); \quad (t > t_1), \quad (4.48)$$

com  $1/2 \leq \gamma_1, \gamma_2 \leq 1$ . As ações nos momentos  $t_0$  e  $t_1$  correspondem, respectivamente, à apresentação do padrão  $\rho$ , e à sua substituição pelo novo padrão  $\nu$ , levantando a questão de se o RNDE responderá adequadamente a essas modificações.

Na Fig. 4.10 ilustramos como a RNDE reage às mudanças anteriormente mencionadas, esboçando as superposições macroscópicas dependentes do tempo [cf. Eq. (4.13)]  $m_\rho(t)$  (círculos pretos) e  $m_\nu(t)$  (linha verde tracejada) versus tempo, para  $\alpha = 0.8$ , considerando quatro diferentes intensidades do estímulo externo, em ordem crescente, conforme mostrado em painéis (a)–(d). Até  $t_0 = 5.0 \times 10^4$  nenhum padrão externo é apresentado ( $\kappa = 0$ ), e nenhuma memória é reconhecida, como esperado. Então, um padrão com 80% ( $\gamma_1 = 0.8$ ) de superposição com o padrão armazenado  $\{\xi^\rho\}$  é apresentado, levando ao início da superposição macroscópica  $m_\rho(t)$  ( $0 \leq m_\rho(t) \leq 2\gamma_1 - 1$ ). Da mesma forma, até  $t_1 = 10^5$  a superposição macroscópica  $m_\nu(t)$  permanece zero, tornando-se diferente de zero apenas para  $t > t_1$ , devido a uma mudança abrupta no estímulo em  $t = t_1$ , onde o padrão  $\rho$  é substituído pelo novo padrão, que apresenta uma superposição  $\gamma_2 = 1.0$  com o padrão armazenado  $\{\xi^\nu\}$ . Assim, a Fig. 4.10 apresenta outro

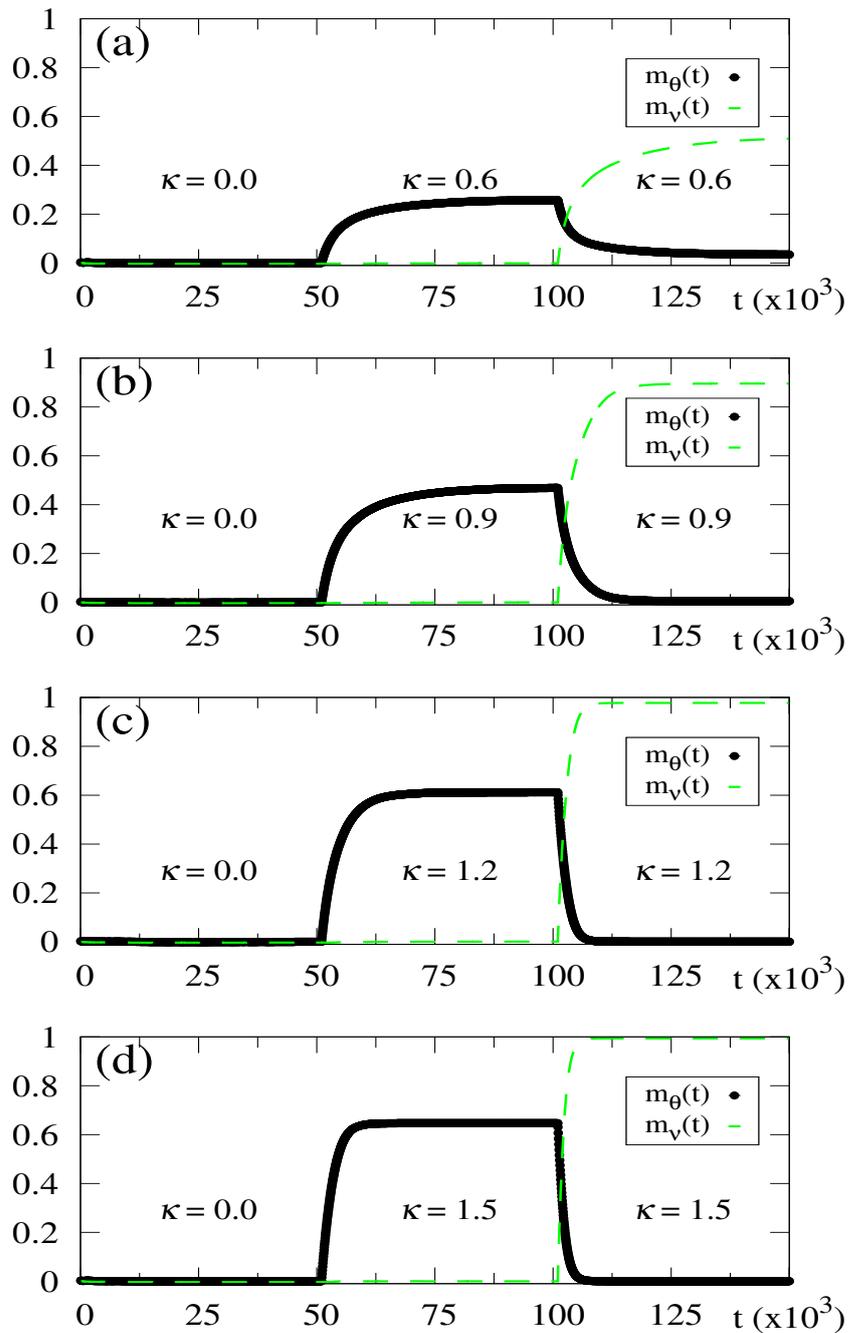


Figura 4.10: Resultados de simulações numéricas mostrando a evolução temporal de  $m_\rho(t)$  e  $m_\nu(t)$ , definidos de acordo com Eq. (4.13), para  $\alpha = 0.8$ . O estímulo externo muda com o tempo, seguindo as Eqs. (4.47) e (4.48), onde os dois tempos relevantes são  $t_0 = 5.0 \times 10^4$  e  $t_1 = 10^5$ . Por tanto, para  $t < t_0$  não há estímulo, ou seja,  $\kappa = 0.0$ , e em  $t = t_0$  um estímulo com uma intensidade  $\kappa$  aparece com uma superposição  $\gamma_1 = 0.8$  com  $\{\xi^\rho\}$ ; então, em  $t = t_1$  ele muda abruptamente apresentando uma superposição  $\gamma_2 = 1.0$  com  $\{\xi^\nu\}$ . A intensidade de  $\eta$  para  $t > t_0$  será: (a)  $\kappa = 0.6$ ; (b)  $\kappa = 0.9$ ; (c)  $\kappa = 1.2$ ; (d)  $\kappa = 1.5$ . A superposição  $m_\nu(t)$  permanece zero para  $t < t_1$ , enquanto  $m_\rho(t)$  torna-se diferente de zero para  $t_0 < t < t_1$ .

atributo importante da RNDE, correspondendo às reações a esta mudança em  $t = t_1$ , resultando em uma diminuição em  $m_\rho(t)$ , junto com um crescimento em  $m_\nu(t)$  ( $0 \leq m_\nu(t) \leq 2\gamma_2 - 1$ ) para  $t > t_1$ .

Alguns efeitos interessantes são mostrados na Fig. 4.10, conforme discutido abaixo. (i) As superposições  $m_\rho(t)$  e  $m_\nu(t)$  aproximam-se de platôs cuja altura aumenta para valores maiores de  $\kappa$ , em acordo com os resultados apresentados nas Figs. 4.5–4.8. (ii) O tempo de resposta às mudanças consideradas, seja no crescimento ou na redução das superposições, está diretamente relacionado à intensidade do estímulo externo  $\kappa$ ; valores maiores (menores) de  $\kappa$  levam a tempos de respostas menores (maiores); esse tipo de comportamento também está de acordo com as reações dos seres vivos. Desta forma, nota-se que para  $\kappa = 0.6$  os limites de tempo longo para as superposições não foram totalmente atingidos, como se pode ver comparando o gráfico de  $m_\nu(t)$  da Fig. 4.10(a) com o resultado de tempo longo da Fig. 4.5(c) em  $\kappa = 0.6$  (ambos para  $\alpha = 0.8$  e  $\gamma = 1.0$ ).

Deve-se lembrar que o intervalo de tempo para a ação do estímulo externo em todos os painéis da Fig. 4.10 é  $5.0 \times 10^4$  (para ambos os padrões  $\rho$  e  $\nu$ ), sendo menor que o tempo usado nas simulações das Figs. 4.5–4.8,  $t^* \simeq 10^2 N$  (ou seja,  $t^* \simeq 10^6$ ), suficiente para que a RNDE atinja seu mínimo local de energia mais próximo. Devido a isso, para os valores menores de  $\kappa$  [cf., por exemplo, Fig. 4.10(a)], ambas superposições  $m_\rho(t)$  e  $m_\nu(t)$  ainda não atingiram seus limites de longo tempo. No entanto, para valores maiores de  $\kappa$ , tempos muito menores são necessários para as superposições  $m_\rho(t)$  e  $m_\nu(t)$  se aproximem de seus limites de tempo longo; particularmente, na Fig. 4.10(d) nota-se que essas superposições atingiram seus valores máximos, a saber,  $m_\rho(t) \rightarrow 2\gamma_1 - 1$  e  $m_\nu(t) \rightarrow 2\gamma_2 - 1$ . (iii) Todos os painéis da Fig. 4.10 são muito ilustrativos, mostrando a capacidade da RNDE de reagir às mudanças no ambiente; no entanto, o intervalo mais relevante para o reconhecimento do padrões deve ser próximo de  $\kappa = 0.9$ , considerado na Fig. 4.10(b), de acordo com os valores de  $\kappa_c$  estimados nas Figs. 4.5(c) e 4.6(c).

Deve-se lembrar que as reações também podem ser consideradas em RNAs, como no modelo de Hopfield padrão para  $\alpha < \alpha_c$ . Nesses casos, isso pode ser obtido alterando o estado inicial  $\{\sigma_i(0)\}$ , o que pode levar a pulos no espaço de fase entre diferentes bacias de atração. Os resultados apresentados anteriormente mostram claramente que a RNDE modifica o padrão reconhecido, de forma suave, de acordo com as mudanças no padrão externo, que é uma carac-

terística comum em seres vivos reais. Em muitas situações reais, os seres vivos precisam reagir a qualquer novo padrão externo apresentado, e esse modelo se comporta exatamente assim.

## Capítulo 5

# Modificações na RNDE: correlação entre os padrões e diluição das sinapses

“O que sabemos é uma gota; o que ignoramos é um oceano.”

– *Isaac Newton.*

**G**RAÇAS as ferramentas de mecânica estatística e a simulação computacional, muitas redes diferentes foram intensamente investigadas, desde o simples perceptron, passando pelas RNA's, até redes neuronais de várias camadas. Características tais como a arquitetura da rede, a natureza de seus acoplamentos (discretos ou contínuos), simetria da força sináptica, dependência da estrutura interna dos dados (correlações) e as consequências do desaparecimento das conexões (diluição), foram de interesse para compreender os diversos modelos de redes neuronais.

Com objetivo de compreender melhor as RNDE, neste capítulo, faremos um análise nos casos em que a RNDE possui padrões correlacionados e quando ela está diluída. Nosso principal objetivo será estudar os efeitos da correlação e diluição no valor de  $\kappa_c$ .

Primeiramente, serão introduzidas correlações internas dentro dos padrões armazenados na RNDE e calcularemos o valor de  $\kappa_c$  adequado. Uma comparação com os resultados obtidos para uma RNDE com padrões descorrelacionados (vide capítulo 4) se faz evidente.

A seguir, um estudo do comportamento de  $\kappa_c$  será feito em uma RNDE assimétrica ( $J_{ij} \neq J_{ji}$ ) diluída. Uma possível maneira de introduzir a assimetria é manter as “regras de aprendizado” de Hebb, mas eliminar algumas das conexões sinápticas.

Cabe ressaltar que, devido à complexidade do Hamiltoniano da RNDE, para os dois casos descritos acima, o estudo dos sistemas será em grande parte numérico.

## 5.1 Reconhecimento de padrões correlacionados

Sabemos que, na maioria das situações realistas, os padrões armazenados  $\{\xi_i^\mu\}$  não são independentes (ver, por exemplo, Löwe, 1998). Na verdade, existem pelo menos duas maneiras sensatas de introduzir correlações entre os padrões. Uma é considerar a correlação espacial, ou seja, escolher os padrões correlacionados em  $i$  mas independentes em  $\mu$ , o que pode ser interessante quando, por exemplo, pensarmos nos padrões como imagens a serem armazenadas. A outra maneira é escolher padrões correlacionados sequencialmente ou semanticamente, o que significa que a dependência agora entra apenas via  $\mu$ . Essa situação pode ser útil como um modelo muito simples para padrões com algum tipo de relação causal, como, por exemplo, filmes.

Os resultados das pesquisas de diversos modelos de redes neuronais tipo Hopfield, com padrões armazenados correlacionados, indicam que em geral, a capacidade de armazenamento aumenta (Amit et al., 1987a; Gardner, 1988).

Nesta seção investigaremos o desempenho da estrutura RNDE quando padrões espacialmente correlacionados são armazenados. Vamos então supor um estímulo externo apresentando uma superposição máxima com uma determinada memória armazenada, digamos a memória  $\rho$ , ou seja,  $\gamma = 1$  na Eq. (4.6). Em seguida, consideramos outro padrão,  $\theta$ , apresentando uma correlação  $b$  ( $0 \leq b \leq 1$ ) com o padrão  $\rho$ ; isso significa que  $\xi_i^\rho = \xi_i^\theta$  para uma fração  $b$  dos índices  $i$ , levando a

$$\frac{1}{N} \sum_{i=1}^N \xi_i^\rho \xi_i^\theta = b - (1 - b) = 2b - 1 \quad (\rho \neq \theta). \quad (5.1)$$

Acima, os limites  $b = 0$  e  $b = 1$  correspondem aos padrões anti-correlacionados e totalmente correlacionados, respectivamente, enquanto as situações mais interessantes ocorrem para  $0 < b < 1$ .

### 5.1.1 Simulação computacional

Na Fig. 5.1 apresentamos resultados de simulações de dois padrões,  $\{\xi_i^\rho\}$  e  $\{\xi_i^\theta\}$ , com um parâmetro de correlação  $b = 0.8$  entre eles, todos os outros padrões não estão correlacionados entre si e com esses dois. Nota-se que as superposições macroscópicas,  $m_\rho$  e  $m_\theta$ , atingem os limites de saturação esperados para valores suficientemente grandes de  $\kappa$ , ou seja,  $m_\rho \rightarrow 1$  (devido à sua superposição máxima com o estímulo externo) e  $m_\theta \rightarrow 0.6$  [segundo cf. Eq. (5.1)]. Similarmente às situações anteriormente investigadas com padrões não correlacionados, o valor ótimo  $\kappa_c$  aumenta para valores crescentes de  $\alpha$ . Em torno de seus valores ótimos correspondentes  $\kappa_c$ , o sistema reconhece ambos os padrões, com  $m_\rho \gtrsim 0.8$  e  $m_\theta$  já perto de seu limite de saturação,  $m_\theta \simeq 0.6$ , para todos os valores de  $\alpha$  considerados, enquanto todas as memórias  $\mu \neq \rho, \theta$  apresentam superposições essencialmente zero (na verdade, da ordem de  $1/\sqrt{N}$ ).

É importante ressaltar que mesmo para  $\alpha = 1.0$ , encontra-se o reconhecimento de ambos os padrões armazenados, indicando que o padrão  $\theta$  apresenta uma superposição significativa com o estímulo externo, como consequência direta de sua correlação com o padrão armazenado  $\rho$ , mostrando um bom desempenho do modelo RNDE para lidar com padrões correlacionados. Como outro resultado interessante, deve-se chamar a atenção para a redução nas estimativas de  $\kappa_c$ , como consequência das correlações, este aspecto é revelado pela comparação das Figs. 4.5(d) ( $\kappa_c = 0.95$ ) e 5.1(d) ( $\kappa_c = 0.70$ ), ambos para  $\gamma = 1$  e  $\alpha = 1$ , para os quais normalmente se nota uma redução de 25% no valor de  $\kappa_c$ .

Os resultados de simulações numéricas de vários padrões, mais precisamente,  $\{\xi_i^\rho\}$ , e outros três padrões correlacionados com  $\{\xi_i^\rho\}$ , são exibidos na Fig. 5.2; os  $(p - 4)$  padrões restantes não estão correlacionados com esses quatro e entre eles. O padrão  $\{\xi_i^\rho\}$  apresenta uma superposição máxima com o estímulo externo ( $\gamma = 1$ ), enquanto os três correlacionados seguem a Eq. (5.1) com  $b = 0.7, 0.8$  e  $0.9$ . Todas as superposições macroscópicas atingem os limites de saturação esperados para valores suficientemente grandes de  $\kappa$ , ou seja,  $m_\rho \rightarrow 1$ , enquanto os três padrões correlacionados se aproximam de seus valores  $(2b - 1)$  correspondentes, de acordo com a Eq. (5.1). Observe que  $\kappa_c \simeq 0.2$  é suficiente para recuperar todos os quatro padrões, com  $m_\rho \gtrsim 0.9$  e as outras três superposições macroscópicas já muito próximas de seus limites de saturação, para todos os valores de  $\alpha$  considerados, tais valores significativos aparecem como consciências diretas das correlações com o padrão  $\{\xi_i^\rho\}$ .

Um aspecto curioso na Fig. 5.2 diz respeito ao fato de que o valor ótimo  $\kappa_c$  aumenta muito

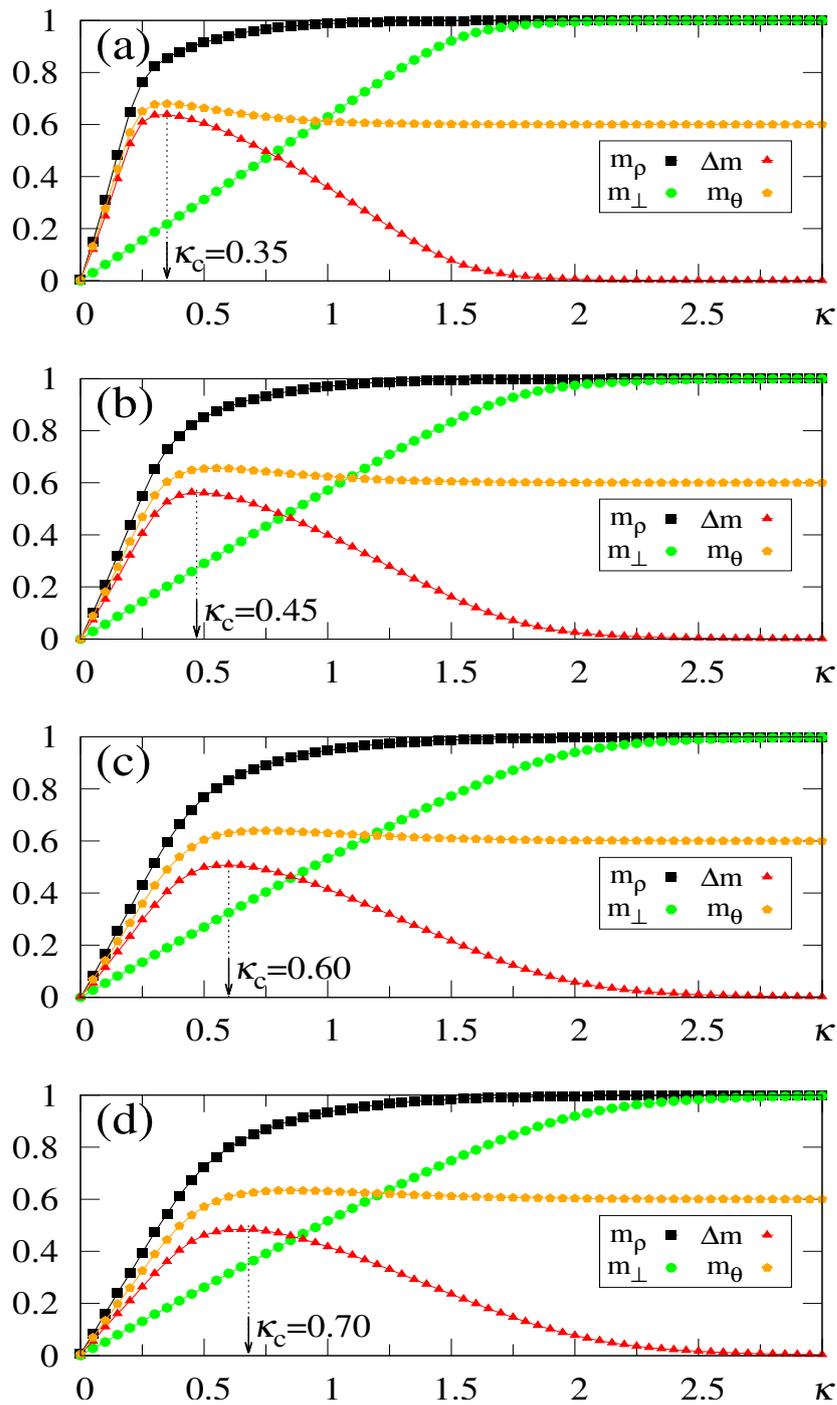


Figura 5.1: Resultados de simulações numéricas de dois padrões correlacionados ( $\{\xi_i^\rho\}$  e  $\{\xi_i^\theta\}$ ), apresentando um parâmetro de correlação  $b = 0.8$  [cf. Eq. (5.1)] entre eles, são plotados versus  $\kappa$ . O padrão  $\rho$  está totalmente correlacionado com o estímulo externo ( $\gamma = 1$ ) e os dados para valores típicos de  $\alpha$  são mostrados: (a)  $\alpha = 0.5$ ; (b)  $\alpha = 0.7$ ; (c)  $\alpha = 0.9$ ; (d)  $\alpha = 1.0$ . Perceba que  $m_\perp$  [cf. Eq. (4.18)], é calculado com um estímulo externo diferente, ortogonal a todas as memórias armazenadas.

mais lentamente com  $\alpha$ , quando comparado com os casos anteriormente investigados, tanto para padrões não correlacionados (ver, por exemplo, Figs. 4.5 e 4.6) como para dois padrões correlacionados (cf. Fig. 5.1). Além de pequenas variações em  $\kappa_c$ , a magnitude de  $\kappa_c$  diminui com o aumento do número de padrões correlacionados, como pode ser visto comparando, por exemplo, as Figs. 4.5(d) ( $\kappa_c = 0.95$ ), 5.1(d) ( $\kappa_c = 0.70$ ), e 5.2(d) ( $\kappa_c = 0.21$ ), todos eles para  $\gamma = 1$  e  $\alpha = 1$ . As presentes simulações sugerem que, para um número suficientemente grande destes padrões,  $\kappa_c$  deve convergir para um pequeno valor finito  $\kappa_c \gtrsim 0$ , para  $\alpha$  fixo; aumentando  $\alpha$ , um aumento lento nos valores de  $\kappa_c$  deve ocorrer. Este último resultado está de acordo com o reconhecimento de padrões semelhantes realizados por seres vivos, onde uma vez que um constituinte de um determinado grupo é reconhecido, todos os membros semelhantes do grupo também são imediatamente, requerendo um pequeno estímulo externo para essa tarefa. Os resultados das Figs. 5.1 e 5.2 ilustram características adicionais importantes do modelo RNDE, apresentando aspectos muito semelhantes às dos seres vivos no reconhecimento de padrões correlacionados.

Também testamos como a RNDE se comporta considerando vários padrões correlacionados no regime de grandes valores de  $\alpha$ , como mostrado na Fig. 5.3. Similarmente ao que aconteceu nas Figs. 5.1 e 5.2, as estimativas de  $\kappa_c$  diminuem para padrões correlacionados. Este efeito pode ser verificado comparando os resultados da Fig. 5.3 (quatro padrões correlacionados) com aqueles da Fig. 4.9 (padrões não correlacionados), por exemplo, no caso  $\alpha = 4$  [Fig. 5.3(b) ( $\kappa_c = 1.0$ ) e Fig. 4.9(a) ( $\kappa_c = 1.8$ )], bem como  $\alpha = 16$  [Fig. 5.3(d) ( $\kappa_c = 2.6$ ) e Fig. 4.9(c) ( $\kappa_c = 3.3$ )], todos eles para  $\gamma = 1$ . Este aspecto está diretamente relacionado às brechas entre as curvas para  $m_\rho$  e  $m_\perp$ , como pode ser visto nos gráficos correspondentes acima mencionados. Nestes casos, nota-se que o valor máximo de  $\Delta m$  normalmente é duplicado nas Figs. 5.3(b) e 5.3(d), quando contrastado com aqueles das Figs. 4.9(a) e 4.9(c), respectivamente, tornando mais fácil calcular os valores de  $\kappa_c$ . Tal aumento no máximo de  $\Delta m$ , junto com a redução nos valores de  $\kappa_c$ , deve produzir um aumento no número total de memórias, uma vez que se amplia o intervalo de valores de  $\alpha$  para o reconhecimento de padrões, esses resultados indicam que o efeito de saturação observado na Fig. 4.9 deve ocorrer para valores ainda maiores de  $\alpha$ , quando se introduzem correlações entre os padrões. Consequentemente, a introdução de um ingrediente tão importante para sistemas reais, ou seja, correlações entre padrões, deverá melhorar ainda mais a capacidade de armazenamento da RNDE.

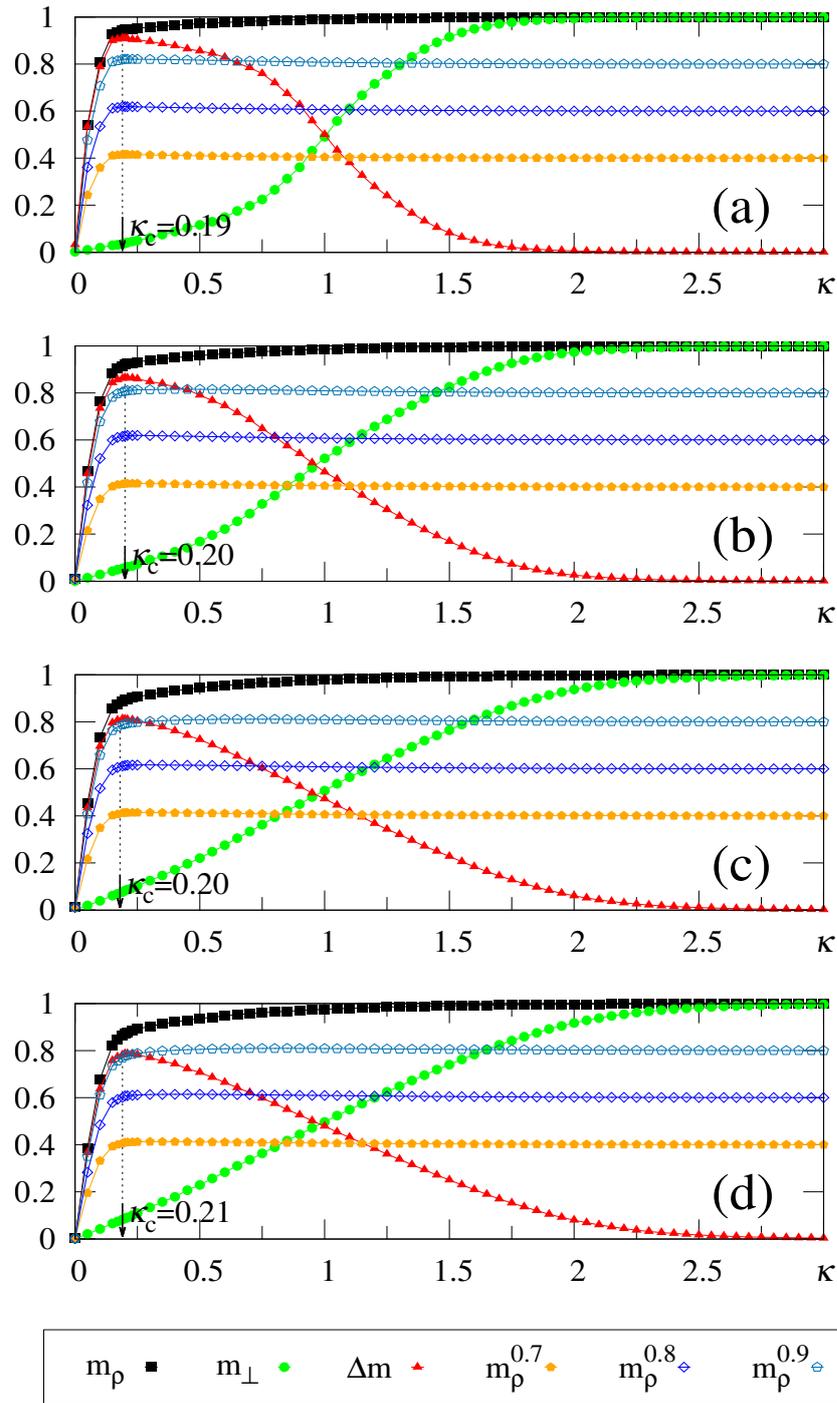


Figura 5.2: Resultados das simulações numéricas para três padrões, correlacionados com o padrão  $\{\xi_i^\rho\}$ , plotados versus  $\kappa$ . O padrão  $\rho$  está totalmente correlacionado com o estímulo externo ( $\gamma = 1$ ) e os dados para valores típicos de  $\alpha$  são mostrados: (a)  $\alpha = 0.5$ ; (b)  $\alpha = 0.7$ ; (c)  $\alpha = 0.9$ ; (d)  $\alpha = 1.0$ . Apresentamos  $b$  como super índice nos três padrões correlacionados com  $\{\xi_i^\rho\}$ . O  $m_\perp$ , é calculado com um estímulo externo diferente, ortogonal a todas as memórias armazenadas, para acharmos um  $\kappa_c$  adequado. Nota-se que  $\kappa_c$  aumenta com  $\alpha$ , embora muito mais devagar do que na Fig. 5.1.

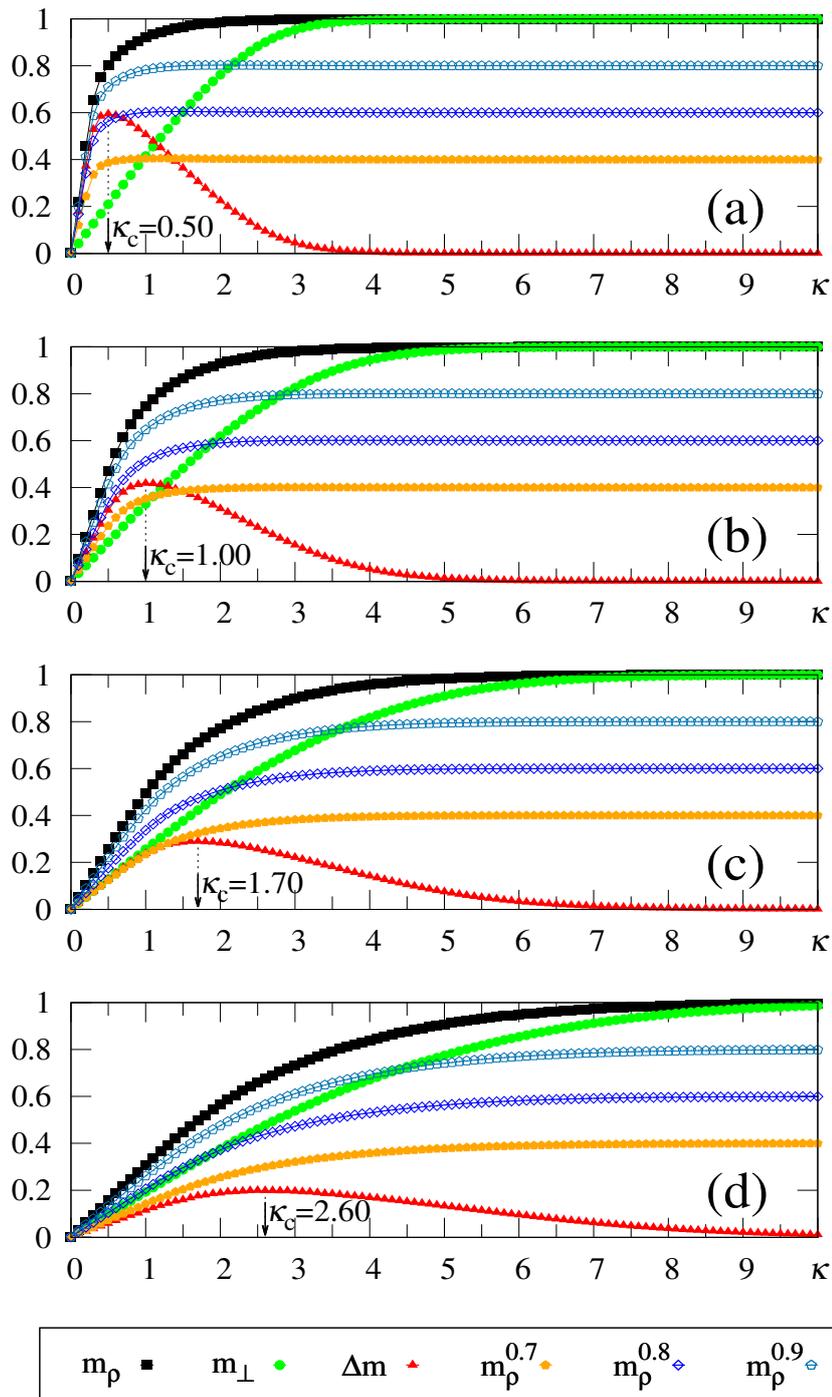


Figura 5.3: Resultados das simulações numéricas para três padrões, correlacionados com o padrão  $\{\xi_i^\rho\}$ , plotados versus  $\kappa$ . O padrão  $\rho$  está totalmente correlacionado com o estímulo externo ( $\gamma = 1$ ) e os dados para valores típicos de  $\alpha$  são mostrados: (a)  $\alpha = 2$ ; (b)  $\alpha = 4$ ; (c)  $\alpha = 8$ ; (d)  $\alpha = 16$ . Apresentamos  $b$  como super índice nos três padrões correlacionados com  $\{\xi_i^\rho\}$ . O  $m_\perp$ , é calculado com um estímulo externo diferente, ortogonal a todas as memórias armazenadas, para acharmos um  $\kappa_c$  adequado. Nota-se que  $\kappa_c$  ainda aumenta com  $\alpha$ .

## 5.2 Reconhecimento em uma rede neuronal diluída

A robustez das redes neuronais artificiais, como sistemas de armazenamento de informação, é um de seus maiores atributos. Em particular, sua tolerância a danos sinápticos tem sido o foco de alguns trabalhos interessantes. Em geral, as redes neuronais mostram uma “degradação graciosa” da memória, muito semelhante aos sintomas da doença de Alzheimer. Derrida et al. (1987) mostram que sob extrema diluição ( $1 - d$  é de  $\mathcal{O}(1)$ , onde  $d$  é a fração do número total de conexões cortadas) o caso das conexões assimétricas é exatamente solucionável. Nesse caso, a capacidade crítica de armazenamento pode ser tão alta quanto  $2/\pi$ . Em um trabalho relacionado (Kree e Zippelius, 1991), essa mesma capacidade crítica de armazenamento deve ser mantida para “diluição forte” ( $1 - d$  é de  $\mathcal{O}(1/N)$ ). A tolerância deste modelo para “diluição fraca” é mostrada como nula para o caso saturado ( $\alpha$  se aproximando de 2). Uma estimativa da capacidade crítica em função da diluição para o caso de conexões simétricas foi feita por Sompolinsky (1986) com base no método usado por Amit et al. (1985a). A aproximação mostrou que a capacidade da rede diluída é ligeiramente maior do que a aproximação ingênua de  $\alpha_c = 0.138(1 - d)$ , onde  $d$  é a fração do número total de conexões cortadas. Esta análise, no entanto, pode não ser válida para dinâmica de temperatura zero. Nesta seção, discutiremos apenas as características de uma RNDE em um caso intermediário entre a diluição fraca e forte ( $0 < d < 1 - N^{-1}$ ), cujo estudo só é possível mediante simulação computacional.

### 5.2.1 RNDE diluída

Nesta seção, apresentamos uma versão diluída não simétrica do modelo RNDE e mostramos que também tem um bom desempenho no reconhecimento de padrões. A eficiência da RNDE será ilustrada, analisando seu desempenho no modelo de Hopfield com diluição de sinapses não simétricas. Vamos então considerar um sistema de  $N$  neurônios  $S_i = \pm 1$ , cujas interações  $J_{ij}$  dependem de  $p$  padrões armazenados. Considerando o mesmo acoplamento sináptico  $J_{ij}$ , usado por Sompolinsky (1987) e Derrida et al. (1987), o qual é dado por

$$J_{ij} = \frac{C_{ij}}{N(1-d)} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad (5.2)$$

onde  $\xi_i^\mu (= \pm 1)$  é o valor do padrão  $\mu$  no sítio  $i$ . As variáveis  $\{C_{ij}\}$ , que obedecem a seguinte distribuição de probabilidade

$$P(C_{ij}) = (1 - d) \delta(C_{ij} - 1) + d \delta(C_{ij}), \quad (5.3)$$

são aleatórias e independentes, e representam a diluição e assimetria do sistema. O parâmetro de diluição  $d \in [0, 1[$ , define-se como o número total de conexões que foram cortadas, normalizado com o número total de conexões de uma rede neuronal totalmente conectada. Note-se que as interações  $J_{ij}$  são não simétricas devido a que para cada par  $(i, j)$ ,  $C_{ij} \neq C_{ji}$ ; consequentemente, nenhum hamiltoniano pode ser definido. Os dois valores extremos para o parâmetro  $d$  correspondem a  $d = 0$  (limite não diluído), enquanto seu valor máximo vem da Eq. (5.2), levando a  $(1 - d) \sim 1/N$ .

Neste modelo, as sinapses são cortadas aleatoriamente, de modo que o número médio de sinapses por neurônio é  $N(1 - d)$ . A normalização de  $J_{ij}$ , equação (5.2), foi escolhida por conveniência, de modo que a magnitude da entrada sináptica total média para um neurônio, em um dos padrões, seja a unidade, independentemente do valor de  $d$ . Isso requer a divisão por um fator  $1 - d$  para compensar as conexões eliminadas.

## 5.2.2 Estimativa de $\kappa$ ótimo

A seguir, para entender melhor os efeitos em  $\kappa_c$  devido à diluição das sinapses, vamos estimar seu valor dentro da abordagem de campo molecular, o procedimento é semelhante ao realizado na Seção 4.2, através de uma análise da contribuição do ruído. Portanto, o campo local no neurônio  $i$  é dado por

$$h_i(t) = \sum_{j \neq i}^N J_{ij} S_j(t) + \kappa \eta_i, \quad (5.4)$$

com  $\{J_{ij}\}$  agora dado pela Eq. (5.2). Então, assumimos que em um tempo  $t$  o estado do neurônio  $\{S_i(t)\}$ , assim como o estímulo externo  $\{\eta_i\}$ , coincidem com o padrão  $\{\xi_i^\rho\}$ , levando a,

$$\begin{aligned} h_i(t) &= \sum_{j \neq i}^N \sum_{\mu=1}^p \frac{C_{ij}}{N(1-d)} \xi_i^\mu \xi_j^\mu \xi_j^\rho + \kappa \xi_i^\rho, \\ &= \sum_{j \neq i}^N \frac{C_{ij}}{N(1-d)} \xi_i^\rho + \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \frac{C_{ij}}{N(1-d)} \xi_i^\mu \xi_j^\mu \xi_j^\rho + \kappa \xi_i^\rho. \end{aligned} \quad (5.5)$$

Como antes, a soma sobre as memórias  $\mu$  foi dividida em duas contribuições, a saber, um sinal produzido pelo padrão  $\mu = \rho$ , e um ruído devido às memórias  $\mu \neq \rho$ ,

$$h_i^{\text{ruído}}(t) = \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \frac{C_{ij}}{N(1-d)} \xi_i^\mu \xi_j^\mu \xi_j^\rho. \quad (5.6)$$

Assumindo que não há correlação entre os padrões  $\mu$  e  $\rho$ , e que os padrões associados a diferentes sítios ( $j \neq i$ ) são independentes, cada contribuição de memória assume os valores  $\pm 1$  com probabilidades iguais. Lembrando que  $\{C_{ij}\}$  também são variáveis independentes, as variáveis aleatórias de ruído  $\{h_i^{\text{ruído}}(t)\}$  não produzem uma Gaussiana para um  $N$  finito, mas eles devem se tornam em gaussianamente distribuídos quando  $N \rightarrow \infty$ . Portanto, considerando valores grandes o suficiente de  $N$  e  $p$ , obtêm-se para a média sobre as variáveis  $\{\xi\}$ 's,

$$\langle h_i^{\text{ruído}}(t) \rangle_\xi \simeq 0, \quad (5.7)$$

enquanto que

$$(h_i^{\text{ruído}}(t))^2 = \frac{1}{[N(1-d)]^2} \sum_{j \neq i}^N \sum_{\mu \neq \rho}^p \sum_{k \neq i}^N \sum_{\nu \neq \rho}^p C_{ij} C_{ik} \xi_i^\mu \xi_j^\mu \xi_j^\rho \xi_i^\nu \xi_k^\nu \xi_k^\rho. \quad (5.8)$$

Visto que  $\langle C_{ij}^2 \rangle = \langle C_{ij} \rangle = 1-d$ , a variância associada com respeito às variáveis  $\{\xi\}$ 's torna-se

$$\langle (h_i^{\text{ruído}}(t))^2 \rangle_\xi \simeq \frac{(N-1)(p-1)}{N^2(1-d)} \simeq \frac{p}{N(1-d)} = \frac{\alpha}{1-d}. \quad (5.9)$$

Similarmente à discussão realizada na Seção 4.2, deve-se escolher  $\kappa_c$  igual à largura da distribuição do ruído,

$$\kappa_c = \langle (h_i^{\text{ruído}}(t))^2 \rangle_\xi^{1/2} \simeq \sqrt{\frac{\alpha}{1-d}}, \quad (5.10)$$

a fim de cancelar (aproximadamente) a contribuição de ruído. De acordo com isso, um parâmetro de diluição  $0 < d < 1$  produz um aumento no valor analítico aproximado para  $\kappa_c$ , em comparação com o caso não diluído  $d = 0$ ; este efeito está de acordo com as simulações numéricas, como será mostrado a seguir.

### 5.2.3 Simulação computacional

A seguir, para simplificar, ilustramos os efeitos da diluição das sinapses para padrões não correlacionados, considerando  $\gamma = 1.0$  na Eq. (4.6)), ou seja, um padrão externo coincidindo precisamente com o armazenado. Os resultados das simulações numéricas são mostrados nas Figs. 5.4 e 5.5, com as quantidades definidas nas Eqs. (4.18), (4.24) e (4.45) sendo plotados versus  $\kappa$ , para valores decrescentes do parâmetro de diluição  $d$ , e duas escolhas típicas de  $\alpha$ , a saber,  $\alpha = 0.5$  [Fig. 5.4] e  $\alpha = 1.0$  [Fig. 5.5]. Nota-se que os limites  $m_\rho = m_\perp = 0$  ( $\kappa = 0$ ) ainda se mantêm; além disso, para valores suficientemente altos de  $\kappa$ , os limites de saturação  $m_\rho \rightarrow 2\gamma - 1$  e  $m_\perp \rightarrow 1$  devem ser aproximados, mesmo para diluições elevadas. Um aspecto importante nessas figuras diz respeito ao fato de que  $\kappa_c$  é claramente identificado por meio de  $\max(\Delta m)$ , mesmo para o maior valor de  $d$  considerado [ $d = 0.7$  nas Figs. 5.4(a) e 5.5(a)]. Além disso, para escolhas fixas do parâmetro de diluição  $d$ ,  $\kappa_c$  aumenta com  $\alpha$ , como já observado para o caso não diluído, enquanto para  $\alpha$  fixo, nota-se que  $\kappa_c$  decresce ao diminuir  $d$ ; esses resultados estão de acordo com a estimativa teórica da Eq. (5.10). No entanto, as discrepâncias entre as estimativas numéricas e analíticas para  $\kappa_c$  são maiores para  $0 < d < 1$ , quando comparadas com aquelas para o limite não diluído ( $d = 0$ ); enquanto as discrepâncias relativas no limite não diluído são normalmente da ordem de 10%, para diluições maiores, por exemplo,  $d = 0.7$ , pode haver discrepâncias relativas ligeiramente maiores que 20%. Similarmente ao grande limite  $\alpha$  do caso não diluído [ver, por exemplo, Fig. 4.9] as curvas para  $\Delta m$  tornam-se mais planas para altas diluições, levando a tais discrepâncias maiores. Além disso, como antes, o resultado analítico da Eq. (5.10) representa superestimativas em relação às numéricas, para todos os casos mostrados nas Figs. 5.4 e 5.5. Valores crescentes de  $\kappa_c$  com a diluição são qualitativamente esperados, uma vez que a diluição das sinapses diminuí os caminhos possíveis para a transmissão de uma determinada informação e, portanto, maiores intensidades do campo externo devem ser necessárias no processo de reconhecimento de padrões.

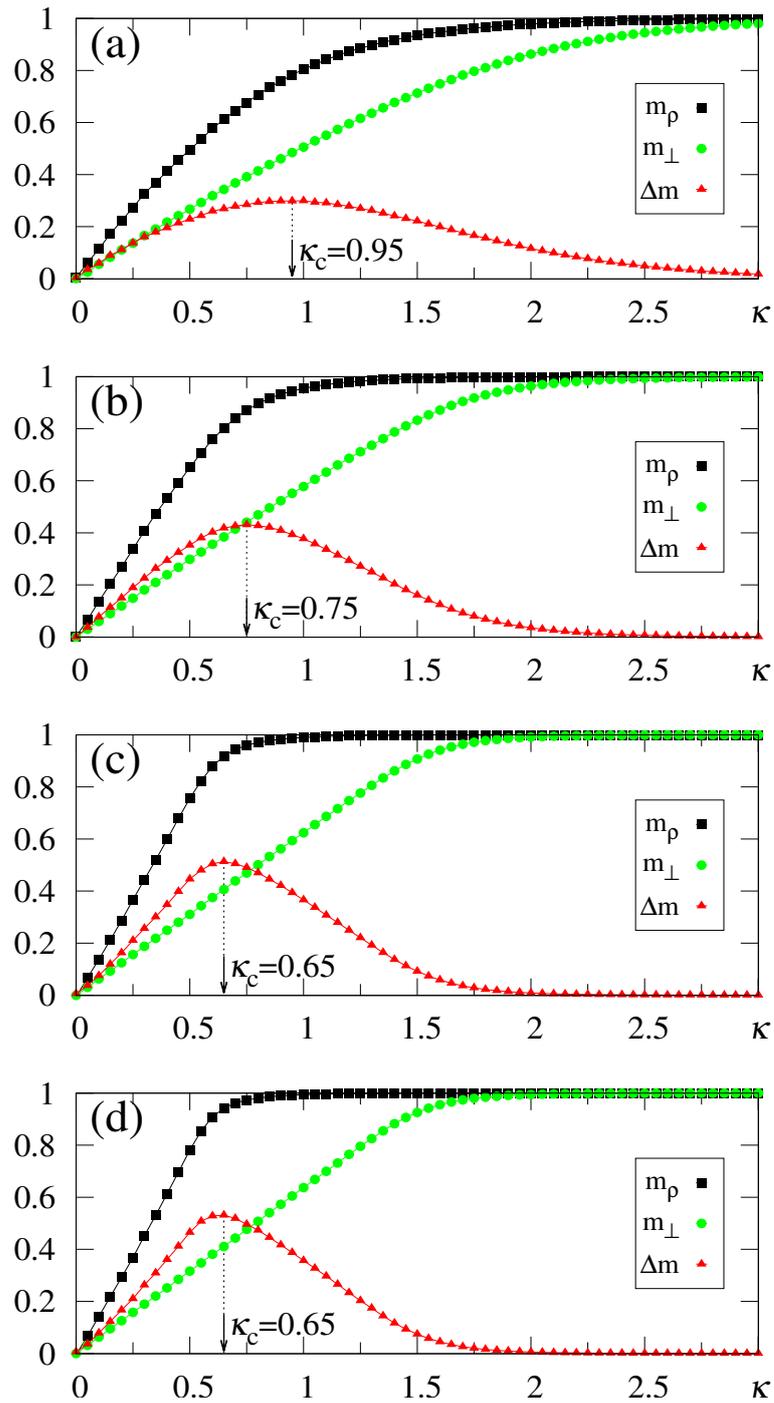


Figura 5.4: Resultados de simulações numéricas para as superposições  $m_\rho$  [cf. Eq. (4.24)],  $m_\perp$  [cf. Eq. (4.18)], e o módulo de sua diferença,  $\Delta m$  [cf. Eq. (4.45)], são plotados versus  $\kappa$ , para  $\gamma = 1.0$ ,  $\alpha = 0.5$ , e valores típicos do parâmetro de diluição  $d$ : (a)  $d = 0.7$ ; (b)  $d = 0.4$ ; (c)  $d = 0.1$ ; (d)  $d = 0.0$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ . As duas superposições mostradas representam, respectivamente, superposições macroscópicas do estado neuronal com um padrão armazenado  $\rho$  ( $m_\rho$ ) e com padrões não armazenados, ortogonais ao armazenado ( $m_\perp$ ).

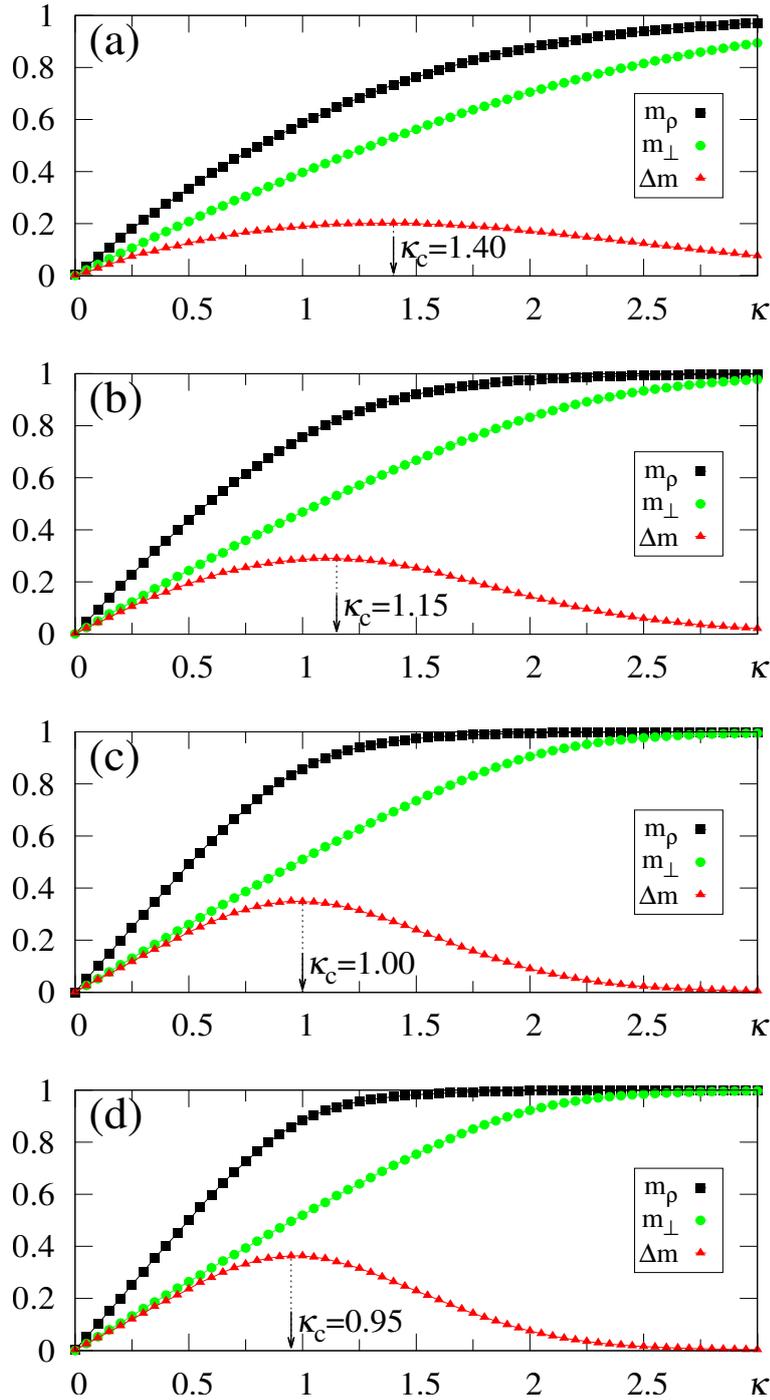


Figura 5.5: Resultados de simulações numéricas para as superposições  $m_\rho$  [cf. Eq. (4.24)],  $m_\perp$  [cf. Eq. (4.18)], e o módulo de sua diferença,  $\Delta m$  [cf. Eq. (4.45)], são plotados versus  $\kappa$ , para  $\gamma = 1.0$ ,  $\alpha = 1.0$ , e valores típicos do parâmetro de diluição  $d$ : (a)  $d = 0.7$ ; (b)  $d = 0.4$ ; (c)  $d = 0.1$ ; (d)  $d = 0.0$ . Em cada caso, o valor máximo de  $\Delta m$  produz a melhor escolha para  $\kappa$ , denotado por  $\kappa_c$ . As duas superposições mostradas representam, respectivamente, superposições macroscópicas do estado neuronal com um padrão armazenado  $\rho$  ( $m_\rho$ ) e com padrões não armazenados, ortogonais ao armazenado ( $m_\perp$ ).

# Capítulo 6

## Conclusões

“Em algum lugar, alguma coisa incrível  
está esperando para ser descoberta.”

– *Carl Sagan.*

COM base no comportamento comum dos seres vivos, propomos uma nova estrutura de rede neuronal, na qual o padrão externo exerce forte influência no processo de reconhecimento de um grande número de padrões armazenados, expresso como  $p = \alpha N$ , onde  $N$  representa o número total de neurônios. Esse estímulo externo, introduzido na forma de um padrão externo fixo, permanece ativo durante todo o processo de reconhecimento, aumentando consideravelmente a capacidade da rede neuronal em reconhecer padrões previamente armazenados.

Em contraste com as redes neuronais atratoras usuais, na ausência do campo externo, as memórias não são atratores fixos da dinâmica, e bacias de atração podem ser geradas para estímulos externos que apresentem superposições macroscópicas significativas com memórias armazenadas.

O ponto crucial da estrutura RNDE consiste em calibrar a intensidade do estímulo externo  $\kappa$  afim de cancelar (o máximo possível) a contribuição do ruído devido a outros padrões, não correlacionados com o padrão externo apresentado. Assim, desenvolvemos uma técnica para calcular este valor ótimo (referido como  $\kappa_c$ ), que consiste em estimar a diferença máxima entre  $m_\rho$ , a superposição macroscópica de um estado neuronal com um padrão armazenado  $\rho$ , e com padrões não armazenados ortogonais ao armazenado,  $m_\perp$ . Mais especificamente, o máximo de

$\Delta m = |m_\rho - m_\perp|$  permite estimar  $\kappa_c$ , e foi demonstrado que, em torno desse valor a superposição  $m_\rho$  apresentou valores significativos, adequados para o reconhecimento de padrões.

Em contraste com as redes neuronais atratoras padrão, as quais tem memórias associadas a mínimos em uma espécie de paisagem de energia, esse procedimento aumenta consideravelmente o número de memórias armazenadas, uma vez que um estímulo correlacionado com um padrão armazenado é apresentado. Ilustramos como a proposta funciona incluindo esta nova contribuição no modelo de Hopfield padrão, conhecido por ser incapaz de reconhecer qualquer outro padrão para  $\alpha \gtrsim 0.14$ . Levando em consideração que cálculos analíticos podem ser realizados para o modelo de Hopfield, comparamos resultados analíticos e numéricos, mostrando a boa concordância entre as duas abordagens.

Na simulação da RNDE, foram considerados 10 000 neurônios que evoluíam assincronamente com  $T = 0$ . As simulações foram levadas a cabo usando um computador com CPU Intel Xeon 2630 V4 com 3.1 GHz e 384 GB de memória RAM. O código gastou um tempo de  $\simeq 20$  min em cada realização, para  $\alpha \leq 1$ . As médias das superposições,  $m_\rho$  e  $m_\perp$ , foram feitas sobre  $10^3$  realizações. O tempo de realização aumenta conforme  $\alpha$  cresce.

O estudo numérico da estrutura RNDE, aplicado à rede de Hopfield, permitiu descobrir que  $\gamma \simeq 0.74$  é o limite de perturbação do padrão armazenado para o reconhecimento. Para esse valor, não é possível escolher um  $\kappa_c$  adequado, já que não podemos distinguir um padrão armazenado de um novo padrão, não armazenado e ortogonal a outros padrões armazenados. Esse reconhecimento acontecerá mesmo para valores de  $\alpha$  maiores que o limite crítico do modelo de Hopfield. Em contraste com o caso Hopfield padrão, aqui não encontramos nenhum valor limite para  $\alpha$ , mas sim um tipo de efeito de saturação para  $\alpha$  suficientemente grande (com  $\gamma = 1.0$ ), no sentido de que as duas superposições  $m_\rho$  e  $m_\perp$  tornam-se muito próximas, levando a  $\Delta m \rightarrow 0$ . Essa saturação ocorre para  $\alpha \gtrsim 40$ .

Como observado neste trabalho, as discrepâncias entre as estimativas numéricas e analíticas de  $\kappa_c$ , são devido, principalmente aos efeitos de tamanho finito de nosso sistema. Uma forma de reduzir estes efeitos, é considerando um sistema com  $N \gg 10^4$ . Infelizmente, o aumento no número de neurônios afeta enormemente o tempo de simulação. Para contornar esse problema, poderia ser usado um código otimizado, como o desenvolvido por Penna e Oliveira (1989; 1990), para simular uma rede com um grande número de neurônios e assim diminuir os efeitos de tamanho finito. Nesse código, não é armazenado o valor dos  $J_{ij}$ , como usualmente é feito,

mas apenas os padrões  $\xi^\mu$  bit a bit em “palavras” inteiras de computador. A beleza do código reside em que o tempo de simulação da rede neuronal, dependerá apenas de operações XOR de bit a bit e do número de sub-redes, não interagentes, em que se dividiu a rede original. Os autores mostraram a surpreendente eficiência do código ao simular uma rede de 10 000 neurônios em um computador com CPU 80286 e 20 MHz, construído nos anos 80.

Apesar do descrito linhas acima, verificamos um aumento significativo na capacidade de reconhecimento da rede neuronal, de forma que nesta aplicação específica, considerando não haver correlações entre padrões armazenados, a faixa de valores de  $\alpha$  é ampliada por um fator  $10^2$ , tipicamente.

O bom desempenho da presente proposta também foi ilustrado em diversas situações, comuns no cotidiano dos seres vivos, conforme descrito a seguir. (i) Sua capacidade de reagir prontamente às mudanças no ambiente externo, reproduzindo uma característica fundamental dos seres vivos, que precisam reagir rapidamente aos novos padrões externos apresentados. (ii) O reconhecimento de padrões correlacionados, mostrando que as correlações levam a uma diminuição nos valores ótimos de  $\kappa_c$ , juntamente com um aumento nos máximos de  $\Delta m$ ; como consequência direta, as simulações numéricas indicam que o intervalo dos valores de  $\alpha$  deve ser aumentado, devido às correlações. Esses resultados estão de acordo com o reconhecimento de padrões semelhantes realizados por seres vivos, para os quais uma vez reconhecido um constituinte de um determinado grupo, todos os membros semelhantes desse grupo são identificados imediatamente, sendo necessário um pequeno estímulo externo para essa tarefa. (iii) Seu funcionamento para sinapses assimétricas e diluídas, mostrando que quando a diluição é incluída, maiores valores do estímulo externo são necessários.

Na RNDE, a capacidade de reagir às mudanças no ambiente externo, foi avaliada analisando as mudanças no valor da superposição  $m_\rho$ . Quando,  $\kappa = 0$ , nenhuma memória será reconhecida,  $m_\rho = 0$ . Se em um tempo  $t_1$ , um padrão com uma superposição  $\gamma_1$  ( $\kappa \neq 0$ ) com o padrão armazenado  $\{\xi^\rho\}$  for apresentada abruptamente, aparecerá uma superposição macroscópica  $m_\rho(t)$  com  $m_\rho \in [0, 2\gamma_1 - 1]$ . Já, se no tempo  $t_2$ , um padrão com uma superposição  $\gamma_2$  ( $\kappa \neq 0$ ) com o padrão armazenado  $\{\xi^\theta\}$  for apresentado ao sistema,  $m_\rho \rightarrow 0$  rapidamente, enquanto  $m_\theta \in [0, 2\gamma_2 - 1]$ . Isso foi verificado mediante simulações Monte Carlo de uma RNDE com  $\alpha = 0.8$ ,  $\gamma_1 = 0.8$  e  $\gamma_2 = 1.0$  para valores usuais de  $\kappa$ .

As simulações computacionais, de uma RNDE com um padrão  $\theta$ , que apresentam uma correlação  $b$  com o padrão  $\rho$ , ou seja,  $\xi_i^\rho = \xi_i^\theta$  para uma fração  $b$  dos índices  $i$  (considerando  $b = 0.8$  e  $\gamma = 1.0$ ), mostram que a aproximação  $\kappa_c \simeq \sqrt{\alpha}$  não é mais válida, nesses casos. Caso tenhamos uma RNDE com muitos padrões correlacionados, como esperado, se o número de padrões correlacionados com  $\{\xi^\rho\}$  aumentar, o valor de  $\kappa_c$  tende a diminuir.

Para uma RNDE diluída ( $\gamma = 1.0$ ) com  $\alpha = 0.5, 1.0$  e cujo parâmetro de diluição  $d$  varia entre 0.0 e 0.7, verificamos que; para dado um  $\alpha$  fixo, o valor de  $\kappa_c$  aumenta à medida que o parâmetro de diluição  $d$ , ou seja, as conexões sinápticas, diminuem. O valor de  $\kappa_c$ , obtido numericamente como o valor máximo de  $\Delta m$  é próximo do valor teórico,  $\kappa_c \simeq \sqrt{\frac{\alpha}{1-d}}$ .

Finalmente, é importante mencionar que este procedimento pode ser implementado em uma grande diversidade de modelos, que possuam sinal e ruído, tais como; redes neuronais oscilatórias, máquinas de Boltzmann e possíveis aplicações em redes neuronais diretas tais como as de aprendizagem profunda (*Deep Learning*) entre outros, utilizando tanto métodos analíticos como numéricos. Além disso, sua aplicação na análise de outras propriedades, distintas do reconhecimento de padrões, torna-se muito atraente.

A presente proposta pode contribuir para reduzir a distância entre os resultados de muitos desses modelos e as características comuns dos seres vivos reais, melhorando seu desempenho e, principalmente, levando a um aumento considerável em sua capacidade de reconhecimento e reação.

# Apêndice A

## Estudo analítico do modelo RNDE

### A.1 Teoria do Campo médio

Usa-se o método das réplicas, que se vale da identidade matemática

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

para efetuar a média configuracional do logaritmo da função de partição

$$\langle \ln Z_{(\beta, h)} \rangle_{\xi} = \lim_{n \rightarrow 0} \frac{1}{n} \left[ \langle Z_{(\beta, h)}^n \rangle_{\xi} - 1 \right] \quad (\text{A.2})$$

a energia livre por neurônio será então

$$\begin{aligned} f(\beta) &= -\frac{1}{\beta} \lim_{N \rightarrow \infty} \frac{1}{N} \langle \ln Z_{(\beta, h)} \rangle_{\xi}, \\ &= -\frac{1}{\beta} \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{nN} \left[ \langle Z_{(\beta, h)}^n \rangle_{\xi} - 1 \right]. \end{aligned}$$

Lembrando, que se considera a comutatividade nos limites  $n$  vezes, a qual é o produto de  $n$  funções de partição, correspondentes a  $n$  sistemas independentes (réplicas idênticas)

$$\begin{aligned} Z^n &= \prod_{a=1}^n Z_{(\{S_i^a\})} = \prod_{a=1}^n \left( \text{Tr}_{\{S_i^a\}} e^{-\beta H^a} \right), \\ &= \text{Tr}_{\{S_i^a\}} \left( \prod_{a=1}^n e^{-\beta H^a} \right) = \text{Tr}_{\{S_i^a\}} \left( e^{-\beta \sum_{a=1}^n H^a} \right). \end{aligned} \quad (\text{A.3})$$

A somatória na exponencial da equação anterior, representa a energia de um sistema de  $n \cdot N$  spins. O traço é feito sobre o conjunto total de variáveis de spin.

O hamiltoniano da  $a$ -ésima réplica,  $H^a$ , num sistema RNDE com padrões descorrelacionados é dado por:

$$\mathcal{H}^a = -\frac{1}{N} \sum_{(ij)} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu S_i^a S_j^a - \sum_{\nu=1}^s h^\nu \sum_{i=1}^N \xi_i^\nu S_i^a - \sum_{i=1}^N \kappa \eta_i S_i^a,$$

onde  $\eta_i$  obedece a seguinte distribuição de probabilidade

$$P(\eta_i) = \gamma \delta(\eta_i - \xi_i^\rho) + (1 - \gamma) \delta(\eta_i + \xi_i^\rho). \quad (\text{A.4})$$

Usando a identidade

$$\sum_{(ij)} \sum_{\mu=1}^p S_i^\mu S_j^\mu = \frac{1}{2} \sum_{\mu=1}^p \left( \sum_{i=1}^N S_i^\mu \right)^2 - \frac{1}{2} Np \quad (\text{A.5})$$

reescrevemos  $H^a$ , obtendo:

$$\mathcal{H}^a = -\frac{1}{2N} \sum_{\mu=1}^p \left( \sum_{i=1}^N \xi_i^\mu S_i^a \right)^2 + \frac{1}{2} p - \sum_{\nu=1}^s h^\nu \sum_{i=1}^N \xi_i^\nu S_i^a - \sum_{i=1}^N \kappa \eta_i S_i^a. \quad (\text{A.6})$$

Substituindo a equação anterior na função de partição  $n$  replicada  $Z^n$ , temos:

$$\mathcal{Z}^n = \text{Tr}_s \left[ \prod_{a=1}^n \exp \left( \frac{\beta}{2N} \sum_{\mu=1}^p \left( \sum_{i=1}^N \xi_i^\mu S_i^a \right)^2 - \frac{1}{2} \beta p + \beta \sum_{\nu=1}^s h^\nu \sum_{i=1}^N \xi_i^\nu S_i^a + \beta \sum_{i=1}^N \kappa \eta_i S_i^a \right) \right]. \quad (\text{A.7})$$

Fazemos  $C_{\mu a} = \frac{1}{N} \sum_i \xi_i^\mu S_i^a$  e  $f_a = \frac{1}{N} \sum_i \eta_i S_i^a$ , para melhorar a manipulação algébrica, obtendo-se:

$$\begin{aligned} \mathcal{Z}^n &= \text{Tr}_s \left[ \prod_{a=1}^n \exp \left( \frac{\beta N}{2} \sum_{\mu=1}^p C_{\mu a}^2 - \frac{1}{2} \beta p + \beta N \sum_{\nu=1}^s h^\nu C_{\nu a} + \beta N \kappa f_a \right) \right], \\ &= e^{-\beta p n / 2} \text{Tr}_s \left[ \prod_{a=1}^n \exp \left( \frac{\beta N}{2} \sum_{\mu=1}^p C_{\mu a}^2 + \beta N \sum_{\nu=1}^s h^\nu C_{\nu a} + \beta N \kappa f_a \right) \right], \\ &= e^{-\beta p n / 2} \text{Tr}_s \left[ \prod_{a=1}^n \left( \prod_{\mu=1}^p \exp \left( \frac{\beta N}{2} C_{\mu a}^2 \right) \cdot \exp \left( \beta N \sum_{\nu=1}^s h^\nu C_{\nu a} + \beta N \kappa f_a \right) \right) \right]. \end{aligned}$$

Para desacoplar os termos quadráticos presentes na equação anterior, se usa a identidade de Hubbard-Stratonovitch,

$$e^{\lambda a^2} = \int \frac{dx}{\sqrt{2\pi}} e^{-x^2/2 + \sqrt{2\lambda} a x}, \quad (\text{A.8})$$

obtendo-se:

$$\mathcal{Z}^n = e^{-\beta p n / 2} \text{Tr}_s \left[ \prod_{a=1}^n \left( \prod_{\mu=1}^p \int \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( -\frac{(y_a^\mu)^2}{2} + \sqrt{\beta N} y_a^\mu C_{\mu a} + \beta N \sum_{\nu=1}^s h^\nu C_{\nu a} + \beta N \kappa f_a \right) \right) \right]$$

Introduzindo na integral o produtório das réplicas, temos

$$Z^n = e^{-\beta pn/2} \text{Tr}_s \int_{-\infty}^{+\infty} \prod_{a=1}^n \prod_{\mu=1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \times \exp \left( -\frac{1}{2} \sum_{a=1}^n \sum_{\mu=1}^p (y_a^\mu)^2 + \sqrt{\beta N} \sum_{a=1}^n \sum_{\mu=1}^p y_a^\mu C_{\mu a} + \beta N \sum_{a=1}^n \sum_{\nu=1}^s h^\nu C_{\nu a} + \beta N \kappa \sum_{a=1}^n f_a \right)$$

É fácil fazer a média de  $Z^n$  na desordem, pois os expoentes da função de partição são funções lineares das variáveis aleatórias  $\xi_i^\mu$ ,  $\mu = s+1, \dots, p$ . Então, tomando a média sobre os padrões aleatórios  $\eta$  e  $\xi$  na equação anterior, obtemos

$$\begin{aligned} \langle Z^n \rangle &= \left\langle e^{-\beta pn/2} \text{Tr}_s \left[ \int \prod_{a=1}^n \prod_{\mu=1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( -\sum_{a=1}^n \sum_{\mu=1}^p \frac{(y_a^\mu)^2}{2} + \beta N \sum_{a=1}^n \sum_{\nu=1}^s h^\nu C_{\nu a} \right. \right. \right. \\ &\quad \left. \left. \left. + \sum_{a=1}^n \sum_{\mu=1}^p \left( \sqrt{\beta N} y_a^\mu + \beta N h^\nu \right) C_{\mu a} + \beta N \kappa \sum_{a=1}^n f_a \right) \right] \right\rangle_{\eta, \xi} \\ &= e^{-\beta pn/2} \text{Tr}_s \int \prod_{a=1}^n \prod_{\mu=1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( -\sum_{a=1}^n \sum_{\mu=s+1}^p \frac{(y_a^\mu)^2}{2} \right) \underbrace{\left\langle \exp \left( \sqrt{\beta N} \sum_{a=1}^n \sum_{\mu=s+1}^p y_a^\mu C_{\mu a} \right) \right\rangle_{\xi^\mu}}_{L_1} \\ &\quad \times \left\langle \exp \left( -\sum_{a=1}^n \sum_{\nu=1}^s \frac{(y_a^\nu)^2}{2} + \sum_{a=1}^n \sum_{\nu=1}^s \left( \sqrt{\beta N} y_a^\nu + \beta N h^\nu \right) C_{\nu a} + \beta N \kappa \sum_{a=1}^n f_a \right) \right\rangle_{\eta, \xi^\nu} \quad (\text{A.9}) \end{aligned}$$

A seguir, se calcula separadamente o termo  $L_1$ , que corresponde à média aleatória dos padrões

"altos"  $\{\xi_i^\mu\}$  ( $\mu = s+1, \dots, p$ ), como:

$$\begin{aligned} L_1 &= \left\langle \exp \left[ (\beta N)^{1/2} \sum_{a=1}^n \sum_{\mu=s+1}^p y_a^\mu C_{\mu a} \right] \right\rangle_{\xi^\mu}, \\ &= \left\langle \exp \left[ (\beta N)^{1/2} \sum_{a=1}^n \sum_{\mu=s+1}^p y_a^\mu \left( \frac{1}{N} \sum_i \xi_i^\mu S_i^a \right) \right] \right\rangle_{\xi^\mu}, \\ &= \prod_{\mu=s+1}^p \prod_{i=1}^N \left\langle \exp \left[ \left( \frac{\beta}{N} \right)^{1/2} \sum_{a=1}^n y_a^\mu S_i^a \xi_i^\mu \right] \right\rangle_{\xi^\mu}, \\ &= \prod_{\mu=s+1}^p \prod_{i=1}^N \int d\xi_i^\mu P(\xi_i^\mu) \exp \left( \sqrt{\frac{\beta}{N}} \sum_{a=1}^n y_a^\mu S_i^a \xi_i^\mu \right), \\ &= \prod_{\mu=s+1}^p \prod_{i=1}^N \cosh \left( \sqrt{\frac{\beta}{N}} \sum_{a=1}^n y_a^\mu S_i^a \right), \\ &= \exp \left[ \sum_{\mu=s+1}^p \sum_{i=1}^N \ln \left( \cosh \left( \sqrt{\frac{\beta}{N}} \sum_{a=1}^n y_a^\mu S_i^a \right) \right) \right]. \end{aligned}$$

onde foi usado o fato que  $\xi_i^\mu$  toma  $\pm 1$  com probabilidade  $1/2$ . Já que  $\ln [\cosh (x)] \approx \ln \left(1 + \frac{1}{2}x^2\right) \approx \frac{1}{2}x^2$ , quando  $x \rightarrow 0$ ,  $L_1$  se reduz a:

$$L_1 \approx \exp \left[ \sum_{\mu=s+1}^p \sum_{i=1}^N \frac{\beta}{2N} \left( \sum_{a=1}^n y_a^\mu S_i^a \right)^2 \right] = \exp \left( \sum_{\mu=s+1}^p \sum_{i=1}^N \frac{\beta}{2N} \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right). \quad (\text{A.10})$$

Substituindo em (A.9), o resultado obtido na equação (A.10), temos:

$$\begin{aligned} \langle Z^n \rangle &= e^{-\beta pm/2} \text{Tr}_s \int \prod_{a,\mu} \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( - \sum_{a=1}^n \sum_{\mu=s+1}^p \frac{(y_a^\mu)^2}{2} \right) \exp \left( \sum_{\mu=s+1}^p \sum_{i=1}^N \frac{\beta}{2N} \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right) \\ &\times \left\langle \exp \left( - \sum_{a=1}^n \sum_{\nu=1}^s \frac{(y_a^\nu)^2}{2} + \sum_{a=1}^n \sum_{\nu=1}^s \left( \sqrt{\beta N} y_a^\nu + \beta N h^\nu \right) C_{\nu a} + \beta N \kappa \sum_{a=1}^n f_a \right) \right\rangle_{\eta, \xi^\nu} \\ &= e^{-\beta pm/2} \text{Tr}_s \int \underbrace{\prod_{a,\mu} \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( - \sum_{a=1}^n \sum_{\mu=s+1}^p \frac{(y_a^\mu)^2}{2} + \frac{\beta}{2N} \sum_{\mu=s+1}^p \sum_{i=1}^N \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right)}_{L_2} \\ &\times \left\langle \exp \left( - \frac{1}{2} \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu)^2 + \sum_{a=1}^n \sum_{\nu=1}^s \left( \sqrt{\beta N} y_a^\nu + \beta N h^\nu \right) C_{\nu a} + \beta N \kappa \sum_{a=1}^n f_a \right) \right\rangle_{\eta, \xi^\nu} \quad (\text{A.11}) \end{aligned}$$

onde  $\langle \dots \rangle_{\xi^\nu}$  denota a média aleatória dos  $s$  “padrões condensados”  $\{\xi_i^\nu\}$ . Fazendo o cálculo da integral dos  $p - s$  termos quadráticos, denotado por  $L_2$ , separadamente, temos:

$$\begin{aligned} L_2 &= \int \prod_{a=1}^n \prod_{\mu=s+1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} \sum_{a=1}^n \sum_{\mu=s+1}^p (y_a^\mu)^2 + \frac{\beta}{2N} \sum_{\mu=s+1}^p \sum_{i=1}^N \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right), \\ &= \int \prod_{a=1}^n \prod_{\mu=s+1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left[ \sum_{\mu=s+1}^p \left( - \frac{1}{2} \sum_{a=1}^n (y_a^\mu)^2 + \frac{\beta}{2N} \sum_{i=1}^N \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu S_i^a S_i^b \right) \right], \\ &= \int \prod_{a=1}^n \prod_{\mu=s+1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left[ \sum_{\mu=s+1}^p \left( - \frac{1}{2} \sum_{a=1}^n \sum_{a=1}^n y_a^\mu y_b^\mu \delta_{ab} + \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n y_a^\mu y_b^\mu \left( \frac{\beta}{N} \sum_{i=1}^N S_i^a S_i^b \right) \right) \right], \end{aligned}$$

Introduzimos a matriz  $\mathbf{K}$  com elementos da forma  $K_{ab} = \delta_{ab} - \frac{\beta}{N} \sum_{i=1}^N S_i^a S_i^b$ , obtendo-se

$$\begin{aligned} L_2 &= \int \prod_{a=1}^n \prod_{\mu=s+1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left[ \sum_{\mu=s+1}^p \left( - \frac{1}{2} \sum_{a=1}^n \sum_{a=1}^n y_a^\mu y_b^\mu \underbrace{\left( \delta_{ab} - \frac{\beta}{N} \sum_{i=1}^N S_i^a S_i^b \right)}_{K_{ab}} \right) \right], \\ &= \int \prod_{a=1}^n \prod_{\mu=s+1}^p \frac{dy_a^\mu}{\sqrt{2\pi}} \exp \left[ \sum_{\mu=s+1}^p \left( - \frac{1}{2} \sum_{a=1}^n \sum_{a=1}^n y_a^\mu y_b^\mu K_{ab} \right) \right]. \end{aligned}$$

**Observação 1:** Integração da exponencial de uma forma quadrática de  $N$  variáveis  $\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\vec{X}^t \cdot C\vec{X}\right) d\vec{X}$  onde  $C$  é uma matriz simétrica definida positiva e  $\vec{X}^t = (X_1, X_2, \dots, X_N)$  é a transposta do vetor  $\vec{X}$ .

Seja  $U$  a matriz ortogonal que diagonaliza  $C$ , ou seja  $U^{-1}CU = D$ . Onde  $D$  é a matriz diagonal e  $U^{-1}\vec{X} = \vec{Y}$ . Então:

$$\begin{aligned} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\vec{X}^t \cdot C\vec{X}\right) d\vec{X} &= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\vec{X}^t \cdot (UU^{-1}CUU^{-1})\vec{X}\right) d\vec{X} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\vec{Y}^t \cdot D\vec{Y}\right) d\vec{Y} \\ &= \frac{(2\pi)^{N/2}}{(\det C)^{1/2}}. \end{aligned}$$

Como a matriz  $D$  é diagonal, a integral pode-se calcular como uma integral gaussiana de uma única variável para cada componente de  $\vec{Y}$  separadamente. Os elementos  $d_1, d_2, \dots, d_N$  da diagonal de  $D$  (autovalores de  $C$ ). Lembre-se que  $\det C = \prod_{j=1}^N d_j$

Usando a observação anterior, a expressão de  $L_2$ , toma a forma:

$$\begin{aligned} L_2 &= \prod_{\mu=s+1}^p \int_{-\infty}^{+\infty} \prod_{a=1}^n \frac{dy_a^\mu}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \sum_{a=1}^n \sum_{a=1}^n y_a^\mu y_b^\mu K_{ab}\right), \\ &= \left[ (2\pi)^{-n/2} \int_{-\infty}^{+\infty} \prod_{a=1}^n dy_a^\mu \exp\left(-\frac{1}{2} \sum_{a=1}^n \sum_{a=1}^n y_a^\mu y_b^\mu K_{ab}\right) \right]^{p-s}, \\ &= \left[ (2\pi)^{-n/2} \int_{-\infty}^{+\infty} d\vec{y}_a^\mu \exp\left(-\frac{1}{2} (\vec{y}_a^\mu)^T K_{ab} \vec{y}_b^\mu\right) \right]^{p-s}, \\ &= \left[ (2\pi)^{-n/2} \frac{(2\pi)^{n/2}}{(\det K)^{1/2}} \right]^{p-s} = (\det K)^{-(p-s)/2}. \end{aligned}$$

onde  $K_{ab} \equiv \delta_{ab} - \beta/N \sum_i S_i^a S_i^b$ . Para  $p$  grande, da ordem de  $N$ , usamos a aproximação  $p - s \approx p$

$$L_2 = (\det \mathbf{K})^{-p/2} = \exp\left[-\frac{p}{2} \text{Tr}(\ln \mathbf{K})\right].$$

**Observação 2:** No cálculo anterior, denota-se de forma explícita  $\text{Tr}(\ln \mathbf{K}) = \sum_{i=1}^n \ln(d_i)$  onde  $d_i$  são os autovalores de  $\mathbf{K}$ .

$$\begin{aligned}
 (\det \mathbf{K})^{-p/2} &= \prod_i (d_i)^{-p/2} = \exp\left(\sum_i \ln(d_i)^{-p/2}\right), \\
 &= \exp\left(-\frac{p}{2} \sum_i \ln(d_i)\right) = \exp\left(-\frac{p}{2} \text{Tr}(\ln D)\right), \\
 &= \exp\left[-\frac{p}{2} \text{Tr}(U^{-1}U \cdot \ln D)\right], \\
 &= \exp\left[-\frac{p}{2} \text{Tr}(U \cdot \ln D \cdot U^{-1})\right], \\
 &= \exp\left[-\frac{p}{2} \text{Tr}(\ln K)\right].
 \end{aligned}$$

Aqui, precisamos de um procedimento complicado para diagonalizar uma matriz  $K$ : de fato, será descoberto que  $K$  é uma das matrizes de réplica e sua diagonalização precisa de alguma concentração, em particular, se alguém aprofundar nas discussões de quebra de simetria de réplicas.

Usando propriedade da delta de Dirac (multidimensional), obtemos

$$\begin{aligned}
 L_2 &= \exp\left[-\frac{p}{2} \text{Tr}(\ln \mathbf{K})\right], \\
 &= \int \prod_{a \neq b} dq_{ab} \exp\left(-\frac{p}{2} \text{Tr}[\ln((1-\beta)\mathbf{I} - \beta\mathbf{Q})]\right) \cdot \prod_{a \neq b} \delta\left(q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b\right).
 \end{aligned}$$

Onde fizemos  $\mathbf{K} = (1-\beta)\mathbf{I} - \beta\mathbf{Q}$ , com  $\mathbf{I}$  denotando uma matriz unitária  $n$  dimensional. Um novo conjunto de variáveis auxiliares  $\mathbf{Q} = [q_{ab}]$  é introduzido, e define-se como:

$$q_{ab} = \begin{cases} \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b & a \neq b \\ 0 & a = b \end{cases} \quad (\text{A.12})$$

**Observação 3:** Função delta de Dirac multidimensional

$$\int f(\vec{x}) \delta^n(\vec{x} - \vec{x}_0) d\vec{x} = f(\vec{x}_0), \quad \text{onde} \quad \delta^n(\vec{x} - \vec{x}_0) = \prod_i \delta(x_i - x_{i0})$$

Além disso, ao introduzir uma representação formal da integral de Fourier da função delta:

$$\delta(x - x_0) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \exp(-r(x - x_0)) dr,$$

para  $n(n-1)\{q_{ab}\}$ , e desprezando o fator multiplicante, encontra-se:

$$\begin{aligned} \exp\left[-\frac{p}{2}\text{Tr}(\ln \mathbf{K})\right] &\propto \int \prod_{a \neq b} dq_{ab} \exp\left(-\frac{p}{2}\text{Tr} \ln \mathbf{K}\right) \cdot \prod_{a \neq b} \int dr_{ab} \exp\left[-r_{ab} \left(q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b\right)\right], \\ &\propto \int \prod_{a \neq b} dr_{ab} \int \prod_{a \neq b} dq_{ab} \exp\left[-\frac{p}{2}\text{Tr} \ln \mathbf{K} - \sum_{a \neq b} r_{ab} \left(q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b\right)\right]. \end{aligned}$$

Reescalando os  $r_{ab} \rightarrow \frac{N\alpha\beta^2}{2}r_{ab}$  para conveniência posterior, obtemos

$$\exp\left[-\frac{p}{2}\text{Tr} \ln \mathbf{K}\right] \propto \int \prod_{a \neq b} dr_{ab} \int \prod_{a \neq b} dq_{ab} \exp\left[-\frac{p}{2}\text{Tr} \ln \mathbf{K} - \frac{N\alpha\beta^2}{2} \sum_{a \neq b} r_{ab} \left(q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b\right)\right].$$

Substituindo a expressão anterior na equação (A.11), obtêm-se

$$\begin{aligned} \langle Z^n \rangle &\propto e^{-\beta p n/2} \text{Tr}_s \left[ \int \prod_{a=1}^n \prod_{\nu=1}^s \frac{dy_a^\nu}{\sqrt{2\pi}} \int \prod_{a \neq b} dr_{ab} \int \prod_{a \neq b} dq_{ab} \right. \\ &\quad \times \exp\left(-\frac{p}{2}\text{Tr} \ln((1-\beta)\mathbf{I} - \beta\mathbf{Q}) - \frac{N\alpha\beta^2}{2} \sum_{a \neq b} r_{ab} \left(q_{ab} - \frac{1}{N} \sum_{i=1}^N S_i^a S_i^b\right)\right) \\ &\quad \left. \times \left\langle \exp\left(-\frac{1}{2} \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu)^2 + \sum_{a=1}^n \sum_{\nu=1}^s (\sqrt{\beta N} y_a^\nu + \beta N h^\nu) C_{\nu a} + \beta N \kappa \sum_{a=1}^n f_a\right) \right\rangle_{\eta, \xi^\nu} \right] \end{aligned}$$

Depois de fazer o reescalamento  $y_a^\mu \rightarrow \sqrt{\beta N} y_a^\mu$ , as variáveis  $y_a^\mu$  se tornam da ordem de 1, garantindo a extensividade da energia livre. Assim, lembrando que  $C_{\mu a} = \frac{1}{N} \sum_i \xi_i^\mu S_i^a$  e  $f_a = \frac{1}{N} \sum_i \eta_i S_i^a$ , podemos escrever nossa expressão completa (A.11) para  $\langle Z^n \rangle$ , como

$$\begin{aligned} \langle Z^n \rangle &\propto e^{-\beta p n/2} \int \prod_{a,\nu} \frac{dy_a^\nu}{\sqrt{2\pi}} \int \prod_{a \neq b} dr_{ab} \int \prod_{a \neq b} dq_{ab} \\ &\quad \times \exp N\beta \left( -\frac{\alpha}{2\beta} \text{Tr} \ln((1-\beta)\mathbf{I} - \beta\mathbf{Q}) - \frac{\alpha\beta}{2} \sum_{a \neq b} r_{ab} q_{ab} - \sum_{a=1}^n \sum_{\nu=1}^s \frac{(y_a^\nu)^2}{2} \right) \\ &\quad \times \left\langle \underbrace{\text{Tr}_s \left[ \exp \sum_{i=1}^N \left( \frac{\alpha\beta^2}{2} \sum_{a \neq b} r_{ab} S_i^a S_i^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu + h^\nu) \xi_i^\nu S_i^a + \beta \kappa \sum_{a=1}^n \eta_i S_i^a \right) \right]}_{L_3} \right\rangle_{\eta, \xi^\nu}. \end{aligned}$$

A última linha da equação anterior ( $\equiv L_3$ ) pode ser representada como uma função dos padrões aleatórios,  $\{\xi_i^\nu\}$ , e as  $n$  réplicas dos spins binomiais  $\{S_i^a\}$ , e pode-se escrever de forma geral como:

$$\begin{aligned} L_3 &= \left\langle \text{Tr}_s \left[ \exp \sum_{i=1}^N F\{\xi_i^\nu, S_i^a\} \right] \right\rangle_{\eta, \xi^\nu} = \left\langle \prod_{i=1}^N \text{Tr}_{S_i^a} [\exp F\{\xi_i^\nu, S_i^a\}] \right\rangle_{\eta, \xi^\nu} \\ &= \left\langle \exp N \left( \frac{1}{N} \sum_{i=1}^N \ln \left( \text{Tr}_{S_i^a} [\exp F\{\xi_i^\nu, S_i^a\}] \right) \right) \right\rangle_{\eta, \xi^\nu}, \end{aligned} \quad (\text{A.13})$$

onde  $\text{Tr}_{S^a}$  indica o traço sobre  $n$  spins binomiais replicados  $\{S^a (= \pm 1)\}$ . Note-se que  $s \ll N$  (com  $N \rightarrow \infty$ ) garante que  $L_3$  tenha a propriedade de auto média (vide apêndice C) e seja representado geralmente como:

$$\frac{1}{N} \sum_{i=1}^N G(\xi_i^\nu) \implies \frac{1}{2^s} \sum_{\xi^\nu = \pm 1} G(\xi^\nu) \equiv \langle G(\xi^\nu) \rangle_{\eta, \xi^\nu},$$

onde consegue-se eliminar a dependência de  $i$  da expressão (A.13), ficando como

$$L_3 = \exp N \left\langle \ln \left( \text{Tr}_{S^a} [\exp F \{\xi^\nu, S^a\}] \right) \right\rangle_{\eta, \xi^\nu}. \quad (\text{A.14})$$

Nota-se que não precisamos da média  $\langle \dots \rangle_{\eta, \xi^\nu}$  em (A.13), porque a automediança da soma em  $i$  na equação (A.14) já realizou toda a média sobre os padrões.

Agora, pode-se escrever toda a expressão para  $\langle Z^n \rangle_{\eta, \xi^\nu}$  como uma integral da exponencial de algo proporcional a  $N$ :

$$\langle Z^n \rangle_{\eta, \xi^\nu} \propto \int \prod_{\nu a} dy_a^\nu \int \prod_{a \neq b} dq_{ab} dr_{ab} \exp(-N\beta\mathcal{F}\{y, q, r\}) \quad (\text{A.15})$$

onde

$$\mathcal{F}\{y, q, r\} = \frac{\alpha n}{2} + \sum_{a=1}^n \sum_{\nu=1}^s \frac{(y_a^\nu)^2}{2} + \frac{\alpha}{2\beta} \text{Tr} \ln((1 - \beta)\mathbf{I} - \beta\mathbf{Q}) + \frac{\alpha\beta}{2} \sum_{a \neq b} r_{ab} q_{ab} - \frac{1}{\beta} \langle \ln Z_0 \rangle_{\eta, \xi^\nu} \quad (\text{A.16})$$

e

$$Z_0 = \text{Tr}_{S^a} \left[ \exp \left( \frac{\alpha\beta^2}{2} \sum_{a \neq b} r_{ab} S^a S^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y_a^\nu + h^\nu) \xi^\nu S^a + \beta\kappa \sum_{a=1}^n \eta S^a \right) \right] \quad (\text{A.17})$$

## A.2 Soluções da simetria de réplicas

Como uma primeira aproximação, permite-se só um pequeno subconjunto dos parâmetros variacionais  $(y_a^\nu, q_{ab}, r_{ab})$  em que a todas as réplicas ou pares de réplicas, respectivamente, se lhes atribui os mesmos valores dos parâmetros, a qual é chamada de “simetria de réplicas” (SR).

$$y_i^\nu = y^\nu, \quad (\text{A.18})$$

$$q_{ab} = q, \quad a \neq b \quad (\text{A.19})$$

$$r_{ab} = r, \quad a \neq b. \quad (\text{A.20})$$

Então  $\mathcal{F}$  será minimizado em relação a estes três parâmetros apenas, além disso, o traço  $\text{Tr}[\ln((1-\beta)\mathbf{I} - \beta\mathbf{Q})]$  é fácil de calcular pois a matriz  $K = (1-\beta)\mathbf{I} - \beta\mathbf{Q}$  tem vetores próprios simples; um vetor com todas as componentes iguais e  $n-1$  vetores cujas componentes somam zero (vide apêndice E). A matriz  $K$  tem os seguintes autovalores:

$$\begin{cases} \text{um autovalor :} & 1 - \beta - (n-1)\beta q, \\ (n-1) \text{ autovalores :} & 1 - \beta + \beta q. \end{cases}$$

e, portanto,

$$\begin{aligned} \text{Tr} \ln [(1-\beta)\mathbf{I} - \beta\mathbf{Q}] &= (n-1) \ln(1-\beta+\beta q) + \ln(1-\beta-(n-1)\beta q) \\ &\simeq n \left[ \ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q} \right] \end{aligned} \quad (\text{A.21})$$

Assim, tendo em conta a equação anterior e depois do apropriado processo de limite na equação (A.16), a densidade de energia livre, pode ser representada usando os parâmetros de ordem de simetria de réplica. Detalhamos este processo a seguir

$$\begin{aligned} f_{SR} &= \lim_{n \rightarrow 0} \frac{\mathcal{F}}{n} \\ &= \frac{\alpha}{2} + \lim_{n \rightarrow 0} \left( \frac{n}{2n} \sum_{\nu=1}^s (y^\nu)^2 + \frac{\alpha}{2\beta n} \text{Tr} \ln [(1-\beta)\mathbf{I} - \beta\mathbf{Q}] + \frac{\alpha\beta}{2n} r q (n-1) n - \frac{1}{\beta n} \langle \ln Z_0 \rangle_{\eta, \xi^\nu} \right) \\ &= \frac{\alpha}{2} + \sum_{\nu=1}^s \frac{(y^\nu)^2}{2} - \frac{\alpha\beta}{2} r q + \frac{\alpha}{2\beta} \lim_{n \rightarrow 0} \left( \frac{1}{n} \text{Tr} [\ln [(1-\beta)\mathbf{I} - \beta\mathbf{Q}]] \right) - \lim_{n \rightarrow 0} \frac{1}{\beta n} \langle \ln Z_0 \rangle_{\eta, \xi^\nu} \\ &= \frac{\alpha}{2} + \sum_{\nu=1}^s \frac{(y^\nu)^2}{2} - \frac{\alpha\beta}{2} r q + \frac{\alpha}{2\beta} \left[ \ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q} \right] - \lim_{n \rightarrow 0} \frac{1}{\beta n} \langle \ln Z_0 \rangle_{\eta, \xi^\nu}. \end{aligned} \quad (\text{A.22})$$

Focando no último termo,  $Z_0$ . Um tem

$$\begin{aligned} Z_0 &= \text{Tr}_{s^a} \left[ \exp \left( \frac{\alpha\beta^2 r}{2} \sum_{a \neq b}^n S^a S^b + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y^\nu + h^\nu) \xi^\nu S^a + \beta\kappa \sum_{a=1}^n \eta S^a \right) \right] \\ &= \text{Tr}_{s^a} \left[ \exp \left( \frac{\alpha\beta^2 r}{2} \left( \left( \sum_{a=1}^n S^a \right)^2 - n \right) + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y^\nu + h^\nu) \xi^\nu S^a + \beta\kappa \sum_{a=1}^n \eta S^a \right) \right] \\ &= e^{-n\alpha\beta^2 r/2} \cdot \text{Tr}_{s^a} \left[ \exp \left( \frac{\alpha\beta^2 r}{2} \left( \sum_{a=1}^n S^a \right)^2 + \beta \sum_{a=1}^n \sum_{\nu=1}^s (y^\nu + h^\nu) \xi^\nu S^a + \beta\kappa \sum_{a=1}^n \eta S^a \right) \right] \end{aligned}$$

e nós usamos a identidade de Hubbard-Stratonovitch, mais uma vez para desacoplar o termo quadrático:

$$\exp \left[ \frac{\alpha\beta^2 r}{2} \left( \sum_{a=1}^n S^a \right)^2 \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz \exp \left[ -\frac{z^2}{2} + \sqrt{\alpha\beta^2 r} \left( \sum_{a=1}^n S^a \right) z \right], \quad (\text{A.23})$$

Após o desacoplamento dos termos quadráticos em  $Z_0$ , tomamos o traço  $\text{Tr}_{S^a}$ , obtendo-se

$$\begin{aligned} Z_0 &= e^{-n\alpha\beta^2 r/2} \frac{1}{\sqrt{2\pi}} \int dz \exp\left(-\frac{z^2}{2}\right) \text{Tr}_{S^a} \left[ \exp \beta \sum_{a=1}^n \left( z\sqrt{\alpha r} + \sum_{\nu=1}^s (y^\nu + h^\nu) \xi^\nu + \kappa\eta \right) S^a \right] \\ &= e^{-n\alpha\beta^2 r/2} \frac{1}{\sqrt{2\pi}} \int dz \exp\left(-\frac{z^2}{2}\right) \left( \text{Tr}_s \left[ \exp \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) S \right] \right)^n \\ &= e^{-n\alpha\beta^2 r/2} \frac{1}{\sqrt{2\pi}} \int dz \exp\left(-\frac{z^2}{2}\right) \left( 2 \cosh \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) \right)^n. \quad (\text{A.24}) \end{aligned}$$

onde  $\vec{m}$ ,  $\vec{h}$  e  $\vec{\xi}$  são vetores  $s$  dimensionais.

Tomando o logaritmo natural na equação (A.24) e como no limite  $n \rightarrow 0$ , a aproximação

$$\ln \int dz e^{-z^2/2} f^n(z) \approx n \int dz e^{-z^2/2} \ln(f^n(z)), \quad (\text{A.25})$$

pode ser usada, obtemos a seguinte expressão para  $\ln Z_0$ :

$$\begin{aligned} \ln Z_0 &= -\frac{n\alpha\beta^2 r}{2} + \ln \left[ \frac{1}{\sqrt{2\pi}} \int dz e^{-z^2/2} \cdot \left( 2 \cosh \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) \right)^n \right] \\ &= -\frac{n\alpha\beta^2 r}{2} + \frac{n}{\sqrt{2\pi}} \int dz e^{-z^2/2} \ln \left[ 2 \cosh \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) \right]. \end{aligned}$$

Usando a expressão anterior, a energia livre dada em (A.22), após o limite fica então, como

$$\begin{aligned} f_{SR} &= \frac{\alpha}{2} + \frac{1}{2} \vec{m}^2 + \frac{\alpha\beta}{2} r (1-q) + \frac{\alpha}{2\beta} \left[ \ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q} \right] \\ &\quad - \frac{1}{\beta} \left\langle \frac{1}{\sqrt{2\pi}} \int dz e^{-z^2/2} \ln \left[ 2 \cosh \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) \right] \right\rangle_{\eta, \xi^\nu} \end{aligned}$$

Lembrando que  $\eta_i$  obedece a distribuição  $P(\eta_i) = \gamma \delta(\eta_i - \xi_i^\rho) + (1-\gamma) \delta(\eta_i + \xi_i^\rho)$ , a densidade de energia livre pode-se reescrever da seguinte forma:

$$\begin{aligned} f_{SR} &= \frac{\alpha}{2} + \frac{\alpha\beta}{2} r (1-q) + \frac{1}{2} \vec{m}^2 + \frac{\alpha}{2\beta} \left[ \ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q} \right] \\ &\quad - \frac{1}{\beta} \left\langle \left\langle \ln \left[ 2 \cosh \beta \left( z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi} + \kappa\eta \right) \right] \right\rangle_{\eta} \right\rangle_z, \quad (\text{A.26}) \end{aligned}$$

onde  $\langle \dots \rangle_z \equiv \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \langle \dots \rangle_{\xi^\nu}$  representa uma média dupla: sobre os padrões recuperados  $\xi^\nu = \pm 1$ , e sobre o ruído multiplicativo gaussiano  $z$ .

### A.3 Equações de campo médio

As equações dos parâmetros de ordem, podem ser determinadas ao minimizar a energia livre  $f_{SR}$ . Para facilitar os cálculos algébricos introduzimos a variável  $\Xi = z\sqrt{\alpha r} + (\vec{m} + \vec{h}) \cdot \vec{\xi}$ .

De  $\partial f_{SR}/\partial m^\nu = 0$ , obtemos

$$\begin{aligned} \frac{\partial f_{SR}}{\partial m^\nu} = 0 &\longrightarrow 0 = m^\nu - \left\langle \left\langle \xi^\nu \tanh \beta (\Xi + \kappa \eta) \right\rangle_\eta \right\rangle_z \\ m^\nu &= \left\langle \left\langle \xi^\nu \tanh \beta (\Xi + \kappa \eta) \right\rangle_\eta \right\rangle_z. \end{aligned} \quad (\text{A.27})$$

Similarmente, de  $\partial f_{SR}/\partial q = 0$ :

$$\begin{aligned} \frac{\partial f_{SR}}{\partial q} = 0 &\longrightarrow 0 = 0 - \frac{\alpha\beta}{2}r + \frac{\alpha}{2\beta} \left[ \frac{\beta}{1 - \beta + \beta q} - \frac{\beta(1 - \beta + \beta q) - \beta q(\beta)}{(1 - \beta + \beta q)^2} \right] \\ r &= \frac{q}{(1 - \beta + \beta q)^2}, \end{aligned} \quad (\text{A.28})$$

e de  $\partial f_{SR}/\partial r = 0$ , obtemos:

$$\begin{aligned} \frac{\partial f_{SR}}{\partial r} = 0 &\longrightarrow 0 = 0 + \frac{\alpha\beta}{2}(1 - q) - \frac{1}{\beta} \left\langle \left\langle \tanh \beta (\Xi + \kappa \eta) \cdot \beta z \frac{\sqrt{\alpha}}{2\sqrt{r}} \right\rangle_\eta \right\rangle_z \\ \frac{\alpha\beta}{2}(1 - q) &= \frac{\sqrt{\alpha}}{2\sqrt{r}} \left\langle \left\langle z \tanh \beta (\Xi + \kappa \eta) \right\rangle_\eta \right\rangle_z \\ \sqrt{\alpha r} \beta (1 - q) &= \left\langle \left\langle z \tanh \beta (\Xi + \kappa \eta) \right\rangle_\eta \right\rangle_z. \end{aligned} \quad (\text{A.29})$$

Com objetivo de reescrever a média envolvida em (A.29), fazemos uma integração por partes:

$$\begin{aligned} \left\langle \left\langle z \tanh \beta (\Xi + \kappa \eta) \right\rangle_\eta \right\rangle_z &= \left\langle \left\langle \int \frac{dz}{\sqrt{2\pi}} \underbrace{\exp\left(-\frac{z^2}{2}\right)}_{v'} \cdot \underbrace{z \tanh \beta (\Xi + \kappa \eta)}_u \right\rangle_\eta \right\rangle_{\xi^\nu} \\ &\begin{cases} u = \tanh \beta (\Xi + \kappa \eta) &\rightarrow u' = \beta \sqrt{\alpha r} \operatorname{sech}^2 \beta (\Xi + \kappa \eta) \\ v' = z \exp\left(-\frac{z^2}{2}\right) &\rightarrow v = -\exp\left(-\frac{z^2}{2}\right) \end{cases} \end{aligned} \quad (\text{A.30})$$

Tendo em conta que  $\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} = 1$  e a definição da média  $\langle \dots \rangle_z$ , podemos reescrever a

equação anterior, obtendo-se

$$\begin{aligned}
\left\langle \left\langle z \tanh \beta (\Xi + \kappa \eta) \right\rangle_{\eta} \right\rangle_z &= \frac{1}{\sqrt{2\pi}} \left\langle -e^{-z^2/2} \cdot \tanh \beta (\Xi + \kappa \eta) \Big|_{-\infty}^{+\infty} \right. \\
&\quad \left. + \int e^{-z^2/2} \beta \sqrt{\alpha r} \operatorname{sech}^2 \beta (\Xi + \kappa \eta) dz \right\rangle_{\eta, \xi^\nu} \\
&= \beta \sqrt{\alpha r} \left\langle \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \cdot [1 - \tanh^2 \beta (\Xi + \kappa \eta)] \right\rangle_{\eta, \xi^\nu} \\
&= \beta \sqrt{\alpha r} \left[ 1 - \left\langle \left\langle \tanh^2 \beta (\Xi + \kappa \eta) \right\rangle_{\eta} \right\rangle_z \right]. \tag{A.31}
\end{aligned}$$

Substituindo (A.31) em (A.29), e após alguns cálculos analíticos, obtemos:

$$\begin{aligned}
\frac{\partial f_{SR}}{\partial r} = 0 &\longrightarrow \sqrt{\alpha r} \beta (1 - q) = \beta \sqrt{\alpha r} \left[ 1 - \left\langle \left\langle \tanh^2 \beta (\Xi + \kappa \eta) \right\rangle_{\eta} \right\rangle_z \right] \\
q - 1 &= -1 + \left\langle \left\langle \tanh^2 \beta (\Xi + \kappa \eta) \right\rangle_{\eta} \right\rangle_z \\
q &= \left\langle \left\langle \tanh^2 \beta (\Xi + \kappa \eta) \right\rangle_{\eta} \right\rangle_z \tag{A.32}
\end{aligned}$$

Recapitulando, obtiveram-se

$$m^\nu = \left\langle \left\langle \xi^\nu \tanh \beta \left( z \sqrt{\alpha r} + \left( \vec{m} + \vec{h} \right) \cdot \vec{\xi} + \kappa \eta \right) \right\rangle_{\eta} \right\rangle_z \tag{A.33}$$

$$r = \frac{q}{(1 - \beta + \beta q)^2} \tag{A.34}$$

$$q = \left\langle \left\langle \tanh^2 \beta \left( z \sqrt{\alpha r} + \left( \vec{m} + \vec{h} \right) \cdot \vec{\xi} + \kappa \eta \right) \right\rangle_{\eta} \right\rangle_z. \tag{A.35}$$

### A.3.1 Soluções em $T = 0$

Agora indo ao núcleo da teoria AGS em  $T = 0$  (baixos níveis de ruído). Tratando especificamente com a solução para a chamada *fase de recuperação ferromagnética (FMR)*, correspondente ao caso em que  $m^\nu \sim$

$$\begin{cases} 1 & \nu = \rho \\ 0 & \nu \neq \rho \end{cases} \quad (\text{nos representaremos ele como } m).$$

Lembrando-se que  $P(\eta) = \gamma \delta(\eta - \xi^\rho) + (1 - \gamma) \delta(\eta + \xi^\rho)$ , e assumindo  $\vec{h} = 0$ , reescrevemos as médias envolvidas nas equações (A.33) e (A.35) denotando o argumento da  $\tanh$  na FMR com  $\Theta = z \sqrt{\alpha r} + m \xi^\rho \pm \kappa \xi^\rho$ .

Introduzindo na integral, a média presente na equação (A.33), temos:

$$\begin{aligned} \langle \xi^\nu \tanh \beta \Theta \rangle_z &= \left\langle \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \cdot \xi^\rho \tanh \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \right\rangle_{\xi^\nu} \\ &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \langle \xi^\rho \tanh \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \rangle_{\xi^\nu} \end{aligned}$$

Perceba que  $\int P(\xi^\nu) d\xi^\nu = 1$ ,  $\forall \nu \neq \rho$  e já que  $\xi^2 = 1$ , a expressão anterior pode ser reescrita da forma

$$\begin{aligned} \langle \xi^\nu \tanh \beta \Theta \rangle_z &\stackrel{(\xi^2=1)}{=} \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \xi^\rho \tanh \beta (z(\xi^\rho)^2 \sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \xi^\rho \tanh \beta \xi^\rho (z\xi^\rho \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \end{aligned}$$

como  $\tanh$  é uma função ímpar, cumpre-se  $\tanh(\xi x) = \xi \tanh(x)$ , a média pode ser escrita como

$$\begin{aligned} \langle \xi^\nu \tanh \beta \Theta \rangle_z &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \cancel{(\xi^\rho)^2}^1 \tanh \beta (z\xi^\rho \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \tanh \beta (z\xi^\rho \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \end{aligned}$$

Fazendo a mudança de variável  $w = z\xi^\rho \rightarrow dw = \xi^\rho dz$  na integral anterior, tendo em conta que  $\xi^\rho = \pm 1$  e os limites de integração mudam com o sinal de  $\xi^\rho$ , temos

$$\begin{aligned} \langle \xi^\nu \tanh \beta \Theta \rangle_z &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \left( \int P(\xi^\rho) \tanh \beta (w\sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \left( \tanh \beta (w\sqrt{\alpha r} + m \pm \kappa) \int P(\xi^\rho) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot \tanh \beta (w\sqrt{\alpha r} + m \pm \kappa). \end{aligned} \quad (\text{A.36})$$

Tomando o limite  $T \rightarrow 0$  (ou  $\beta \rightarrow \infty$ ) da expressão (A.36)

$$\begin{aligned} \langle \xi^\nu \tanh \beta \Theta \rangle_z &\stackrel{\beta \rightarrow \infty}{\simeq} \int \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot \text{sgn}(w\sqrt{\alpha r} + m \pm \kappa) \\ &\approx \int \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot [\text{sgn}(w\sqrt{\alpha r} + m \pm \kappa) + 1 - 1] \\ &= \int \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot [\text{sgn}(w\sqrt{\alpha r} + m \pm \kappa) + 1] - 1. \end{aligned} \quad (\text{A.37})$$

Usando a definição da função sinal,

$$\text{sgn}(w\sqrt{\alpha r} + m \pm \kappa) = \begin{cases} +1 & w\sqrt{\alpha r} + m \pm \kappa > 0 \\ -1 & w\sqrt{\alpha r} + m \pm \kappa < 0 \end{cases} \quad \text{ou equivalentemente} \quad \begin{cases} +1 & w > -\frac{m \pm \kappa}{\sqrt{\alpha r}} \\ -1 & w < -\frac{m \pm \kappa}{\sqrt{\alpha r}} \end{cases} \quad (\text{A.38})$$

a integral do lado direito de (A.37), toma a seguinte forma:

$$\begin{aligned}
 \langle \xi^\nu \tanh \beta \Theta \rangle_z &\approx -1 + \int_{-\infty}^{-\frac{m \pm \kappa}{\sqrt{\alpha r}}} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot [\operatorname{sgn}(w\sqrt{\alpha r} + m \pm \kappa) + 1] \\
 &\quad + \int_{-\frac{m \pm \kappa}{\sqrt{\alpha r}}}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot [\operatorname{sgn}(w\sqrt{\alpha r} + m \pm \kappa) + 1] \\
 &\approx -1 + 2 \int_{-\frac{m \pm \kappa}{\sqrt{\alpha r}}}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2}. \tag{A.39}
 \end{aligned}$$

Fazemos uso da seguinte troca de variável

$$x = \frac{w}{\sqrt{2}} \rightarrow dx = \frac{dw}{\sqrt{2}} \begin{cases} w = -\frac{m \pm \kappa}{\sqrt{\alpha r}} \rightarrow x = -\frac{m \pm \kappa}{\sqrt{2\alpha r}} \\ w \rightarrow \infty \rightarrow x \rightarrow \infty \end{cases}$$

no resultado obtido em (A.39), a média toma a seguinte forma

$$\begin{aligned}
 \langle \xi^\nu \tanh \beta \Theta \rangle_z &\approx -1 + 2 \int_{-\frac{m \pm \kappa}{\sqrt{2\alpha r}}}^{+\infty} \frac{dx}{\sqrt{\pi}} e^{-x^2} \\
 &\approx \frac{2}{\sqrt{\pi}} \int_{-\frac{m \pm \kappa}{\sqrt{2\alpha r}}}^0 dx e^{-x^2} + \frac{2}{\sqrt{\pi}} \int_0^{+\infty} dx e^{-x^2} - 1 \\
 &\approx \frac{2}{\sqrt{\pi}} \int_{-\frac{m \pm \kappa}{\sqrt{2\alpha r}}}^0 dx e^{-x^2}
 \end{aligned}$$

devido à simetria da função, escrevemos:

$$\langle \xi^\nu \tanh \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \rangle_z \approx \frac{2}{\sqrt{\pi}} \int_0^{\frac{m \pm \kappa}{\sqrt{2\alpha r}}} dx e^{-x^2}$$

e como  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dz e^{-z^2}$ , escrevemos a expressão para a média, como

$$\langle \xi^\nu \tanh \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \rangle_z \approx \operatorname{erf}\left(\frac{m \pm \kappa}{\sqrt{2\alpha r}}\right). \tag{A.40}$$

De forma análoga ao cálculo anterior, desenvolvemos uma expressão para a média presente na equação (A.35). Considera-se novamente  $\Theta = z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho$ . Obtendo-se

$$\begin{aligned}
 \langle \tanh^2 \beta \Theta \rangle_z &= \left\langle \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \cdot \tanh^2 \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \right\rangle_{\xi^\nu} \\
 &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \langle \tanh^2 \beta (z\sqrt{\alpha r} + m\xi^\rho \pm \kappa\xi^\rho) \rangle_{\xi^\nu},
 \end{aligned}$$

Como  $\int P(\xi^\nu) d\xi^\nu = 1$ ,  $\forall \nu \neq \rho$  e já que  $\xi^2 = 1$ , reescrevemos a expressão anterior da forma

$$\begin{aligned} \langle \tanh^2 \beta \Theta \rangle_z &\stackrel{(\xi^2=1)}{=} \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \tanh^2 \beta (z(\xi^\rho)^2 \sqrt{\alpha r} + m \xi^\rho \pm \kappa \xi^\rho) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \tanh^2 \beta \xi^\rho (z \xi^\rho \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \end{aligned}$$

como a função  $\tanh^2$  é par, usamos a propriedade  $\tanh^2(\xi x) = \tanh^2(x)$ , obtendo-se

$$\langle \tanh^2 \beta \Theta \rangle_z = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left( \int P(\xi^\rho) \tanh^2 \beta (z \xi^\rho \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right).$$

Fazendo uso da mudança de variável  $w = z \xi^\rho \rightarrow dw = \xi^\rho dz$  na integral da equação anterior, tendo em conta que  $\xi^\rho = \pm 1$  e que os limites de integração mudam de acordo ao sinal de  $\xi^\rho$ , temos

$$\begin{aligned} \langle \tanh^2 \beta \Theta \rangle_z &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \left( \int P(\xi^\rho) \tanh^2 \beta (w \sqrt{\alpha r} + m \pm \kappa) d\xi^\rho \right) \\ &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \left( \tanh^2 \beta (w \sqrt{\alpha r} + m \pm \kappa) \int P(\xi^\rho) d\xi^\rho \right) \quad \xrightarrow{1} \\ &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \tanh^2 \beta (w \sqrt{\alpha r} + m \pm \kappa) \end{aligned} \quad (\text{A.41})$$

Usando a seguinte relação  $1 - \tanh^2 x = \text{sech}^2 x$ , a integral em (A.41), toma a forma

$$\begin{aligned} \langle \tanh^2 \beta \Theta \rangle_z &= \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot [\tanh^2 \beta (w \sqrt{\alpha r} + m \pm \kappa) - 1 + 1] \\ &= 1 - \int \frac{dw}{\sqrt{2\pi}} e^{-w^2/2} \cdot \text{sech}^2 \beta (w \sqrt{\alpha r} + m \pm \kappa). \end{aligned} \quad (\text{A.42})$$

Se fizermos a seguinte substituição algébrica,

$$x = \beta (w \sqrt{\alpha r} + m \pm \kappa) \rightarrow dx = \beta \sqrt{\alpha r} dw \text{ e } w = \frac{x - (m \pm \kappa) \beta}{\beta \sqrt{\alpha r}}$$

no lado direito da equação (A.42), obtemos

$$\begin{aligned} \langle \tanh^2 \beta \Theta \rangle_z &= 1 - \frac{1}{\sqrt{2\pi}} \int \frac{dx}{\beta \sqrt{\alpha r}} \exp \left[ -\frac{(x - (m \pm \kappa) \beta)^2}{2\beta^2 \alpha r} \right] \text{sech}^2(x) \\ &= 1 - \frac{1}{\beta \sqrt{2\pi \alpha r}} \int dx \exp \left[ -\frac{x^2 - 2x(m \pm \kappa) \beta}{2\beta^2 \alpha r} \right] \exp \left[ -\frac{(m \pm \kappa)^2 \beta^2}{2\beta^2 \alpha r} \right] \text{sech}^2(x) \\ &= 1 - \frac{1}{\beta \sqrt{2\pi \alpha r}} \exp \left[ -\frac{(m \pm \kappa)^2}{2\alpha r} \right] \int dx \exp \left[ -\frac{x^2 - 2x(m \pm \kappa) \beta}{2\beta^2 \alpha r} \right] \text{sech}^2(x) \end{aligned}$$

No limite  $\beta \rightarrow \infty$ , é possível considerar a seguinte aproximação  $\exp\left[-\frac{x^2-2x(m\pm\kappa)\beta}{2\beta^2\alpha r}\right] \rightarrow 1$ , em consequência a média toma a seguinte forma:

$$\begin{aligned} \langle \tanh^2 \beta \Theta \rangle_z &\approx 1 - \frac{1}{\beta\sqrt{2\pi\alpha r}} \exp\left[-\frac{(m\pm\kappa)^2}{2\alpha r}\right] \int dx \operatorname{sech}^2(x) \\ &\approx 1 - \frac{1}{\beta\sqrt{2\pi\alpha r}} \exp\left[-\frac{(m\pm\kappa)^2}{2\alpha r}\right] \cdot \tanh(x) \Big|_{-\infty}^{+\infty} \\ &\approx 1 - \frac{1}{\beta} \sqrt{\frac{2}{\pi\alpha r}} \exp\left[-\frac{(m\pm\kappa)^2}{2\alpha r}\right] \end{aligned} \quad (\text{A.43})$$

Com ajuda dos resultados obtidos em (A.40) e (A.43), reescrevemos as equações de campo médio (A.33), (A.35) e (A.34) em  $T = 0$ ,

$$m = \gamma \operatorname{erf}\left(\frac{m+\kappa}{\sqrt{2\alpha r}}\right) + (1-\gamma) \operatorname{erf}\left(\frac{m-\kappa}{\sqrt{2\alpha r}}\right) \quad (\text{A.44})$$

$$r = \frac{q}{(1-\beta + \beta q)^2} \quad (\text{A.45})$$

$$q = 1 - \frac{1}{\beta} \sqrt{\frac{2}{\pi\alpha r}} \left[ \gamma e^{-(m+\kappa)^2/2\alpha r} + (1-\gamma) e^{-(m-\kappa)^2/2\alpha r} \right] \quad (\text{A.46})$$

É possível condensar as três equações anteriores em só duas. Para isso, se substitui (A.46) na equação (A.45) e tendo em conta que  $\beta \rightarrow \infty$  em um sistemas sem ruído, temos

$$r \left( 1 - \sqrt{\frac{2}{\pi\alpha r}} \left[ \gamma e^{-(m+\kappa)^2/2\alpha r} + (1-\gamma) e^{-(m-\kappa)^2/2\alpha r} \right] \right)^2 \simeq 1.$$

Fazendo a troca de variáveis  $m = y\sqrt{2\alpha r}$  e  $\kappa = x\sqrt{2\alpha r}$  na equação anterior, e tomando a raiz quadrada, acha-se a seguinte expressão:

$$\sqrt{r} - \sqrt{\frac{2}{\pi\alpha}} \left[ \gamma e^{-(y+x)^2} + (1-\gamma) e^{-(y-x)^2} \right] \approx 1. \quad (\text{A.47})$$

Usando essa mesma troca de variáveis na equação (A.44), tem-se

$$y\sqrt{2\alpha r} = \gamma \operatorname{erf}(y+x) + (1-\gamma) \operatorname{erf}(y-x) \quad (\text{A.48})$$

Isolando  $\sqrt{r}$  na equação (A.48) e substituindo a expressão resultante em (A.47), temos

$$\frac{\gamma \operatorname{erf}(y+x) + (1-\gamma) \operatorname{erf}(y-x)}{y\sqrt{2\alpha}} - \sqrt{\frac{2}{\pi\alpha}} \left[ \gamma e^{-(y+x)^2} + (1-\gamma) e^{-(y-x)^2} \right] = 1,$$

da qual, isolando  $y$  obtemos finalmente a equação

$$y = \frac{\gamma \operatorname{erf}(y+x) + (1-\gamma) \operatorname{erf}(y-x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} [\gamma e^{-(y+x)^2} + (1-\gamma) e^{-(y-x)^2}]}$$

Usando adequadamente  $m = y\sqrt{2\alpha r}$  e  $\kappa = x\sqrt{2\alpha r}$  em (A.44), obtemos

$$m = \gamma \operatorname{erf}(y+x) + (1-\gamma) \operatorname{erf}(y-x)$$

Observa-se destas duas equações, os seguintes casos particulares:

- Ao fazer  $\gamma = 1$ , obtêm-se

$$\begin{aligned} y &= \frac{\operatorname{erf}(y+x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-(y+x)^2}} \\ m &= \operatorname{erf}(y+x), \end{aligned}$$

recupera-se a forma das equações inicialmente obtidas no modelo AGS, que correspondem ao caso do reconhecimento de um padrão previamente armazenado.

- Se além de  $\gamma = 1$ , consideramos que  $y+x \rightarrow x$ , obtemos

$$\begin{aligned} y &= \frac{\operatorname{erf}(x)}{\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-x^2}} \\ m &= \operatorname{erf}(x), \end{aligned}$$

recuperando-se as equações no caso do reconhecimento de um padrão não armazenado (ortogonal).

Para que as equações anteriores sejam idênticas as expressões originais, basta fazer  $\kappa = h$  (ou  $x = h/\sqrt{2\alpha r}$ ).

# Apêndice B

## Algoritmo da RNDE

A seguir, apresentamos uma versão resumida do algoritmo usado na simulação da RNDE.

---

**Algoritmo 1:** Pseudocódigo de uma RNDE

---

**Dados:**  $\kappa, \alpha, \gamma, N, Nexp, Passos$

**Resultado:** Superposições  $m_\rho, m_\perp, \Delta m$

1 **início**

2     **para**  $k \leftarrow 1$  **até**  $Nexp$  **faça**

3         Escolher  $S(0), \xi^\mu, \eta$ ;

4         Calcular  $J_{ij}$ ;

5         **para**  $t \leftarrow 1$  **até**  $Passos$  **faça**

6             **para**  $i \leftarrow 1$  **até**  $N$  **faça**

7                 Calcular campo  $h(i)$ ;

8                 **se**  $h_i(t) > 0$  **então**

9                      $S_i(t) = +1$ ;

10                 **senão**

11                      $S_i(t) = -1$ ;

12                 **fim**

13             **fim**

14         **fim**

15     **fim**

16 **fim**

---

# Apêndice C

## Auto média

Extensos estudos sobre sistemas aleatórios temperados têm sido feitos na mecânica estatística (Ver, por exemplo Brout, 1959; Monari et al., 1999). Um problema importante que surge em tais sistemas é a questão da auto-média. Em termos gerais, uma propriedade ( $X$ ) de um sistema é auto-média, se a “maioria” das realizações da aleatoriedade tiver o mesmo valor de  $X$  no limite termodinâmico. Uma das definições de auto-média mais usada em estudos numéricos é a seguinte. Diz-se que o sistema é auto-mediável em relação à propriedade  $X$  se

$$\frac{\langle X_n^2 \rangle - \langle X_n \rangle^2}{\langle X_n \rangle^2} \rightarrow 0, \quad (\text{C.1})$$

quando o tamanho ( $n$ ) do sistema vai para o infinito.

Aqui, os colchetes angulares  $\langle \dots \rangle$ , indicam médias sobre as realizações da aleatoriedade temperada do sistema e  $X_n$  é o valor da propriedade  $X$  quando o sistema tem tamanho  $n$  (Vide Binder e Young, 1986; Wiseman e Domany, 1998).

O fato de que a energia livre possui a propriedade de auto-média foi provado para diversos modelos, como; sistemas de spin com interações de curto alcance (van Hemmen e Palmer, 1982) e de longo alcance (van Enter e van Hemmen, 1983), para modelos de auto deslocamento de adsorção de polímero (Orlandini et al., 1999) e localização de polímero em uma interface (Van Rensburg et al., 2001).

### C.1 Auto-média da energia livre

Note-se que é possível escrever a soma sobre sítios dentro do expoente em (A.13), como

$$\frac{1}{N} \sum_i G(\xi_i) = \frac{1}{N} \sum_{\{\xi\}} \phi(\xi) G(\xi), \quad (\text{C.2})$$

onde  $\sum_{\{\xi\}}$ , indica uma soma sobre todas as possíveis configurações que podemos ter de  $\xi^\mu$ ,  $\mu = 1, \dots, p$ , e a função  $\phi(\xi)$  indica o número de vezes que uma destas configurações aparece na soma em  $i$ . A probabilidade de ocorrer uma destas configurações é

$$P(\xi) = \frac{1}{2^p}, \quad (\text{C.3})$$

e a probabilidade de a configuração escolhida ocorrer  $n$  vezes, ou seja  $\phi(\xi) = n$  é

$$P(\xi, n) = \frac{N!}{(N-n)!n!} \left(\frac{1}{2^p}\right)^n \left(1 - \frac{1}{2^p}\right)^{N-n}. \quad (\text{C.4})$$

Assim, o valor médio e a dispersão de  $n$  são, respetivamente

$$\langle n \rangle = \frac{N}{2^p}$$

e

$$\langle n^2 \rangle - \langle n \rangle^2 = \frac{N}{2^p} \left(1 - \frac{1}{2^p}\right),$$

de modo que, no limite  $N \rightarrow \infty$ , temos

$$\frac{\sqrt{\langle n^2 \rangle - \langle n \rangle^2}}{\langle n \rangle} \rightarrow 0.$$

Este resultado indica que neste limite, esta distribuição é fortemente centrada em torno de seu valor médio o que nos permite escrever

$$\frac{1}{N} \sum_i G(\xi_i) \simeq \frac{1}{2^p} \sum_{\{\xi\}} G(\xi) = \langle G(\xi) \rangle_\xi. \quad (\text{C.5})$$

Nesta situação, dizemos que as variáveis  $\xi_i^\mu$  são auto-mediáveis, pois sua soma sobre sítios é equivalente à média sobre as configurações. Observe-se que a condição para que isto ocorra é que

$$\lim_{N \rightarrow \infty} \frac{2^p}{N} = 0, \quad (\text{C.6})$$

que é justamente o caso se  $p$  é finito.

# Apêndice D

## O ruído no modelo de Hopfield

Ao considerar no modelo de Hopfield, como regra de aprendizado a prescrição de Hebb (1949), dada por  $J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu$ , o campo molecular que age no neurônio  $i$ , na ausência de campo externo, terá a forma

$$\begin{aligned} h_i(\xi^\nu) &= \frac{N-1}{N} \xi_i^\nu + \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \\ &= S + R. \end{aligned}$$

O termo  $S$  é chamado de sinal e o termo  $R$ , que envolve a soma dupla, de ruído. Observe que o estado  $\xi_i^\nu$  pode ser desestabilizado se o termo de ruído se torna maior que o termo sinal.

Nós assumimos que os padrões memorizados são aleatórios e descorrelacionados (Vide Amit, 1992, pág. 175). A distribuição de probabilidade das componentes de  $\xi^\nu$  é dada por

$$P(\xi_i^\mu) = \frac{1}{2} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)]. \quad (\text{D.1})$$

A seguir, calculamos o primeiro e segundo momento do termo de ruído,

$$R = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (\text{D.2})$$

Todas as médias envolvidas, serão realizadas sobre todos os padrões aprendidos.

Ao calcular o primeiro momento, temos

$$\langle R \rangle = \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \rangle_{\{\xi_j\}}, \quad (\text{D.3})$$

onde  $\langle \dots \rangle_{\{\xi_j\}}$  indica uma média sobre todos os padrões  $\{\xi_j\}$ .

Fazendo uso de (D.1), é possível demonstrar a seguinte expressão:

$$\langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \rangle_{\{\xi_j\}} = \xi_i^\nu \delta_{\mu\nu}, \quad (\text{D.4})$$

que nos permite, reduzir a soma dupla presente na equação (D.3), obtendo-se

$$\begin{aligned} \langle R \rangle &= \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\nu \delta_{\mu\nu} \\ &= 0. \end{aligned} \quad (\text{D.5})$$

O segundo momento do termo de ruído, para  $p$  padrões aprendidos, é dado por:

$$\begin{aligned} \langle R^2 \rangle &= \left\langle \frac{1}{N} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \xi_i^\mu \xi_j^\mu \xi_j^\nu \times \frac{1}{N} \sum_{k \neq i}^N \sum_{\sigma \neq \nu}^p \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \right\rangle \\ &= \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{k \neq i}^N \sum_{\sigma \neq \nu}^p \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \rangle_{\{\xi_j, \xi_k\}}, \end{aligned} \quad (\text{D.6})$$

onde  $\langle \dots \rangle_{\{\xi_j, \xi_k\}}$  representa uma média sobre os padrões  $\{\xi_j\}$  e  $\{\xi_k\}$ . Ao separar o somatório para  $k = j$  e  $k \neq j$ , na equação anterior, obtêm-se:

$$\langle R^2 \rangle = \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \xi_i^\sigma \xi_j^\sigma \xi_j^\nu \rangle_{\{\xi_j\}} + \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{k \neq i, j}^N \sum_{\sigma \neq \nu}^p \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \rangle_{\{\xi_j, \xi_k\}}. \quad (\text{D.7})$$

Novamente, com o auxílio de (D.1), é possível demonstrar a seguinte expressão:

$$\langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \rangle_{\{\xi_j, \xi_k\}} = \begin{cases} \xi_i^\mu \xi_i^\sigma \delta_{\mu\sigma} & j = k \\ \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \rangle_{\{\xi_j\}} \langle \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \rangle_{\{\xi_k\}} & j \neq k. \end{cases} \quad (\text{D.8})$$

Usando (D.8) é possível reescrever os somatórios do lado direito de (D.7), da seguinte forma:

$$\langle R^2 \rangle = \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\mu \xi_i^\sigma \delta_{\mu\sigma} + \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \langle \xi_i^\mu \xi_j^\mu \xi_j^\nu \rangle_{\{\xi_j\}} \cdot \sum_{k \neq i, j}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \langle \xi_i^\sigma \xi_k^\sigma \xi_k^\nu \rangle_{\{\xi_k\}} \quad (\text{D.9})$$

Com auxílio de (D.4), obtemos a seguinte expressão:

$$\begin{aligned} \langle R^2 \rangle &= \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\mu \xi_i^\sigma \delta_{\mu\sigma} + \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\nu \delta_{\mu\nu} \cdot \sum_{k \neq i, j}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\nu \delta_{\sigma\nu} \\ &= \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p (\xi_i^\mu)^2 + \frac{1}{N^2} \sum_{j \neq i}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\nu \delta_{\mu\nu} \cdot \sum_{k \neq i, j}^N \sum_{\mu \neq \nu}^p \sum_{\sigma \neq \nu}^p \xi_i^\nu \delta_{\sigma\nu} \\ &= \frac{(N-1)(p-1)}{N^2}. \end{aligned}$$

---

Assim, variância do termo de ruído será:

$$\begin{aligned}\sigma_R^2 &= \langle R^2 \rangle - \langle R \rangle^2 \\ &= \frac{(N-1)p}{N^2}.\end{aligned}$$

A correspondente desvio padrão, no limite termodinâmico ( $N \rightarrow \infty$ ), toma a forma

$$\sigma_R = \sqrt{\frac{(N-1)p}{N^2}} \approx \sqrt{\frac{p}{N}} \quad (\text{D.10})$$

# Apêndice E

## Autovalores e autovetores de $\mathbf{K}$

A matriz  $\mathbf{K} = [K_{ab}]$  com elementos da forma

$$K_{ab} \equiv \delta_{ab} - \beta/N \sum_i S_i^a S_i^b, \quad (\text{E.1})$$

sob a condição de simetria de réplicas, toma a seguinte forma simples

$$\mathbf{K} = (1 - \beta) \mathbf{I} - \beta \mathbf{Q} = \begin{pmatrix} 1 - \beta & -\beta q & \dots & -\beta q \\ -\beta q & 1 - \beta & \dots & -\beta q \\ \vdots & \vdots & \ddots & \vdots \\ -\beta q & -\beta q & \dots & 1 - \beta \end{pmatrix}. \quad (\text{E.2})$$

Devido a simplicidade desta matriz, é possível calcular seus autovalores e autovetores.

### E.1 Autovalores

Com objetivo de calcular os autovalores desta matriz, fazemos uso da definição da equação característica,

$$|\mathbf{K} - \lambda \mathbf{I}| = 0 \rightarrow \left| \begin{pmatrix} 1 - \beta - \lambda & -\beta q & \dots & -\beta q \\ -\beta q & 1 - \beta - \lambda & & -\beta q \\ \vdots & & \ddots & \vdots \\ -\beta q & -\beta q & \dots & 1 - \beta - \lambda \end{pmatrix} \right| = 0.$$

Para o cálculo do determinante da matriz envolvida na equação anterior, se faz uso das seguintes operações elementares:

1. Somam-se todas as colunas da matriz à coluna 1, e obtemos

$$\begin{aligned}
 |\mathbf{K} - \lambda\mathbf{I}| &= \left| \begin{pmatrix} 1 - \beta - (n-1)\beta q - \lambda & -\beta q & \dots & -\beta q \\ 1 - \beta - (n-1)\beta q - \lambda & 1 - \beta - \lambda & & -\beta q \\ \vdots & & \ddots & \vdots \\ 1 - \beta - (n-1)\beta q - \lambda & -\beta q & \dots & 1 - \beta - \lambda \end{pmatrix} \right| \\
 &= (1 - \beta - (n-1)\beta q - \lambda) \left| \begin{pmatrix} 1 & -\beta q & \dots & -\beta q \\ 1 & 1 - \beta - \lambda & & -\beta q \\ \vdots & & \ddots & \vdots \\ 1 & -\beta q & \dots & 1 - \beta - \lambda \end{pmatrix} \right|.
 \end{aligned}$$

2. Adiciona-se,  $\beta q$  vezes a coluna 1 a todas as outras colunas, obtendo-se uma matriz triangular, cujo determinante é dado pelo produto dos elementos da diagonal principal. Chegando assim ao seguinte resultado:

$$\begin{aligned}
 |\mathbf{K} - \lambda\mathbf{I}| &= (1 - \beta - (n-1)\beta q - \lambda) \left| \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 - \beta + \beta q - \lambda & & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 0 & \dots & 1 - \beta + \beta q - \lambda \end{pmatrix} \right| \\
 &= (1 - \beta - (n-1)\beta q - \lambda) (1 - \beta + \beta q - \lambda)^{n-1}.
 \end{aligned}$$

Observa-se que temos um valor próprio  $\lambda_1 = 1 - \beta - (n-1)\beta q$  de multiplicidade 1 e outro valor próprio  $\lambda_2 = 1 - \beta + \beta q$  de multiplicidade  $n-1$ .

## E.2 Autovetores

Para achar os vetores próprios correspondentes aos autovalores da matriz  $\mathbf{K}$ , resolve-se a equação  $(\mathbf{K} - \lambda\mathbf{I}) \vec{v} = \vec{0}$  para cada um dos autovalores obtidos.

**E.2.1** Para  $\lambda_1 = 1 - \beta - (n - 1) \beta q$ 

Ao substituir  $\lambda_1 = 1 - \beta - (n - 1) \beta q$  na equação que define os autovetores, temos

$$(\mathbf{K} - \lambda_1 \mathbf{I}) \vec{v} = \vec{0} \rightarrow \begin{bmatrix} (n-1)\beta q & -\beta q & \dots & -\beta q \\ -\beta q & (n-1)\beta q & & -\beta q \\ \vdots & & \ddots & \vdots \\ -\beta q & -\beta q & \dots & (n-1)\beta q \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Seja  $l_i$  a  $i$ -ésima linha, fazemos as seguintes operações elementares; a cada linha  $l_i$  se subtrai a linha  $l_{i+1}$ , obtendo-se;

$$\begin{bmatrix} n\beta q & -n\beta q & 0 & \dots & 0 \\ 0 & n\beta q & -n\beta q & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & n\beta q & -n\beta q \\ -\beta q & -\beta q & -\beta q & \dots & (n-1)\beta q \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (\text{E.3})$$

A representação matricial em (E.3) é equivalente ao sistema de equações

$$\begin{cases} n\beta q x_1 - n\beta q x_2 = 0 & x_1 = x_2 \\ n\beta q x_2 - n\beta q x_3 = 0 & x_2 = x_3 \\ \vdots & \iff \\ n\beta q x_{n-2} - n\beta q x_{n-1} = 0 & x_{n-2} = x_{n-1} \\ n\beta q x_{n-1} - n\beta q x_n = 0 & x_{n-1} = x_n, \end{cases}$$

cuja solução proporciona o vector próprio associado a  $\lambda_1$ . Esse vector não nulo é da forma

$$\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = x_n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

Observe-se que este vector próprio tem todas as componentes iguais.

### E.2.2 Para $\lambda_2 = 1 - \beta + \beta q$

Após substituir  $\lambda_2 = 1 - \beta + \beta q$  na equação que define os autovetores da matriz  $\mathbf{K}$ , temos

$$(\mathbf{K} - \lambda\mathbf{I}) \vec{v} = \vec{0} \rightarrow \begin{bmatrix} -\beta q & -\beta q & \dots & -\beta q \\ -\beta q & -\beta q & & -\beta q \\ \vdots & & \ddots & \vdots \\ -\beta q & -\beta q & \dots & -\beta q \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Subtrai-se a cada linha  $l_j$  a primeira linha  $l_1$ , obtendo muitas linhas cheias de zero;

$$\begin{bmatrix} -\beta q & -\beta q & \dots & -\beta q \\ 0 & 0 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

que se cumpre somente se

$$-\beta q x_1 - n\beta q x_2 - \dots - \beta q x_{n-1} - n\beta q x_n = 0 \iff x_1 = -x_2 - x_3 - \dots - x_{n-1} - x_n.$$

Em consequência, o vector próprio associado a  $\lambda_2$ , não é único

$$\vec{v} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -x_2 - x_3 - \dots - x_{n-1} - x_n \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = x_2 \underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{v}_1} + x_3 \underbrace{\begin{bmatrix} -1 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{v}_2} + x_n \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\vec{v}_{n-1}}$$

Podemos concluir que o valor próprio  $\lambda_2 = 1 - \beta + \beta q$  de multiplicidade  $n - 1$ , possui  $n - 1$  vetores próprios cujas componentes somam zero.

# Referências Bibliográficas

- M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Dover publications New York, 1972.
- Y. Abu-Mostafa and J.S. Jacques. Information capacity of the hopfield model. *IEEE Transactions on Information Theory*, 31(4):461–464, 1985.
- B. Aghoghovwia. Córtex cerebral. <https://www.kenhub.com/pt/library/anatomia/cortex-cerebral>, 2020.
- D.J. Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1992. ISBN 9780511623257.
- D.J. Amit, H. Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007, 1985a.
- D.J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55(14):1530, 1985b.
- D.J. Amit, H. Gutfreund, and H. Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293, 1987a.
- D.J. Amit, H. Gutfreund, and H. Sompolinsky. Statistical mechanics of neural networks near saturation. *Annals of physics*, 173(1):30–67, 1987b.
- E. Bienenstock, F. Soulié, and G. Weisbuch. *Disordered Systems and Biological Organization: Proceedings of the NATO Advanced Research Workshop on Disordered Systems and Biological Organization held at Les Houches, 1985*, volume 20. Springer Science & Business Media, 2012.

- K. Binder and A.P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern physics*, 58(4):801, 1986.
- R. Brout. Statistical mechanical theory of a random ferromagnetic system. *Physical Review*, 115(4):824, 1959.
- N. Brunel. Is cortical connectivity optimized for storing information? *Nature neuroscience*, 19(5):749, 2016.
- D. Bruno. Plasticity in neural networks. [https://thebrain.mcgill.ca/flash/i/i\\_07/i\\_07\\_cl/i\\_07\\_cl\\_tra/i\\_07\\_cl\\_tra.html](https://thebrain.mcgill.ca/flash/i/i_07/i_07_cl/i_07_cl_tra/i_07_cl_tra.html), 2002.
- L.A.V. Carvalho. Redes neuronais e a tradição conexionista da inteligência artificial. *Relatório Interno*, 1988.
- A. Crisanti, D.J. Amit, and Gutfreund H. Saturation level of the hopfield model for neural network. *Europhysics Letters*, 2(4):337, 1986.
- E. Curado, N. Melgar, and F. Nobre. Neural network under external stimulus: Improving storage capacity and reactions. *Physica A: Statistical Mechanics and its Applications*, 564:125507, 2021.
- I.N. Da Silva, Spatti, and R.A. Flauzino. *Redes Neurais Artificiais para Engenharia e Ciencias Aplicadas*. Artliber. ISBN 9788588098534.
- C. Darwin. *On the origin of species by means of natural selection*. Jhon Murray, 1859.
- C. Darwin and A. Wallace. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*, 3(9):45–62, 1858.
- N.G. de Bruijn. *Asymptotic Methods in Analysis*. Bibliotheca mathematica. Dover Publications, 1981. ISBN 9780486642215.
- B. Derrida and J.P. Nadal. Learning and forgetting on asymmetric, diluted neural networks. *Journal of statistical physics*, 49(5-6):993–1009, 1987.

- B. Derrida, E. Gardner, and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhysics Letters*, 4(2):167, 1987.
- V. Dotsenko. *An introduction to the theory of spin glasses and neural networks*, volume 54. World Scientific, 1995.
- S.F. Edwards and P.W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.
- A. Engel, H. Englisch, and A. Schütte. Improved retrieval in neural networks with external fields. *Europhysics Letters*, 8(4):393, 1989.
- L.V. Fausett. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, 1994. ISBN 9780133341867.
- J.A. Feldman, M.A. Fanty, and N.H. Goddard. Computing with structured neural networks. *IEEE Computer*, 21(3):91–103, 1988.
- V. Folli, M. Leonetti, and G. Ruocco. On the maximum storage capacity of the hopfield model. *Frontiers in computational neuroscience*, 10:144, 2017.
- J.F. Fontanari. *Mecânica estatística do processo de aprendizado em redes neurais*. PhD thesis, Universidade de São Paulo, 1994.
- J.F. Fontanari and R. Köberle. Information storage and retrieval in synchronous neural networks. *Physical Review A*, 36(5):2475, 1987.
- J.F. Fontanari and R. Koberle. Enhancing the learning of a finite number of patterns in neural networks. *Journal of Physics A: Mathematical and General*, 21(4):L253, 1988.
- J.F. Fontanari and R. Köberle. Information processing in synchronous neural networks. *Journal de Physique*, 49(1):13–23, 1988.
- J.F. Fontanari and W.K. Theumann. On the storage of correlated patterns in hopfield’s model. *Journal de Physique*, 51(5):375–386, 1990.
- W.J. Freeman. The hebbian paradigm reintegrated: local reverberations as internal representations. *Behavioral and brain sciences*, 18(4):631–631, 1995.

- E. Gardner. The space of interactions in neural network models. *Journal of physics A: Mathematical and general*, 21(1):257, 1988.
- E. Gardner and B. Derrida. Optimal storage properties of neural network models. *Journal of Physics A: Mathematical and general*, 21(1):271, 1988.
- T. Geszti. *Physical models of neural networks*. World Scientific, 1990. ISBN 9810200129.
- R.J. Glauber. Time-dependent statistics of the ising model. *Journal of mathematical physics*, 4(2):294–307, 1963.
- M.B. Gordon. Memory capacity of neural networks learning within bounds. *Journal de Physique*, 48(12):2053–2058, 1987.
- D.O. Hebb. *The organization of behavior*. new york, 1949.
- D.O. Hebb. Neuropsychology: Retrospect and prospect. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 37(1):4, 1983.
- S. Herculano. O universo de 1.5 kg. <https://www12.senado.leg.br/institucional/programas/primeira-infancia/artigos/artigos-ano-2014/o-universo-de-1.5-kg-suzana-herculano-ano-2014>, 2014.
- J.A. Hertz. *Introduction to the theory of neural computation*. CRC Press, 2018. ISBN 0201515601.
- J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- J. Hubbard. Calculation of partition functions. *Physical Review Letters*, 3(2):77, 1959.
- Queensland Brain Institute. Lobes of the brain. <https://qbi.uq.edu.au/brain/brain-anatomy/lobes-brain>, 2005.
- W. Kinzel. Learning and pattern recognition in spin glass models. *Zeitschrift für Physik B Condensed Matter*, 60(2-4):205–213, 1985.

- S. Kirkpatrick and D. Sherrington. Infinite-ranged models of spin-glasses. *Physical Review B*, 17(11):4384, 1978.
- E. Koscielny-Bunde. Pattern recognition in damaged hopfield networks parallel processing in neural systems and computers ed r eckrniller et a. *N. Holland: Elsevier Science Publishers*, pages 459–462, 1990.
- W. Krauth, J.P. Nadal, and M. Mezard. The roles of stability and symmetry in the dynamics of neural networks. *Journal of Physics A: Mathematical and General*, 21(13):2995, 1988.
- R. Kree and A. Zippelius. Asymmetrically diluted neural networks. In *Models of neural networks*, pages 193–212. Springer, 1991.
- W.A. Little. The existence of persistent states in the brain. In *From High-Temperature Superconductivity to Microminiature Refrigeration*, pages 145–164. Springer, 1974.
- M. Löwe. On the storage capacity of hopfield models with correlated patterns. *The Annals of Applied Probability*, 8(4):1216–1250, 1998.
- W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- R. McEliece, E. Posner, E. Rodemich, and S. Venkatesh. The capacity of the hopfield associative memory. *IEEE transactions on Information Theory*, 33(4):461–482, 1987.
- M. Mézard, G. Parisi, and M. Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- P. Monari, A.L. Stella, C.O. Vanderzande, and E. Orlandini. Sequence randomness and polymer collapse transitions. *Physical review letters*, 83(1):112, 1999.
- P.M.C. Oliveira. *Computing Boolean Statistical Models*. World Scientific, 1991.
- E. Orhan. The hopfield model. Technical report, Technical report, NYU, 2014.
- E. Orlandini, M.C. Tesi, and S.G. Whittington. A self-avoiding walk model of random copolymer adsorption. *Journal of Physics A: Mathematical and General*, 32(3):469, 1999.

- F. PazmAndi and T. Geszti. Relative stability in the dynamics of a two-pattern neural net. *Journal of Physics A: Mathematical and General*, 22(23):5117, 1989.
- Thadeu Josino Pereira Penna. *Aplicações Multidisciplinares de Física Estatística*. PhD thesis, Universidade Federal Fluminense, 2009.
- T.J.P. Penna and P.M.C. Oliveira. Simulations with a large number of neurons. *Journal of Physics A: Mathematical and General*, 22(14):L719, 1989.
- T.J.P. Penna and P.M.C. Oliveira. Enhancing retrieval of low-activity patterns in neural networks. *Europhysics Letters*, 11(3):191, 1990.
- P. Peretto. Collective properties of neural networks: a statistical physics approach. *Biological cybernetics*, 50(1):51–62, 1984.
- P. Peretto. On learning rules and memory storage abilities of asymmetrical neural networks. *Journal de Physique*, 49(5):711–726, 1988.
- P. Peretto. *An introduction to the modeling of neural networks*, volume 2. Cambridge University Press, 1992.
- P. Peretto and J.J. Niez. Collective properties of neural networks. In *Disordered systems and biological organization*, pages 171–185. Springer, 1986.
- W. Pitts and W.S. McCulloch. How we know universals the perception of auditory and visual forms. *The Bulletin of mathematical biophysics*, 9(3):127–147, 1947.
- S. Raschka. Machine learning faq. <https://sebastianraschka.com/faq/docs/diff-perceptron-adaline-neuralnet.html>, 2013.
- D.E. Rumelhart and J.L. McClelland. *Parallel Distributed Processing: Foundations*. A Bradford book. MIT Press, 1986. ISBN 9780262680530.
- F. Schiller. The hidden structure: A scientific biography of camillo golgi. *Bulletin of the History of Medicine*, 75(1):156–157, 2001.
- D. Sherrington. Neural networks: the spin glass approach. In *North-Holland Mathematical Library*, volume 51, pages 261–291. Elsevier, 1993.

- D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35 (26):1792, 1975.
- T. Soares and F. Fagundes. Aplicacao de redes neurais artificiais para o diagnostico de patologias traumato ortopedicas dos membros inferiores. page 10. ENCOINFO, 2012.
- R. Sollacher and G.A.O. Huaiien. Towards real-world applications of online learning spiral recurrent neural networks. *Journal of Intelligent Learning Systems and Applications*, 1(01): 1, 2009.
- H. Sompolinsky. Neural networks with nonlinear synapses and a static noise. *Physical Review A*, 34(3):2571, 1986.
- H. Sompolinsky. The theory of neural networks: The hebb rule and beyond. In *Heidelberg colloquium on glassy dynamics*, pages 485–527. Springer, 1987.
- H. Sompolinsky, A. Crisanti, and H.J. Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.
- D.A. Stariolo, E.M.F. Curado, and F.A. Tamarit. Distribution of eigenvalues of ensembles of asymmetrically diluted hopfield matrices. *Journal of Physics A: Mathematical and General*, 29(15):4733, 1996.
- A. Storkey. Increasing the capacity of a hopfield network without sacrificing functionality. In *International Conference on Artificial Neural Networks*, pages 451–456. Springer, 1997.
- F. Tanaka and S.F. Edwards. Analytic theory of the ground state properties of a spin glass. i. ising spin glass. *Journal of Physics F: Metal Physics*, 10(12):2769, 1980.
- A. Treves and D.J. Amit. Metastable states in asymmetrically diluted hopfield networks. *Journal of Physics A: Mathematical and General*, 21(14):3155, 1988.
- A.C.D. van Enter and J.L. van Hemmen. The thermodynamic limit for long-range random systems. *Journal of statistical physics*, 32(1):141–152, 1983.
- J.L. Van Hemmen and R.G. Palmer. The replica method and solvable spin glass model. *Journal of Physics A: Mathematical and General*, 12(4):563, 1979.

- J.L. van Hemmen and R.G. Palmer. The thermodynamic limit and the replica method for short-range random systems. *Journal of Physics A: Mathematical and General*, 15(12):3881, 1982.
- E.J. Van Rensburg, A. Rechnitzer, M.S. Causo, and S.G. Whittington. Self-averaging sequences in the statistical mechanics of random copolymers. *Journal of Physics A: Mathematical and General*, 34(33):6381, 2001.
- G. Wainrib and J. Touboul. Topological and dynamical complexity of random neural networks. *Physical review letters*, 110(11):118101, 2013.
- G. Weisbuch and F. Fogelman-Soulié. Scaling laws for the attractors of hopfield networks. *Journal de Physique Lettres*, 46(14):623–630, 1985.
- P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907.
- S. Wiseman and E. Domany. Self-averaging, distribution of pseudocritical temperatures, and finite size scaling in critical disordered systems. *Physical Review E*, 58(3):2938, 1998.
- B. Yegnanarayana. *Artificial Neural Networks*. PHI Learning, 2009. ISBN 9788120312531.