UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA E QUÍMICA DE SÃO CARLOS

PROCESSAMENTO DE INFORMAÇÕES
EM REDES DE
NEURÔNIOS SÍNCRONAS
JOSÉ FERNANDO FONTANARI

Tese apresentada ao Instituto de
Física e Química de São Carlos,
USP, para a obtenção do título
de Doutor em Ciências (Física
Básica)

Orientador :

Prof. Dr. Roland Köberle

OK

São Carlos

1 9 8 8

BIBLIOTECA DO INSTITUTO DE FÍSICA E QUÍMICA DE SÃO CARLOS - USP
FÍSICA

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE _____
José Fernando Fontanari APRESENTADA AO INSTITUTO DE
FÍSICA E QUÍMICA DE SÃO CARLOS, DA UNIVERSIDADE DE SÃO PAULO,
EM 26 DE maio DE 1988

Comissão Julgadora:

Prof. Dr. Roland Küberle (orientador)

Prof. Dr. José Nelson Onuchic

Prof. Dr. José Roberto Drugowich de Felício

Prof. Dr. Walter Karl Theumann

Prof. Dr. Paulo Murilo Castro de Oliveira

P A R A D E B O R A H B R A S I L B A L B Ã O

(1 9 6 3 - 1 9 8 7)

A G R A D E Ç O

Aos **meus pais** pela compreensão e confiança com que sempre pude contar.

Aos meus irmãos - **César, Juliano e Valeska** - pelo apoio e carinho, em especial ao Juliano pelo estímulo e pelas inúmeras discussões sobre neurofisiologia.

Ao Professor **Walter K. Theumann** cuja dedicação e profissionalismo tenho tomado como modelo.

Ao Professor **Roland Köberle** pelo estimulante convívio e colaboração que tornou possível este trabalho.

Ao **Drugo** e à **Ana Sílvia** pela amizade.

Aos **meus colegas do IF-UFRGS** por aqueles quatro anos inesquecíveis.

Aos **colegas de pós-graduação e Professores do DFCM-IFQSC** pelo ótimo clima de trabalho que caracteriza este Departamento.

À **Vera e Irene** pela datilografia.

À **FAPESP e ao CNPq** pelo apoio financeiro.

PROCESSAMENTO DE INFORMAÇÕES
EM REDES DE
NEURÔNIOS SÍNCRONAS

" É mais fácil usar vidros de spin
para entender o cérebro do que o
cérebro para entender vidros de
spin "

M. Virasoro

RESUMO

Vidros de spins são sistemas extremamente complexos caracterizados por um número enorme de estados estáveis e metaestáveis. Se identificarmos cada um desses estados com uma informação memorizada, esses sistemas podem ser utilizados como memórias associativas ou endereçáveis por conteúdo. O modelo de vidro de spins passa então a ser chamado de rede de neurônios.

Neste trabalho estudamos a termodinâmica e alguns aspectos dinâmicos de uma rede de neurônios com processamento paralelo ou síncrono - o Modelo de Little de memória associativa - no regime em que o número de informações memorizadas p cresce como $p = \alpha N$, onde N é o número de neurônios.

Usando a teoria simétrica em relação às réplicas obtemos o diagrama de fases no espaço de parâmetros do modelo no qual incluímos um termo de autointeração dos neurônios. A riqueza do diagrama de fases que possui uma superfície de pontos tricríticos é devida à competição entre os dois regimes assintóticos da dinâmica síncrona: pontos fixos e ciclos de período dois.

A B S T R A C T

Spin glasses are very complex systems characterized by a huge number of stable and metastable states. If we identify each state with a memorized information then spin glasses may be used as associative or content addressable memories. This spin glass model is then called a neural network.

In this work we study the thermodynamics and some dynamical aspects of a neural network with parallel or synchronous processing - Little's model of associative memory - in the regime where the number of memorized informations p grows as $p = \alpha N$, where N is the number of neurons.

Using the replica symmetric theory we determine the phase diagram in the space of the model's parameters, in which we include a neural self interaction term. The richness of the phase diagram which possesses a surface of tricritical points is due to the competition between the two asymptotic dynamical behaviours of the synchronous dynamics : fixed points and cycles of length two.

Í N D I C E

CAPÍTULO	I -	INTRODUÇÃO	1
CAPÍTULO	II -	MODELOS DE MEMÓRIA ASSOCIATIVA	8
		II.A. - O neurônio de McCulloch-Pitts	10
		II.B. - O Modelo de Hopfield	17
		II.C. - O Modelo de Little	21
		II.D. - A memória	28
CAPÍTULO	III -	A TERMODINÂMICA DO MODELO DE LITTLE	40
		III.A.- O Hamiltoniano	42
		III.B.- A energia livre	48
		III.C.- O ansatz simetria de réplicas	58
		III.D.- Os estados estacionários para $T=0$..	65
		III.E.- Os estados estacionários para $\alpha=0$..	75
		III.F.- O diagrama de fases no espaço(J_0, α, T)	79
		III.G.- A quebra de simetria de réplicas ...	89
CAPÍTULO	IV -	A DINÂMICA DO MODELO DE LITTLE À TEMPERATURA ZERO	95
		IV.A. - A aproximação de Kinzel	97
		IV.B. - Os ciclos-2	107
		IV.C. - A autointeração	113
		IV.D. - A correlação de recuperação remanes- cente	119
CAPÍTULO	V -	CONCLUSÃO	123
APÊNDICE	A -	O limite $n \rightarrow 0$ no ansatz simetria de réplicas	133
APÊNDICE	B -	A estabilidade das soluções ferromagnéticas frente a quebra de simetria das réplicas ...	135
REFERÊNCIAS	-		142

CAPÍTULO I

INTRODUÇÃO

I. INTRODUÇÃO

Entender o funcionamento do cérebro e projetar computadores inspirados na sua arquitetura são os ambiciosos objetivos de um modismo que nos últimos cinco anos tem agitado uma boa parte da comunidade física, em particular a especializada em Mecânica Estatística: a Teoria das Redes de Neurônios.

Entretanto, já há 40 anos atrás John Von Neumann [1] , estudando sistemas computacionais de lógica Booleana, previa que as teorias da computação viriam a assemelhar-se em muitos aspectos com a termodinâmica:

"De fato, numerosas indicações nos levam a crer que esse novo sistema de lógica formal evoluirá em direção a uma outra disciplina... termodinâmica, principalmente na forma concebida por Boltzmann que é a parte da Física teórica que mais se aproxima ... da manipulação e medida da "informação"."

Logo a seguir, Cragg e Temperley (1954) [2] propuseram uma analogia entre sistemas de neurônios interagentes

e sistemas de dipolos magnéticos acoplados. Na sua versão mais simples um dipolo magnético é um elemento que pode assumir dois estados - para "cima" ou para "baixo" - caso estiver paralelo ou antiparalelo à direção de um campo magnético externo. Similarmente, o modelo de neurônios formais proposto por McCulloch e Pitts (1943) [3] também representava o neurônio por uma variável de dois estados - ativo ou passivo - dependendo se o neurônio está ou não emitindo um pulso elétrico. Ainda, em um sistema de neurônios, cada neurônio escolhe seu estado de acordo com a soma dos sinais elétricos que ele recebe dos outros neurônios similarmente a um átomo magnético que alinha seu momento na direção do campo resultante criado pelos seus vizinhos em um material magnético.

O trabalho de Cragg e Temperley, publicado em uma revista médica, praticamente não teve repercussão na comunidade física.

Somente em 1974 Little [4], independentemente, retomou essa analogia. Sua idéia era relacionar a correlação temporal em redes de neurônios com a correlação espacial de um sistema de spins de Ising [5] (dipolos magnéticos). De fato, esses problemas são formalmente idênticos: tanto a existência de ordem de longo alcance no sistema de spins como o aparecimento de estados persistentes na rede de neurônios estão associados com a degenerescência do maior autovalor da matriz de transferência dos dois problemas.

Entretanto havia dificuldades insuperáveis: no Modelo de Ising as interações entre os spins são simétricas, isto é, o efeito do spin j sobre o spin i é idêntico ao do spin i sobre o spin j , enquanto que as interações entre os neurônios são sabidamente assimétricas. Como consequência, a matriz de transferência da rede de neurônios não é diagonalizável. Ainda, a constatação de que um neurônio pode excitar ou inibir outros neurônios implica na existência de interações positivas e negativas originando então o efeito conhecido como frustração que complica extremamente a análise mesmo para o problema com interações simétricas.

Essa abordagem teve vida curta e não produziu resultados significativos.

A presente fase, iniciada por Hopfield [6] em 1982, caracteriza-se por simplificações radicais tanto na formulação das questões como na descrição do sistema. Por exemplo, questões complexas como os processos de aprendizado, abstração, condicionamento, etc... não são abordadas nesta etapa inicial. Os esforços são todos dirigidos para a solução do seguinte problema:

Como escolher as interações (sinápses) acoplando N neurônios formais de modo que a rede possa memorizar uma coleção arbitrária de padrões (informações)? (Um padrão é dito memorizado quando atua como um atrator no comportamento dinâmico da rede de neurônios).

Assim, principiando com uma configuração inicial de atividade (estados dos neurônios) induzida por algum estímulo externo, a rede evoluirá de acordo com sua dinâmica interna até atingir o estado final, por definição um atrator ou estado persistente. Se as interações foram escolhidas corretamente esse estado será um dos padrões memorizados e, portanto, a rede funciona como uma memória endereçável por conteúdo que pode reconhecer estímulos incompletos e restaurar a informação perdida.

No nível de construção do modelo, Hopfield levou a simplificação ao extremo de considerar as interações entre os neurônios simétricas. Embora biologicamente insustentável, essa hipótese abriu o caminho para a aplicação dos métodos de mecânica estatística no estudo de redes de neurônios. De fato, a dinâmica de uma rede de neurônios simétrica minimiza uma função "energia" e, portanto, podemos estudar os estados estacionários da dinâmica olhando para os estados que minimizam a energia da rede de neurônios.

Outra inovação introduzida foi quanto ao processo dinâmico da rede. Até então, a dinâmica utilizada era síncrona, isto é, só é permitido aos neurônios evoluírem em tempos que sejam múltiplos inteiros de um período arbitrário τ . Assim, todos os neurônios evoluem simultaneamente. Hopfield propôs uma dinâmica assíncrona, na qual apenas um neurônio evolui a cada unidade de tempo τ . Essa é a dinâmi

ca de Monte-Carlo-Glauber [7] utilizada em simulações de sistemas de spins.

Posto dessa forma e lembrando o efeito de frustração devido às interações positivas e negativas entre os neurônios vemos que o problema de redes de neurônios assemelha-se em muitos aspectos ao problema do vidro de spin de Ising [8,9]. De fato, em duas notáveis contribuições Amit, Gutfreund e Sompolinsky (1985) [10, 11] utilizando técnicas desenvolvidas no estudo de vidros de spin, resolveram analiticamente o problema de redes de neurônios na forma concebida por Hopfield. A partir de então os principais participantes na exploração desses modelos de memória têm sido os físicos teóricos especialistas em vidros de spin.

Neste trabalho nos propomos a estudar o problema de redes de neurônios simétricas segundo a forma concebida por Little na qual a rede evolui através de uma dinâmica síncrona. Usaremos as mesmas técnicas empregadas por Amit et al. no estudo de redes simétricas assíncronas. A tese é organizada como segue:

a) Capítulo II- neste capítulo apresentamos ao leitor o problema de redes de neurônios através de uma rápida descrição neurofisiológica do neurônio e de seu modelo mais simples proposto por McCulloch e Pitts. As diferenças entre os Modelos de Hopfield (assíncrono) e de Little (sín-

crono) são enfatizados e discutimos a implementação da memória em uma rede de neurônios.

b) Capítulo III - este capítulo é dedicado ao estudo da termodinâmica do Modelo de Little. Uma das motivações desse estudo é o fato da dinâmica síncrona possuir dois diferentes comportamentos no regime de equilíbrio: pontos fixos e ciclos de período 2. Derivamos o Hamiltoniano do Modelo de Little e mostramos então como a mecânica estatística de equilíbrio pode descrever sistemas que apresentem comportamento cíclico. Utilizando o formalismo das réplicas desenvolvido no estudo de vidros de spin [8,9] obtemos a energia livre do modelo e as equações dos parâmetros de ordem que descrevem os estados estacionários da dinâmica. A solução dessas equações é apresentada em forma de diagramas de fases, onde estão delimitadas as regiões de existência das diferentes soluções no espaço de parâmetros do modelo.

c) Capítulo IV - a fim de tentar esclarecer algumas questões que não puderam ser satisfatoriamente tratadas com o cálculo termodinâmico, apresentamos neste capítulo um estudo numérico da dinâmica síncrona. Inicialmente estudamos uma aproximação, proposta por Kinzel [12], que torna o problema da dinâmica solúvel analiticamente. Nas simulações numéricas damos especial ênfase ao papel desempenhado por um termo de autointeração dos neurônios e à natureza e estrutura dos ciclos de período 2.

d) Capítulo V - este capítulo é reservado para a apresentação de nossas conclusões e de alguns comentários sobre as perspectivas futuras do estudo de redes de neurônios.

Há também dois apêndices onde apresentamos detalhes de alguns cálculos realizados no Capítulo III.

CAPÍTULO II

MODELOS DE MEMÓRIA ASSOCIATIVA

II. MODELOS DE MEMÓRIA ASSOCIATIVA

Um dos grandes desafios da Psicologia, Neurobiologia e Linguística é entender como as informações memorizadas são **codificadas, armazenadas e recuperadas** no cérebro humano. Essa questão também desperta interesse em áreas mais aplicadas como a Inteligência Artificial e Ciências da Computação.

Abordaremos esse problema propondo um modelo de memória cujas características imitem tanto quanto possível às do cérebro humano. Tal modelo deve satisfazer os seguintes requisitos:

a) Seus componentes básicos devem ser inspirados nos elementos neurofisiológicos - neurônios e sinapses.

b) O modelo deve ser capaz de classificar - de um grande número de estímulos ele deverá produzir um número muito menor de respostas. Por exemplo, todos os quadrados devem ser reconhecidos como quadrados independentemente de translações ou rotações que não alterem sua forma.

c) A resposta do modelo deve ser dada em não mais que algumas dezenas de passos do processo dinâmico desencadeado pelo estímulo. Isso é necessário para que o modelo seja compatível com a escala de tempo neurofisiológica: processos neuronais são caracterizados por tempos da ordem de

milisegundos e nossas reações ocorrem em frações de segundo.

d) O modelo deve ser insensível a danos aleatórios - morte de neurônios, destruição de sinapses, ruído externo,...

Além de satisfazer a todos esses requisitos, o modelo deverá prover uma localização para a memória bem como sua relação com os parâmetros neurofisiológicos - neurônios e sinapses.

As secções seguintes são dedicadas à formulação de um modelo que satisfaça as exigências acima.

II.A. O NEURÔNIO DE MCCULLOCH-PITTS

O sistema nervoso humano contém aproximadamente 10^{12} neurônios. Cada neurônio é uma unidade independente composta de três partes: o corpo ou soma, o axônio e os dendritos. Estes últimos são extensões do soma. Um neurônio comunica-se com os outros neurônios através de sinais elétricos que se propagam pelo seu axônio e ramificações deste. Essas ramificações conectam-se quimicamente com os dendritos de aproximadamente 10^4 neurônios através das sinapses. Na figura 1 mostramos o diagrama esquemático de uma célula nervosa típica.

Fig.1- Diagrama esquemático de um neurônio.

Basicamente toda a atividade neural é elétrica. Na ausência de estímulos, o interior do neurônio está a um potencial de -70 mV em relação ao exterior. Dizemos então que o neurônio está no estado passivo. Essa diferença de potencial (ddp) é devida às diferentes concentrações de íons de sódio (Na^+) e potássio (K^+) entre o interior e o exterior do neurônio.

Quando o neurônio é estimulado observam-se variações da ddp através de sua membrana: ou a ddp torna-se menos negativa (despolarização) ou mais negativa (hiperpolariz

zação).

No caso de hiperpolarização o neurônio permanecerá passivo. Entretanto, se a despolarização atingir um certo valor limiar (≈ -60 mV) inicia-se um processo violentíssimo, relacionado com a mudança da permeabilidade iônica da membrana, de curta duração (≈ 1 ms), capaz de levar a ddp através da membrana a $+30$ mV. Dizemos então que o neurônio está no estado ativo. O pulso gerado por essa rápida variação do potencial é conhecido como potencial de ação.

Após ter produzido um potencial de ação o neurônio entra em um período refratário (≈ 2 ms) permanecendo inativo, indiferente aos estímulos que receber neste intervalo de tempo.

O potencial de ação propaga-se até as sinápses onde estimula a liberação de uma substância química, o neurotransmissor, capaz de induzir uma despolarização ou hiperpolarização na membrana do neurônio receptor. O neurotransmissor atua alterando a permeabilidade iônica da membrana neuronal. O potencial produzido por esse distúrbio é chamado de potencial pós-sináptico (PPS) e pode ser excitatório caso o neurotransmissor tenha produzido uma despolarização ou inibitório no caso de hiperpolarização. Assim o processo descrito repete-se novamente.

O efeito de vários PPS's emitidos por diferentes

neurônios que chegam quase simultaneamente ao soma do neurônio receptor é aproximadamente aditivo. Assim, é a soma desses potenciais que deve exceder o limiar de -60 mV para tornar o neurônio ativo. Caso contrário, o neurônio permanecerá passivo e a ddp através de sua membrana decairá ao seu valor de equilíbrio (-70 mV) quase que exponencialmente. É devido a esse decaimento que o efeito de PPS's que não chegam ao mesmo tempo ao soma não é aditivo.

Ignorando completamente a dinâmica interna do neurônio, McCulloch e Pitts (1943) [3] propuseram modelá-lo por uma variável discreta de dois estados correspondendo aos estados passivo e ativo descritos acima. Assim, considerando uma rede de N neurônios, representamos o estado do neurônio i , $i=1, \dots, N$, por um spin de Ising: $S_i = +1$ (ativo) ou $S_i = -1$ (passivo).

Um neurônio pode inibir ou ativar outros neurônios. O efeito do neurônio j sobre o neurônio i é descrito pela variável J_{ij} que pode ser positiva (ativa o neurônio i) ou negativa (inibe o neurônio i). Vemos então que J_{ij} desempenha o papel da constante de acoplamento no Modelo de Ising. O módulo de J_{ij} nos dá uma medida de quanto o pulso emitido pelo neurônio j contribui para o potencial da membrana do neurônio i . Esse potencial V_i , é dado pela soma de todos os potenciais de ação que chegam ao neurônio i quase que simultaneamente:

$$V_i = \sum_{j=1}^N J_{ij} (S_j + 1) \quad (\text{II.1})$$

Os J_{ij} 's podem ser considerados como elementos de uma matriz quadrada de dimensão $N \times N$ - a matriz sináptica.

No momento, façamos a hipótese de que a soma dos potenciais de ação V_i seja idêntica a soma dos potenciais pós-sinápticos. Assim, o neurônio i irá tornar-se ativo sempre que V_i for maior que um dado limiar U_i :

$$S_i(t+\tau) = \begin{cases} \text{sign} \left(\sum_{j=1}^N J_{ij} S_j(t) - \alpha_i \right) \\ S_i(t) \quad \text{se} \quad \sum_{j=1}^N J_{ij} S_j(t) = \alpha_i \end{cases} \quad (\text{II.2})$$

onde

$$\alpha_i = U_i - \sum_{j=1}^N J_{ij} \quad (\text{II.3})$$

e τ é uma unidade de tempo arbitrária a ser definida posteriormente.

O estado de cada neurônio irá mudar com o tempo de acordo com a equação (II.2) e dependerá somente dos estados dos neurônios conectados a ele no tempo imediatamente anterior. Os neurônios não possuem memória de estados anteriores a τ (processo Markoviano).

A dinâmica da rede de neurônios não é completamente descrita pela equação (II.2): é ainda necessário especificar como vamos incrementar a variável de tempo t . Tratem-se de dois casos limites:

a) t é incrementado sempre que um neurônio é testado, isto é, sempre que escolhemos um valor de i na equação (II.2). Essa dinâmica é usualmente referida como dinâmica assíncrona ou sequencial e corresponde exatamente ao limite de temperatura nula da dinâmica de Monte Carlo usada em simulações de sistemas de spins. O neurônio a ser testado pode ser escolhido aleatoriamente entre os N neurônios ou então segundo alguma regra determinística.

b) t é incrementado somente após termos testado todos os N neurônios. Essa dinâmica, dita síncrona ou paralela, foi escolhida por McCulloch e Pitts para governar a evolução do modelo de neurônios formais.

A escala natural para medir o tempo t deve ser o passo da dinâmica, definido como o tempo necessário para que os N neurônios sejam testados. Com isso fixamos o valor de τ na equação (II.2): se $t=1$ equivale a um passo da dinâmica então $\tau = 1/N$ para a dinâmica sequencial e $\tau = 1$ para a dinâmica paralela.

Na figura 2 temos um sistema de dois neurônios S_1 e S_2 com $J_{12} = J_{21} = +1$, $J_{11} = J_{22} = 0$ e $\alpha_1 = \alpha_2 = 0$

que evoluem a partir da mesma configuração inicial mas com dinâmicas distintas. A dinâmica sequencial (a) leva o sistema a um ponto fixo e a paralela (b) a um ciclo de período 2.

		a		b	
$t \backslash i$		1	2	1	2
0		+1	-1	+1	-1
1		-1	-1	-1	+1
2		-1	-1	+1	-1

Fig.2- Comparação entre as dinâmicas
a) sequencial e b) paralela.

A dinâmica real de um sistema de neurônios deve ser algo intermediário entre os extremos descritos acima. Uma situação possível de ocorrer é a formação de grupos de neurônios sincronizados, isto é, que evoluem segundo a dinâmica paralela, enquanto que os grupos evoluem sequencialmente.

Por fim devemos definir o que entendemos por estímulo e resposta do modelo. O estímulo deve ser algo capaz de desencadear o processo dinâmico do modelo. Assim, nada mais natural do que identificá-lo com a configuração inicial dos neurônios, $\{S_i(t=0)\}$.

Dado um estímulo, a dinâmica do modelo seleciona

uma trajetória no espaço dos 2^N estados possíveis do sistema. A cada momento o sistema passa por um desses 2^N estados. Portanto, é natural identificar a resposta do modelo como a repetição persistente de um estado ou sequência de estados. Em outras palavras, a resposta do modelo será seu estado estacionário, $\{S_i(t=\infty)\}$, no caso de atingir um ponto fixo ou, no caso de ciclos, a sequência de estados que compõem o ciclo.

Finalizando essa seção observemos que o fato dos pulsos elétricos propagarem-se apenas em um sentido no axônio determina uma forte assimetria na matriz sináptica. De fato, é comum a situação em que o neurônio j influencia o neurônio i , $J_{ij} \neq 0$, não ocorrendo a recíproca, $J_{ji} = 0$.

II.B. O MODELO DE HOPFIELD

A assimetria da matriz sináptica é uma das maiores dificuldades a enfrentar no estudo de sistemas neurais. Esses sistemas são estudados quase que exclusivamente através de simulações numéricas [13]. Isso se deve ao comportamento extremamente complexo da equação (II.2) que, dependendo da configuração inicial (estímulo), pode atingir um regime caótico, cíclico de período qualquer ou um ponto fixo, independente da dinâmica ser síncrona ou assíncrona.

O modelo de neurônios formais de McCulloch- Pitts já possui algumas super-simplificações como a eliminação total da dinâmica interna do neurônio (período refratário, decaimento exponencial de potenciais menores que os limiares,...). Em 1982, Hopfield foi mais além e escolheu a matriz sináptica como sendo simétrica [6]. Biologicamente isso é insustentável. Entretanto, podemos considerar essa hipótese como a primeira etapa de uma estratégia usualmente bem sucedida na abordagem de problemas complexos: uma vez entendido o comportamento do modelo com interações simétricas podemos procurar quais os aspectos que mudam com a introdução da assimetria nas interações. De fato, mais tarde mostrou-se que essa hipótese não parece ser muito restritiva [14].

O Modelo de Hopfield é definido pela equação (II.2) onde escolhemos $J_{ij} = J_{ji}$ e a evolução temporal assíncrona ou sequencial. Vamos agora mostrar alguns resultados importantes sobre esse modelo.

Teorema: A dinâmica do Modelo de Hopfield minimiza a seguinte quantidade, dita energia do Modelo de Hopfield:

$$E = - \frac{1}{2} \sum_{i,j} J_{ij} S_i S_j + \sum_i \alpha_i S_i \quad (\text{II.4})$$

Prova: Como a cada incremento τ em t apenas um neurônio é testado, digamos S_1 , escrevemos a variação de energia, $\Delta E = E(t+\tau) - E(t)$, como:

$$\Delta E = - \left[\sum_{j \neq 1} J_{1j} S_1(t+\tau) S_j(t) - \alpha_1 S_1(t+\tau) - \right.$$

$$\left. \sum_{j \neq 1} J_{1j} S_1(t) S_j(t) + \alpha_1 S_1(t) \right]$$

$$\Delta E = - (S_1(t+\tau) - S_1(t)) \left[\sum_{j \neq 1} J_{1j} S_j(t) - \alpha_1 \right].$$

Usando a equação (II.2) e a identidade $|x| = x \operatorname{sign} x$ obtemos

$$\Delta E = - |h_1(t)| + S_1(t) h_1(t) \quad (\text{II.5})$$

onde
$$h_1(t) \equiv \sum_{j \neq 1} J_{1j} S_j(t) - \alpha_1.$$

Uma vez que $S_1(t) h_1(t) \leq |h_1(t)|$ temos $\Delta E \leq 0$ o que prova o teorema.

Corolário: Os estados estacionários da dinâmica do Modelo de Hopfield são pontos fixos.

Prova: Os estados estacionários devem satisfazer a condição $\Delta E = 0$. Isso só ocorre se $S_1(t)$ estiver na dire-

ção do "campo molecular" $h_i(t)$, isto é, se possuírem o mesmo sinal. Mas neste caso, a equação (II.2) nos dá $S_i(t+\tau)=S_i(t)$ de modo que o estado do sistema não muda, caracterizando assim um ponto fixo da dinâmica.

A analogia com o Modelo de Ising à temperatura nula fica então completa. Supondo por simplicidade que $J_{ij} > 0$ vejamos como esse modelo satisfaz a maior parte dos requisitos apresentados no início do capítulo. O requisito (a) é satisfeito por construção. Dado qualquer estímulo o sistema relaxará em um dos dois estados: todos os neurônios ativos (+1) ou todos passivos (-1). Isso satisfaz (b) de um modo um tanto radical. O sistema é robusto à diluição de neurônios ou à adição ou destruição de acoplamentos (sinapses) o que satisfaz (d). O requisito (c) é satisfeito pois bastam algumas dezenas de trocas de estado por spin para o sistema relaxar em um dos dois estados fundamentais.

A analogia correta não é com o Modelo de Ising ferromagnético, mas sim com o vidro de spin de Ising pois, como já foi observado, existem sinapses excitatórias ($J_{ij} > 0$) e inibitórias ($J_{ij} < 0$). Introduz-se assim o efeito de frustração no modelo. Discutiremos esse ponto, juntamente com a questão da localização da memória na seção II.D.

II.C. O MODELO DE LITTLE

Até esse momento tratamos apenas de modelos determinísticos, isto é, fixados os parâmetros e dada a configuração inicial a evolução do sistema está completamente determinada⁺. Entretanto existe um elemento estocástico nos sistemas biológicos cujo efeito é fazer com que a soma dos potenciais de ação não seja idêntica a soma dos potenciais pós-sinápticos. Isso é devido às variações aleatórias da quantidade de neurotransmissor liberado nas sinapses pelos potenciais de ação.

Assim é possível que um neurônio permaneça inativo mesmo se o potencial de ação V_i for maior que o limiar U_i ou então que se torne ativo se V_i for menor que U_i .

Para descrever esta situação Little propôs que a evolução do modelo fosse governada por uma função de probabilidade [4]. Assim, cada neurônio "escolheria" o seu estado no tempo $t+\tau$ de acordo com a probabilidade.

$$P(S_i(t+\tau)) = [1 + \exp(-\beta S_i(t+\tau) h_i(t))]^{-1} \quad (\text{II.6})$$

onde

$$h_i(t) = \sum_j J_{ij} S_j(t) - \alpha_i \quad (\text{II.7})$$

⁺ Não estamos considerando o efeito estocástico de escolhermos aleatoriamente o neurônio a ser testado na dinâmica assíncrona.

O parâmetro $\beta^{-1} \equiv T$ simula o ruído nas sinapses e pode ser identificado com a temperatura que atua em um sistema de spins levando, então, a analogia com o Modelo de Ising à temperaturas finitas. Observe que no limite $T \rightarrow 0$ a equação (II.6) se reduz a equação determinística (II.2).

Essa generalização foi proposta originalmente para o modelo de neurônios formais de McCulloch-Pitts cuja evolução é síncrona. O mesmo pode ser feito para o Modelo de Hopfield que passa então a chamar-se Modelo de Hopfield generalizado.

Por simplicidade vamos nos referir ao modelo de dinâmica síncrona como Modelo de Little e ao de dinâmica assíncrona como Modelo de Hopfield, independente da existência ou não de ruído sináptico.

A seguir vamos mostrar para o Modelo de Little a $T=0$ resultados similares aos obtidos para o Modelo de Hopfield na secção anterior.

Teorema: A dinâmica a $T=0$ do Modelo de Little com interações simétricas, $J_{ij} = J_{ji}$, minimiza a seguinte quantidade, dita energia do Modelo de Little a $T=0$:

$$E = - \sum_i \left| \sum_j J_{ij} S_j - \alpha_i \right| + \sum_i \alpha_i S_i \quad (\text{II.8})$$

Prova: A cada incremento τ em t todos os neurônios são testados. Daí:

$$E(t) = - \sum_i S_i(t+\tau) \left[\sum_j J_{ij} S_j(t) - \alpha_i \right] + \sum_i \alpha_i S_i(t) \quad (\text{II.9a})$$

$$E(t+\tau) = - \sum_i \left| \sum_j J_{ij} S_j(t+\tau) S_i(t) - \alpha_i S_i(t) \right| + \sum_i \alpha_i S_i(t+\tau) \quad (\text{II.9b})$$

onde usamos a identidade $|x| = x \operatorname{sign} x$ e a equação (II.2) para obter (II.9a). Em (II.9b) usamos simplesmente que $|S_i(t)| = 1$. Assim

$$\begin{aligned} \Delta E &= E(t+\tau) - E(t) = \sum_{i,j} J_{ij} S_i(t) S_j(t+\tau) - \\ &\quad \sum_i \alpha_i S_i(t) - \sum_i \left| \sum_j J_{ij} S_i(t) S_j(t+\tau) - \alpha_i S_i(t) \right| \end{aligned} \quad (\text{II.10})$$

e portanto $\Delta E \leq 0$.

Corolário: Os estados estacionários da dinâmica a $T=0$ do Modelo de Little com interações simétricas são pontos fixos ou ciclos de período 2.

Prova: Façamos $\alpha_i=0$ por simplicidade. Os estados estacionários devem satisfazer $\Delta E=0$. Assim, usando a equação (II.10) e os truques já convencionais para eliminar o valor absoluto e a função sinal, obtemos

$$\sum_i S_i(t+\tau) \left[\sum_j J_{ij} (S_j(t) - S_j(t+2\tau)) \right] = 0 \quad (\text{II.11})$$

O corolário estará provado quando mostrarmos que a única solução da equação (II.11) é

$$S_i(t) = S_i(t+2\tau) \quad i=1, \dots, N \quad (\text{II.12})$$

Para isso, usemos o método de demonstração por redução ao absurdo. Seja C_M um conjunto de M neurônios que não satisfaçam a relação $S_i(t) = S_i(t+2\tau)$. Assim

$$S_k(t) = -S_k(t+2\tau) \quad k \in C_M \quad (\text{II.13})$$

Reescrevemos (II.11) como:

$$\sum_k S_k(t+2\tau) - \sum_{i=1}^N J_{ik} S_i(t+\tau) = 0$$

$$\sum_k \left| \sum_{i=1}^N J_{ik} S_i(t+\tau) \right| = 0$$

$$\sum_{i=1}^N J_{ik} S_i(t+\tau) = 0 \quad (\text{II.14})$$

onde $k \in C_M$. Mas $\sum_{i=1}^N J_{ik} S_i(t+\tau) = h_k(t+\tau)$ é justamente o campo molecular ou potencial atuando sobre o neurônio k no tempo $t+\tau$. Como $h_k(t+\tau)=0$ o neurônio k permanecerá no mesmo estado:

$$S_k(t+2\tau) = S_k(t+\tau) \quad k \in C_M. \quad (\text{II.15})$$

Como, por hipótese, o sistema está no regime de equilíbrio podemos utilizar a igualdade $E(t+\tau)=E(t+2\tau)$ para obter a equação

$$\sum_i S_i(t+2\tau) \left[\sum_j J_{ij} (S_j(t+\tau) - S_j(t+3\tau)) \right] = 0. \quad (\text{II.16})$$

Seja $C_{M'}$ um conjunto de M' neurônios que não satisfaçam $S_i(t+\tau) = S_i(t+3\tau)$ com $1 < M' < N$. Daí,

$$S_{k'}(t+\tau) = - S_{k'}(t+3\tau) \quad k' \in C_{M'}. \quad (\text{II.17})$$

Raciocínio idêntico ao anterior nos leva à rela-

ção

$$S_{k'}(t+3\tau) = S_{k'}(t+2\tau) = -S_{k'}(t+\tau) \quad k' \in C_{M'}. \quad (\text{II.18})$$

Por outro lado, as equações (II.13) e (II.15) dão

$$S_k(t+2\tau) = S_k(t+\tau) = -S_k(t) \quad k \in C_M. \quad (\text{II.19})$$

Portanto $C_M \cap C_{M'} = \emptyset$. Assim os neurônios pertencentes a C_M devem satisfazer $S_k(t+\tau) = S_k(t+3\tau)$. Daí

$$S_k(t+3\tau) = S_k(t+2\tau) = S_k(t+\tau) = -S_k(t) \quad k \in C_M. \quad (\text{II.20})$$

Repetindo esse raciocínio para as relações $E(t+2\tau) = E(t+3\tau), \dots, E(t+(n-1)\tau) = E(t+n\tau)$ concluímos que

$$S_k(t+n\tau) = \dots = S_k(t+2\tau) = S_k(t+\tau) = -S_k(t) \quad k \in C_M \quad (\text{II.21})$$

onde n é um inteiro arbitrário.

A equação (II.21) claramente contradiz nossa hipótese inicial de equilíbrio pois o sistema nunca volta ao seu estado inicial $\{S_i(t)\}$.

Assim $C_M = \emptyset$ e portanto (II.12) é a única solução

de (II.11).

Na figura 2 já tivemos a oportunidade de observar um ciclo-2 com J_{ij} simétrico.

Apresentaremos agora um resultado simples, mas de importância fundamental para os cálculos a serem realizados no capítulo seguinte.

Suponhamos que o sistema esteja em um ciclo-2 no qual N_2 neurônios participam do ciclo e N_1 estão fixos. Vamos mostrar que os subsistemas S_1 contendo os neurônios fixos e S_2 contendo os oscilantes são independentes.

Tomemos um neurônio qualquer em S_1 . O campo sobre ele varia no tempo devido a contribuição dos neurônios em S_2 . Porém uma vez que ele está fixo o sinal do campo resultante sobre ele deve permanecer constante. Isso só é possível se o módulo do campo constante produzido pelos neurônios em S_1 for maior que o módulo do campo variável proveniente de S_2 . Assim, podemos ignorar a contribuição devida a S_2 .

Tomemos agora um neurônio qualquer em S_2 . Os neurônios em S_1 contribuem com um campo constante no tempo, entretanto o campo resultante está mudando de sinal a cada etapa de tempo para que o neurônio oscile. Isso só é possível se o módulo do campo devido a S_2 for maior que o devido a S_1 . Podemos então ignorar a contribuição devida a S_1 .

Portanto, o sistema terá o mesmo comportamento se eliminarmos todas as ligações entre neurônios pertencentes a diferentes subsistemas. Podemos então estudar apenas os casos em que todos os neurônios estejam fixos (pontos fixos) ou oscilando (ciclo-2 completo). Qualquer outra situação é uma combinação trivial desses dois casos.

II.D. A MEMÓRIA

Tratemos agora da questão da localização e implementação da memória nos modelos propostos nas secções anteriores.

Para abordarmos essa questão seguiremos a seguinte sugestão feita por Donald Hebb [15] em 1949:

"Quando o axônio da célula A está próximo o suficiente para excitar a célula B e repetidamente ou persistentemente contribui para que B torne-se ativa, ocorre algum processo de crescimento ou mudança metabólica em uma ou ambas as células tal que a eficiência de A em ativar B seja aumentada".

Decorre dessa sugestão que o aprendizado se dá nas sinapses cuja eficiência é modificada de acordo com a atividade dos neurônios pré e pós-sinápticos. Assim, obtemos de

uma só vez duas informações fundamentais:

- a) a memória está localizada nas sinapses
- b) existe uma relação entre a memória e os parâmetros neuro-fisiológicos - a eficiência das sinapses é função da atividade neural.

As informações memorizadas são recuperadas pela sua **associação** com o estímulo ou então endereçadas pelo **conteúdo** do estímulo mais do que por um endereço físico como no caso de computadores convencionais.

Apliquemos essas idéias aos nossos modelos. Restringir-nos-emos ao Modelo de Hopfield a $T=0$ embora todos os resultados a serem derivados a seguir sejam válidos também para o Modelo de Little.

Primeiramente notemos que a eficiência do neurônio j em ativar ou inibir o neurônio i é medida pelo módulo da constante de acoplamento J_{ij} .

A comunicação com o sistema é feita através de linguagem binária: os estímulos e as respostas são seqüências de $+1$'s e -1 's. As informações memorizadas também devem ter essa forma. Para distingui-las dos estados do modelo $\{S_i = \pm 1\}$ vamos representá-las por $\{\xi_i^\mu = \pm 1\}$ onde $\mu=1, \dots, p$ rotula p diferentes informações ou padrões.

Para $J_{ij} > 0$ (Ising ferromagnético) qualquer estímulo produz uma das respostas $\{S_i = +1\}$ ou $\{S_i = -1\}$. Portanto esse modelo possui apenas um padrão memorizado, $\{\xi_i = +1\}$. Não consideramos o padrão $\{\xi_i = -1\}$ pois este relaciona-se com $\{\xi_i = +1\}$ através de uma simples transformação de simetria do modelo.

Suponha agora que desejamos memorizar um único padrão, diferente do ferromagnético, digamos $\{\xi_i\}$. Escolhendo

$$J_{ij} = \begin{cases} \frac{1}{N} \xi_i \xi_j & i \neq j \\ 0 & i = j \end{cases} \quad (\text{II.22})$$

obtemos o Modelo de Mattis [16] cujos estados fundamentais são $\{S_i = \xi_i\}$ e $\{S_i = -\xi_i\}$.

A generalização de (II.22) é conhecida como a prescrição de Cooper [17] para a memorização de p padrões:

$$J_{ij} = \begin{cases} \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} & i \neq j \\ 0 & i = j \end{cases} \quad (\text{II.23})$$

A construção da matriz sináptica é considerada como um processo de aprendizagem. A equação (II.23) é, atualmente, referida como regra de aprendizagem de Hebb. Observe

a localidade dessa regra: a interação entre os neurônios i e j depende apenas de variáveis dos sítios i e j , em concordância com a sugestão de Hebb.

Em analogia com o vocabulário utilizado em Psicologia denominaremos de **protótipos** os padrões memorizados.

Analiseemos agora o desempenho do Modelo de Hopfield com interações dadas pela prescrição de Cooper, equação (II.23). Em um modelo de memória associativa ou endereçável por conteúdo os protótipos devem ser pontos fixos atratores da dinâmica.

O primeiro teste do modelo deve ser quanto a estabilidade dos protótipos. Escolhamos os limiares U_i tal que $\alpha_i = 0$ na equação (II.3). A seguir façamos a hipótese de que os p protótipos sejam descorrelacionados, isto é:

$$\sum_{i=1}^N \xi_i^\mu \xi_i^\nu = N \delta_{\mu\nu} . \quad (\text{II.24})$$

Se o estado do sistema no instante t for um dos protótipos, digamos $\{\xi_i^l\}$, reescrevemos o "campo molecular" no sítio i , equação (II.7), como:

$$h_i(t) = \sum_j J_{ij} S_j(t) = N^{-1} \sum_{j \neq i} \sum_{\mu} \xi_i^\mu \xi_j^\mu \xi_j^l = \xi_i^l (1 + \delta_i) \quad (\text{II.25})$$

onde $\delta_i = N^{-1} \sum_{j \neq i} \sum_{\mu \neq \ell} \xi_i^\mu \xi_i^\ell \xi_j^\mu \xi_j^\ell$ é um passeio aleatório de passo $1/N$ com média zero e variância $(\delta_i^2)^{1/2}$ onde

$$\begin{aligned} \delta_i^2 &= N^{-2} \sum_{j \neq i} \sum_{k \neq i} \sum_{\mu \neq \ell} \sum_{\nu \neq \ell} \xi_i^\mu \xi_j^\mu \xi_j^\ell \xi_i^\nu \xi_k^\nu \xi_k^\ell \\ &= N^{-2} \sum_{j \neq i} \sum_{k \neq i} \sum_{\mu \neq \ell} \sum_{\nu \neq \ell} \delta_{kj} \delta_{\mu\nu} \\ &= \frac{(p-1)(N-1)}{N^2} \approx P/N. \end{aligned}$$

Assim, se $\alpha = P/N$ for pequeno, tal que $|\delta_i| < 1$ para qualquer neurônio teremos

$$S_i(t+\tau) = \text{sign}(h_i(t)) = \text{sign}(\xi_i^\ell) = \xi_i^\ell \quad i=1, \dots, N$$

(II.26)

A razão entre o número de protótipos e o número de neurônios (α) é conhecida como capacidade de armazenagem do modelo. O argumento acima mostra que deve existir algum α crítico acima do qual o modelo não é útil como memória associativa.

A seguir faremos uma estimativa numérica do valor limite de α , α_c , abaixo do qual o modelo é capaz de recuperar os protótipos. Em outras palavras, vamos calcular o número de informações que podem ser eficientemente memorizadas

por um sistema de N neurônios.

Escolhamos o estado inicial do sistema como sendo o protótipo $\{\xi_i^\ell\}$ e deixemos o sistema evoluir até atingir o estado de equilíbrio $\{S_i^{eq}\}$. Uma medida de capacidade do sistema de recuperar o protótipo $\{\xi_i^\ell\}$ é dada pela correlação de recuperação:

$$m^\ell = N^{-1} \sum_i S_i^{eq} \cdot \xi_i^\ell \quad (\text{II.27})$$

A figura 3 mostra a dependência de $m = \langle m^\ell \rangle_\xi$ com α para sistemas com 50, 100 e 200 neurônios. Aqui $\langle \dots \rangle_\xi$ indica uma média sobre $3 \cdot 10^3$ realizações de protótipos des-correlacionados tomados como configurações iniciais do sistema.

Fig.3- Simulações mostrando m como função de α .

Observe que para $\alpha < 0.2$ existe um platô onde o sistema consegue recuperar os protótipos com um erro menor que 5%. Esse erro de recuperação é definido pela equação

$$d^l = \frac{1 - m^l}{2}, \quad d = \langle d^l \rangle_\xi \quad (\text{II.28})$$

e indica a fração de neurônios que estão em estados diferentes nas configurações $\{S_i^{eq}\}$ e $\{\xi_i^l\}$.

A extrapolação para $N \rightarrow \infty$ de simulações com $N \leq 3 \cdot 10^3$ indica que ocorre uma transição descontínua para $\alpha = \alpha_c = 0.145 \pm 0.01$: a correlação de recuperação passa descontinuamente de $m \approx 0.97$ para $m \approx 0.45$ em $\alpha = \alpha_c$ [11]. Portanto, para $\alpha < 0.15$ o erro na recuperação dos protótipos é menor que 1.5%.

Os estados de equilíbrio para $\alpha \leq \alpha_c$ nunca são idênticos aos protótipos exceto para $\alpha = 0$, vamos denominá-los de estados de recuperação ou ferromagnéticos.

Para ilustrar o fato de que os estados de recuperação são pontos fixos atratores da dinâmica mostramos na figura 4 a evolução do sistema quando o estímulo é obtido invertendo-se 30% dos bits de um dos protótipos. O estado $S_i = +1$ é representado por um asterisco enquanto que $S_i = -1$ não é mostrado [12].

Fig.4- Evolução do sistema com $\alpha = 0.03$.

Mesmo para $\alpha < \alpha_c$ o estímulo deve estar suficientemente próximo a um dos protótipos para recuperá-lo, caso contrário o sistema atinge um dos estados vidro de spin com $m < 0.45$. Quanto mais próximo α estiver de α_c , mais próximo o estímulo deve estar de um dos protótipos a fim de recuperá-lo.

A deterioração catastrófica da memória do modelo quando o número de protótipos excede o valor crítico $\alpha_c N$ não possui paralelo em sistemas biológicos. Esses sistemas são capazes de memorizar continuamente novas informações através de um processo de esquecimento das memórias antigas, garantindo assim que o número de informações memorizadas nunca exceda o valor crítico. Mezard et al [18] implementaram esse processo de esquecimento no Modelo de Hopfield evitando assim a degradação da memória para $\alpha > \alpha_c$.

A questão natural levantada neste momento é se sistemas biológicos como o cérebro humano possuem uma capacidade limitada de memorização. Pelo menos três diferentes aspectos da operação de memorização do cérebro humano foram identificados [19].

Primeiramente há um sistema de memória que mantém a imagem detalhada, por uns poucos décimos de segundo, da informação sensorial enviada por algum órgão sensitivo. Esse sistema é conhecido por depósito de informação sensorial.

Um segundo sistema de memória mantém a informação já codificada por alguns segundos. É o estágio onde guardamos as informações que necessitamos temporariamente ou que estamos organizando para memorizar permanentemente. Esse é o sistema de memória de trabalho ou de curta duração. Experiências mostram que podemos memorizar entre 5 e 9 itens separados nesse sistema de memória, independente da natureza dos itens (números, letras, frases,...). Assim, esse sistema de memória possui uma capacidade de memorização limitada.

O terceiro sistema de memória é o de longa duração. É nesse sistema que são armazenadas as lembranças de nossas experiências. Sua capacidade de armazenagem é praticamente ilimitada.

Assim nosso modelo poderia descrever tanto a memória de curta como a de longa duração. Para a primeira apenas

uns poucos neurônios são necessários pois o número máximo de informações é pequeno enquanto que para a última provavelmente sejam necessários os 10^{12} neurônios disponíveis.

Podemos também perguntar o que ocorre quando o modelo memoriza repetidas vezes o mesmo padrão. Essa questão é abordada na referência [20] usando técnicas a serem desenvolvidas no capítulo seguinte. Encontramos que quando um padrão é memorizado mais do que 3 vezes ocorre uma mudança qualitativa no seu processo de esquecimento: a memória desse padrão degrada-se continuamente até desaparecer à medida que o sistema memoriza mais e mais padrões. Lembre que a memória de um padrão memorizado apenas uma vez desaparece descontinuamente em $\alpha = \alpha_c$.

Já foi mencionado que os estados de equilíbrio do sistema para $0 < \alpha < \alpha_c$ - estados de recuperação - não são idênticos aos protótipos. Argumentos baseados na Teoria de Probabilidades [21] mostram que os protótipos são pontos fixos atratores da dinâmica somente se

$$P < \frac{N}{2 \log N} \quad N \gg 1. \quad (\text{II.29})$$

Até esse momento estivemos tratando de protótipos descorrelacionados. A prescrição de Cooper não é muito eficiente para memorizar padrões correlacionados: o modelo apre

senta grande dificuldade em distinguir os protótipos e, para que isso ocorra, é necessário reduzir enormemente o número de protótipos do sistema. Esse problema é abordado na referência [22].

Existem prescrições menos realísticas do ponto de vista biológico (por exemplo, não locais) para a matriz sináptica em que é possível armazenar padrões correlacionados [23, 24], mas todas elas carecem da simplicidade da prescrição de Cooper.

Finalizando esse capítulo vamos dirigir nossa atenção para um problema prático no uso dos Modelos de Hopfield ou Little como memórias associativas [25]. Suponha que desejamos conhecer o conteúdo da memória de um desses modelos. O modo de recuperar os protótipos é estimular o modelo com padrões próximos aos protótipos. Entretanto, sem conhecimento prévio dos protótipos como saberemos se a resposta do sistema é um estado de recuperação (estímulo próximo a um protótipo) ou um estado vidro de spin (estímulo distante de todos os protótipos)?

A resposta seria simples se o modelo respondesse com diferentes comportamentos dinâmicos (pontos fixos ou ciclos) dependendo se o estímulo estivesse próximo ou não a um dos protótipos. Assim, produzindo estímulos aleatórios e observando a resposta do modelo poderíamos extrair o conteúdo

de sua memória. Na referência [26] apresentamos um modelo cujo desempenho é exatamente o descrito neste parágrafo.

CAPÍTULO III

A TERMODINÂMICA DO MODELO DE LITTLE

III. A TERMODINÂMICA DO MODELO DE LITTLE

Os estados estacionários do Modelo de Little com interações simétricas são governados por um Hamiltoniano derivado pela primeira vez por Peretto [27]. Entretanto, esses estados podem ser ciclos de período 2 e o desafio que surge é como uma abordagem de mecânica estatística de equilíbrio poderia descrever esses estados. Como veremos, a solução para essa questão surge de uma duplicação natural do espaço de fase do sistema.

Na secção III.A apresentamos a derivação de Peretto do Hamiltoniano do Modelo de Little e propomos um Hamiltoniano efetivo para esse modelo no qual aparece explicitamente a duplicação do espaço de fase acima mencionado.

As secções seguintes são devotadas ao estudo da termodinâmica do Modelo de Little. As técnicas utilizadas são as mesmas empregadas por Amit et al [11] em um estudo análogo do Hamiltoniano do Modelo de Hopfield

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

onde J_{ij} é dada pela prescrição de Cooper, equação (II.23).

III.A. O HAMILTONIANO

A evolução de um sistema formado por N unidades elementares i , $i=1, \dots, N$, no espaço de fase gerado pelos graus de liberdade x_i e seus momentos associados p_i é governada por um Hamiltoniano $H(\{x, p\})$ através das equações de movimento canônicas. Se N for muito grande (limite termodinâmico) essa abordagem é impraticável e inútil. Entretanto essa é a situação ideal para uma abordagem estatística na qual estudamos a evolução da função densidade de probabilidade $\rho(\{x, p\}, t)$ em lugar das trajetórias individuais das unidades que compõem o sistema. $\rho d\Omega$ nos dá a probabilidade do sistema estar no elemento de volume $d\Omega$, centrado no ponto $\{x, p\}$ do espaço de fase, no tempo t . No equilíbrio ρ é dada pela distribuição de Gibbs

$$\rho(\{x, p\}) = \rho(\{x, p\}, t=\infty) =$$

$$Z^{-1} \exp(-\beta H(\{x, p\})) \quad (\text{III.2})$$

onde Z é uma constante de normalização, a função de partição,

$$Z = \sum_{\{x, p\}} \exp(-\beta H(\{x, p\})) \quad (\text{III.3})$$

e β é um parâmetro do sistema.

Em um sistema de spins, a configuração de spins S_i , $i=1, \dots, N$, no tempo t desempenha o papel do par x_i, p_i da discussão acima.

Nosso objetivo é obter o Hamiltoniano (se existir) do Modelo de Little. Para isso vamos derivar uma equação de evolução para a função densidade de probabilidade baseada na dinâmica do modelo apresentada no capítulo anterior. Impondo condições para que a solução de equilíbrio tenha a forma (III.2) poderemos identificar o Hamiltoniano do modelo.

Seja $W(I, \tau/J, 0)$ a probabilidade para que no tempo $t = \tau$ o sistema esteja no estado I , dado que ele se encontrava no estado J no tempo imediatamente anterior $t=0$. τ é a nossa unidade de tempo. A probabilidade $\rho(I, (n+1)\tau)$ de o sistema estar no estado I no tempo $t = (n+1)\tau$, n inteiro, é dada então pela equação

$$\begin{aligned} \rho(I, (n+1)\tau) = & \rho(I, n\tau) + \sum_{J \neq I} W(I, \tau/J, 0) \rho(J, n\tau) \\ & - \sum_{K \neq I} W(K, \tau/I, 0) \rho(I, n\tau) \end{aligned} \quad (\text{III.4})$$

No equilíbrio temos $\rho(I, (n+1)\tau) = \rho(I, n\tau)$. Impondo que as probabilidades de transição W obedeçam o princípio do balanço detalhado

$$W(J, \tau/I, 0) \exp(-\beta H(I)) = W(I, \tau/J, 0) \exp(-\beta H(J)) \quad (\text{III.5})$$

obtemos que $\rho(I, t=\infty)$ é dada pela distribuição de Gibbs, equação (III.2).

Omitindo a dependência em t , reescrevemos a equação (III.5) como

$$\frac{W(I/J)}{W(J/I)} = \frac{F(I)}{F(J)} \quad (\text{III.6})$$

onde $F(I)$ relaciona-se com o Hamiltoniano do modelo através da equação

$$H(I) = -\beta^{-1} \log F(I) \quad (\text{III.7})$$

Usando a probabilidade de um neurônio disparar ou não, equação (II.6), vamos obter uma expressão explícita para as probabilidades de transição W em termos do estado do sistema $I = \{S_i\}$ e dos parâmetros β , J_{ij} e a_i .

Na dinâmica síncrona todos os neurônios evoluem si

multaneamente. Assim, o estado escolhido por um dado neurônio independe dos estados escolhidos pelos outros neurônios, isto é, as probabilidades (II.6) são descorrelacionadas. Podemos, então, escrever a probabilidade de transição $W(I/J)$ como um produto das probabilidades que governam a evolução de cada neurônio

$$W(I/J) = \prod_{i=1}^N [1 + \exp(-\beta S_i(I) h_i(J))]^{-1} \quad (\text{III.8})$$

onde
$$h_i(J) = \sum_j J_{ij} S_j(J) - \alpha_i$$

Após alguma manipulação algébrica escrevemos

$$W(I/J) = \frac{\exp(-\beta H(I/J))}{\sum_K \exp(-\beta H(K/J))} \quad (\text{III.9})$$

com

$$H(I/J) = \sum_{i,j} J_{ij} S_i(I) S_j(J) - \sum_i \alpha_i S_i(I) \quad (\text{III.10})$$

Agora é imediato constatar que as probabilidades de transição, eq. (III.9), só satisfarão a eq. (III.6) se $J_{ij} = J_{ji}$.

Supondo então que as interações entre os neurônios sejam simétricas obtemos

$$\frac{W(I/J)}{W(J/I)} = \frac{\exp(\beta \sum_i \alpha_i S_i(I)) \prod_i 2 \cosh \beta (\sum_j J_{ij} S_j(I) + \alpha_i)}{\exp(\beta \sum_i \alpha_i S_i(J)) \prod_i 2 \cosh \beta (\sum_j J_{ij} S_j(J) + \alpha_i)}$$

$$= \frac{F(I)}{F(J)} \quad (\text{III.11})$$

onde usamos a identidade

$$\prod_i^N 2 \cosh S_i X_i = \sum_{I=\{\sigma_i\}} \exp(\sum_i \sigma_i X_i)$$

$$\sigma_i, S_i = \pm 1 \quad (\text{III.12})$$

Assim, as equações (III.7) e (III.11) nos dão o Hamiltoniano do Modelo de Little

$$H(I) = - \sum_i \alpha_i S_i(I)$$

$$- \frac{1}{\beta} \sum_i \log 2 \cosh \beta (\sum_j J_{ij} S_j(I) + \alpha_i)$$

$$(\text{III.13})$$

Note a complexidade desse Hamiltoniano quando comparado com o do modelo de Hopfield, eq. (III.1). Entretanto, sua termodinâmica é idêntica à do Hamiltoniano contendo as variáveis adicionais σ_i

$$H(S, \sigma) = - \sum_{ij} J_{ij} S_i \sigma_j - \sum_i \alpha_i (S_i + \sigma_i)$$

$$S_i, \sigma_i = \pm 1 \quad (\text{III.14})$$

pois possuem a mesma função de partição

$$Z = \text{Tr}_{S_i} \exp(-\beta H(S)) = \text{Tr}_{S_i} \text{Tr}_{\sigma_i} \exp(-\beta H(S, \sigma)) \quad (\text{III.15})$$

onde introduzimos a notação $\text{Tr}_{S_i} = \sum_{\{S_i\}}$. Essa igualdade é mostrada facilmente através da identidade (III.12).

Denominaremos $H(S, \sigma)$ de Hamiltoniano efetivo do Modelo de Little.

A fim de obtermos uma interpretação física para a variável de spin σ_i , tomemos o limite $\beta \rightarrow \infty$ da equação (III.13):

$$\begin{aligned} H(t) &= - \sum_i | \sum_j J_{ij} S_j(t) + \alpha_i | - \sum_i \alpha_i S_i(t) \\ &= - \sum_{ij} J_{ij} S_j(t) S_i(t+1) \\ &\quad - \sum_i \alpha_i (S_i(t) + S_i(t+1)) \end{aligned} \quad (\text{III.16})$$

onde usamos $|x| = x \text{ sign } x$ e a equação de evolução determinística (II.2) para eliminar a função sinal.

Comparando as equações (III.14) e (III.16) constatamos que as configurações $\{S_i\}$ e $\{\sigma_i\}$ representam estados do sistema em tempos consecutivos. Percebemos então que mecânica estatística de equilíbrio, em que não há referência a uma variável temporal, descreve um sistema contendo ciclos (para cuja caracterização o tempo é essencial) aumentando n vezes o número de graus de liberdade, onde n é o comprimento do ciclo.

III.B. A ENERGIA LIVRE

Obtido o Hamiltoniano efetivo do Modelo de Little dediquemo-nos agora a estudar sua termodinâmica.

As interações entre os neurônios são dadas pela prescrição de Cooper

$$J_{ij} = \begin{cases} N^{-1} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} & i \neq j \\ J_0 & i = j \end{cases} \quad (\text{III.17})$$

onde introduzimos um termo de autointeração dos neurônios. Note que a contribuição desse termo ao Hamiltoniano (III.14) não é trivial. De acordo com a nossa interpretação das variáveis S_i e σ_i esperamos que $J_0 < 0$ ($J_0 > 0$) favoreça ciclos-2 (pontos fixos).

Vamos supor que os ξ_i^μ 's sejam variáveis aleatórias estatisticamente independentes e fixas para fins de médias térmicas, distribuídas de acordo com a equação

$$P\{\xi_i^\mu\} = \prod_{i,\mu} p(\xi_i^\mu) = 2^{-Np} \prod_{i,\mu} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)] \quad (\text{III.18})$$

Suporemos também que o número de protótipos (p) seja proporcional ao número de neurônios do sistema:

$$p = \alpha N \quad (\text{III.19})$$

Uma vez que a interação (III.17) é de alcance infinito podemos, em princípio, resolver o modelo exatamente. Por simplicidade, escolhamos os limiares U_i na eq.(II.3) de modo a termos $u_i = 0$.

Observe que para $p \rightarrow \infty$ os $J_{ij}(i/j)$ são distribuídos por uma Gaussiana pelo teorema do limite central. Entretanto, existem correlações entre os sítios que tornam esse problema diferente do Modelo de Sherrington-Kirkpatrick (SK) [9]. Por exemplo o laço de frustração

$$\langle \sum_{j \neq k} J_{ij} J_{jk} J_{ki} \rangle_0 = \alpha \quad j,k/i$$

não é nulo como acontece no Modelo SK.

Conforme já vimos, os parâmetros de ordem que caracterizam as fases onde o sistema atua como uma memória associativa são as correlações entre os protótipos e os estados de equilíbrio do sistema

$$m^\mu = N^{-1} \sum_i \xi_i^\mu \langle S_i \rangle_T \quad n^\mu = N^{-1} \sum_i \xi_i^\mu \langle \sigma_i \rangle_T$$

$$\mu=1, \dots, p \quad (\text{III.20})$$

onde $\langle \dots \rangle_T$ indica uma média térmica. Se m^k e n^k são não nulos então o sistema reconheceu o protótipo $\{\xi^k\}$.

Em termos desses parâmetros escrevemos a energia por neurônio, no limite $T \rightarrow 0$, como

$$E = - \sum_{\mu=1}^p m^\mu n^\mu + (\alpha - J_0) \frac{1}{N} \sum_i S_i \sigma_i \quad (\text{III.21})$$

Como mostrado na secção II.C podemos dividir o sistema em dois subsistemas S_1 e S_2 contendo N_1 neurônios fixos e N_2 neurônios oscilantes respectivamente. Assim, a energia para $T=0$ fica escrita como

$$E = E_1 + E_2 \quad (\text{III.22a})$$

onde

$$E_1 = - \sum_{\mu} (m_1^\mu)^2 + (\alpha - J_0) \frac{N_1}{N} \quad (\text{III.22b})$$

$$E_2 = + \sum_{\mu} (m_2^{\mu})^2 - (\alpha - J_0) \frac{N_2}{N} \quad (\text{III.22c})$$

com
$$m_1^{\mu} = \frac{1}{N} \sum_{i \in S_1} S_i \xi_i^{\mu} \quad e$$

$$m_2^{\mu} = \frac{1}{N} \sum_{i \in S_2} S_i \xi_i^{\mu}$$

Uma vez que esses subsistemas são independentes, podemos tratá-los separadamente. Ambas as energias E_1 e E_2 devem ser finitas. A condição de que E_1 seja finita implica na impossibilidade de descrevermos, através de um cálculo termodinâmico, estados que possuam correlações macroscópicas com todos os protótipos. Esses estados, cujo número cresce exponencialmente com N , são ditos estados de não equilíbrio e desempenham um papel fundamental na dinâmica do sistema [28]. Voltaremos a esse ponto no capítulo seguinte. Por outro lado, vemos que as soluções que minimizam E_2 devem possuir $m_2^{\mu} = 0$.

Essa subdivisão em dois subsistemas aparece naturalmente em nossos cálculos: as equações que descrevem os estados estacionários do sistema para $T=0$ possuem duas classes de soluções correspondendo aos pontos fixos e ciclos-2 completos (todos os neurônios participam do ciclo).

Trataremos então de estados que possuam correlações finitas apenas com um número finito s de protótipos.

A grandeza de interesse a ser calculada é a energia livre média por neurônio

$$f = - \lim_{N \rightarrow \infty} \frac{1}{N \beta} \langle \log Z \rangle_{\xi} \quad (\text{III.23a})$$

onde

$$Z = \text{Tr}_{\{S, \sigma\}} \exp(\beta \sum_{ij} J_{ij} S_i \sigma_j) \quad (\text{III.23b})$$

e $\langle \dots \rangle_{\xi}$ indica uma média sobre as realizações dos protótipos. Para calcularmos essa média usamos o método das réplicas [8]

$$\langle \log Z \rangle_{\xi} = \lim_{n \rightarrow 0} \frac{\langle Z^n \rangle_{\xi} - 1}{n} \quad (\text{III.24})$$

onde Z^n é a função de partição replicada n vezes com n positivo e inteiro. Como usual, suporemos válida a extensão para n real e a troca dos limites $N \rightarrow \infty$ e $n \rightarrow 0$.

Introduzindo as variáveis replicadas $S_i^{\rho}, \sigma_i^{\rho}$, $\rho = 1, \dots, n$, escrevemos $\langle Z^n \rangle_{\xi}$ como

$$\langle Z^n \rangle_{\xi} = \text{Tr}_{\{S_i^{\rho}, \sigma_i^{\rho}\}} \exp[\beta (J_0 - \alpha) \sum_{\rho} S_i^{\rho} \sigma_i^{\rho}] Z^{\mu} Z^{\nu} \quad (\text{III.25})$$

onde

$$Z^\phi = \int \frac{d\mathbf{m}_\rho^\theta d\mathbf{n}_\rho^\theta d\mathbf{l}_\rho^\theta}{(2\pi)^{3/2}} \exp \left\{ -\frac{1}{2} \sum_{\theta \rho} (m_\rho^\theta)^2 + (n_\rho^\theta)^2 + (l_\rho^\theta)^2 \right\} +$$

$$+ (\beta/N)^{1/2} \sum_{\theta \rho i} \xi_i^\theta \left[(l_\rho^\theta + i m_\rho^\theta) \sigma_i^\rho + (l_\rho^\theta + i n_\rho^\theta) S_i^\rho \right] \right\} \quad (III.26)$$

com $Z^\phi = Z^\nu$ para $\phi < s$ e $Z^\phi = Z^\mu$ para $\phi > s$.

Tratemos primeiramente de Z^μ . Calculando a média sobre ξ_i^μ obtemos o fator

$$\% = \exp \left\{ \sum_{i\mu} \log \cosh \left[(\beta/N)^{1/2} \sum_{\rho} [(l_\rho^\mu + i m_\rho^\mu) \sigma_i^\rho + (l_\rho^\mu + i n_\rho^\mu) S_i^\rho] \right] \right\} \quad (III.27)$$

Expandindo o log cosh e mantendo termos de ordem N^{-1} obtemos

$$\% = \exp \left\{ \beta/2N \sum_{i\mu} \left[\sum_{\rho} (l_\rho^\mu + i m_\rho^\mu) \sigma_i^\rho + (l_\rho^\mu + i n_\rho^\mu) S_i^\rho \right]^2 \right\} \quad (III.28)$$

Obteríamos esse mesmo resultado se considerássemos os ξ_i^μ 's distribuídos por uma Gaussiana de média zero e variância unitária. De fato, esse ruído Gaussiano gerado pelas correlações de ordem $N^{-1/2}$ é o responsável pela deterioração da capacidade de memorização do modelo.

Uma vez que pretendemos descrever os estados de equilíbrio do sistema através de um número finito de parâmetros de ordem será necessário eliminar os parâmetros que de-

pendam do índice μ . Para isso devemos efetuar as integrações em m_{ρ}^{μ} , n_{ρ}^{μ} e ℓ_{ρ}^{μ} na expressão de Z^{μ} . Obtemos após um cálculo tedioso, mas sem grandes dificuldades:

$$\begin{aligned}
 Z^{\mu} = \int & \left[\prod_{\rho\sigma} dr_{\rho\sigma} d\tilde{r}_{\rho\sigma} dp_{\rho\sigma} d\tilde{p}_{\rho\sigma} dq_{\rho\sigma} d\tilde{q}_{\rho\sigma} \right] \times \\
 & \times \exp \left\{ - \frac{N\alpha\beta^2}{2} \left[\beta^{-2} \log \det [(\underline{1} - \beta \underline{Q})^2 - \beta^2 \underline{R} \underline{P} + \right. \right. \\
 & \left. \left. \sum_{\rho\sigma} (r_{\rho\sigma} \tilde{r}_{\rho\sigma} + p_{\rho\sigma} \tilde{p}_{\rho\sigma} + q_{\rho\sigma} \tilde{q}_{\rho\sigma}) - \right. \right. \\
 & \left. \left. \frac{1}{N} \sum_{i\rho\sigma} (\tilde{p}_{\rho\sigma} \sigma_i^{\rho} \sigma_i^{\sigma} + \tilde{r}_{\rho\sigma} S_i^{\rho} S_i^{\sigma} + \tilde{q}_{\rho\sigma} S_i^{\rho} \sigma_i^{\sigma}) \right] \right\}
 \end{aligned}
 \tag{III.29}$$

onde $\underline{Q}$, $\underline{P}$, $\underline{R}$ e $\underline{1}$ são matrizes simétricas, $n \times n$, dadas por

$$\begin{aligned}
 Q_{\rho\sigma} &= q_{\rho\sigma} & P_{\rho\sigma} &= p_{\rho\sigma} & R_{\rho\sigma} &= r_{\rho\sigma} & 1_{\rho\sigma} &= \delta_{\rho\sigma}
 \end{aligned}
 \tag{III.30}$$

com $p_{\rho\sigma}$, $\tilde{p}_{\rho\sigma}$, $r_{\rho\sigma}$, $\tilde{r}_{\rho\sigma}$ definidos apenas para $\rho \neq \sigma$. Omitimos também constantes multiplicativas que não alteram o limite termodinâmico.

Tratemos agora de Z^{ν} . A fim de obtermos um limite termodinâmico bem definido façamos a seguinte reescala:

$$\begin{pmatrix} m_{\rho}^v \\ n_{\rho}^v \\ \ell_{\rho}^v \end{pmatrix} + (N\beta)^{1/2} \begin{pmatrix} \tilde{m}_{\rho}^v \\ \tilde{n}_{\rho}^v \\ \tilde{\ell}_{\rho}^v \end{pmatrix} \quad (\text{III.31})$$

Assim,

$$Z^v = \int \prod_{\rho} dm_{\rho}^v dn_{\rho}^v d\ell_{\rho}^v \exp \left\{ -N\beta \left[\frac{1}{2} \sum_{\rho} (\ell_{\rho}^v)^2 - (m_{\rho}^v)^2 - (n_{\rho}^v)^2 \right] - \right. \\ \left. - \frac{1}{N} \sum_{i \neq \rho} \xi_i^v [(\ell_{\rho}^v - m_{\rho}^v) \sigma_i^{\rho} + (\ell_{\rho}^v - n_{\rho}^v) S_i^{\rho}] \right\} \quad (\text{III.32})$$

A fim de eliminarmos a dependência no índice i usemos a identidade

$$\text{Tr}_{S_i} \exp \left\{ \sum_i F(\xi_i^v) S_i \right\} = \\ = \exp \left\{ N \left[\frac{1}{N} \sum_i \log \text{Tr}_S \exp(F(\xi_i^v) S) \right] \right\} \quad (\text{III.33})$$

Entretanto, pode-se mostrar que grandezas que dependem apenas de um número finito de ξ_i^v 's são auto-médiantes, isto é, no limite termodinâmico as flutuações em torno do valor médio são suprimidas [29, 30]. Podemos então, substituir as somas $\frac{1}{N} \sum_i$ por médias sobre os ξ_i^v 's. Daí

$$\text{Tr}_S \exp \left\{ \sum_i F(\xi_i^v) S \right\} =$$

$$= \exp \{ N \langle \log \text{Tr}_S \exp[F(\xi^\nu)S] \rangle_{\xi^\nu} \} \quad (\text{III.34})$$

Juntando as expressões de Z^μ e Z^ν obtemos

$$\langle Z^\mu \rangle_\xi = \int \prod_{\nu\rho} d\mathbf{m}_\rho^\nu d\mathbf{n}_\rho^\nu d\mathbf{l}_\rho^\nu \int \prod_{\rho\sigma} d\mathbf{r}_{\rho\sigma} d\tilde{\mathbf{r}}_{\rho\sigma} d\mathbf{p}_{\rho\sigma} d\tilde{\mathbf{p}}_{\rho\sigma} d\mathbf{q}_{\rho\sigma} d\tilde{\mathbf{q}}_{\rho\sigma} \times$$

$$\times \exp \{ N\beta \left[-\frac{1}{2} \sum_{\nu\rho} (\mathbf{l}_\rho^\nu)^2 - (\mathbf{m}_\rho^\nu)^2 - (\mathbf{n}_\rho^\nu)^2 - \right.$$

$$\left. \frac{\alpha}{2\beta} \log \det [(1 - \beta \underline{Q})^2 - \beta^2 \underline{R} \underline{P}] + (J_0 - \alpha) \sum_{\rho} q_{\rho\rho} - \right.$$

$$\left. \frac{\alpha\beta}{2} \sum_{\rho\sigma} (r_{\rho\sigma} \tilde{\mathbf{r}}_{\rho\sigma} + p_{\rho\sigma} \tilde{\mathbf{p}}_{\rho\sigma} + q_{\rho\sigma} \tilde{\mathbf{q}}_{\rho\sigma}) \right.$$

$$\left. + \beta^{-1} \langle \log \text{Tr}_{S,\sigma} \exp \beta H_\xi \rangle_{\xi^\nu} \right\} \quad (\text{III.35})$$

onde

$$H_\xi = \frac{\alpha\beta}{2} \sum_{\rho\sigma} (\tilde{\mathbf{p}}_{\rho\sigma} \cdot \sigma^\rho \sigma^\sigma + \tilde{\mathbf{r}}_{\rho\sigma} \cdot \mathbf{S}^\rho \mathbf{S}^\sigma + \tilde{\mathbf{q}}_{\rho\sigma} \cdot \mathbf{S}^\rho \sigma^\sigma) +$$

$$\vec{\xi} \cdot \sum_{\rho} [(\vec{\ell}_\rho - \vec{m}_\rho) \sigma^\rho + (\vec{\ell}_\rho - \vec{n}_\rho) \mathbf{S}^\rho]$$

$$(\text{III.36})$$

com a notação $\vec{m} \cdot \vec{\xi} = \sum_{\nu} m^\nu \cdot \xi^\nu$.

No limite $N \rightarrow \infty$ o integrando é dominado pelo seu ponto de sela, fornecendo a seguinte energia livre por neurônio

$$f = \lim_{n \rightarrow 0} \frac{1}{n} \text{Min} \left\{ \frac{1}{2} \sum_{\rho} (\vec{\ell}_{\rho}^2 - \vec{m}_{\rho}^2 - \vec{n}_{\rho}^2) + \frac{\alpha}{2\beta} \Delta' + \right. \\ \left. + \frac{1}{2} \alpha \beta \sum_{\rho, \sigma} (r_{\rho\sigma} \tilde{r}_{\rho\sigma} + p_{\rho\sigma} \tilde{p}_{\rho\sigma} + q_{\rho\sigma} \tilde{q}_{\rho\sigma}) - \right. \\ \left. \beta^{-1} \langle \log \text{Tr}_{S, \sigma} \exp \beta H_{\xi} \rangle_{\xi v} \right\} \quad (\text{III.37})$$

com

$$\Delta' = \log \det [(1 - \beta \underline{Q})^2 - \beta^2 \underline{P} \underline{R}] \quad (\text{III.38})$$

Os estados estacionários do sistema são descritos pelos parâmetros de ordem que minimizam a expressão entre as chaves na eq. (III.37). Entretanto, encontrar o mínimo dessa expressão para n geral é tarefa praticamente impossível, de forma que nos limitaremos a propor uma solução particular, a mais simples possível, e então testar a sua viabilidade física.

III.C. O ANSATZ SIMETRIA DE REPLICAS

Supondo que as réplicas sejam indistinguíveis, propomos a seguinte parametrização

$$\left. \begin{aligned} p_{\rho\sigma} &= p & r_{\rho\sigma} &= r & q_{\rho\sigma} &= q_1 \\ \tilde{p}_{\rho\sigma} &= \tilde{p} & \tilde{r}_{\rho\sigma} &= \tilde{r} & \tilde{q}_{\rho\sigma} &= \tilde{q}_1 \end{aligned} \right\} \rho \neq \sigma$$

$$\begin{aligned} q_{\rho\rho} &= q_0 & \tilde{q}_{\rho\rho} &= \tilde{q}_0 \\ m^\rho &= m & n^\rho &= n & \ell^\rho &= \ell \end{aligned} \quad (\text{III.39})$$

com a qual podemos calcular explicitamente o limite $n \rightarrow 0$ na equação (III.37). Dediquemo-nos então a esse cálculo.

A parametrização (III.39) permite-nos escrever os elementos da matriz

$$\underline{E} = (1 - \beta \underline{Q})^2 - \beta^2 \underline{R} \cdot \underline{P} \quad (\text{III.40})$$

que aparece na expressão de Δ' como

$$E_{\alpha\beta} = E_0 \delta_{\alpha\beta} + E_1 (1 - \delta_{\alpha\beta}) \quad (\text{III.41})$$

onãe

$$E_0 = (1 - \beta q_0)^2 - \beta^2 (1 - pr + q_1^2) \quad (\text{III.42a})$$

$$E_1 = 2\beta [\beta q_1 (q_0 - q_1) - q_1 - \frac{\beta}{2} (r(1-p) + p(1-r))] \quad (\text{III.42b})$$

$\underline{E}$ possui um autovalor igual a $E_0 + (n-1)E_1$ e $n-1$ autovalores iguais a $E_0 - E_1$. Portanto

$$\det \underline{E} = (E_0 - E_1)^{n-1} [E_0 + (n-1)E_1] \quad (\text{III.43})$$

e

$$\Delta \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \Delta' = \lim_{n \rightarrow \infty} \frac{1}{n} \log \det \underline{E} = \log(E_0 - E_1) +$$

$$\frac{E_1}{E_0 - E_1} \quad (\text{III.44})$$

Usando os valores de E_0 e E_1 dados pelas equações (III.42a-b) obtemos:

$$\Delta = \log [(1-Q)^2 - PR] - \beta \frac{2q_1(1-Q) + rP + pR}{(1-Q)^2 - PR} \quad (\text{III.45})$$

onde

$$Q = \beta(q_0 - q_1) \quad P = \beta(1-p) \quad R = \beta(1-r) \quad (\text{III.46})$$

Lembrando que $r_{\rho\sigma}$, $\tilde{r}_{\rho\sigma}$, $p_{\rho\sigma}$, $\tilde{p}_{\rho\sigma}$ são definidos apenas para $\rho \neq \sigma$ é fácil verificar que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\rho \neq \sigma} (\tilde{r}_{\rho\sigma} r_{\rho\sigma} + \tilde{p}_{\rho\sigma} p_{\rho\sigma} + \tilde{q}_{\rho\sigma} q_{\rho\sigma}) =$$

$$\tilde{q}_0 q_0 - \tilde{q}_1 q_1 - \tilde{p} p - \tilde{r} r \quad (\text{III.47})$$

Finalmente, devemos calcular o termo

$$\% = \lim_{n \rightarrow 0} \frac{1}{n} \langle \log \text{Tr}_{S^{\rho} \sigma^{\rho}} \exp \beta H_{\xi} \rangle_{\xi} \quad (\text{III.48})$$

Usando a parametrização (III.39), reescrevemos H_{ξ} como

$$\begin{aligned} H_{\xi} = \frac{\alpha\beta}{2} [- n(\tilde{p}+r) + (\tilde{p}-\tilde{q}_1/2) (\sum_{\rho} \sigma^{\rho})^2 + (\tilde{r}-\tilde{q}_1/2) (\sum_{\rho} S^{\rho})^2 \\ + \tilde{q}_1/2 (\sum_{\rho} S^{\rho} + \sigma^{\rho})^2 + (\tilde{q}_0 - \tilde{q}_1) \sum_{\rho} S^{\rho} \sigma^{\rho} + \\ + \vec{\xi} \cdot (\vec{\ell} - \vec{m}) \sum_{\rho} \sigma^{\rho} + \vec{\xi} \cdot (\vec{\ell} - \vec{n}) \sum_{\rho} S^{\rho}] \quad (\text{III.49}) \end{aligned}$$

Linearizando os termos quadráticos nas variáveis de spin através de transformações Gaussianas podemos calcular facilmente os traços sobre S^{ρ} e σ^{ρ} . Apresentamos o resultado final (Apêndice A):

$$\begin{aligned} \% = -\frac{\alpha\beta^2}{2} (\tilde{p}+\tilde{r}-\tilde{q}_0 + \tilde{q}_1) + \langle \log 2 [\cosh \beta \Xi_1 + \\ e^{-\alpha\beta\tilde{Q}} \cosh \beta \Xi_2] \rangle \quad (\text{III.50}) \end{aligned}$$

onde

$$\begin{aligned} \Xi_1 = (2\alpha \tilde{q}_1)^{1/2} Z_1 + [\alpha (\tilde{p}-\tilde{q}_1/2)]^{1/2} Z_2 + \\ [\alpha (\tilde{r}-\tilde{q}_1/2)]^{1/2} Z_3 + \vec{\xi} \cdot (2\vec{\ell} - \vec{m} - \vec{n}) \quad (\text{III.51a}) \end{aligned}$$

$$\Xi_2 = [\alpha (\tilde{p} - \tilde{q}_1/2)]^{1/2} Z_2 - [\alpha (\tilde{r} - \tilde{q}_1/2)]^{1/2} Z_3 +$$

$$\xi \cdot (\vec{m} - \vec{n}) \quad (\text{III.51b})$$

$$\tilde{Q} = \beta (\tilde{q}_0 - \tilde{q}_1) \quad (\text{III.51c})$$

e $\langle \dots \rangle$ indica uma média sobre os ξ^v 's e sobre três variáveis Gaussianas Z_1, Z_2, Z_3 de média zero e variância unitária.

Podemos então escrever a energia livre média por neurônio simétrica em relação às réplicas como

$$f = \frac{1}{2} (\vec{\ell}^2 - \vec{m}^2 - \vec{n}^2) + \frac{\alpha}{2\beta} [\log [(1-Q)^2 - PR] -$$

$$\beta \frac{2q_1(1-Q) + rP + pR}{(1-Q)^2 - PR} +$$

$$+ \frac{1}{2} \alpha \beta [\tilde{p}(1-p) + \tilde{r}(1-r) + \tilde{q}_1(1-q_1) - \tilde{q}_0(1-q_0) +$$

$$+ (\alpha - J_0)q_0 - \beta^{-1} \log 2 [\cosh \beta \Xi_1 +$$

$$e^{-\alpha \beta \tilde{Q}} \cosh \beta \Xi_2] \rangle \quad (\text{III.52})$$

Variando-se f com relação aos parâmetros de ordem obtemos as equações que descrevem os estados estacionários do sistema. O significado físico dos parâmetros de ordem sai das equações de ponto de sela. Temos:

- i) As correlações macroscópicas entre estados de equilíbrio e protótipos

$$m^v = \langle \xi^v \langle S \rangle_T \rangle_{\xi^v}$$

$$n^v = \langle \xi^v \langle \sigma \rangle_T \rangle_{\xi^v} \quad (\text{III.53})$$

$$l^v = m^v + n^v \quad \text{com } v < s$$

- ii) Os parâmetros de ordem do tipo Edwards-Anderson que medem a "rigidez" dos neurônios

$$q_1 = \frac{1}{N} \langle \sum_i \langle S_i \rangle_T \langle \sigma_i \rangle_T \rangle_{\xi}$$

$$r = \frac{1}{N} \langle \sum_i \langle \sigma_i \rangle_T^2 \rangle_{\xi} \quad (\text{III.54})$$

$$p = \frac{1}{N} \langle \sum_i \langle S_i \rangle_T^2 \rangle_{\xi}$$

- iii) A correlação entre estados do sistema em tempos consecutivos

$$q_0 = \frac{1}{N} \langle \sum_i \langle S_i \sigma_i \rangle_T \rangle_{\xi} \quad (\text{III.55})$$

iv) As médias quadradas das correlações entre os estados de equilíbrio e os p-s protótipos

$$\tilde{q}_1 = \frac{2}{\alpha N^2} \left\langle \sum_{ij\mu} \xi_i^\mu \xi_j^\mu \left\langle S_i \right\rangle_T \left\langle \sigma_j \right\rangle_T \right\rangle_{\xi^\mu}$$

$$\tilde{q}_0 = \frac{2}{\alpha N^2} \left\langle \sum_{ij\mu} \xi_i^\mu \xi_j^\mu \left\langle S_i \sigma_j \right\rangle_T \right\rangle_{\xi^\mu}$$

$$\tilde{p} = \frac{1}{\alpha N^2} \left\langle \sum_{ij\mu} \xi_i^\mu \xi_j^\mu \left\langle \sigma_i \right\rangle_T \left\langle \sigma_j \right\rangle_T \right\rangle_{\xi^\mu}$$

$$\tilde{r} = \frac{1}{\alpha N^2} \left\langle \sum_{ij\mu} \xi_i^\mu \xi_j^\mu \left\langle S_i \right\rangle_T \left\langle S_j \right\rangle_T \right\rangle_{\xi^\mu} \quad (\text{III.56})$$

com $\mu > s$.

Esses parâmetros de ordem satisfazem o seguinte sistema de equações:

$$\left\{ \begin{matrix} n^v \\ m^v \end{matrix} \right\} = \left\langle \xi^v \left[\sinh \beta_{\equiv_1} \pm e^{-\alpha \beta \tilde{Q}} \sinh \beta_{\equiv_2} \right] / D \right\rangle \quad (\text{III.57a})$$

$$l^v = m^v + n^v \quad v=1, \dots, s \quad (\text{III.57b})$$

$$\left(\begin{matrix} p \\ r \end{matrix} \right) = \left\langle \left(\sinh \beta_{\equiv_1} \pm e^{-\alpha \beta \tilde{Q}} \sinh \beta_{\equiv_2} \right)^2 / D^2 \right\rangle \quad (\text{III.57c})$$

$$q_0 = \left\langle \left(\cosh \beta_{\equiv_1} - e^{-\alpha \beta \tilde{Q}} \cosh \beta_{\equiv_2} \right) / D \right\rangle \quad (\text{III.57d})$$

$$q_1 = \langle (\sinh^2 \beta \Xi_1 - e^{-2\alpha\beta\tilde{Q}} \sinh^2 \beta \Xi_2) / D^2 \rangle \quad (\text{III.57e})$$

$$\tilde{r} = [p(1-Q)^2 + 2q_1 p(1-Q) + rP^2] / \tilde{D}^2 \quad (\text{III.57f})$$

$$\tilde{p} = [r(1-Q)^2 + 2q_1 r(1-Q) + pR^2] / \tilde{D}^2 \quad (\text{III.57g})$$

$$\tilde{q}_1/2 = [q_1(1-Q)^2 + (rP + pR)(1-Q) + q_1 RP] / \tilde{D}^2 \quad (\text{III.57h})$$

$$\tilde{Q} = \beta (\tilde{q}_0 - \tilde{q}_1) = 2 \{ [Q(1-Q) + RP] / \tilde{D} + J_0/\alpha \} \quad (\text{III.57i})$$

onde

$$D = \cosh \beta \Xi_1 + e^{\alpha\beta\tilde{Q}} \cosh \beta \Xi_2 \quad (\text{III.57j})$$

$$\tilde{D} = (1-Q)^2 - RP \quad (\text{III.57k})$$

Observe que essas equações são invariantes à troca $p \leftrightarrow r$ ou $m^v \leftrightarrow n^v$. Essa simetria é consequência da simetria $S \leftrightarrow \sigma$ do Hamiltoniano efetivo. Assim, $p=r$ e $m^v = n^v$ sempre será solução do sistema de equações (III.57). A questão é se existe uma outra solução com $p \neq r$ ou $m^v \neq n^v$ correspondendo então à quebra espontânea da simetria $S \leftrightarrow \sigma$.

Se essa quebra espontânea de simetria ocorrer de fato, deveremos poder observá-la no limite $T \rightarrow 0$. Entretanto, sabemos que nesse limite o sistema pode ser dividido em dois subsistemas independentes S_1 e S_2 . O primeiro, correspon-

te aos pontos fixos, não quebra a simetria $S \leftrightarrow \sigma$: $m_1^v = n_1^v$ e $p_1 = r_1$. O segundo, correspondente aos ciclos-2 completos, aparentemente quebra essa simetria: $m_2^v = -n_2^v$, porém, em um ciclo completo a energia é minimizada por estados em que $m_2^v = -n_2^v = 0$. Lembrando que p_2 e r_2 são quadráticos em $\langle S \rangle_T$ e $\langle \sigma \rangle_T$ teremos sempre $p_2 = r_2$. Assim, no sistema completo a quebra espontânea da simetria $S \leftrightarrow \sigma$ também não ocorre: $m^v = n^v$ e $p = p_1 + p_2 = r_1 + r_2 = r$.

Nas secções seguintes resolveremos explicitamente o sistema (III.57) nos casos limites $T=0$ e $\alpha=0$ e poderemos então comprovar a validade desse argumento.

III.D. OS ESTADOS ESTACIONÁRIOS PARA $T=0$

O sinal de $\tilde{Q}$ desempenha um papel crucial no estudo do limite $\beta \rightarrow \infty$ das equações dos parâmetros de ordem. Analisemos os três casos possíveis.

a) $\tilde{Q} > 0$

Temos $e^{-\alpha\beta\tilde{Q}} \rightarrow 0$ de modo que o sistema (III.57) é reescrito como

$$m^v = n^v = \lim_{\beta \rightarrow \infty} \langle \xi^v \tanh \beta \xi_1 \rangle \quad v=1, \dots, s$$

$$p = r = q_1 = \lim_{\beta \rightarrow \infty} \langle \tanh^2 \beta \xi_1 \rangle, \quad q_0 = 1$$

$$\tilde{q}_1/2 = \tilde{p} = \tilde{r} = q_1/(1-2c)^2 \quad (\text{III.58a})$$

$$\tilde{Q} = 2c/(1-2c) + 2J_0/\alpha > 0 \quad (\text{III.58b})$$

onde

$$c \equiv \lim_{\beta \rightarrow \infty} \beta(1-q_1) \quad e \quad \Xi_1 = (2\alpha\tilde{q}_1)^{1/2} Z + 2\vec{\xi} \cdot \vec{m}$$

Esse caso corresponde a pontos fixos da dinâmica, como podemos constatar pelo resultado $q_0=1$.

A fim de calcularmos os limites $\beta \rightarrow \infty$ usemos as seguintes identidades

$$\int_{-\infty}^{\infty} \frac{\tilde{\alpha} z}{(2\pi)^{1/2}} e^{-\frac{1}{2}z^2} \operatorname{tgh} \beta(Az+B) \xrightarrow{\beta \rightarrow \infty} \\ (2/\pi)^{1/2} \int_0^{B/A} \tilde{\alpha} z e^{-\frac{1}{2}z^2} \equiv \operatorname{erf}(B/\sqrt{2} A)$$

$$\int_{-\infty}^{\infty} \frac{\tilde{\alpha} z}{(2\pi)^{1/2}} e^{-\frac{1}{2}z^2} \operatorname{tgh}^2 \beta(Az+B) \xrightarrow{\beta \rightarrow \infty} \\ 1 - \frac{1}{\beta A} (2/\pi)^{1/2} e^{-B^2/2A} \quad (\text{III.59})$$

com as quais é imediato obter

$$m^v = \langle \xi^v \operatorname{erf} [\vec{m} \cdot \vec{\xi} / (\alpha \tilde{q}_1)^{1/2}] \rangle_{\xi^v} \quad (\text{III.60a})$$

$$c = (\pi \alpha \tilde{q}_1)^{-1/2} < \exp[-(\vec{m} \cdot \vec{\xi})^2 / \alpha \tilde{q}_1] >_{\xi} \quad (\text{III.60b})$$

$$\tilde{q}_1 = 2/(1-2c)^2 \quad (\text{III.60c})$$

Essas equações são idênticas às obtidas por Amit et al [11] para o limite $T \rightarrow 0$ do Modelo de Hopfield. Isso já era esperado uma vez que a equação (III.13) pode ser reescrita, no limite $T \rightarrow 0$, como

$$H = - \sum_{ij} J_{ij} S_i(t) S_j(t+1) = - \sum_{ij} J_{ij} S_i(t) S_j(t) \quad (\text{III.61})$$

se o estado estacionário corresponder a um ponto fixo. Seguindo Amit et al apresentaremos agora uma descrição do comportamento do sistema a $T=0$.

Neste limite a energia livre é dada por

$$E(T=0) = -\vec{m}^2 + \alpha (1 - \tilde{q}_1/2) - J_0 \quad (\text{III.62})$$

Estando interessados nas propriedades de memória associativa do sistema devemos estudar as soluções que correspondem a estados com correlação finita com apenas um dos protótipos:

$$m^v = m \delta_{v\ell} \quad (\text{III.63})$$

Esses estados, ditos ferromagnéticos, caracterizam a região do espaço de parâmetros α , J_0 e T onde o sistema reconhece e recupera os protótipos. As três equações (III.60a-c) podem então ser reduzidas a uma equação para a variável $y = m/(\alpha \tilde{q}_1)^{1/2}$:

$$y = \text{erf}(y)/[(2\alpha)^{1/2} + 2/\pi^{1/2} e^{-y^2}] \quad (\text{III.64})$$

A equação (III.64) sempre possui a solução $y=m=0$, correspondendo aos estados viôro de spin, os quais não possuem correlação finita com nenhum dos protótipos. Para $\alpha > \alpha_c \approx 0.138$ essas são as únicas soluções. Entretanto, para $\alpha < \alpha_c$ aparecem 2p soluções ferromagnéticas altamente correlacionadas com os protótipos (o fator 2 surge devido a simetria $\xi_i^\mu \rightarrow -\xi_i^\mu$ de J_{ij}). Por exemplo, para $\alpha = \alpha_c$ temos

$$m = 0.967$$

correspondendo a um erro menor que 1.5% na recuperação dos protótipos. Esse erro diminui com α , chegando a zero para $\alpha = 0$.

A solução viôro de spin pode ser obtida analiticamente. Eliminando c das equações (III.60b) e (III.60c) obtemos

$$\tilde{q}_1 = 2 [1 + (2/\pi\alpha)^{1/2}]^2 \quad (\text{III.65})$$

com a qual calculamos a energia da fase viôro

$$E_{\text{VS}}(T=0) = -2 [1/\pi + (2\alpha/\pi)^{1/2}] - J_0 \quad (\text{III.66})$$

Na região $\alpha_f < \alpha < \alpha_c$ com $\alpha_f \approx 0.051$ a energia da solução viôro é menor que a da solução ferromagnética, embora os estados ferromagnéticos sejam dinamicamente estáveis. A situação é invertida para $\alpha < \alpha_f$

Existem soluções que descrevem estados correlacionados com diversos protótipos. De fato, esses estados são combinações lineares dos protótipos. Essas soluções existem apenas para $\alpha < 0.03$ e são sempre meta-estáveis, isto é, sua energia é maior que a das soluções ferromagnéticas.

A condição $\tilde{Q} > 0$ não limita a região de existência das soluções associadas a pontos fixos ($q_0=1$). De fato, para essas soluções sempre teremos $\varepsilon_2 = 0$ e, portanto, basta satisfazermos a condição

$$| \varepsilon_1 | > \alpha | \tilde{Q} | \quad (\text{III.67})$$

para obtermos $q_0=1$ na eq.(III.57d).

b) $\tilde{Q} < 0$

Nesse caso temos $e^{-\alpha\beta\tilde{Q}} \rightarrow \infty$. Ficamos então com o seguinte sistema a resolver:

$$m^v = -n^v = \lim_{\beta \rightarrow \infty} \langle \xi^v \tanh \beta \Xi_2 \rangle \quad v=1, \dots, s$$

$$p = r = -q_1 = \lim_{\beta \rightarrow \infty} \langle \tanh^2 \beta \Xi_2 \rangle, \quad q_0 = -1$$

$$\tilde{p} = \tilde{r} = -\tilde{q}_1/2 = p/(1+2c)^2 \quad (\text{III.68a})$$

$$\tilde{Q} = -2c/(1+2c) + 2J_0/\alpha < 0 \quad (\text{III.68b})$$

onde

$$c = \lim_{\beta \rightarrow \infty} \beta (1-p)$$

e

$$\Xi_2 = (2\alpha\tilde{p})^{1/2} (z_2 - z_3) + 2\vec{\xi} \cdot \vec{n}.$$

O resultado $q_0 = -1$ mostra que as soluções desse sistema estão associadas a ciclos de período 2 em que todos os neurônios mudam seus estados a cada etapa de tempo.

Embora essas equações pareçam mais complicados que as do caso anterior devido a integrais duplas da forma

$$\% = \int_{-\infty}^{\infty} dz_2 \int_{-\infty}^{\infty} dz_3 e^{-\frac{1}{2}(z_2^2 + z_3^2)} g(z_2 - z_3) \quad (\text{III.69})$$

podemos utilizar a transformação de Jacobiano unitário

$$u = z_2 - z_3 \quad v = \frac{1}{2}(z_2 + z_3) \quad (\text{III.70})$$

para escrever ξ em termos de uma integral apenas

$$\xi = \pi^{1/2} \int_{-\infty}^{\infty} du e^{-u^2/4} g(u). \quad (\text{III.71})$$

Com essa observação e as identidades (III.59) reescrevemos o sistema (III.68a) como

$$m^v = - \langle \xi^v \text{erf} [\vec{m} \cdot \vec{\xi} / (2\alpha\tilde{p})^{1/2}] \rangle_{\xi} \quad v=1, \dots, s. \quad (\text{III.72a})$$

$$c = (2\pi\alpha\tilde{p})^{-1/2} \langle \exp[-(\vec{m} \cdot \vec{\xi})^2 / 2\alpha\tilde{p}] \rangle \quad (\text{III.72b})$$

$$\tilde{p} = 1/(1+2c)^2. \quad (\text{III.72c})$$

Multiplicando a equação (III.72a) por m^v e somando sobre v é fácil concluir que a única solução dessa equação é $m^v=0$, $v=1, \dots, s$, recuperando então a simetria $m^v=n^v$.

Assim, é imediato obter a solução viário de spin:

$$c = \frac{1}{2} [(\pi\alpha/2)^{1/2} - 1]^{-1}$$

$$\tilde{p} = [1 - (2/\pi\alpha)^{1/2}]^2 \quad (\text{III.73})$$

com energia

$$E_{\text{VS}}(T=0) = -\alpha(1-\tilde{p}) + J_0 = -2\left[-\frac{1}{\pi} + \left(\frac{2\alpha}{\pi}\right)^{1/2}\right] + J_0. \quad (\text{III.74})$$

Uma vez que c é definido positivo pela eq.(III.72b), essa solução só existe para $\alpha > 2/\pi$.

Por outro lado, relaxando a condição implícita de que c seja finito encontramos uma fase paramagnética com $p=0$, $\tilde{p}=0$ e $c \rightarrow \infty$. Essa solução é consistente com as equações (III.72 b-c) e a definição de c . Sua energia é sempre menor que a da solução vidro:

$$E(T=0) = -\alpha + J_0. \quad (\text{III.75})$$

Como no caso (a), a condição $\tilde{Q} < 0$ não limita a região de existência das soluções vidro de spin com $q_0 = -1$. De fato, essas soluções são obtidas sempre que a condição

$$|\varepsilon_2| - \tilde{Q} > |\varepsilon_1| \quad (\text{III.76})$$

for satisfeita. Entretanto, na fase paramagnética $\varepsilon_1 = \varepsilon_2 = 0$ e $\tilde{Q} < 0$ deve ser satisfeito para obtermos $q_0 = -1$. Nessa fase essa condição é escrita como

$$J_0 < \alpha/2$$

(III.77)

c) $\tilde{Q} = 0$

Este caso pode ser mapeado no caso $\tilde{Q} > 0$, dando resultados idênticos aos do caso (a).

Podemos então obter o diagrama de fases do Modelo de Little no plano $T=0$. Uma fase predomina sobre a outra se possuir a menor energia. Assim, as curvas de transição apresentadas no diagrama da figura 5 são obtidas igualando-se as energias das fases envolvidas na transição. Essas transições são descontínuas, isto é, os parâmetros de ordem que caracterizam as diferentes fases não são iguais sobre a curva de transição.

Fig.5- Diagrama de fases para $T=0$.

A curva α_v separando as fases vidro de spin (VS) e paramagnética (P) é obtida igualando-se as equações (III.66) e (III.75):

$$J_0 = \alpha/2 - (2\alpha/\pi)^{1/2} - 1/\pi \quad \alpha \geq 0.051 \quad (\text{III.78})$$

enquanto que α_f , separando a fase vidro da ferromagnética (F) é independente de J_0 :

$$\alpha_f \approx 0.051 \quad J_0 \geq -0.473 \quad (\text{III.79})$$

Para $0 \leq \alpha \leq 0.051$ e $-0.5 < J_0 \leq -0.473$ temos uma região de contato entre as fases P e F. A curva de transição entre essas fases é obtida igualando-se as equações (III.62) e (III.75).

A curva α_r delimita a região onde existem soluções ferromagnéticas. Temos

$$\alpha_r \approx 0.138 \quad (\text{III.80})$$

praticamente independente de J_0 exceto na vizinhança de $J_0 = -1$.

III.E. OS ESTADOS ESTACIONÁRIOS PARA $\alpha = 0$

A importância deste limite, no qual o sistema memoriza um número finito de protótipos, é devida a podermos calcular sua energia livre exatamente, sem a necessidade das réplicas. Para $J_0 = 0$ é possível mostrar que os estados que minimizam a energia livre são de fato os estados estacionários da dinâmica e vice-versa [10].

Tomando o limite $\alpha \rightarrow 0$ na expressão da energia livre, eq.(III.52), obtemos

$$f = \vec{m} \cdot \vec{n} - J_0 - \beta^{-1} \langle \log 2 [\cosh \beta(\vec{m} + \vec{n}) \cdot \vec{\xi} + e^{-2\beta J_0} \sinh \beta(\vec{m} - \vec{n}) \cdot \vec{\xi}] \rangle \quad (\text{III.81})$$

onde usamos $\vec{l} = \vec{m} + \vec{n}$ e $\lim_{\alpha \rightarrow 0} \alpha \tilde{q}_0 = 2J_0/\alpha$, as equações (III.57b) e (III.57i) respectivamente. Note que os parâmetros de ordem da fase vidro de spin não aparecem na equação (III.81).

Os parâmetros de ordem m^v e n^v satisfazem as equações

$$\frac{n^v}{m^v} = \langle \frac{\sinh \beta(\vec{m} + \vec{n}) \cdot \vec{\xi} \pm e^{-2\beta J_0} \sinh \beta(\vec{m} - \vec{n}) \cdot \vec{\xi}}{\cosh \beta(\vec{m} + \vec{n}) \cdot \vec{\xi} + e^{-2\beta J_0} \cosh \beta(\vec{m} - \vec{n}) \cdot \vec{\xi}} \rangle \quad (\text{III.82})$$

com as quais obtemos facilmente

$$(\vec{n}-\vec{m})^2 = < 2 e^{-2\beta J_0} \frac{(\vec{m}-\vec{n}) \cdot \vec{\xi} \sinh \beta (\vec{n}-\vec{m}) \cdot \vec{\xi}}{\cosh \beta (\vec{m}+\vec{n}) \cdot \vec{\xi} + e^{-2\beta J_0} \cosh \beta (\vec{m}-\vec{n}) \cdot \vec{\xi}} > \quad (\text{III.83})$$

Uma vez que o lado esquerdo é positivo enquanto que o lado direito é negativo concluímos que a única solução possível é

$$\vec{m} = \vec{n}. \quad (\text{III.84})$$

No que segue vamos estudar apenas as soluções ferro magnéticas, $m^v = m \delta^{v\ell}$, com m dado por

$$m = \frac{\sinh 2\beta m}{\cosh 2\beta m + e^{-2\beta J_0}} \quad (\text{III.85})$$

Expandindo essa equação em potências de m e eliminando a solução paramagnética, $m=0$, obtemos:

$$m^2 \approx \frac{3}{4} Q_0 T^2 \frac{2/Q_0 - T}{3/Q_0 - T} \quad (\text{III.86})$$

onde

$$Q_0 = 1 + e^{-2\beta J_0}.$$

Assim m^2 anula-se na temperatura crítica T_c definida pela equação

$$T_c = 2 / (1 + e^{-2J_0/T_c}) \quad (\text{III.87})$$

com a condição

$$1 + e^{-2J_0/T_c} < 3. \quad (\text{III.88})$$

Essa transição de 2ª ordem entre as fases F e P muda para uma transição de 1ª ordem quando a condição (III.88) deixa de ser satisfeita. Isso ocorre no ponto tricrítico

$$T_{PTC} = 2/3 \quad J_0^{PTC} = -1/3 \log 2 \quad (\text{III.89})$$

onde uma nova solução da equação (III.85) com energia menor que a da solução (III.86) aparece.

A curva de transição de 1ª ordem é obtida numericamente igualando-se as energias livres das fases P e F. Essa curva toca o eixo $T=0$ no ponto $J_0=-0.5$.

Esses resultados são mostrados no diagrama de fases da figura 6, juntamente com a curva T_r que delimita a região onde existem soluções ferromagnéticas.

Fig.6- Diagrama de fases para $\alpha = 0$.

A fim de obtermos informações sobre os ciclos calculamos o parâmetro q_0 :

$$q_0 = \frac{\cosh 2\beta m - e^{-2\beta J_0}}{\cosh 2\beta m + e^{-2\beta J_0}} \quad (\text{III.90})$$

Na fase P, $m=0$, temos

$$q_0 = \tanh \beta J_0 \quad (\text{III.91})$$

de modo que, para $J_0 < 0$, esta fase corresponde aos ciclos-2 observados na dinâmica do Modelo de Little.

III.F. O DIAGRAMA DE FASE NO ESPAÇO (J_0, α, T)

Podemos classificar as soluções encontradas na análise dos limites $T=0$ e $\alpha=0$ em duas classes:

i) soluções simétricas

$$n^v = m^v, \quad p=r=q_1, \quad \tilde{p}=\tilde{r}=\tilde{q}_1/2 \quad (\text{III.92})$$

ii) soluções anti-simétricas

$$n^v = -m^v = 0, \quad p=r=-q_1, \quad \tilde{p}=\tilde{r}=-\tilde{q}_1/2 \quad (\text{III.93})$$

Acreditamos que essas sejam as únicas soluções possíveis do sistema (III.57). Por outro lado, as soluções anti-simétricas existem apenas para $T=0$. Isso pode ser facilmente constatado usando a relação $\tilde{p}=\tilde{r}=-\tilde{q}_1/2$ na expressão de Ξ_1 , equação (III.51a):

$$\Xi_1 = (2\alpha\tilde{q}_1)^{1/2} z_1 + (-\alpha\tilde{q}_1)^{1/2} (z_2 + z_3) \quad (\text{III.94})$$

Resultando, então, parâmetros de ordem complexos, sem significado físico. Esse problema não ocorre para $T=0$ pois os termos onde Ξ_1 aparece anulam-se nesse limite.

Assim, para $T>0$, somente as soluções simétricas existem. Restringiremos nossa análise às soluções ferromagnéticas

ticas uma vez que nosso interesse original é estudar as propriedades de memória associativa do modelo.

As soluções ferro-simétricas satisfazem as seguintes equações

$$m = \left\langle \frac{\sinh \beta \Xi_1}{\cosh \beta \Xi_1 + e^{-\alpha \beta \tilde{Q}}} \right\rangle \quad (\text{III.95a})$$

$$q_1 = \left\langle \frac{\sinh^2 \beta \Xi_1}{(\cosh \beta \Xi_1 + e^{-\alpha \beta \tilde{Q}})^2} \right\rangle \quad (\text{III.95b})$$

$$q_0 = \left\langle \frac{\cosh \beta \Xi_1 - e^{-\alpha \beta \tilde{Q}}}{\cosh \beta \Xi_1 + e^{-\alpha \beta \tilde{Q}}} \right\rangle \quad (\text{III.95c})$$

$$\tilde{q} = 2q_1 / (1 - Q - Q_1)^2 \quad (\text{III.95d})$$

$$\tilde{Q} = 2(1 - Q) / [(1 - Q)^2 - Q_1^2] - 2(1 - J_0 / \alpha) \quad (\text{III.95e})$$

onde

$$\tilde{Q} = \beta (\tilde{q}_0 - q_1), \quad Q = \beta (q_0 - q_1), \quad Q_1 = \beta (1 - q_1)$$

$$\langle \dots \rangle = \int_{-\infty}^{\infty} \frac{dz}{(2\pi)^{1/2}} e^{-\frac{1}{2}z^2} (\dots)$$

$$\Xi_1 = (2\alpha \tilde{q}_1)^{1/2} z + 2m$$

A energia livre dessas soluções é

$$f = m^2 + \frac{\alpha}{2\beta} \Delta + (\alpha - J_0)q_0 + \frac{\alpha\beta}{2} [2\tilde{q}_1(1-q_1) - \tilde{q}_0(1-q_0)] - \beta^{-1} \langle \log 2 [\cosh \beta_{=1} + e^{-\alpha\beta\tilde{Q}}] \rangle \quad (\text{III.96})$$

com

$$\Delta = \log [(1-Q)^2 - Q_1^2] - 2\beta q_1 / [1 - Q - Q_1].$$

Dessas equações obtemos as seguintes fases e transições:

a) Fase Paramagnética

Essa fase é caracterizada por $q_0, \tilde{q}_0 \neq 0$ e todos os outros parâmetros iguais a zero. Ficamos então com as seguintes equações:

$$q_0 = \tanh(\alpha\beta^2 \tilde{q}_0/2) \quad (\text{III.97a})$$

$$\tilde{q}_0 = 2[q_0(1 - \beta q_0) + \beta] / [(1 - \beta q_0)^2 - \beta^2] + 2J_0/\alpha\beta \quad (\text{III.97b})$$

No limite de alta temperatura obtemos

$$q_0 \approx 1/T(\alpha/T^2 + J_0) \quad (\text{III.98})$$

indicando então que o sistema perde a memória do estado imediatamente anterior $t-\tau$ somente no limite $T \rightarrow \infty$. Se $\alpha = J_0 = 0$ isso ocorre na temperatura crítica $T_0 = 1$.

Para baixa temperatura vemos pela eq. (III.97a) que $q'_0 + 1$ como $e^{-cte/T}$, tal que podemos escrever

$$\beta \tilde{q}_0 \approx -1 + 2J_0/\alpha \quad (\text{III.99})$$

implicando

$$q_0 = \text{sign}(-1 + 2J_0/\alpha). \quad (\text{III.100})$$

Esse resultado está de acordo com a nossa discussão sobre a ocorrência de ciclos a $T=0$.

b) Fase de vidro de spin

Todos os parâmetros de ordem são finitos, exceto m .

A equação da superfície de transição de 2ª ordem $T_V(\alpha, J_0)$ entre as fases P e V é obtida expandindo-se as equações (III.95b) e (III.95d) em potências de q_1 :

$$\tilde{q}_1 \approx 2q_1 / (1 - \beta - \beta q_0)^2 \quad (\text{III.101a})$$

$$q_1 \approx 2\alpha\beta^2 \tilde{q}_1 / [1 + \exp(-\alpha\beta^2 \tilde{q}_0)]^2. \quad (\text{III.101b})$$

Eliminando $\tilde{q}_1$ e descartando a solução paramagnética $q_1=0$, obtemos

$$2\alpha^{1/2}\beta = [1 + \exp(-\alpha\beta^2 \tilde{q}_0)] (1 - \beta q_0 - \beta) \quad (\text{III.102})$$

onde $\beta \equiv 1/T_V$ e $q_0, \tilde{q}_0$ são dados pelas equações (III.97a-b).

Para $\alpha \ll 1$ e $J_0=0$ temos

$$T_V \approx 1 + \alpha^{1/2} \quad (\text{III.103})$$

Por outro lado, sabemos que para $T=0$ essa transição é de 1ª ordem. Deve existir, então, uma linha tricrítica (LTC) sobre a superfície (III.102) separando esses dois diferentes comportamentos críticos. Para obtermos a equação da LTC notemos que, na vizinhança dessa linha, coexistem duas soluções das equações da fase VS. Cada solução está associada a um diferente tipo de transição com a fase P. Exatamente sobre a LTC essas soluções são idênticas, ambas com $q_1=0$. Assim, impondo que as equações (III.97a-b) tenham raízes duplas e que a LTC pertença à superfície (III.102) calculamos numericamente essa curva. A figura 7 mostra as projeções da

LTC nos planos $T=0$ e $J_0=0$. Em concordância com os resultados da secção III.E, essa curva toca o plano $\alpha=0$ no ponto $T=2/3$ e $J_0 = -\frac{1}{3} \log 2$. Observemos ainda que a LTC existe apenas na região $J_0 < 0$.

Fig.7- Projeções da linha tricrítica (LTC) nos planos $T=0$ (linha tracejada) e $J_0=0$ (linha contínua).

Como usual, a transição da 1ª ordem entre as fases P e VS é obtida igualando-se as energias livres dessas duas fases.

c) A fase ferromagnética

Resolvendo numericamente as equações (III.95) obtemos a superfície $T_F(\alpha, J_0)$ abaixo da qual os estados ferromagnéticos, com todos os parâmetros de ordem finitos, começam a aparecer como estados localmente estáveis, isto é, me-

taestáveis. Esses estados possuem um papel importante na dinâmica do sistema: inicialmente o sistema atinge um desses estados, mas após um certo tempo, o ruído térmico acaba por levar o sistema à fase vidro de spin. Entretanto, abaixo da superfície $T_f(\alpha, J_0)$ os estados ferromagnéticos tornam-se mínimos globais da energia livre. Os resultados são mostrados na figura 8.

Para $J_0=0$, obtemos um diagrama de fases muito similar ao do Modelo de Hopfield, exceto que para $\alpha \geq 3.71$ aparece a fase paramagnética com $q_0 < 0$, a qual associamos com a existência de ciclos no Modelo de Little. Na fase paramagnética tradicional que aparece para altas temperaturas temos $q_0 > 0$.

Fig. 8-Diagrama de fases para $J_0=0$.

d) O diagrama de fases próximo a $T=1$ no plano $J_0=0$

Nosso objetivo é estudar o comportamento das curvas $T_f(\alpha)$ e $T_r(\alpha)$ na vizinhança de $T=1$. Essas curvas estão associadas a transições descontínuas de modo que os parâmetros de ordem não são pequenos nas suas vizinhanças exceto próximo a $T=1$ onde temos $m \approx 0$ de acordo com a seção III.E.

Expandindo as equações (III.95) em potências de m, α e $t (\equiv 1-T)$ e eliminando a solução $m=0$ obtemos:

$$t = \frac{1}{3} m^2 - \frac{\alpha}{2} (\tilde{q}_0 - 2\tilde{q}_1)$$

$$q_1 = m^2 + \frac{1}{2} \alpha \tilde{q}_1$$

$$q_0 = m^2 + \frac{1}{2} \alpha \tilde{q}_0$$

$$\tilde{q}_1 = 2q_1 / (2q_1 - q_0 - t)^2$$

$$\tilde{q}_0 = (4q_1 - q_0 - t) / (2q_1 - q_0 - t)^2 \quad (\text{III.104})$$

Escrevendo todos os parâmetros de ordem em termos de m ficamos com a equação

$$g(y) \equiv y^4 - 2\hat{\tau} y^3 + y^2 + 2\hat{\tau} y + 1 = 0 \quad (\text{III.105})$$

onde

$$y = \frac{2}{3} \frac{m^2}{\alpha^{1/2}} \quad \text{e} \quad \hat{\tau} = \frac{t}{\alpha^{1/2}} .$$

Essa equação ou possui duas ou então nenhuma raiz positiva. O desaparecimento simultâneo dessas duas soluções nos dá $T_f(\alpha)$. Calculamos então o valor de $\hat{\tau}$ para o qual $g(y) = \frac{dg}{dy}(y) = 0$. O resultado é

$$T_f(\alpha) \approx 1 - 1.73 \alpha^{1/2} \quad (1 - 1.95 \alpha^{1/2}). \quad (\text{III.106})$$

onde incluímos o resultado do mesmo cálculo no Modelo de Hopfield [11] entre parênteses. Assim os estados ferromagnéticos são ligeiramente menos sensíveis ao ruído no Modelo de Little.

A fim de estudarmos o comportamento de $T_f(\alpha)$ devemos expandir a energia livre da fase ferromagnética em potências de m , α e t . Obtemos $T_f(\alpha)$ calculando o valor de $\hat{\tau}$ para o qual a energia livre da fase ferro $f(m)$ iguale a energia livre da fase vidro $f(0)$. Com a mesma notação anterior escrevemos essa condição como

$$f(m) - f(0) = \frac{\alpha}{2} \left(\log y + \frac{y^2}{2} - \frac{3}{y^2 - 1} + \frac{1}{2y^2} + 4 \hat{\tau} + \frac{1}{2} \right) = 0$$

(III.107)

Resolvendo essa equação juntamente com a eq. (III.105) obtemos

$$T_f(\alpha) \approx 1 - 2.40 \alpha^{1/2} (1 - 2.60 \alpha^{1/2}) \quad (\text{III.108})$$

indicando então que os estados ferromagnéticos tornam-se mínimos globais da energia livre a uma temperatura levemente superior à do Modelo de Hopfield (entre parênteses).

e) O Limite $J_0 \rightarrow \infty$

Mantendo α e β fixos tomamos o limite $J_0 \rightarrow \infty$ nas equações (III.95), obtendo

$$m = \langle \tanh \beta \Xi_1 \rangle$$

$$q_1 = \langle \tanh^2 \beta \Xi_1 \rangle$$

$$q_0 = 1$$

$$\tilde{q}_1 = 2q_1 / [1 - 2\beta(1 - q_1)]^2 \quad (\text{III.109})$$

onde $\Xi_1 = (2\alpha \tilde{q}_1)^{1/2} z + 2m$.

Essas equações reduzem-se às do Modelo de Hopfield se fizermos a reescala $T_{\text{Little}} = 2 T_{\text{Hopfield}}$. Esse resultado é razoável pois, uma vez que as configurações S e σ estão

alinhadas, o ruído térmico deverá agora mudar os estados dos dois spins S_i e σ_i em cada sítio.

Do ponto de vista dinâmico esse limite é trivial e desinteressante: o sistema congela na configuração inicial. Entretanto, a conclusão permanece importante: uma autointeração positiva e finita aumenta a estabilidade dos estados ferromagnéticos frente ao ruído térmico.

III.G. A QUEBRA DE SIMETRIA DAS RÉPLICAS

O modo tradicional de mostrar que o método das réplicas dá soluções não físicas para os estados estacionários do sistema a baixas temperaturas é calcular a entropia desses estados para $T=0$.

$$S = - \left. \frac{\partial f}{\partial T} \right|_{T=0} \quad (\text{III.110})$$

onde f é a energia livre do sistema. Para as soluções viário de spin do Modelo SK obtemos $S = - \frac{1}{2\pi} \approx -0.16$ indicando então que algo está errado com o método das réplicas.

Inicialmente pensou-se que o problema estava na troca dos limites termodinâmico ($N \rightarrow \infty$) e das réplicas ($n \rightarrow 0$) ou na extensão para n real [31]. Entretanto, de Almeida e Thouless [32] mostraram que o problema estava na escolha

do ansatz simetria de réplicas como solução das equações de ponto de sela. De fato, um estudo da estabilidade das soluções simétricas em relação as réplicas no Modelo SK mostrou que essas soluções são instáveis a quebra de simetria das réplicas (QSR) e, portanto, não minimizam a energia livre, equação (III.37).

A interpretação física da QSR e também o ansatz atualmente aceito como correto são devidos a Parisi [33]. Na fase vidro de spin existe um número infinito de estados termodinâmicos puros $\vec{m}^\alpha = (\langle S_1 \rangle_\alpha, \dots, \langle S_N \rangle_\alpha)$ indexados por diferentes índices de réplicas. A média térmica $\langle \dots \rangle_\alpha$ é realizada no espaço de fase gerado pela QSR. Se definirmos uma correlação $q_{\alpha\beta}$ entre os estados α e β por

$$q_{\alpha\beta} = \frac{1}{N} \sum_i \langle S_i \rangle_\alpha \langle S_i \rangle_\beta$$

então a termodinâmica do sistema é dada pela distribuição

$$P(q) = \lim_{n \rightarrow 0} \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} \delta(q - q_{\alpha\beta}),$$

que pode ser calculada usando o método das réplicas com a QSR [34].

A correlação $q_{\alpha\beta}$ define uma distância $d_{\alpha\beta}$ entre os estados α e β :

$$\bar{d}_{\alpha\beta} = (1-q_{\alpha\beta})/2$$

Com respeito a essa distância os estados vídros de spin possuem uma estrutura ultramétrica, isto é, quaisquer três estados α, β, γ formam um triângulo isósceles ($\bar{d}_{\alpha\beta} = \bar{d}_{\beta\gamma} > \bar{d}_{\alpha\gamma}$, após rearmarmos os índices). A estrutura ultramétrica é característica de sistemas organizados hierarquicamente. De fato, o ansatz de Parisi é baseado em uma estrutura hierárquica para a matriz $n \times n$ $\underline{Q}$, cujos elementos são os $q_{\alpha\beta}$. Mais, sua proposta de QSR é iterativa: a cada quebra de simetria a entropia torna-se menos negativa. Assim, tornou-se usual medir o efeito da QSR pelo valor absoluto da entropia. Quanto menos negativa menos quebras de réplicas serão necessárias para atingir o valor esperado $S=0$.

Nesta secção vamos calcular a entropia para $T=0$ das diversas soluções encontradas na secção III.D e então testar a estabilidade das soluções ferromagnéticas frente a QSR obtendo assim o análogo da linha de Almeida-Thouless que limita a região onde o ansatz simetria de réplicas é estável.

Primeiramente tratemos das soluções ferromagnéticas. Sua energia livre é dada pela equação (III.96). Efetuando a derivação e tomando o limite $T \rightarrow 0$ lembrando que

$$c = \lim_{T \rightarrow 0} (1-q_1)/T$$

obtemos

$$S_f = - \frac{\alpha}{2} [\log (1-2c) + 2c/(1-2c)] \quad (\text{III.111})$$

onde c é dado pelas equações (III.60). Entretanto, é fácil ver por essas equações que $2c < 1$ e, portanto, a entropia dessas soluções é sempre negativa. Para $\alpha = \alpha_c \approx 0.138$ temos $S = -1.4 \cdot 10^{-3}$ e S vai a zero exponencialmente quando $\alpha \rightarrow 0$: para $\alpha \ll 1$ temos $\tilde{q}_1 \approx 2$, $m \approx 1$ implicando $2c \approx (2/\pi\alpha)^{1/2} e^{-1/2\alpha}$, daí

$$S_f \approx - \frac{1}{2\pi} e^{-1/\alpha} \quad (\text{III.112})$$

Esses resultados são indicações de que o efeito da QSR é muito pequeno na fase ferromagnética.

Na fase vidro de spin podemos obter uma expressão analítica para S :

$$S_v = - \frac{\alpha}{2} \left[(2/\pi\alpha)^{1/2} - \log (1 + (2/\pi\alpha)^{1/2}) \right]. \quad (\text{III.113})$$

Temos $S_v \approx -0.07$ para $\alpha = \alpha_c$ e $S_v \rightarrow -\frac{1}{2\pi} \approx -0.16$ quando $\alpha \rightarrow \infty$. Observe então que a QSR é mais forte nessa fase do que na ferromagnética.

Tratemos agora das soluções antisimétricas ($q_0 = -1$) as quais associamos os ciclos-2 do Modelo de Little. Um cálculo direto nos leva a expressão

$$S_A = - \frac{\alpha}{2} [\log(1+2c) + 2c/(1+2c)] \quad (\text{III.114})$$

com $c = \lim_{T \rightarrow 0} (1-p)/T$. Na fase vidro de spin temos $2c = [(\frac{\pi\alpha}{2})^{1/2} - 1]^{-1}$ e, portanto

$$S_V = - \frac{\alpha}{2} [(2/\pi\alpha)^{1/2} - \log(1 - (2/\pi\alpha)^{1/2})], \alpha > \frac{2}{\pi} \quad (\text{III.115})$$

Vemos daí que $S_V \rightarrow -\infty$ para $\alpha \rightarrow \frac{2}{\pi}$ e $S_V \rightarrow -\infty$ para $\alpha \rightarrow \infty$. Esse comportamento bizarro mostra que essa solução não deve ser física (lembre que existe apenas para $T=0$) ou então que a QSR é extremamente forte.

Na fase paramagnética temos $c \rightarrow \infty$ e portanto $S \rightarrow -\infty$ para todos os valores de α , exceto $\alpha=0$ no qual obtemos $S=\log 2$. No Modelo SK a entropia da fase paramagnética também diverge para $T=0$. Uma vez que, nesse modelo, todos os parâmetros de ordem são nulos na fase paramagnética a QSR não deve sanar o problema da entropia. Voltaremos a esse ponto no capítulo V.

No apêndice B calculamos o análogo da linha de Almeida-Thouless para os estados ferromagnéticos. Essa curva é mostrada na figura 9, no plano $J_0=0$ juntamente com a do Modelo de Hopfield, para fins de comparação. Abaixo de $T_Q(\alpha)$ as soluções ferromagnéticas simétricas em relação as réplicas são instáveis a QSR.

Fig.9- A linha de Almeida-Thouless para os Modelos de Little (linha cheia) e de Hopfield (linha tracejada) no plano $J_0=0$.

Assim, o efeito da QSR é levemente maior no modelo de Little. A adição de um termo de autointeração positivo aumenta a região de instabilidade.

$$T_Q(J_0 > 0) > T_Q(J_0=0)$$

enquanto que uma autointeração negativa diminui essa região

$$T_Q(J_0 < 0) < T_Q(J_0=0)$$

embora acabe por destruir os estados ferromagnéticos.

Esse efeito provavelmente está relacionado com o fato de que $J_0 > 0$ aumenta o número de estados meta-estáveis do sistema (capítulo IV). Assim, a superfície de energia onde cada estado meta-estável corresponde a um vale torna-se mais complexa, diminuindo as chances do ansatz simetria de réplicas, no qual existe apenas um vale, ser a solução correta.

CAPÍTULO IV

A DINÂMICA DO MODELO DE LITTLE À TEMPERATURA ZERO

IV. A DINÂMICA DO MODELO DE LITTLE À TEMPERATURA ZERO

A razão pela qual o estudo da dinâmica à temperatura zero é não trivial e extremamente interessante está ligada a existência de um número enorme de estados metaestáveis. Esses estados metaestáveis são responsáveis pelos fenômenos de remanescência e extrema sensibilidade às condições iniciais.

Neste capítulo apresentamos ao leitor alguns fenômenos que não puderam ser observados no cálculo termodinâmico, especialmente os devidos ao termo de autointeração. Para isso, utilizamos uma aproximação que torna a dinâmica solúvel analiticamente e então realizamos simulações numéricas no modelo exato (secções IV.A e IV.C). Na secção IV.B descrevemos algumas experiências numéricas com as quais procuramos um melhor entendimento do fenômeno ciclo-2. Finalmente, na secção IV.D relatamos alguns resultados existentes na literatura sobre o interessante fenômeno da correlação de recuperação remanescente.

Na referência [35], Gardner et al realizaram um extenso estudo abordando outros aspectos da dinâmica do Modelo de Little à temperatura zero o qual recomendamos ao leitor a fim de completar sua visão do problema.

IV.A. A APROXIMAÇÃO DE KINZEL

A dinâmica do Modelo de Little à temperatura zero é governada pela equação (II.2)

$$S_i(t+1) = \text{sign} \left(J_0 S_i(t) + \frac{1}{N} \sum_{j \neq i} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu S_j(t) \right) \quad i=1, \dots, N \quad (\text{IV.1})$$

onde fizemos $\alpha_i=0$, $\tau=1$ e usamos a prescrição de Cooper com o termo de autointeração, equação (III.17), para escrever as interações entre os neurônios. Todos os neurônios evoluem si multaneamente.

Nosso objetivo é estudar a evolução de uma configuração $\{S_i(t)\}$ que possui correlação macroscópica - $O(1)$ - com apenas um dos protótipos $\{\xi_i^1\}$ e correlações microscópicas - $O(N^{-\frac{1}{2}})$ com os $p-1$ restantes.

Primeiramente, multipliquemos (IV.1) por $\frac{1}{N} \xi_i^1$ e so memos sobre i :

$$m(t+1) = \frac{1}{N} \sum_i m_i(t) \text{sign} (J_0 + m(t)m_i(t) + \delta_i(t)) \quad (\text{IV.2})$$

onde

$$m(t) = \frac{1}{N} \sum_i m_i(t) = \frac{1}{N} \sum_i \xi_i^1 S_i(t) \quad (\text{IV.3})$$

e

$$\delta_i(t) = \frac{1}{N} \sum_{j \neq i} \sum_{\mu=1}^{\mu} \xi_i^{\mu} \xi_j^{\mu} S_j(t) S_i(t). \quad (\text{IV.4})$$

Seja n o número de sítios nos quais $S_i(t) = -\xi_i^1$ ou $m_i(t) = -1$. Reescrevemos (IV.2) como

$$m(t+1) = \frac{N-n(t)}{N} \text{sign}(J_0 + m(t) + \delta_i(t)) - \frac{n(t)}{N} \text{sign}(J_0 - m(t) + \delta_i(t)). \quad (\text{IV.5})$$

Se os N sítios forem descorrelacionados, isto é, não possuírem ancestrais comuns, a probabilidade de que, entre os N sítios, n sejam tais que $S_i(t) = -\xi_i^1$ é

$$P_n = \binom{N}{n} \frac{(1-m)^n (1+m)^{N-n}}{2^N} \quad (\text{IV.6})$$

onde $m = m(t)$, $n = n(t)$ e $\binom{N}{n} = \frac{N!}{n!(N-n)!}$.

Obviamente essa suposição é falsa: em uma rede onde todos os neurônios estão conectados entre si, os ancestrais, são comuns a todos os sítios. Entretanto, Derrida et al [36] mostraram que se diluirmos a rede, eliminando aleatoriamente e assimetricamente uma parte das conexões, até atingirmos uma conectividade média C que satisfaça

$$C \ll \log N \quad (\text{IV.7})$$

os C sítios conectados ao sítio i não possuirão ancestrais comuns. Assim, a aproximação de Kinzel [12], dada pela equação (IV.6), é exata para o modelo diluído assimétrica e aleatoriamente no regime da equação (IV.7).

Outra situação onde a aproximação de Kinzel é exata é o modelo recozido ("annealed") : a cada etapa de tempo geramos uma nova realização dos protótipos $\{\xi_i^\mu, \mu=1, \dots, p\}$ completamente descorrelacionada da anterior. Assim, a probabilidade de encontrarmos n sítios com $S_i(t) = -\xi_i^1$ independe da configuração dos neurônios em tempos anteriores a t .

Voltemos a equação (IV.6). O valor médio $\bar{n}$ e a dispersão em torno da média $\overline{(\Delta n)^2}$ são dados por

$$\bar{n} = \frac{1-m}{2} N \quad (\text{IV.8a})$$

$$\overline{(\Delta n)^2} = \frac{1-m^2}{4} N \quad (\text{IV.8b})$$

Assim as flutuações de n/N em torno do valor médio $\bar{n}/N$ são $O(N^{-\frac{1}{2}})$ e podem ser desprezadas no limite termodinâmico. Podemos então substituir n/N na equação (IV.6) pelo seu valor médio.

O termo δ_i , equação (IV.4), já foi discutido na secção (II.D) : é um passeio aleatório de $(N-1)(p-1)$ passos de comprimento $1/N$. Para $N, p \gg 1$ podemos aproximar a distribui-

ção binomial de δ_i por uma Gaussiana de média nula e variância $\overline{\delta_i^2} = (p-1)/N \approx \alpha$:

$$P(\delta_i) = p(\delta) = (2\pi\alpha)^{-\frac{1}{2}} e^{-\delta^2/2\alpha}. \quad (\text{IV.9})$$

Usando

$$\int_{-\infty}^{\infty} d\delta p(\delta) \text{sign}(x+\delta) \equiv \langle \text{sign}(x+\delta) \rangle_{\delta} = p(\delta > -x) - p(\delta < -x) \quad (\text{IV.10})$$

escrevemos

$$\langle \text{sign}(J_0 + m(t) + \delta) \rangle_{\delta} = \text{erf}\left(\frac{J_0 + m(t)}{\sqrt{2\alpha}}\right) \quad (\text{IV.11a})$$

$$\langle \text{sign}(J_0 - m(t) + \delta) \rangle_{\delta} = \text{erf}\left(\frac{J_0 - m(t)}{\sqrt{2\alpha}}\right) \quad (\text{IV.11b})$$

de modo que, mediando a equação (IV.5) sobre n e δ , obtemos

$$m(t+1) = \frac{1+m(t)}{2} \text{erf}\left(\frac{J_0 + m(t)}{\sqrt{2\alpha}}\right) - \frac{1-m(t)}{2} \text{erf}\left(\frac{J_0 - m(t)}{\sqrt{2\alpha}}\right) \quad (\text{IV.12})$$

A seguir estudaremos os pontos fixos dessa equação:

$$m(t+1) = m(t) = m. \quad (\text{IV.13})$$

Observe que o fato de $m(t)$ atingir um ponto fixo não significa que a configuração dos neurônios $\{S_i(t)\}$ também tenha

atingido um ponto fixo. De fato, uma vez que a equação (IV.12) descreve a dinâmica dos modelos recozido e diluído assimétrico esperamos que esse ponto fixo configuracional nunca seja alcançado. Lembramos que os resultados da secção (II.C) garantem pontos fixos e ciclos-2 como os únicos regimes de equilíbrio da dinâmica síncrona somente para o modelo com interações simétricas.

A solução paramagnética $m=0$ é sempre um ponto fixo da equação (IV.12). Expandindo o lado direito dessa equação em potências de m e eliminando a solução $m=0$ obtemos a equação da curva de transição de 2ª ordem entre as fases ferromagnética ($m \neq 0$) e paramagnética ($m=0$) :

$$1 - \operatorname{erf}(J_0/\sqrt{2\alpha}) - (2/\pi\alpha)^{\frac{1}{2}} e^{-J_0^2/2\alpha} = 0. \quad (\text{IV.14})$$

Essa transição termina em um ponto tricrítico de coordenadas

$$\alpha_{\text{ptc}} = 0.263 \quad J_0^{\text{ptc}} = -0.253 \quad (\text{IV.15})$$

quando então a transição passa a ser descontínua conforme mostra a figura 10.

Fig. 10- Diagrama de fases do Modelo de Little a $T=0$ na aproximação de Kinzel.

Observe a semelhança desse diagrama com o da figura 6. O parâmetro α desempenha um papel análogo ao da temperatura no modelo não diluído.

Na figura 11 mostramos a correlação de recuperação crítica m_c sobre as curvas de transição da figura anterior.

Fig. 11- A correlação de recuperação crítica como função de α .

Embora $m_c=0$ para $J_0 > J_0^{\text{PTC}}$, a autointeração positi

va efetivamente melhora as propriedades de recuperação do modelo. Por exemplo, a figura 12 mostrando m vs. α para vários valores de J_0 indica claramente que, para α fixo, o valor da correlação de recuperação aumenta com J_0 . Entretanto, quanto maior J_0 , maior será o tempo necessário para $m(t)$ alcançar seu valor de equilíbrio m .

Fig. 12- A correlação de equilíbrio em função de α para $J_0 = -0.5, -0.2, 0$ e 0.2 .

Outra informação que podemos obter facilmente no esquema da aproximação de Kinzel é a correlação entre configurações do sistema em tempos consecutivos:

$$q_0(t+1) = \frac{1}{N} \sum_i S_i(t) S_i(t+1) = \frac{1}{N} \sum_i \text{sign} (J_0 + m(t) + \delta_i(t)) \quad (\text{IV.16})$$

Seguindo um procedimento análogo ao usado no cálculo de $m(t+1)$ obtemos

$$q_0(t+1) = \frac{1+m(t)}{2} \text{erf}\left(\frac{J_0+m(t)}{\sqrt{2\alpha}}\right) + \frac{1-m(t)}{2} \text{erf}\left(\frac{J_0-m(t)}{\sqrt{2\alpha}}\right). \quad (\text{IV.17})$$

Na figura 13 mostramos o valor de equilíbrio de q_0 calculado no ponto fixo de $m(t)$ em função de α para diversos valores de J_0 .

Fig. 13- A correlação entre configurações em tempos consecutivos como função de α para $J_0 = -1.0, -0.5, -0.2, 0$ e 0.2 .

Fica claro que para $\alpha \neq 0$ o ponto fixo de $m(t)$ não corresponde a um ponto fixo configuracional ($q_0=1$). No limite $\alpha \rightarrow \infty$ temos $q_0 \rightarrow 0$ independente de J_0 .

Uma vez que o papel de α na aproximação de Kinzel é semelhante ao da temperatura no modelo exato é interessan-

te estudar o limite $\alpha \rightarrow 0$ das equações (IV.12) e (IV.17). Temos

$$m(t+1) = \frac{1+m(t)}{2} \text{sign}(J_0+m(t)) - \frac{1-m(t)}{2} \text{sign}(J_0-m(t)) \quad (\text{IV.18a})$$

$$q_0(t+1) = \frac{1+m(t)}{2} \text{sign}(J_0+m(t)) + \frac{1-m(t)}{2} \text{sign}(J_0-m(t)) \quad (\text{IV.18b})$$

O regime de equilíbrio depende fortemente da condição inicial $m(0)$. Identificamos três diferentes comportamentos assintóticos :

$$\text{i) } |m(0)| > J_0 \rightarrow m = \text{sign}(m(0)) , \quad q_0 = 1 \quad (\text{IV.19a})$$

$$\begin{aligned} \text{ii) } |m(0)| < J_0 \quad \{ m = & \begin{aligned} & m(0), \quad q_0 = 1 \quad \text{se } J_0 > 0 \\ & (m(0), -m(0)), \quad q_0 = -1 \quad \text{se } J_0 < 0 \end{aligned} \end{aligned} \quad (\text{IV.19b})$$

No caso (i) o sistema evolui da condição inicial para o atrator mais próximo, $\{\xi_i^1\}$ ou $\{-\xi_i^1\}$. No caso (ii) o sistema não evolui : ou permanece congelado na configuração inicial ou entra em um ciclo-2, oscilando entre a configuração inicial e sua oposta.

Separando essas regiões estão as retas $|m(0)| = J_0$, sobre as quais identificamos as seguintes soluções :

$$\text{i) } m(0) = J_0 \rightarrow m = \begin{cases} 1 & \text{se } J_0 > 0 \\ -\frac{1}{3} & \text{se } J_0 < 0 \end{cases} \quad (\text{IV.20a})$$

$$\text{ii) } m(0) = -J_0 \rightarrow m = \begin{cases} -1 & \text{se } J_0 > 0 \\ \frac{1}{3} & \text{se } J_0 < 0 \end{cases} \quad (\text{IV.20b})$$

onde usamos $\text{sign}(0) = 0$. Esses resultados são apresentados na figura 14 sob a forma de um diagrama de fluxos.

Fig. 14- Diagrama de fluxo mostrando as bacias de atração para $\alpha = 0$.

Entretanto as fases "rígidas" que aparecem para $|m(0)| < J_0$ são extremamente sensíveis ao ruído, de forma que qualquer valor finito de α é suficiente para desestabilizá-las. Na figura 15 mostramos o diagrama de fluxos para $\alpha = 0.01$ onde vemos então que a região dos ciclos-2 passa a fazer parte da bacia de atração da solução paramagnética. A região dos pontos fixos com $m = m(0)$ é dividida igualmente entre as bacias de atração do protótipo $\{\xi_i^1\}$ e de seu oposto $\{-\xi_i^1\}$.

Fig. 15- Diagrama de fluxo mostrando as bacias de atração para $\alpha = 0.01$.

IV.B. OS CICLOS-2

A principal informação obtida no cálculo termodinâmico do capítulo anterior sobre os ciclos de período 2 (ciclos-2) a $T=0$ é que esses ciclos são completos, isto é, todos os neurônios participam do ciclo mudando seus estados a cada etapa de tempo. Obviamente, no limite termodinâmico a caracterização de um ponto fixo ou ciclo completo não é afetada se um número finito de neurônios oscilar ou permanecer fixo, respectivamente.

Entretanto, essa informação não é confiável : a entropia dos ciclos completos (S) é negativa e divergente para $\alpha \neq 0$. No caso de um número finito de protótipos, $\alpha = 0$, o método das réplicas não é necessário e obtemos $S = \log 2$ como esperado.

O objetivo desta secção é dar ao leitor uma idéia da real natureza dos ciclos observados nas simulações da dinâmica do Modelo de Little. Devido as nossas limitações computacionais que não nos permitiriam simular, em tempo aceitável, sistemas com $N > 300$ nossos resultados não podem ser considerados conclusivos. De fato, Amit et al [11] mostraram que, próximo a uma região de transição de fase descontínua, o comportamento do sistema com $N > 10^3$ difere radicalmente do comportamento para $N \leq 500$.

Nossa primeira experiência é a seguinte : dado que o estado inicial do sistema é um dos protótipos desejamos obter a probabilidade P_c de que o regime estacionário seja um ciclo. É claro que P_c está diretamente relacionado com o tamanho da bacia de atração dos ciclos.

Fazendo $N=50$ e $J_0=0$ na equação (IV.1) e mediando sobre $5 \cdot 10^3$ configurações iniciais escolhidas entre os protótipos obtemos P_c em função de α , figura 16.

Fig. 16- A probabilidade do regime estacionário ser um ciclo-2 como função de α .

Observe a predominância dos ciclos com o aumento de α . A barra de erro nesta figura e na seguinte representa uma medida típica das flutuações em torno dos pontos apresentados.

Na figura 17 apresentamos o comportamento de P_c , para $\alpha=0.5$ e $J_0=0$ fixos, quando N varia entre 20 e 300.

Fig. 17- A probabilidade do regime estacionário ser um ciclo-2 como função de N^{-1} .

Esses resultados permanecem essencialmente os mesmos se, no lugar de protótipos, escolhermos a configuração inicial aleatoriamente.

Assim, no limite $N \rightarrow \infty$ podemos esperar o completo domínio dos ciclos-2 como estados estacionários da dinâmica síncrona. Para entender esse surpreendente resultado devemos estudar a estrutura dos ciclos com mais cuidado. Nas experiências descritas acima não nos importamos em medir o número de neurônios que participam do ciclo. Trataremos de eliminar essa lacuna

próxima experiência.

Seja η a razão entre o número de neurônios que participam do ciclo e o número total de neurônios N . A relação de η com o parâmetro q_0 do capítulo III é

$$\eta = (1 - q_0)/2. \quad (\text{IV.21})$$

Desejamos calcular a probabilidade $P(\eta)$ de que o estado estacionário seja um ciclo-2 em que $N\eta$ neurônios oscilem. Lembremos que a teoria simétrica em relação as réplicas desenvolvida no capítulo III prevê a seguinte forma para $P(\eta)$:

$$p(\eta) = A\delta(\eta) + B\delta(\eta-1) \quad (\text{IV.22})$$

sem especificar as amplitudes A e B .

Na figura 18 mostramos $P(\eta)$ para $N=100, 200$ e 300 . Os parâmetros $\alpha=0.5$ e $J_0=0$ são mantidos fixos. À medida que N cresce sobrevivem apenas ciclos com $\eta \ll 1$ embora a probabilidade de ocorrer pontos fixos $P(\eta=0)$ diminua.

O pico em $\eta=0$ não é conectado ao restante do gráfico pois $P(\eta=1/N)=0$, isto é, a probabilidade de apenas um neurônio oscilar é nula. É fácil entender isso : se a autointeração é nula e os $N-1$ neurônios estão fixos então o campo resultante sobre o neurônio oscilante é constante de modo que esse neurô-

nio deverá alinhar-se na direção de seu campo.

Fig. 18- Histogramas da fração η de neurônios oscilantes para $J_0 = 0$ e $\alpha = 0.5$.

Interpretamos esses resultados do seguinte modo :no limite termodinâmico apenas um número finito de neurônios oscilam de forma que $P(\eta)$ está de acordo com a teoria das réplicas :

$$P(\eta) = \delta(\eta). \quad (\text{IV.23})$$

Entretanto a probabilidade de encontrar um estado em que nenhum neurônio esteja oscilando é nula.

Essa conclusão permanece válida se mantivermos $J_0 = 0$ e variarmos α . Assim a fase paramagnética que aparece para $\alpha > 3.71$ no diagrama da figura 8 deve ser um artefato do método

das réplicas.

Vejamos agora o comportamento do modelo quando fazemos J_0 negativo. Mantendo $\alpha=0.5$ e $J_0=-0.5$ fixos construímos os histogramas da figura 19 para $N=50$ (a) e $N=100$ (b).

Fig. 19- Histogramas da fração η de neurônios oscilantes para $J_0 = -0.5$ e $\alpha=0.5$.

Interpretamos esse curioso resultado que indica uma predominância dos ciclos incompletos, $0 < \eta < 1$, na dinâmica como um efeito de N finito. De fato, o diagrama da figura 5 mostra que a transição descontínua entre as fases vidro de spin ($q_0=+1$) e paramagnética ($q_0=-1$) ocorre para $J_0=-0.63$, muito próximo do valor de J_0 usado nessa experiência. No limite $N \rightarrow \infty$, esperamos que $p(\eta)$ fique concentrada em $\eta=0$ ou $\eta=1$.

Repetindo a mesma experiência para $J_0=-1.0$ e $\alpha = 0.5$ encontramos

$$P(\eta) = \delta(\eta-1)$$

(IV.24)

com desprezíveis flutuações em torno de $\eta=1$.

IV.C. A AUTOINTERAÇÃO

Há duas razões para darmos uma atenção especial ao papel desempenhado pelo termo de autointeração J_0 na dinâmica do Modelo de Little.

Primeira, a aproximação de Kinzel (secção IV.A) prevê um aumento da capacidade de armazenagem crítica α_c para $J_0 > 0$: o mesmo poderia ocorrer para o Modelo de Little exato.

A segunda razão é motivada pela observação do seguinte interessante fenômeno que ocorre na fase vidro de spin, $\alpha > \alpha_c$. Se a configuração inicial é próxima a um dos protótipos, o sistema evolui inicialmente em direção ao protótipo e então afasta-se em direção a um dos atratores da fase vidro de spin [35]. Um comportamento similar é observado na fase ferromagnética quando a configuração inicial não é suficientemente próxima a um dos protótipos [37]. Assim, existe uma capacidade latente de recuperação de informações que poderia ser aproveitada. Para isso, devemos incluir no modelo uma espécie de termo de atrito que permita a dinâmica ficar retida em um desses estados de recuperação próximos aos protótipos. Esse termo po

de ser facilmente implementado introduzindo-se uma autointeracção positiva em cada neurônio :

$$J_{ii} = J_0 > 0. \quad (\text{IV.25})$$

Até então, o termo de autointeração positivo era conhecido pelo seu efeito negativo sobre o desempenho do modelo : o tamanho das bacias de atração dos estados ferromagnéticos é severamente reduzido [24]. Isso se deve ao fato de $J_0 > 0$ criar estados metaestáveis espúrios, cada qual com sua própria bacia de atração, que competem com os estados ferromagnéticos.

Por outro lado, há um aumento de estabilidade dos estados ferromagnéticos devido ao trivial fato de que o atrito gerado pela autointeração dificulta ou até impede ao sistema evoluir da configuração inicial (escolhida um estado ferromagnético para testar a estabilidade).

Entretanto, para que esses efeitos negativos sejam observados é necessário que existam estados ferromagnéticos. O que vamos mostrar a seguir, através de simulações numéricas, é que escolhendo valores apropriados para J_0 podemos criar estados ferromagnéticos onde eles não existiam, isto é, na fase vidro de spin. Aumentando J_0 além desses valores começamos a observar os efeitos prejudiciais descritos acima.

Nas figuras 20 mostramos a correlação de recupera-

ção de equilíbrio $m = m(t=\infty)$, equação (IV.3), como função da correlação de recuperação inicial

$$m_0 = \frac{1}{N} \langle \sum_i \xi_i^\mu S_i(t=0) \rangle_\xi \quad (\text{IV.26})$$

para $\alpha = 0.1, 0.3$ e 0.4 e vários valores de J_0 . A média $\langle \dots \rangle_\xi$ é realizada sobre $\approx 10^3$ configurações iniciais e $\approx 10^2$ realizações dos protótipos. Procuramos isolar o efeito de J_0 repetindo sempre a mesma experiência com diferentes valores de J_0 .

Para $\alpha=0.1$, figura 20.a, o sistema é capaz de lembrar, isto é, recuperar os protótipos, se $m_0 \gtrsim 0.4$ uma vez que m é muito próximo a 1. Como esperado, para termos de atrito grandes, por exemplo, $J_0 > 1$ o sistema não consegue evoluir muito além da configuração inicial tornando-se inútil como memória associativa. O melhor desempenho

Fig. 20.a - A correlação de equilíbrio como função da correlação inicial m_0 para $\alpha = 0.1$.

na recuperação dos protótipos foi obtido para $J_0 = 0.25$. A

linha tracejada indica $m=m_0$.

Escolhendo valores apropriados para J_0 o sistema é capaz de lembrar mesmo para $\alpha=0.3$ e $\alpha=0.4$ como mostram as figuras 20.b e 20.c.

Fig. 20.b - ídem para $\alpha=0.3$ Fig. 20.c - ídem para $\alpha=0.4$

Na figura 21 mostramos que, para um dado α , existe um valor ótimo de J_0 que maximiza a correlação de recuperação de equilíbrio de equilíbrio.

Fig. 21 - A correlação de recuperação de equilíbrio como função de J_0 para $m_0 = 0.6$.

Para comprovar que o papel de J_0 é justamente reter o sistema em um estado ferromagnético nas primeiras etapas de

tempo mostramos na figura 22 o tempo de relaxação normalizado τ_r/τ_0 contra J_0 . Aqui $\tau_r = \tau_r(J_0)$ e $\tau_0 = \tau_r(J_0=0)$. Vemos então que para $J_0 > 0$ e $\alpha > 0.14$ o sistema alcança o equilíbrio muito mais rapidamente do que para $J_0 = 0$. Acreditamos que o aumento do tempo de relaxação para $\alpha < 0.14$ seja devido a competição entre os estados ferromagnéticos já existentes e os gerados pela autointeração.

Fig. 22 - A dependência do tempo de relaxação τ_r/τ_0 com J_0 .

Observamos qualitativamente nas figuras 20 e 21 como a autointeração positiva aumenta de fato a capacidade de armazenagem do modelo em concordância com a aproximação de Kinzel. A dificuldade em medirmos quantitativamente esse efeito está em que o método usual utilizado na medida de α_c , figura 3, é baseado na análise da estabilidade dos protótipos. Entretanto, $J_0 > 0$ afeta essa estabilidade de um modo que nada tem a ver com o aumento da capacidade de armazenagem do mode-

lo como comentado anteriormente.

Voltemos agora nossa atenção para o efeito do termo de autointeração negativo, $J_0 < 0$. Tanto a aproximação de Kinzel quanto o cálculo termodinâmico mostram que esse termo é prejudicial para as propriedades de recuperação do modelo.

Esperamos que para $|J_0| \gg 1$ a dinâmica congele em um ciclo-2 entre a configuração inicial $\{S_i(t=0)\}$ e seu oposto $\{-S_i(t=0)\}$. Como praticamente todos os 2^N estados são estados estacionários da dinâmica nesse regime, o modelo é inútil como memória associativa. A figura 23 ilustra esse fenômeno.

Fig. 23 - A correlação de recuperação de equilíbrio contra J_0 para $m_0 = 1(0)$ e $0.4 (\square)$.

Para $J_0 > -1$ o comportamento é o previsto pelo cálculo termodinâmico. Entretanto para $J_0 < -1$ começam a aparecer efeitos dinâmicos que originam ciclos-2 com $m \neq 0$.

Finalmente, devemos enfatizar que os efeitos da autointeração descritos aqui são puramente dinâmicos e, portanto, não observáveis no cálculo termodinâmico. No Modelo de Hopfield isso é visto claramente : o termo de autointeração contribui como uma constante aditiva ao Hamiltoniano (III. 1) e então não afeta a termodinâmica do modelo. Já no Modelo de Little a situação não é tão clara assim, pois, J_0 contribui de uma forma não trivial ao Hamiltoniano (III.14). Por que então os estados de recuperação descritos nessa secção não foram observados no cálculo termodinâmico do capítulo III ?

Uma das limitações do cálculo termodinâmico está em não poder descrever estados metaestáveis que não estejam separados de outros estados ou dos estados fundamentais por barreiras de energia livre da ordem do tamanho do sistema, $O(N)$. Acreditamos que a separação energética entre os estados de recuperação gerados pela autointeração e os estados vidros de spin seja $O(N^x)$ com $x < 1$, respondendo então a pergunta acima.

Os estados criados pela autointeração são conhecidos como estados de não equilíbrio justamente por serem invisíveis ao cálculo termodinâmico. Na secção seguinte veremos mais um exemplo desses estados.

IV.D. A CORRELAÇÃO DE RECUPERAÇÃO REMANESCENTE

Uma vez que a correlação de recuperação remanescente dos modelos de redes de neurônios possui a mesma natureza da magnetização remanescente do modelo de vidro de spin de Sherrington-Kirkpatrick, vamos, por simplicidade, utilizar esse modelo para apresentar o problema ao leitor.

No equilíbrio térmico a magnetização de um vidro de spin é zero pois os momentos magnéticos apontam em direções aleatórias. Entretanto, se primeiramente alinharmos os dipolos aplicando um campo magnético e então desligarmos o campo, o sistema relaxará para um estado de magnetização finita m_r que decairá lentamente para zero. Portanto, m_r está associado a existência de um enorme espectro de estados metaestáveis e correspondentes tempos de relaxação.

O Modelo SK possui exatamente esse comportamento : não existem estados estáveis com magnetização finita para T finito. Entretanto, esses estados de não equilíbrio tornam-se estáveis para $T=0$ e, portanto, nesse limite podemos obter uma magnetização remanescente finita.

Kinzel [38] mostrou que m_r é uma quantidade autome-diante, isto é, as flutuações entre diferentes amostras desaparecem no limite termodinâmico, dada por

$$m_r = 0.14 \pm 0.01. \quad (\text{IV.27})$$

Ainda, qualquer estado inicial relaxará com probabi

lidade um (no limite $N \rightarrow \infty$) para m_r . A bacia de atração do estado fundamental $m=0$ é praticamente nula. Isso é facilmente entendido desde que o número de estados de não equilíbrio cresce exponencialmente com N [39,40]. Esses estados possuem correlações finitas com qualquer uma das configurações do sistema.

Voltemos agora para as redes de neurônios. Uma vez que a medida de m_r exige recursos computacionais além dos disponíveis, apresentaremos a seguir os poucos resultados disponíveis na literatura, todos para $J_0=0$.

Em redes de neurônios também encontramos um efeito de remanescência devido a existência de um número exponencial em N de estados de não equilíbrio com correlações finitas com qualquer uma das configurações da rede, inclusive os protótipos. Esses estados dominam completamente a dinâmica para $\alpha > \alpha_c$ ofuscando o estado fundamental da fase vidro de spin com $m=0$.

A correlação de recuperação remanescente é uma função de α , $m_r(\alpha)$. Para a dinâmica sequencial (Hopfield), Amit et al [11] obtiveram que $m_r(\alpha)$ é uma função decrescente de α :

$$0.4 < m_r(\alpha) < 0.14 \quad 0.15 < \alpha < \infty \quad . \quad (IV.28)$$

O mesmo comportamento é encontrado na dinâmica paralela (Little), embora com pequenas diferenças quantitativas. Por exemplo, Gardner et al [35] obtiveram

$$m_r(\alpha=\infty) = 0.23 \pm 0.02. \quad (\text{IV.29})$$

A correlação de recuperação remanescente no limite $\alpha \rightarrow \infty$ é idêntica à magnetização remanescente do Modelo SK (com dinâmica sequencial ou paralela). Isso também é válido para a entropia. Entretanto não é claro como esse limite destrói as correlações entre sítios que distinguem o problema das redes de neurônios do Modelo SK.

CAPÍTULO V

CONCLUSÃO

V. CONCLUSÃO

Nesta tese estudamos as propriedades de armazenagem e recuperação de informações de um modelo de memória associativa com processamento paralelo: O Modelo de Little. Damos especial ênfase ao estudo da termodinâmica desse modelo embora alguns aspectos puramente dinâmicos também tenham sido tratados.

Nossos principais resultados referentes a termodinâmica do Modelo de Little são descritos a seguir.

A dinâmica síncrona do Modelo de Little com interações simétricas possui dois diferentes comportamentos assintóticos: pontos fixos e ciclos de período 2 (ciclos- 2). Esses comportamentos são naturalmente descritos por um Hamiltoniano equação (III.14) onde aparece explicitamente uma duplicação do espaço de fase do modelo capaz de acomodar os ciclos - 2. A contribuição de um termo de autointeração J_0 a esse Hamiltoniano não é trivial.

As propriedades de recuperação termodinâmicas dos Modelos de Little e Hopfield são idênticas para $T = 0$ e $J_0 > -1$. Para $T \neq 0$, o Modelo de Little é ligeiramente mais robusto a erros causados pelo ruído térmico. Essa robustez aumenta à medida que aumentamos o valor de J_0 .

O termo de autointeração negativa favorece os ci-

clos - 2. Esses ciclos não estão relacionados com as propriedades de recuperação do modelo as quais estão associadas apenas aos pontos fixos. Para $J_0 < -1$ a fase ferromagnética desaparece totalmente. A competição entre ciclos - 2 e pontos fixos é responsável pela riqueza do diagrama de fases do modelo dando origem a uma superfície de pontos tricríticos.

Com respeito ao aspecto dinâmico do Modelo de Little nosso principal resultado é a observação numérica de que o termo de autointeração positiva cria estados ferromagnéticos de não equilíbrio. Esses estados são criados até mesmo na fase vidro de spin ($\alpha > \alpha_c$) e, portanto, a autointeração positiva produz um aumento efetivo da capacidade de armazenagem do modelo.

Nossos resultados para $J_0 = 0$ são qualitativamente similares aos do Modelo de Hopfield cujo processamento é sequencial. A principal diferença entre esses resultados é a existência de uma fase paramagnética (P) para baixas temperaturas no Modelo de Little, a qual associamos os ciclos - 2 observados na dinâmica paralela.

Se por um lado essa fase P é bem-vinda ao explicar os ciclos - 2 da dinâmica paralela, por outro nos trouxe problemas insolúveis: sua entropia é negativa e diverge, juntamente com a energia, para $-\infty$ no limite $T \rightarrow 0$.

Esse comportamento é comum em modelos de vidro de

spins tratados com o método das réplicas : o mesmo ocorre com a fase P do modelo SK. Entretanto, nesse modelo essa fase não física pode ser descartada pois ela co-existe com uma fase vidro de spin (VS) também de entropia negativa mas, pelo menos, finita. Sabemos que a quebra de simetria das réplicas (QSR) pode sanar o problema da entropia da fase VS, embora não afete em nada o da fase P.

A situação é mais grave no nosso modelo, pois a fase P é nossa única opção na região $J_0 \ll -1$ onde uma solução com $q_0 < 0$ (ciclos - 2) é esperada. A fase VS com $q_0 < 0$ não existe para $\alpha < \frac{2}{\pi}$ conforme discutido na secção III.D.

Tentamos estudar a fase P através de simulações da dinâmica do sistema para $T = 0, \alpha < \frac{2}{\pi}$ e $J_0 < 0$. Essa tentativa foi frustrada pela existência de um número enorme de estados metaestáveis de não equilíbrio que impedem o sistema de atingir os estados termodinâmicos, no caso a suposta fase P.

Para elevarmos nosso problema a uma situação equivalente ao do Modelo SK fazamos o seguinte voto de fé: a QSR pode extender a região de existência da fase VS de forma a podermos descartar a fase P como no Modelo SK. Naturalmente, isso poderia ser verificado (e então deixaria de ser um voto de fé) usando o ansatz de Parisi para a QSR. Porém, tal cálculo, além de extremamente complexo, foge ao objetivo desta tese que é estudar as propriedades de recuperação e armazenagem de informa

mações do modelo, sobre as quais a fase P nada tem a dizer.

O desempenho dos Modelos de Hopfield e Little de memória associativa ainda não pode ser considerado satisfatório. De fato, na referência [22] mostramos a ineficiência desses modelos em distinguir dois padrões correlacionados. Qualquer modelo de memória associativa que clame por alguma realidade biológica deve ser capaz de armazenar e recuperar informações correlacionadas.

Um modelo proposto por Personnaz et al [23] e estudado analiticamente por Kanter e Sompolinsky [24] é capaz de armazenar um conjunto qualquer de padrões linearmente independentes, correlacionados ou não. Entretanto, o preço a pagar é alto : foi necessário abrir mão da localidade das interações entre os neurônios, isto é, J_{ij} depende das interações entre os outros $N - 2$ sítios da rede. Além de não ser aceitável biologicamente, esse modelo é extremamente difícil de implementar em hardware (realização prática do modelo).

Por outro lado, a complexa estrutura ultramétrica (hierárquica) do espaço gerado pelos estados fundamentais do Modelo SK [34] vem já há algum tempo inspirando tentativas de criar um modelo capaz de memorizar padrões correlacionados hierarquicamente através de interações locais [41,42,43]. Há um forte apelo biológico nesse tipo de correlação : primeiramente o modelo memoriza um padrão central e a seguir começa a

memorizar variações e variações das variações cada vez mais refinadas deste padrão central. Esse parece ser o mecanismo de armazenagem de informações de memórias biológicas. Entretanto, todos os modelos que seguem essa linha caracterizam-se por uma tamanha complexidade que mesmo a implementação em software (simulações numéricas) é muito difícil.

Recentemente, Gutfreund [44] propôs um modelo local, relativamente simples, capaz de memorizar padrões hierarquicamente correlacionados sem utilizar a estrutura ultramétrica dos estados da fase vidro do Modelo SK. Esse modelo é uma generalização de um modelo proposto por Amit et al [45] no qual os protótipos $\{\xi_i^\mu\}$ obedecem a distribuição

$$p(\xi_i^\mu) = \frac{1}{2}(1+a) \delta(\xi_i^\mu - 1) + \frac{1}{2}(1-a) \delta(\xi_i^\mu + 1) \quad (V.1)$$

onde $-1 < a < 1$. Assim, existe uma correlação efetiva entre os protótipos :

$$\langle \xi_i^\mu \xi_i^\nu \rangle = a^2 \quad \mu \neq \nu. \quad (V.2)$$

Em ambos modelos é necessário modificar o processo dinâmico restringindo o espaço de fase apenas aos estados que satisfaçam a condição :

$$\frac{1}{N} \sum S_i = a. \quad (V.3)$$

Esse tipo de vínculo não possui paralelo em sistemas biológicos.

Com uma filosofia de trabalho um tanto diferente da que vem sido empregada até esse momento, Diederich e Oppen [46] propuseram um modelo local capaz de memorizar sem erros de recuperação tanto padrões descorrelacionados como correlacionados. Nesse modelo cada padrão é aprendido (observe a mudança de vocabulário) apresentando-o repetidas vezes como um estímulo à rede de neurônios. O processo de aprendizado terminará quando todos os protótipos satisfizerem a condição de estabilidade :

$$\xi_i^u h_i > 0 \quad (V.4)$$

onde h_i é o campo molecular sobre o neurônio i . Essa situação é alcançada através de um algoritmo - algoritmo de aprendizado - que modifica as interações J_{ij} a cada etapa do processo até obter um conjunto de sinápses ideal que garanta a estabilidade dos protótipos.

Observe que as interações J_{ij} são obtidas através de um processo de treinamento da rede para realizar a tarefa de armazenar e reconhecer um dado conjunto de padrões. Não existe uma expressão analítica para os J_{ij} como nos modelos descritos anteriormente e, portanto, o estudo analítico das

propriedades de equilíbrio do sistema não é mais possível.

O Modelo de Diederich e Oppper pertence a uma classe de sistemas inteligentes que prometem uma verdadeira revolução na área da Inteligência Artificial : os computadores neurais ou de 6ª geração.

Esses sistemas operam com um alto grau de paralelismo garantindo um processamento muito mais veloz do que o obtido atualmente em super-computadores convencionais. Por outro lado, os computadores neurais não são programáveis : eles são treinados para realizar tarefas específicas. O processo de treinamento ou aprendizado tem sido o principal objeto de estudo e discussão no estágio atual das pesquisas em computadores neurais.

O objetivo é encontrar um algoritmo de aprendizado geral capaz de gerar o conjunto de conexões J_{ij} que habilita o sistema a resolver um dado problema conhecendo apenas alguns exemplos de soluções do problema. O sistema então deveria ser capaz de generalizar, isto é, obter soluções para situações que não foram apresentados no período de treinamento.

Existem vários desses algoritmos em teste atualmente [47,48,49] . O que mais se destaca é o "back-propagation" criado por Rumelhart et al [48] . Para enfatizarmos o poder desse algoritmo mencionamos uma de suas aplicações conhecida como Netalk, uma rede de neurônios que transforma texto escri

to em palavras faladas [50] . É claro que nunca teremos certaza de que a rede vá responder corretamente quando apresentarmos itens que não foram incluídos no processo de treinamento. No entanto Netalk teve uma eficiência da ordem de 90% ao pronunciar palavras que nunca tinha visto antes.

Deve ficar claro que os computadores neurais nunca substituirão os computadores convencionais, senhores absolutos de confiabilidade e precisão. A principal aplicação dos computadores neurais é a solução de problemas de otimização onde devemos escolher uma solução, a melhor possível, entre um número enorme de opções possíveis que podem até ser conflitantes [51].

Finalizando, vamos agora enfatizar explicitamente a nossa contribuição à teoria das redes de neurônios :

Os resultados da secção II.C referentes ao Modelo de Little são originais embora exista na literatura uma demonstração alternativa do teorema que limita o período dos ciclos no Modelo de Little com interações simétricas [52].

O capítulo III representa nossa principal contribuição à teoria das redes de neurônios. Versões compactas deste capítulo podem ser encontradas na referência [53].

As secções IV.A e IV.B apresentam resultados originais enquanto que o tema da secção IV.C é abordado na referência [54]. A secção IV.D é uma compilação de dados existentes

na literatura.

Para evitar que essa tese se tornasse demasiadamente extensa deixamos de incluir os problemas tratados nas referências [20], [22] e [26] uma vez que esses problemas não estão diretamente relacionados com o Modelo de Little.

APÊNDICE A - O LIMITE $n \rightarrow 0$ NO ANSATZ SIMETRIA DE RÉPLICAS

Apresentamos neste apêndice os detalhes do cálculo do limite $n \rightarrow 0$ da equação (III.48) com o qual obtemos a equação (III.50).

Temos

$$\% = \frac{1}{n} \langle \log \text{Tr}_{S^\rho, \sigma^\rho} \exp \beta H_\xi \rangle_\xi \quad (\text{A.1})$$

onde H_ξ é dado pela equação (III.49).

Linearizando os termos quadráticos nas variáveis de spin:

$$\begin{aligned} \exp \left[\frac{\alpha\beta^2}{2} \left(\tilde{p} - \frac{\tilde{q}_1}{2} \right) \left(\sum_\rho \sigma^\rho \right)^2 \right] &= \int_{-\infty}^{\infty} D(z_1) \exp \left[z_1 \alpha^{\frac{1}{2}} \beta \left(\tilde{p} - \frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} \sum_\rho \sigma^\rho \right] \\ \exp \left[\frac{\alpha\beta^2}{2} \left(\tilde{r} - \frac{\tilde{q}_1}{2} \right) \left(\sum_\rho S^\rho \right)^2 \right] &= \int_{-\infty}^{\infty} D(z_2) \exp \left[z_2 \alpha^{\frac{1}{2}} \beta \left(\tilde{r} - \frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} \sum_\rho S^\rho \right] \\ \exp \left[\frac{\alpha\beta^2}{2} \frac{\tilde{q}_1}{2} \left(\sum_\rho S^\rho + \sigma^\rho \right)^2 \right] &= \int_{-\infty}^{\infty} D(z_3) \exp \left[z_3 \alpha^{\frac{1}{2}} \beta \left(\frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} \sum_\rho S^\rho + \sigma^\rho \right] \end{aligned} \quad (\text{A.2})$$

onde $D(z_i) = \frac{dz_i}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} z_i^2}$, $i = 1, 2, 3$, reescrevemos % como

$$\begin{aligned} \% &= \frac{1}{n} \langle \log \text{Tr}_{S^\rho, \sigma^\rho} \int_{-\infty}^{\infty} D(z_1) D(z_2) D(z_3) \exp \left\{ -n \frac{\alpha\beta^2}{2} (\tilde{p} + \tilde{r}) + \right. \\ &+ \frac{\alpha\beta^2}{2} (\tilde{q}_0 - \tilde{q}_1) \sum_\rho S^\rho \sigma^\rho + \beta \alpha^{\frac{1}{2}} \left[z_1 \left(\tilde{p} - \frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} + z_3 \left(\frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} + \vec{\xi} \cdot (\vec{\ell} - \vec{n}) \right] \sum_\rho \sigma^\rho \\ &+ \beta \alpha^{\frac{1}{2}} \left[z_2 \left(\tilde{r} - \frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} + z_3 \left(\frac{\tilde{q}_1}{2} \right)^{\frac{1}{2}} + \vec{\xi} \cdot (\vec{\ell} - \vec{m}) \right] \sum_\rho S^\rho \left. \right\} \rangle_\xi. \end{aligned} \quad (\text{A.3})$$

Agora o cálculo dos traços torna-se trivial:

$$\begin{aligned} \mathfrak{z} = & \frac{1}{n} \left\langle \log \int_{-\infty}^{\infty} D(z_1) D(z_2) D(z_3) \exp \left\{ - \frac{n\alpha\beta^2}{2} (\tilde{p} + \tilde{r} + \tilde{q}_1 - \tilde{q}_0) + \right. \right. \\ & \left. \left. n \log 2 \left[\cosh \beta_{\Xi_1} + e^{-\alpha\beta\tilde{Q}} \cosh \beta_{\Xi_2} \right] \right\} \right\rangle_{\xi} \end{aligned} \quad (\text{A.4})$$

onde usamos

$$\text{Tr}_{S,\sigma} \exp [A_1\sigma + A_2S + A_3S\sigma] = 2e^{A_3} [\cosh(A_1 + A_2) + e^{-2A_3} \cosh(A_1 - A_2)]$$

e $\Xi_1, \Xi_2, \tilde{Q}$ são dados pelas equações (III.51a-c).

Expandido a exponencial para n pequeno e tomando o limite $n \rightarrow 0$ obtemos a equação (III.50):

$$\mathfrak{z} = - \frac{\alpha\beta^2}{2} (\tilde{p} + \tilde{r} + \tilde{q}_1 - \tilde{q}_0) + \left\langle \log 2 \left[\cosh \beta_{\Xi_1} + e^{-\alpha\beta\tilde{Q}} \cosh \beta_{\Xi_2} \right] \right\rangle_{\xi} \quad (\text{A.5})$$

APÊNDICE B - A ESTABILIDADE DAS SOLUÇÕES FERROMAGNÉTICAS FRENTE
A QUEBRA DE SIMETRIA DAS RÉPLICAS (QSR)

Considerando apenas as soluções simétricas, equação (III.92), reescrevemos a energia livre, equação (III.37), como

$$f = \lim_{n \rightarrow 0} \frac{1}{n} \text{Min} \left\{ + \sum_{\rho} \vec{m}_{\rho}^2 + \frac{1}{2} \alpha \beta \sum_{\rho} \tilde{q}_{\rho\rho} q_{\rho\rho} + \alpha \beta \sum_{\rho \neq \sigma} \tilde{q}_{\rho\sigma} q_{\rho\sigma} + \frac{\alpha}{2\beta} \Delta' \right. \\ \left. + (\alpha - J_0) \sum_{\rho} q_{\rho\rho} - \beta^{-1} \langle \log \text{Tr}_{S^{\rho}} \text{Tr}_{\sigma^{\rho}} \exp (\beta H_{\xi}) \rangle_{\xi} \right\} \quad (\text{B.1})$$

onde

$$\Delta' = \log \det \left[(1 - \beta \underline{Q}_0 - \beta \underline{Q})^2 - \beta^2 \underline{Q}^2 \right] \quad (\text{B.2})$$

$$H_{\xi} = \alpha \beta \sum_{\rho \neq \sigma} \tilde{q}_{\rho\sigma} (\sigma^{\rho} \sigma^{\sigma} + S^{\rho} S^{\sigma} + \frac{1}{2} S^{\rho} \sigma^{\sigma}) + \frac{\alpha \beta}{2} \sum_{\rho} \tilde{q}_{\rho\rho} S^{\rho} \sigma^{\rho} + \xi \cdot \sum_{\rho} \vec{m}_{\rho} (S^{\rho} + \sigma^{\rho}) \quad (\text{B.3})$$

com

$$(\underline{Q}_0)_{\rho\sigma} = \delta_{\rho,\sigma} q_{\rho\rho} \quad (\underline{Q})_{\rho\sigma} = \delta_{\rho,\sigma} + (1 - \delta_{\rho,\sigma}) q_{\rho\sigma}.$$

Examinemos agora a questão se o ansatz simetria de réplicas (ASR) minimiza de fato a expressão entre chaves da equação (B.1). O procedimento que deveríamos seguir para estudar a estabilidade das soluções obtidas com o ASR frente a QSR seria escrever

$$\vec{m}_\rho = \vec{m} + \delta\vec{m}_\rho \quad (B.4a)$$

$$q_{\rho\sigma} = q_1 + \delta q_{\rho\sigma} \quad \tilde{q}_{\rho\sigma} = \tilde{q}_1 + \delta\tilde{q}_{\rho\sigma} \quad \rho \neq \sigma \quad (B.4b)$$

$$q_{\rho\rho} = q_0 + \delta q_{\rho\rho} \quad \tilde{q}_{\rho\rho} = \tilde{q}_0 + \delta\tilde{q}_{\rho\rho} \quad (B.4c)$$

e então expandir a energia livre até segunda ordem nas variações $\delta\vec{m}_\rho$, $\delta q_{\rho\sigma}$, etc.. A seguir calcularíamos os autovalores da matriz de estabilidade cujos elementos são os coeficientes das combinações quadráticas dessas variações. O ASR seria estável se todos os autovalores fossem positivos.

Infelizmente, devido ao número excessivo de parâmetros de ordem esse procedimento é impraticável. Entretanto, podemos fazer algumas simplificações de modo a tornar o problema tratável.

Primeiramente, lembremos que no trabalho original de Almeida - Thouless os elementos de matriz associados à variação $\delta\vec{m}_\rho$ dão sempre autovalores positivos. O autovalor que torna-se negativo à baixas temperaturas indicando a instabilidade do ASR está associado apenas ao coeficiente do termo $\delta q_{\rho\sigma} \delta q_{\gamma\delta}$. Façamos então $\delta\vec{m}_\rho = 0$.

A seguir observemos que os parâmetros $q_{\rho\rho}$ e $\tilde{q}_{\rho\rho}$ não são parâmetros de ordem da fase vidro de spin, uma vez que não se anulam na fase paramagnética. Assim, não devem possuir uma estrutura complexa como ocorreria se a QSR tivesse algum efeito

sobre eles. Desse modo temos $\delta q_{\rho\rho} = \delta \tilde{q}_{\rho\rho} = 0$.

Com essas simplificações, a matriz de estabilidade de dimensão $n(n-1) \times n(n-1)$ tem a seguinte estrutura

$$C = \begin{pmatrix} A^{\alpha\beta, \gamma\delta} & D^{\alpha\beta, \gamma\delta} \\ D^{\alpha\beta, \gamma\delta} & B^{\alpha\beta, \gamma\delta} \end{pmatrix} \quad \begin{matrix} \alpha+\beta \\ \gamma+\delta \end{matrix} \quad (B.5)$$

com

$$A^{\gamma\delta, \alpha\beta} \equiv \frac{\partial^2 (nf)}{\partial q_{\gamma\delta} \partial q_{\alpha\beta}} = -2\alpha\beta \left[1+\beta-\frac{\beta}{2} (q_{\alpha\alpha} + q_{\beta\beta}) \right] \cdot \left[1+\beta-\frac{\beta}{2} (q_{\gamma\gamma} + q_{\delta\delta}) \right] \\ \times (1-\beta\Theta)^{-1}_{\beta\gamma} (1-\beta\Theta)^{-1}_{\delta\alpha} \quad (B.6)$$

$$B^{\gamma\delta, \alpha\beta} \equiv \frac{\partial^2 (nf)}{\partial \tilde{q}_{\gamma\delta} \partial \tilde{q}_{\alpha\beta}} = -\frac{\alpha^2 \beta^3}{8} \{ \langle S^\alpha S^\beta S^\gamma S^\delta \rangle + 4 \langle S^\alpha S^\beta S^\gamma S^\delta \rangle + \langle S^\alpha S^\beta S^\gamma S^\delta \rangle \\ + 2 \langle S^\alpha S^\beta S^\gamma S^\delta \rangle - 2 (\langle S^\alpha S^\beta \rangle + \langle S^\alpha S^\beta \rangle)^2 \} \quad (B.7)$$

$$D^{\gamma\delta, \alpha\beta} \equiv \frac{\partial^2 (nf)}{\partial \tilde{q}_{\gamma\delta} \partial q_{\alpha\beta}} = \alpha\beta \sigma_{\alpha\gamma} \sigma_{\beta\delta} \quad (B.8)$$

onde Θ é uma matriz cujos elementos são

$$\Theta_{\alpha\alpha} = 2 q_{\alpha\alpha} + \beta(1-q_{\alpha\alpha}^2) \quad (B.9a)$$

$$\Theta_{\alpha\beta} = 2 q_{\alpha\beta} \left(1+\beta-\frac{\beta}{2} (q_{\alpha\alpha} + q_{\beta\beta}) \right) \quad \alpha+\beta \quad (B.9b)$$

e as médias $\langle \dots \rangle$ são tomadas com o Hamiltoniano (B.3) escri-

to com o ASR.

A matriz $\tilde{A}$ possui três diferentes tipos de elementos calculados com o ASR:

$$A^{\alpha\beta, \beta\alpha} \equiv P_A = - 2\alpha\beta (1+\beta-\beta q_0)^2 \lambda_{\alpha\alpha}^2 \quad (B.10a)$$

$$A^{\alpha\beta, \beta\delta} \equiv Q_A = - 2\alpha\beta (1+\beta-\beta q_0)^2 \lambda_{\alpha\alpha} \lambda_{\beta\alpha} \quad (B.10b)$$

$$A^{\alpha\beta, \gamma\delta} \equiv R_A = - 2\alpha\beta (1+\beta-\beta q_0)^2 \lambda_{\beta\alpha}^2 \quad (B.10c)$$

onde

$$\lambda_{\alpha\alpha} \equiv (\tilde{1}-\beta \tilde{Q})_{\alpha\alpha}^{-1} = \frac{4\beta q_1(1+\beta-\beta q_0) + (1-\beta q_0)^2 - \beta^2}{[2\beta q_1(1+\beta-\beta q_0) + (1-\beta q_0)^2 - \beta^2]^2} \quad (B.11a)$$

$$\lambda_{\alpha\beta} \equiv (\tilde{1}-\beta \tilde{Q})_{\alpha\beta}^{-1} = \frac{2\beta q_1(1+\beta-\beta q_0)}{[2\beta q_1(1+\beta-\beta q_0) + (1-\beta q_0)^2 - \beta^2]^2} \quad (B.11b)$$

A matriz $\tilde{B}$ também possui uma estrutura semelhante:

$$B^{\alpha\beta, \alpha\beta} \equiv P_B = - \frac{\alpha^2 \beta^3}{8} (3 + 4 \langle S_{\sigma}^{\beta} S_{\sigma}^{\beta} \rangle + \langle S_{\sigma}^{\alpha} S_{\sigma}^{\alpha} S_{\sigma}^{\beta} S_{\sigma}^{\beta} \rangle - 8 \langle S_{\sigma}^{\alpha} S_{\sigma}^{\beta} \rangle^2) \quad (B.12a)$$

$$B^{\alpha\beta, \alpha\delta} \equiv Q_B = - \frac{\alpha^2 \beta^3}{8} (7 \langle S_{\sigma}^{\alpha} S_{\sigma}^{\beta} \rangle + \langle S_{\sigma}^{\alpha} S_{\sigma}^{\alpha} S_{\sigma}^{\beta} S_{\sigma}^{\delta} \rangle - 8 \langle S_{\sigma}^{\alpha} S_{\sigma}^{\beta} \rangle^2) \quad (B.12b)$$

$$B^{\alpha\beta, \gamma\delta} \equiv R_B = - \alpha^2 \beta^3 (\langle S_{\sigma}^{\alpha} S_{\sigma}^{\beta} S_{\sigma}^{\gamma} S_{\sigma}^{\delta} \rangle - \langle S_{\sigma}^{\alpha} S_{\sigma}^{\beta} \rangle^2). \quad (B.12c)$$

Testaremos a estabilidade do modo "replicon" no qual temos

$$\delta \tilde{q}_{\rho\sigma} = x \delta q_{\rho\sigma} \equiv x \eta_{\rho\sigma} . \quad (B.13)$$

Segundo de Almeida - Thouless o autovalor relevante para a análise de estabilidade frente a QSR está associado a autovetores simétricos em relação a troca de todos, exceto dois, índices de réplicas. Esses autovetores apropriadamente normalizados são

$$\eta_{\rho\sigma} = \eta \quad \rho, \sigma \neq 1, 2$$

$$\eta_{1\rho} = \eta_{2\rho} = \frac{1}{2} (3 - n) \eta \quad \rho \neq 1, 2$$

$$\eta_{12} = \eta_{21} = \frac{1}{2} (2 - n) (3 - n) \eta$$

$$\eta_{\rho\rho} = 0 \quad (B.14)$$

Assim, a equação de autovalores é dada por

$$\sum_{\gamma\delta} (A^{\rho\sigma, \gamma\delta} + x D^{\rho\sigma, \gamma\delta}) \eta_{\gamma\delta} = \lambda \eta \quad \rho, \sigma \neq 1, 2$$

$$\sum_{\gamma\delta} (x B^{\rho\sigma, \gamma\delta} + D^{\rho\sigma, \gamma\delta}) \eta_{\gamma\delta} = \lambda x n \quad \rho, \sigma \neq 1, 2 . \quad (B.15)$$

Usando os valores obtidos para os elementos das matrizes $\underline{A}$, $\underline{B}$ e $\underline{D}$ e tomando o limite $n \rightarrow 0$ obtemos os seguintes re-

sultados

$$\begin{aligned}\sum_{\gamma\delta} A^{\rho\sigma,\gamma\delta} \eta_{\gamma\delta} &= (P_A - 2Q_A + R_A) n \\ \sum_{\gamma\delta} B^{\rho\sigma,\gamma\delta} \eta_{\gamma\delta} &= 2 (P_B - 2Q_B + R_B) n \\ \sum_{\gamma\delta} D^{\rho\sigma,\gamma\delta} \eta_{\gamma\delta} &= \alpha\beta\eta\end{aligned}\tag{B.16}$$

com os quais reescrevemos a equação de autovalores:

$$\begin{aligned}- \frac{2\alpha\beta (1+\beta - \beta q_0)^2}{[2\beta q_1(1+\beta - \beta q_0) + (1-\beta q_0)^2 - \beta^2]^2} + \alpha\beta x &= \lambda \\ - \alpha^2\beta^3 \chi x + \alpha\beta &= \lambda x\end{aligned}\tag{B.17}$$

onde

$$\chi = \langle 2 [1 - \sinh^2 \beta \Xi_1 / D^2]^2 + e^{-\alpha\beta\tilde{Q}} / D [3 \sinh^2 \beta \Xi_1 / D^2 + e^{-\alpha\beta\tilde{Q}} / D - 3] \rangle$$

$$D = \cosh \beta \Xi_1 + e^{-\alpha\beta\tilde{Q}}$$

$$\tilde{Q} = \beta(\tilde{q}_0 - \tilde{q}_1) \quad \Xi_1 = (2\alpha\tilde{q}_1)^{\frac{1}{2}} z + 2\vec{m} \cdot \vec{\xi} \quad \langle \dots \rangle = \int_{-\infty}^{\infty} \frac{dz}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}z^2} (\dots)$$

Eliminando x obtemos uma equação quadrática para $\tilde{\lambda} = \lambda/\alpha\beta$ cujas soluções são

$$\tilde{\lambda}_{\pm} = -\frac{1}{2} (U+V) \pm \left[\frac{1}{4} (U+V)^2 + 1 - UV \right]^{\frac{1}{2}}\tag{B.18}$$

onde

$$V = \alpha \beta^2 \chi$$

$$U = \frac{1}{2} (1 + \beta - \beta q_0)^2 [2\beta q_1 (1 + \beta - \beta q_0) + (1 - \beta q_0)^2 - \beta^2]^{-2}$$

A altas temperaturas temos $\lambda_- < 0$ e $\lambda_+ > 0$. O autovalor λ_- não troca de sinal nunca e, segundo Amit et al, seu sinal deve ser corrigido por uma escolha apropriada do contorno de integração na aplicação do método do ponto de sela. O autovalor importante é λ_+ que troca de sinal em

$$UV = 1 \tag{B.19}$$

assinalando a instabilidade frente a QSR. A equação (B.19) deve ser resolvida simultaneamente com as equações dos parâmetros de ordem a fim de obtermos o análogo da linha de Almeida-Thouless para o Modelo de Little.

REFERÊNCIAS

- [1] VON NEUMANN, J., Collected Works, Vol.5, 304 (NY: Pergamon, 1963)
- [2] CRAGG, B.G. e TEMPERLEY, H.H.V., Electroencephalog. clin. Neurophys. 6 85 (1954)
- [3] McCULLOCH, W.S. e PITTS, W.A., Bull.Math.Biophys. 5 115 (1943)
- [4] LITTLE, W.A., Math. Biosci. 19 101 (1974)
- [5] ISING, E., Z. Physik. 31 253 (1925)
- [6] HOPFIELD, J.J., Proc. Nat. Acad. Sci. USA 79 2554 (1982)
- [7] GLAUBER, R.J., J. Math. Phys. 4 294 (1963); BINDER, K. em "Fundamental Problems in Statistical Mechanics V", Cohen EGD ed. (North Holland, Amsterdam, 1980)
- [8] EDWARDS, S.F. e ANDERSON, P.W., J.Phys. F 5 965 (1975)
- [9] KIRKPATRICK, S. e SHERRINGTON, D.J., Phys. Rev. B 17 4384 (1978)
- [10] AMIT, D.J., GUTFREUND, H. e SOMPOLINSKY, H., Phys. Rev. A 32 1007 (1985)
- [11] AMIT, D.J., GUTFREUND, H. e SOMPOLINSKY, H., Phys. Rev. Lett. 55 1530 (1985); Ann. Phys. 173 30 (1987)
- [12] KINZEL, W., Z. Phys. B 60 205 (1985)
- [13] CLARK, J.W., RAFELSKI, J. e WINSTON, J.V., Phys. Reports 123 215 (1985)
- [14] AMIT, D.J., "The Properties of Models of Simple Neural Networks", em Morgenstein, I. e Van Hemmen, J.L. (eds) : Heidelberg Symposium on Glassy Dynamics (Berlin : Springer-Verlag, 1987)

- [15] HEBB, D.O., The Organization of Behaviour, 62 (NY: Wiley, 1949)
- [16] MATTIS, D.C., Phys. Lett. A 56 421 (1976)
- [17] COOPER, L.N., "A Possible Organization of Animal Memory and Learning" em Lundqvist, B. e Lundqvist, S. (eds) : Proceeding of the Nobel symposium on collective properties of physical systems (NY: Academic Press, 1973)
- [18] MEZARD, M., NADAL, J.P. e TOULOUSE, G., J. Physique 47 1457 (1986)
- [19] LINDSAY, P.H. e NORMAN, D.A., Human Information Processing (NY: Academic Press, 1977)
- [20] FONTANARI, J.F. e KOBERLE, R., J.Phys. A 21 L253 (1988)
- [21] WEISBUCH, G. e FOGELMAN-SOULIÉ, F., J.Physique Lett. 46 L623 (1985)
- [22] FONTANARI, J.F. e KOBERLE, R., J.Phys.A 21 em impressão
- [23] PERSONNAZ, L., GUYON, I. e DREYFUS, G., Phys. Rev. A 34 4217 (1986)
- [24] KANTER, I. e SOMPOLINSKY, H., Phys. Rev. A 35 380 (1987)
- [25] PARISI, G., J. Phys. A 19 L675 (1986)
- [26] FONTANARI, J.F. e KOBERLE, R., J. Phys. A 21 L259 (1988)
- [27] PERETTO, P., Biol. Cybern. 50 51 (1984)
- [28] GARDNER, E., J. Phys. A 19 L1047 (1986)
- [29] VAN HEMMEN, J.L., Phys. Rev. Lett. 49 409 (1982)
- [30] PROVOST, J.P. e VALLEE, G., Phys. Rev. Lett. 50 598 (1983)

- [31] VAN HEMMEN, J.L. e PALMER, R.G., J. Phys.A 15 3881
(1982)
- [32] DE ALMEIDA, J.R.L. e THOULESS, D.J., J. Phys. A 11
983 (1978)
- [33] PARISI, G., Phys. Rev. Lett. 50 1946 (1983)
- [34] MEZARD, M., PARISI, G., SOURLAS, N., TOULOUSE, G. e
VIRASORO, M., J. Physique 45 843 (1984)
- [35] GARDNER, E., DERRIDA, B. e MOTTISHAW, P., J. Physique
48 741 (1987)
- [36] DERRIDA, B., GARDNER, E. e ZIPPELIUS, A., Europhys.
Lett. 4 167 (1987)
- [37] VAN HEMMEN, J.L., Springer Lectures Notes in Physics
275 547 (1987)
- [38] KINZEL, W., Phys. Rev. B 33 5086 (1986)
- [39] DE DOMINICIS, C., GABAY, M., GAREL, T. e ORLAND, H.,
J. Physique 41 923 (1980)
- [40] BRAY, A.J. e MOORE, M.A., J. Phys. C 13 L469 (1980)
- [41] TOULOUSE, G., DEHAENE, S., CHANGEUX, J.P., Proc.Nat.
Acad. Sci. USA 83 1695 (1986)
- [42] DOTSENKO, V.S., J. Phys. C 18 L1017 (1985)
- [43] PARGA, N. e VIRASORO, M.A., J. Physique 47 1857
(1986)
- [44] GUTFREUND, H., Phys. Rev. A 37 570 (1988)
- [45] AMIT, D.J., GUTFREUND, H. e SOMPOLINSKY, H., Phys.
Rev. A 35 2293 (1987)
- [46] DIEDERICH, S. e OPPER, M., Phys. Rev. Lett. 58 949
(1986)
- [47] SEJNOWSKI, T.J., KIENKER, P.K. e HINTON, G.E., Physi-
ca D 22 260 (1986)

- [48] RUMELHART, D.E., HINTON, G.E. e WILLIAMS, R.J., Nature 323 533 (1986)
- [49] PATARNELLO, S. e CARNEVALI, P., Europhys. Lett. 4 503 (1987)
- [50] SEJNOWSKI, T.J. e ROSENBERG, C.R., The Johns Hopkins University Electrical Engineering and Computer Science Technical Report JHU/EECS - 86/01 (Baltimore, Maryland 1986)
- [51] HOGG, T. e HUBERMAN, B.A., Phys. Reports 156 227 (1987)
- [52] FRUMKIN, A. e MOSES, E., Phys. Rev. A 34 714 (1986)
- [53] FONTANARI, J.F. e KOBERLE, R., Phys. Rev. A 36 2475 (1987), J. Physique 49 13 (1988)
- [54] FONTANARI, J.F. e KOBERLE, R., "Retrieval via non-equilibrium states in neural networks" preprint IFQSC/1988 a aparecer no J. Phys. A, Letters.