

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

**Propriedades cognitivas em redes de
neurônios não-monótonos**

Amanda de Azevedo Lopes

Porto Alegre, Rio Grande do Sul, Brasil

Novembro de 2013

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA

Propriedades cognitivas em redes de neurônios não-monótonos

Trabalho de conclusão de curso submetido ao Instituto de Física da UFRGS como parte dos pré-requisitos para a obtenção do grau de Bacharel em Física: Pesquisa Básica. Realizado sob a orientação do Prof. Rubem Erichsen Junior

Porto Alegre, Rio Grande do Sul, Brasil

Novembro de 2013

Agradecimentos

Ao professor Rubem por todo o apoio e conselhos dados. À minha família, por toda compreensão, apoio e carinho que recebi. Aos meus amigos por todo o companherismo durante o curso, com certeza estes cinco anos foram mais alegres junto de vocês.

Resumo

Entender a origem das propriedades cognitivas do cérebro é uma das grandes questões científicas da atualidade. Em especial, o estudo da capacidade de memória associativa é extremamente relevante para alcançar o entendimento de tais propriedades. Tentando entender tal processo, elaborou-se um breve apanhado histórico sobre redes neurais, foi feita uma revisão do modelo de Hopfield e foi proposta uma alteração na função de ativação da rede de Hopfield estocástica, a fim de determinar a capacidade crítica de aprendizado desta rede. Esta alteração foi estudada em uma rede recorrente totalmente conexa e a função de ativação escolhida foi uma não-monótona. Através de solução numérica pelo método de Newton-Raphson este sistema pode ser resolvido e ao variar um parâmetro de não-monotonicidade da rede, a capacidade pode ser analisada. Um aumento na capacidade da rede é observado em um certo intervalo de valores deste parâmetro de não-monotonicidade.

Abstract

To understand the origin of the brain cognitive properties is one of the major scientific issues of our time. In particular, the study of associative memory is extremely relevant to achieve the understanding of these properties. Trying to understand this process, a short review on neural networks and a review on the Hopfield model were made and an alteration in the activation function of the stochastic Hopfield network was proposed in order to establish the learning capacity of this network. This modification was studied in a totally connected recurrent network and solved through numeric integration. Varying a non monotonicity network parameter the capacity can be analysed. A network capacity increase is observed for a certain range of this parameter.

Sumário

1	Introdução	11
2	Modelo de Hopfield	13
2.1	O modelo	13
2.1.1	Rede com um padrão	15
2.1.2	Rede com p padrões	16
2.2	Capacidade de armazenamento	16
2.3	Redes estocásticas	17
2.3.1	Capacidade de armazenamento da rede estocástica	19
3	Rede não-monótona	25
3.1	Capacidade da rede não-monótona	26
3.1.1	Capacidade da rede no limite de temperatura zero	29
4	Resultados	31
5	Conclusões	35
	Referências	37

1 Introdução

O entendimento da origem das propriedades cognitivas do cérebro, como a capacidade de memória associativa, é um dos grandes desafios da pesquisa científica atual. Para enfrentar este desafio, diferentes áreas científicas, como física, biologia e medicina têm unido esforços. Especificamente na área da física destaca-se, entre outros assuntos, o estudo de modelos elementares de neurônios e sinapses. O cérebro é um sistema altamente complexo, composto de aproximadamente 10^{11} neurônios de vários tipos. Os neurônios são as unidades básicas de processamento de informação no sistema nervoso e podem realizar certas tarefas de computação, como reconhecimento de padrões, de forma muito mais rápida do que o mais rápido computador existente.

Uma rede neural é um modelo que pretende descrever o modo como o cérebro executa uma tarefa ou função de interesse. Estamos, entretanto, muito longe de entender toda a dinâmica do cérebro e sua capacidade de computação. O uso de redes neurais proporciona a compreensão de algumas propriedades, como capacidade de aprendizado, adaptatividade e tolerância a falhas. Um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural. Um modelo de neurônio é formado por um conjunto de sinapses, que conecta os neurônios com um determinado peso para esta conexão, pela soma ponderada com os pesos sinápticos dos sinais de entrada (inputs) e por uma função de ativação para limitar a amplitude do sinal de saída (output) do neurônio.

Um dos primeiros modelos de neurônios foi proposto por McCulloch e Pitts em 1943 (1). McCulloch e Pitts mostraram, que uma rede com um número suficiente de "neurônios" com conexões sinápticas devidamente ajustadas e operando sincronizadamente, em princípio, poderia computar qualquer função que possa ser calculável.

Chamado de Linear Threshold Function ou neurônio de McCulloch-Pitts, o modelo de neurônio proposto por McCulloch e Pitts computa uma soma ponderada dos valores de entrada (input) dos outros neurônios e devolve o valor de saída (output) zero ou um, dependendo se a soma está acima ou abaixo de um limiar de disparo.

$$n_i(t + 1) = \Theta\left(\sum_j w_{ij}n_j(t) - \mu_i\right). \quad (1.1)$$

A função n_i retorna 1 ou 0, representando os estados "disparando" e "não disparando" ou "quiescente", respectivamente. O tempo t é discreto, e $\Theta(x)$ é a função de Heaviside ou função degrau. w_{ij} representa o peso das sinapses entre o neurônio j e o neurônio i . μ_i é o valor do limiar do neurônio i , a soma ponderada precisa alcançar ou exceder o limiar

para o neurônio poder disparar. Porém, a função degrau $\Theta(x)$ poderia ser substituída por outra função não-linear, que determina a complexidade do neurônio.

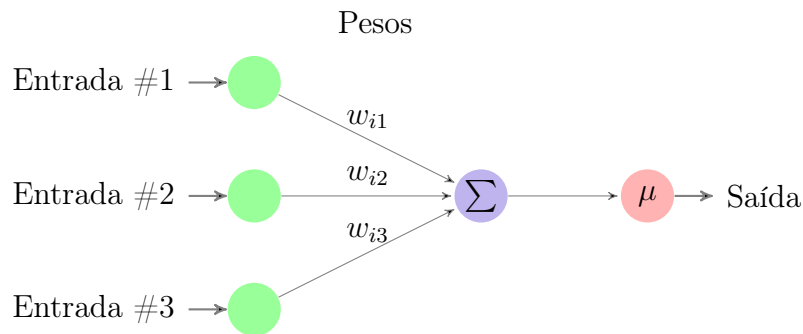


Figura 1 – Diagrama esquemático de um neurônio McCulloch-Pitts

O próximo grande desenvolvimento em redes neurais foi feito com a publicação do livro de Hebb em 1949 (2). No livro *The Organization of Behaviour* foi apresentada pela primeira vez uma regra de aprendizado fisiológico para modificação sináptica. Hebb propôs que os pesos das conexões entre os neurônios sejam ajustados de acordo com as correlações pré e pós-sinápticas. Se o neurônio pré-sináptico influencia no disparo do neurônio pós-sináptico, ou seja, se eles disparam simultaneamente, a conexão entre estes neurônios será fortalecida e se o neurônio pós-sináptico disparar antes de o neurônio pré-sináptico disparar, estes neurônios são ativados de forma não sincronizada e a conexão entre eles será diminuída.

O livro de Hebb incentivou o desenvolvimento de modelos de aprendizado, sistemas adaptativos e memória associativa. Memória associativa significa uma memória localizável por conteúdo, no qual o conteúdo completo da informação é recuperado a partir de uma informação inicial incompleta. Ou seja, diferentes valores de entrada poderiam ativar uma mesma resposta correta da rede, se estes fossem suficientemente similares. A principal função deste tipo de memória é de recuperar um padrão armazenado na memória em resposta a uma versão incompleta ou com ruído do padrão armazenado. Assim, uma memória localizável por conteúdo é capaz de corrigir erros, no sentido de que a memória pode apagar informações inconsistentes nas informações fornecidas a ela. Algumas das aplicações de memórias associativas são em reconhecimento e reconstrução de imagens. (3)

Em 1982, Hopfield (4) usou a idéia de uma função de energia para formular um novo modo de entender a computação feita por redes recorrentes com conexões sinápticas simétricas. Além disso, ele estabeleceu o isomorfismo entre esse tipo de rede e o modelo de Ising usado em mecânica estatística. Assim, Hopfield, usando o modelo de Ising, conseguiu analisar o caso de redes recorrentes com conexões simétricas garantindo, assim, sua convergência para uma condição estável (5). O modelo de Hopfield será estudado no próximo capítulo.

2 Modelo de Hopfield

Neste capítulo será dado enfoque ao modelo de Hopfield, que é o modelo base deste trabalho.

2.1 O modelo

Uma rede de Hopfield pode ser descrita da seguinte forma: Um conjunto de N neurônios de McCulloch-Pitts interconectados aleatoriamente com valores de ativação $+1$ (disparando) e -1 (quiescente). As sinapses podem ser atualizadas seguindo a regra de Hebb, com conexões simétricas e pode-se definir que um neurônio não pode estar conectado a ele mesmo, ou seja, define-se que $w_{ii} = 0$. Desta forma, pode-se escrever uma equação para a dinâmica da rede

$$S_i = \text{sgn} \left(\sum_j w_{ij} S_j - \theta_i \right). \quad (2.1)$$

S_i é o estado do neurônio i , w_{ij} corresponde ao peso sináptico que conecta o neurônio i ao j e θ é o valor do potencial limite. Para esta rede que será estudada, o termo θ_i é definido como zero. A função sinal é definida da forma

$$\text{sgn}(x) = \begin{cases} 1 & \text{se } x \geq 0; \\ -1 & \text{se } x < 0. \end{cases} \quad (2.2)$$

Ou seja, o estado de um neurônio S_i pode assumir os valores $+1$ ou -1 . Um padrão ξ_i é um vetor de estados, cujo tamanho é igual ao número de neurônios N da rede. Para que um padrão seja armazenado é necessário que este seja estável.

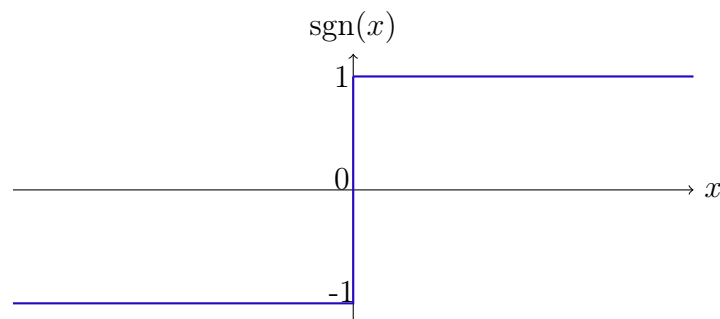


Figura 2 – Função $\text{sgn}(x)$

A atualização da rede pode ser feita de duas formas: sincronizada, atualizando todos os neurônios simultaneamente a cada passo de tempo ou não sincronizada, atualizando

um neurônio por passo de tempo. O caso não sincronizado é escolhido por praticidade para simulação e um neurônio da rede é atualizado a cada passo de tempo através da regra de dinâmica da rede (2.1).

Há duas fases de operação para a rede de Hopfield enquanto memória localizável por conteúdo: uma fase de armazenamento e uma fase de recuperação. Na fase de armazenamento, é fornecido para a rede um conjunto de valores de entrada que se quer memorizar. Os pesos sinápticos, conforme dito anteriormente, são definidos pela regra de Hebb,

$$w_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \quad (2.3)$$

onde p é o número de padrões que se quer memorizar. A constante de proporcionalidade é definida como $\frac{1}{N}$ por simplicidade. Conforme mencionado anteriormente, w_{ii} é definido como zero, para todo i .

Na fase de recuperação, uma entrada de teste é fornecida para a rede. Esta entrada pode ser uma versão com ruído do padrão armazenado, ou apenas uma entrada incompleta. A recuperação de informação segue conforme a regra de dinâmica da rede (2.1). Se num dado instante de tempo a soma dos produtos dos potenciais dos neurônios pré-sinápticos com os pesos sinápticos for maior que zero, o neurônio i mudará seu valor para $+1$, ou continuará nesse estado, caso já estivesse nele. Se a soma desses produtos for menor que zero, o neurônio i mudará seu valor para -1 , ou continuará nesse estado, caso já estivesse nele. A atualização não sincronizada da rede é feita de forma aleatória e ela segue até que o sistema não tenha mais mudança alguma para registrar. O sistema estará agora em um estado estável, ou em um ponto fixo no espaço de fases do sistema (5).

No seu artigo em 1982, Hopfield definiu uma função de energia

$$H = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N w_{ij} S_i S_j . \quad (2.4)$$

Esta foi uma das contribuições mais importantes deste artigo. Hopfield resalta que devido a regra de dinâmica da rede, a função de energia H é uma função monotonicamente decrescente. A rede será atualizada até que ela venha a convergir para um mínimo (local) de energia (4). Ou seja, os pontos mínimos locais da função de energia são os atratores, ou os pontos fixos do sistema.

Para caracterizar as propriedades do modelo, estuda-se o problema genérico de um conjunto de padrões aleatórios. Considera-se que os padrões podem assumir os valores $+1$ e -1 com probabilidades iguais. Primeiro é preciso testar se os pesos sinápticos propostos são aceitáveis. Para tal, a estabilidade dos padrões memorizados é testada e é verificado se a rede corrige pequenos erros nestes padrões conforme ela evolui.

2.1.1 Rede com um padrão

Antes de estudar o caso mais genérico com p padrões aleatórios, estuda-se o caso com apenas um padrão ξ_i . A rede é definida como no início do capítulo e a condição para a estabilidade da rede é

$$\text{sgn} \left(\sum_j w_{ij} \xi_j \right) = \xi_i \quad (\text{para todo } i) \quad (2.5)$$

o que de acordo com a regra (2.1) não causará nenhuma mudança na rede. Pode-se verificar que esta regra é estável ao substituir-se nesta equação a regra de Hebb para os pesos sinápticos, equação (2.3).

Aqui, a analogia com o modelo de Ising pode ser feita da seguinte forma: O spin está orientado para "cima" se $S_i = +1$ e para "baixo" se $S_i = -1$. Ou seja, um neurônio disparando corresponde a um spin para "cima" e um neurônio inativo a um spin para "baixo". As intensidades das conexões na rede de Hopfield correspondem à intensidade das interações de troca em um imã. A entrada de rede em um neurônio corresponde ao campo atuando em um spin, com um campo externo representando um limiar.

Mesmo que a entrada da rede seja diferente do padrão armazenado, a rede ainda será capaz de retornar o padrão armazenado se o número de bits errados inicial for menor que a metade do número total de bits, pois serão suprimidos pelos bits corretos na soma para a entrada da rede. A entrada da rede, que é o campo local no neurônio i , pode ser escrita como

$$h_i = \sum_j w_{ij} S_j. \quad (2.6)$$

e aplicando a regra de Hebb nos pesos sinápticos w_{ij} tem-se

$$h_i = \frac{1}{N} \sum_j \xi_i \xi_j S_j. \quad (2.7)$$

Se a maioria dos bits são iguais, $S_j = \xi_j$, o estado do neurônio é igual ao estado estável do padrão armazenado, ou seja, $S_j \xi_j = 1$. Sabendo que $\xi_j \xi_j = 1$, tem-se

$$h_i = \frac{1}{N} \sum_j \xi_i \xi_j \xi_j = \frac{1}{N} \sum_j \xi_i = \xi_i. \quad (2.8)$$

Logo, $\text{sgn}(h_i)$ é

$$\text{sgn}(h_i) = \text{sgn}(\xi_i) = \xi_i, \quad (2.9)$$

ou seja, este padrão será recuperado. A rede corrigirá os erros conforme o desejado e pode-se dizer que ξ_i é um atrator.

Se mais da metade do número total de bits de entrada estiver errada $S_j \neq \xi_j$, ou seja, $S_j \xi_j = -1$

$$h_i = \frac{1}{N} \sum_j \xi_i (-1) = -\xi_i. \quad (2.10)$$

A rede convergirá para $-\xi_i$, que é um outro atrator da rede, um estado reverso. O espaço de configurações para este caso estará dividido entre duas bacias de atração.

2.1.2 Rede com p padrões

No caso com mais de um padrão, pode-se tomar w_{ij} como uma superposição de termos como em (2.3), uma para cada padrão. A condição de estabilidade (2.5) pode ser reescrita de um padrão para p padrões:

$$\text{sgn}(h_i^\nu) = \xi_i^\nu \quad (2.11)$$

onde a entrada de rede h_i^ν no neurônio i no padrão ν é

$$h_i^\nu \equiv \sum_j w_{ij} \xi_j^\nu = \frac{1}{N} \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (2.12)$$

Separa-se a soma em μ para o caso em que $\mu = \nu$,

$$h_i^\nu = \xi_i^\nu + \frac{1}{N} \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (2.13)$$

Se o termo com o somatório fosse zero, da equação (2.11) poderia ser diretamente concluído que o padrão ν é estável, uma vez que $h_i^\nu = \xi_i^\nu$. Este termo com o somatório é chamado de termo cruzado (crosstalk term). O padrão também pode ser chamado de estável se este termo cruzado for pequeno o suficiente, ou seja, se o valor de p for finito, já que N é considerado tendendo a infinito. Se sua magnitude do termo cruzado for menor que 1, este termo não mudará o sinal de h_i^ν e a condição de estabilidade será mantida. Neste cenário, todos os padrões armazenados serão estáveis. Se uma destas entradas for fornecida para o sistema, a rede permanecerá neste estado. Isto também é válido se os valores de entrada forem uma versão com ruído dos padrões armazenados, pois a rede será capaz de corrigi-los para o padrão estável mais próximo do padrão de entrada. Assim, é possível concluir que os padrões armazenados são realmente atratores do sistema e o sistema funciona como uma memória de conteúdo endereçável, conforme havia sido previsto (3).

2.2 Capacidade de armazenamento

Na última seção foram estudadas condições de estabilidade para a rede de Hopfield com um padrão e com p padrões. Agora, será analisada a capacidade da rede.

Define-se uma quantidade C_i^ν

$$C_i^\nu \equiv -\xi_i^\nu \frac{1}{N} \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu. \quad (2.14)$$

que é o termo cruzado multiplicado por $-\xi_i^\nu$. A estabilidade pode ser analisada da seguinte forma: se C_i^ν é negativo, o termo cruzado tem o mesmo valor de ξ_i^ν e não modificará o sinal

de h_i^ν . Se C_i^ν é positivo e menor que 1, o sinal de h_i^ν ainda será mantido e o padrão ainda será estável. Se h_i^ν é positivo e maior que 1, o sinal de h_i^ν mudará e o neurônio i no padrão ν será instável. Se o sistema for iniciado neste estado ξ_i^ν , o sistema não permanecerá nele.

Sabendo-se destas condições, pode-se estimar a probabilidade de qualquer bit escolhido ser instável, calculando qual a probabilidade do termo cruzado ser maior do que 1.

$$P_{\text{error}} = \text{Prob}(C_i^\nu > 1) \quad (2.15)$$

A medida que tenta-se aumentar o número de padrões armazenados na rede, a probabilidade de a rede retornar mais erros também deve aumentar. Ao definir um critério de performance aceitável para a rede pode-se calcular uma capacidade máxima de armazenamento da rede, p_{max} . A probabilidade de qualquer bit ser instável dependerá do número de neurônios N na rede e da quantidade de padrões p que se quer armazenar. Assume-se que ambos são grandes o suficiente quando comparados com 1, por simplicidade matemática e porque geralmente este é o caso. Os padrões ξ_j^μ tem igual probabilidade de serem +1 ou -1, para cada j e μ . Tendo definido isto, o termo C_i^ν é $1/N$ vezes a soma de Np termos aleatórios. Isto segue uma distribuição binomial com média zero e variância $\sigma^2 = p/N$, mas ao assumir que N e p são grandes o suficiente, por aproximação, pode-se usar uma distribuição gaussiana com média zero e mesma variância (3). Então tem-se que P_{error} é

$$\begin{aligned} P_{\text{error}} &= \frac{1}{\sqrt{2\pi}\sigma} \int_1^\infty e^{-x^2/2\sigma^2} dx \\ &= \frac{1}{2} [1 - \text{erf}(1/\sqrt{2\sigma^2})] = \frac{1}{2} [1 - \text{erf}(\sqrt{N/2p})] \end{aligned} \quad (2.16)$$

onde a função erro $\text{erf}(x)$ é definida por

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du. \quad (2.17)$$

O critério de instabilidade definido determina a estabilidade inicial dos padrões, ou seja, a quantidade de bits errados que há inicialmente. Dependendo da quantidade de informação errada inicialmente, o sistema pode evoluir de forma a propagar este erro pela rede e convergir para um estado diferente do padrão inicial, que se queria encontrar. Se isto ocorre, a memória será inutilizável, pois não será possível recuperar informação correta desta rede.

2.3 Redes estocásticas

Voltando à analogia com sistemas magnéticos, se a temperatura não for muito baixa, flutuações térmicas surgirão e tenderão a inverter a orientação dos spins. Os spins estarão influenciados pelo campo e pelas flutuações térmicas, um tentando alinhar os spins e o outro tentando romper o alinhamento. As flutuações decrescem em temperaturas muito

baixas e praticamente desaparecem no zero absoluto, onde se recupera a rede de Hopfield determinística. Em temperaturas muito altas as flutuações térmicas dominam e o campo perde sua influência sob o spin.

As flutuações térmicas no estado dos neurônios podem ser descritas através de uma regra estocástica:

$$S_i = \begin{cases} 1 & \text{com probabilidade } g(h_i); \\ -1 & \text{com probabilidade } 1 - g(h_i); \end{cases} \quad (2.18)$$

e a função $g(h_i)$ depende da temperatura e pode ser descrita da forma,

$$g(h) = f_\beta(h) \equiv \frac{1}{1 + \exp(-2\beta h)}. \quad (2.19)$$

O parâmetro β está relacionado com uma pseudo-temperatura T por

$$\beta \equiv \frac{1}{T}. \quad (2.20)$$

Como $1 - f_\beta(h) = f_\beta(-h)$, pode-se escrever a regra dinâmica em uma forma simétrica

$$\text{Prob}(S_i = \pm 1) = f_\beta(\pm h_i) = \frac{1}{1 + \exp(\mp 2\beta h_i)}. \quad (2.21)$$

Em um problema com muitos spins interagindo é possível aproximar o comportamento estocástico desta rede. Usa-se uma aproximação de campo médio, que consiste em substituir o campo local h_i em cada neurônio j , que tem flutuações, por um campo médio $\langle h_i \rangle$

$$\langle h_i \rangle = \left\langle \sum_j w_{ij} S_j \right\rangle = \sum_j w_{ij} \langle S_j \rangle \quad (2.22)$$

No caso com um único spin em um campo magnético externo constante, $S = \pm 1$ e o campo $h = h^{ext}$,

$$\begin{aligned} \langle S \rangle &= \text{Prob}(+1) \cdot (+1) + \text{Prob}(-1) \cdot (-1) \\ &= \frac{1}{1 + e^{-2\beta h}} - \frac{1}{1 + e^{2\beta h}} = \frac{e^{\beta h}}{e^{\beta h} + e^{-\beta h}} - \frac{e^{-\beta h}}{e^{-\beta h} + e^{\beta h}} \\ &= \tanh \beta h. \end{aligned} \quad (2.23)$$

Para o caso com vários spins é possível calcular $\langle S \rangle$ de forma análoga ao caso com um spin (2.23)

$$\langle S_i \rangle = \tanh(\beta \langle h_i \rangle) = \tanh\left(\beta \sum_j w_{ij} \langle S_j \rangle\right). \quad (2.24)$$

Com esta aproximação de campo médio, as variáveis que antes eram estocásticas, agora são determinísticas, sendo agora possível lidar com este sistema de equações.

Retorna-se a rede de Hopfield e aplica-se a aproximação de campo médio, primeiramente no caso em que $p \ll N$. Em analogia com (2.24) e usando a regra de Hebb, as equações de campo médio são

$$\langle S_i \rangle = \tanh \left(\frac{\beta}{N} \sum_{j,\mu} \xi_i^\mu \xi_j^\mu \langle S_j \rangle \right). \quad (2.25)$$

Por hipótese é possível tomar $\langle S_i \rangle$ proporcional a um dos padrões armazenados, $m = \frac{1}{N} \sum_i \xi_i^\nu \langle S_i \rangle$, onde m é uma magnetização média da rede,

$$m = \frac{1}{N} \sum_i \xi_i^\nu \tanh \left(\frac{\beta}{N} \sum_{j,\mu} \xi_i^\mu \xi_j^\mu \langle S_j \rangle \right) \quad (2.26)$$

e assim como foi feito em (2.13), pode-se separar o termo em que $\mu = \nu$, resultando em um termo proporcional a ξ_i^ν e um outro termo proporcional ao termo cruzado entre ξ_i^ν e os outros padrões.

$$\begin{aligned} m &= \frac{1}{N} \sum_i \tanh \left(\frac{\beta}{N} \sum_j \xi_i^\nu \xi_j^\nu \langle S_j \rangle + \frac{\beta}{N} \sum_{j,\mu \neq \nu} \xi_i^\mu \xi_j^\mu \langle S_j \rangle \right) \\ &= \frac{1}{N} \sum_i \tanh \left(m\beta\xi_i^\nu + \frac{\beta}{N} \sum_{j,\mu \neq \nu} \xi_i^\mu \xi_j^\mu \langle S_j \rangle \right) \end{aligned}$$

Como foi considerado que o número de padrões armazenados é muito menor que o número de neurônios, o termo cruzado pode ser desprezado. Usando a antissimetria da tangente hiperbólica pode-se reescrever a equação acima como

$$m = \frac{1}{N} \sum_i \xi_i^\nu \tanh(\beta m) \quad (2.27)$$

que pode ser resolvida graficamente, como uma função da temperatura (ver figura 2.12 em (3)). Quando a temperatura é pequena, há a solução trivial e duas soluções não triviais, que são estáveis para pequenas perturbações em m . Então sabe-se que os estados de memória serão estáveis para temperaturas menores que 1 e a temperatura crítica T_c será 1 para a rede estocástica com $p \ll N$.

Após analisar o caso de uma rede em que $p \ll N$, é possível estudar o caso em que p é da mesma ordem que N , o que, conforme comentado em (3) torna a análise consideravelmente mais difícil. Uma análise de campo médio para p da mesma ordem de N foi feita pela primeira vez por Amit, Gutfreund e Sompolinsky (6), (7).

2.3.1 Capacidade de armazenamento da rede estocástica

Nesta seção será vista a capacidade de armazenamento de uma rede estocástica com o número de padrões armazenados p da ordem do número de neurônios da rede N . Pode-se iniciar esta análise com as equações de campo médio encontradas para a rede com $p \ll N$

(2.25), porém agora o termo cruzado não pode ser desprezado, já que p é da ordem de N . De modo análogo ao feito na seção anterior é possível escrever uma expressão para a magnetização, onde o índice ν indicará o padrão a que se refere

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \langle S_i \rangle. \quad (2.28)$$

A magnetização média está sendo calculada para todos os padrões e não apenas para o que será recuperado. Pode-se considerar que o padrão número 1 será o recuperado, então m_1 terá ordem de grandeza unitária

$$m_{\nu=1} = \frac{1}{N} \sum_i \xi_i^{\nu=1} \langle S_i \rangle = \frac{1}{N} \sum_i m \xi_i^\nu \xi_i^\nu = m$$

enquanto os outros m_ν com $\nu \neq 1$ serão da ordem de $\frac{1}{\sqrt{N}}$. Define-se um parâmetro de carga α , que é a razão do número de padrões que se tenta armazenar pelo número total de neurônios na rede,

$$\alpha = \frac{p}{N}. \quad (2.29)$$

Também é útil definir uma média quadrada do *overlap* da configuração do sistema com os padrões não-recuperados

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1} m_\nu^2. \quad (2.30)$$

O fator $\frac{1}{\alpha}$ é incluído para ajustar o valor da média em 1, tendo considerado que no limite em que N é grande pode-se ignorar qualquer diferença entre p e $p - 1$. Usando r para reescrever (2.28), obtém-se

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \tanh \left(\beta \sum_\mu \xi_i^\mu m_\mu \right) \quad (2.31)$$

considerando $\nu \neq 1$, separa-se os termos em que $\mu = 1$ e $\mu = \nu$ e faz-se uso da antissimetria da tangente hiperbólica $\tanh(-x) = -\tanh(x)$, uma vez que os $\xi_i^\nu = \pm 1$, resultando

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right]. \quad (2.32)$$

Analisando os termos no argumento da função \tanh pode-se notar que o primeiro termo m_1 , que é o padrão que se quer recuperar, é de ordem 1. O último termo $\sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu$ também é grande, pois há em torno de p termos nele.

Porém, o segundo termo $\xi_i^\nu \xi_i^1 m_\nu$ tem ordem de $1/\sqrt{N}$ e é possível expandir a tangente hiperbólica em série de Taylor para esse termo:

$$m_\nu = \frac{\beta}{N} \sum_i \left\{ 1 - \tanh^2 \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \right\} m_\nu \quad (2.33)$$

O termo $\sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu$ na equação acima, m_μ , $\mu \neq 1$ são os pequenos *overlaps* entre o campo médio e os padrões não recuperados, assumindo que estes são variáveis aleatórias independentes com média zero e variância $\alpha r/p$, como foi sugerido na definição de r em (2.30). Os produtos $\xi_i^\mu \xi_i^1$ são aleatórios e independentes de m_μ .

O teorema do limite central diz que a média de um número suficiente grande de variáveis aleatórias, cada uma com um valor esperado e uma média bem definidos, pode ser aproximado por uma distribuição normal (8).

Define-se então uma variável x , que é um dos termos em (2.33)

$$x = \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \quad (2.34)$$

e calcula-se a média desta variável,

$$\langle x \rangle = \frac{1}{N} \sum_i \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu = 0 \quad (2.35)$$

bem como a variância, que havia sido definida anteriormente em (2.30),

$$\sigma^2 = \alpha r = \langle x^2 \rangle - \langle x \rangle^2 = \langle x^2 \rangle. \quad (2.36)$$

Calcula-se a média de x^2 para encontrar a variância,

$$\begin{aligned} \langle x^2 \rangle &= \frac{1}{N} \sum_i \sum_{\mu \neq 1, \nu \neq 1} \xi_i^\mu \xi_i^\nu m_\mu m_\nu \\ &= \frac{1}{N} \sum_i \left[\sum_{\mu \neq 1} m_\mu^2 + \sum_{\mu \neq \nu \neq 1} \xi_i^\mu \xi_i^\nu m_\mu m_\nu \right] \\ &= \frac{1}{N} \sum_i \sum_{\mu \neq 1} m_\mu^2 = \alpha r \end{aligned} \quad (2.37)$$

Ou seja, considerando-se o teorema do limite central, a média $1/N \sum_i$ sobre os neurônios é a média sobre um "ruído" gaussiano $\sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu$ de variância αr . Então, pode-se reduzir a equação (2.33) a

$$m_\nu = \frac{N^{-1} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right]}{1 - \beta(1 - q)} \quad (2.38)$$

onde q foi definido como

$$q = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tanh^2 \left[\beta \left(m_1 + \sqrt{\alpha r} z \right) \right]. \quad (2.39)$$

Sabendo-se m_ν , pode-se calcular r . Elevando (2.38) ao quadrado,

$$\begin{aligned} m_\nu^2 &= \left[\frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_{ij} \xi_i^\nu \xi_i^1 \xi_j^\nu \xi_j^1 \tanh \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \\ &\quad \times \tanh \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_j^\mu \xi_j^1 m_\mu \right) \right] \end{aligned} \quad (2.40)$$

e substituindo em r ,

$$r = \frac{1}{\alpha} \left[\frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_{\nu \neq 1} \sum_{ij} \xi_i^\nu \xi_i^1 \xi_j^\nu \xi_j^1 \tanh [\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)] \\ \times \tanh [\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_j^\mu \xi_j^1 m_\mu)] \quad (2.41)$$

Com um pouco de álgebra é possível reescrever r em uma forma mais compacta,

$$r = \left[\frac{1}{1 - \beta(1 - q)} \right]^2 q. \quad (2.42)$$

pode-se definir uma variável C como,

$$C \equiv \beta(1 - q) \quad (2.43)$$

e notando a seguinte identidade

$$1 - q = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} (1 - \tanh^2[\beta(m_1 + \sqrt{\alpha r} z)]) \\ = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\alpha r} \beta} e^{-\frac{m_1^2}{2\alpha r}}; \quad (2.44)$$

e a forma mais compacta de r será

$$r = \frac{q}{(1 - C)^2}. \quad (2.45)$$

O processo feito para encontrar m_ν pode ser repetido para encontrar m_1 . Partindo da equação (2.31) com $\nu = 1$, de forma análoga aos cálculos anteriores é encontrado

$$m_1 = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tanh [\beta(m_1 + \sqrt{\alpha r} z)]. \quad (2.46)$$

As equações (2.39), (2.45) e (2.46) podem ser resolvidas numericamente para m_1 , r e q .

É possível examinar em detalhes as equações para $T \rightarrow 0$ (ou $\beta \rightarrow \infty$). Neste limite pode-se considerar $q \rightarrow 1$, já que $\tanh(\infty) \rightarrow 1$, e usar a seguinte identidade (como pode ser visto em (3))

$$\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh[\beta(az + b)] \\ \xrightarrow{T \rightarrow 0} \int \frac{dz}{\sqrt{2\pi}} \operatorname{sgn}[az + b] \\ = \operatorname{erf} \left(\frac{b}{\sqrt{2a}} \right). \quad (2.47)$$

Fazendo uso desta identidade é possível reescrever m_1 como

$$m_1 = \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right). \quad (2.48)$$

e pode-se escrever m_1 sem o subscrito 1,

$$m = \operatorname{erf} \left(\frac{m}{\sqrt{2\alpha r}} \right). \quad (2.49)$$

E reescrevendo para $T \rightarrow 0$ as equações de C e r calculadas anteriormente,

$$C = \sqrt{\frac{2}{\pi\alpha r}} e^{-\frac{m^2}{2\alpha r}} \quad (2.50)$$

$$r = \frac{1}{(1 - C)^2}. \quad (2.51)$$

A capacidade da rede pode ser encontrada resolvendo estas três equações (2.49), (2.50) e (2.51), usando $y = \frac{m}{\sqrt{2\alpha r}}$. Fazendo esta substituição em (2.49)

$$\operatorname{erf}(y) = m,$$

em (2.50)

$$C = \frac{2y}{\sqrt{\pi}} \frac{1}{m} e^{-y^2}$$

e reescrevendo C em função de r

$$C = 1 - \frac{1}{r^2}.$$

Reorganizando as equações em função de y , tem-se a equação secular

$$y \left(\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-y^2} \right) = \operatorname{erf}(y) \quad (2.52)$$

que pode ser resolvida graficamente. (ver a referência (3) para solução gráfica). Ao resolver este sistema, um valor crítico de α_c é encontrado,

$$\alpha_c = 0.138.$$

Também é possível observar que há uma descontinuidade na magnetização quando $\alpha = \alpha_c$, de valores $m \approx 0.97$ a magnetização cai a zero. Ou seja, há uma descontinuidade entre uma memória que funciona bem com poucos erros com $\alpha < \alpha_c$ e uma memória inutilizável com $\alpha > \alpha_c$.

3 Rede não-monótona

Neste capítulo será analisada a dinâmica de uma rede completamente conexa com uma função de ativação não monótona. A motivação para tal análise está em estudos que mostram que redes neurais em camadas com uma função de transferência não-monótona apresentam ganho na capacidade de armazenamento em relação a rede análoga monótona (9),(10),(11). Não é possível comparar diretamente os resultados obtidos para a capacidade crítica α_c entre os diferentes modelos: rede completamente conexa, rede diluída e rede em camadas, pois a definição desta quantidade é diferente para cada caso. Entretanto é possível esperar que uma rede completamente conexa apresente um aumento na capacidade de armazenamento tal como na rede em camadas. Também é possível comparar redes de mesma arquitetura com funções de ativação diferentes, como as redes monótonas e as redes não-monótonas. Uma maneira de introduzir a não-monotonicidade na rede é através da função de ativação, que pode ser definida como

$$\langle S_i \rangle = -\tanh\left(\beta\left(\sum_{\mu} \xi_i^{\mu} m_{\mu} + \theta\right)\right) + \tanh\left(\beta\sum_{\mu} \xi_i^{\mu} m_{\mu}\right) - \tanh\left(\beta\left(\sum_{\mu} \xi_i^{\mu} m_{\mu} - \theta\right)\right) \quad (3.1)$$

que foi escolhida por semelhança com as equações de campo médio para uma função não-monótona.

Os parâmetros β , ξ_i^{μ} , r e m_{μ} são os mesmos definidos anteriormente. O parâmetro θ introduz a não-monotonicidade da rede. Se este parâmetro for grande o suficiente, recupera-se a rede monótona; se for zero, recupera-se o inverso da rede monótona. Na

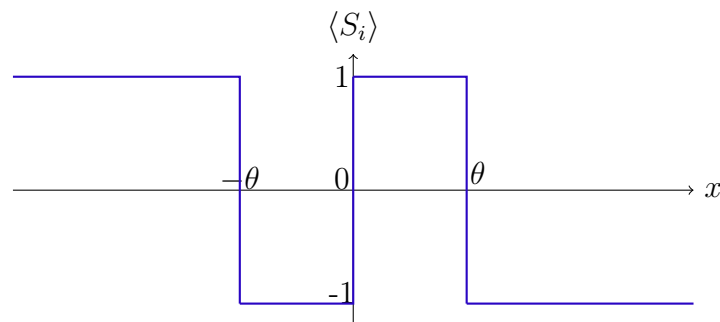


Figura 3 – Função não-monótona de $\langle S_i \rangle$ com $\beta \rightarrow \infty$

figura 3 há um gráfico da função usada para $\langle S_i \rangle$ no limite em que $\beta \rightarrow \infty$. Conforme foi visto para a rede monótona, no limite em que $\beta \rightarrow \infty$ a tangente hiperbólica tende a função sinal.

3.1 Capacidade da rede não-monótona

No início do capítulo foi definida uma função não-monótona para a função de ativação da rede, com o intuito de verificar se, usando este tipo de função, é possível conseguir um ganho na capacidade de armazenamento. A capacidade da rede com esta função pode ser calculada de forma análoga ao caso monótono e os cálculos são iniciados como na seção 2.3.1.

Em (2.28) a magnetização média foi calculada para todos os padrões, agora será calculada a magnetização usando a nova definição das equações de campo médio $\langle S_i \rangle$ em (2.28)

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \left[-\tanh \left(\beta \left(\sum_\mu \xi_i^\mu m_\mu + \theta \right) \right) + \tanh \left(\beta \sum_\mu \xi_i^\mu m_\mu \right) - \tanh \left(\beta \left(\sum_\mu \xi_i^\mu m_\mu - \theta \right) \right) \right] \quad (3.2)$$

Conforme feito anteriormente, o padrão 1 é escolhido como o de recuperação. Considera-se que $\nu \neq 1$, separa-se os termos com $\mu = 1$ e $\mu = \nu$ e, novamente, a antissimetria da tangente hiperbólica é usada, colocando ξ_i^1 em evidência

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \left[-\tanh \left(\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta \xi_i^1 \right) \right) + \tanh \left(\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right) - \tanh \left(\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta \xi_i^1 \right) \right) \right]. \quad (3.3)$$

A equação (3.3) tem forma análoga a (2.32), porém com mais termos devido a alteração nas equações de campo médio. Analisando os termos no argumento da \tanh em (3.3) tem-se que: m_1 tem ordem 1, pois é o padrão que quer-se recuperar; a soma $\sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu$ também é grande, pois há em torno de p termos nela. Pode-se definir θ de maneira que o último termo, $\pm \theta \xi_i^1$ também seja grande. O termo $\xi_i^\nu \xi_i^1 m_\nu$ tem ordem de $1/\sqrt{N}$ e como este valor é pequeno, pode-se expandir a tangente hiperbólica para m_ν em série de Taylor em torno de zero:

$$m_\nu = -\frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \left\{ \tanh \left[\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta \xi_i^1 \right) \right] - \tanh \left[\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] + \tanh \left[\beta \left(m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta \xi_i^1 \right) \right] \right\} - \frac{\beta}{N} \sum_i m_\nu \left\{ \left[1 - \tanh^2 \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta \xi_i^1 \right) \right] \right] - \left[1 - \tanh^2 \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \right] + \left[1 - \tanh^2 \left[\beta \left(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta \xi_i^1 \right) \right] \right] \right\}. \quad (3.4)$$

Pode-se definir q_σ

$$q_\sigma = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh^2[\beta(m_1 + \sqrt{\alpha r}z + \sigma\theta\xi_i^1)] \quad (3.5)$$

e reescrever a equação (3.4) para m_ν

$$m_\nu = \left\{ \frac{1}{1 + \beta(1 - q_+ + q_0 - q_-)} \right\} \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \left[-\tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta\xi_i^1)) \right. \\ \left. + \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)) - \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta\xi_i^1)) \right]. \quad (3.6)$$

A variável aleatória ξ_i^1 pode ser eliminada do termo $\xi_i^1\theta$, sem alterar as equações, uma vez que ξ_i^1 pode ser $+1$ ou -1 e estamos considerando os casos $\pm\theta$,

$$m_\nu = \left\{ \frac{1}{1 + \beta(1 - q_+ + q_0 - q_-)} \right\} \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \left[-\tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta)) \right. \\ \left. + \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)) - \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta)) \right]. \quad (3.7)$$

Ao elevar m_ν ao quadrado, só sobram os termos em que $i = j$,

$$m_\nu^2 = \left\{ \frac{1}{1 + \beta(1 - q_+ + q_0 - q_-)} \right\}^2 \frac{1}{N^2} \sum_i \left\{ -\tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta)) \right. \\ \left. + \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)) \right. \\ \left. - \tanh(\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta)) \right\}^2 \quad (3.8)$$

e expandindo o termo ao quadrado dentro do somatório,

$$\left\{ \right\}_i^2 = \tanh^2[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta)] + \tanh^2[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)] \\ + \tanh^2[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta)] \\ - 2 \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta)] \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)] \\ - 2 \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu)] \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta)] \\ + 2 \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu + \theta)] \tanh[\beta(m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu - \theta)]$$

é possível reescrever $\frac{1}{N} \sum_i \left\{ \right\}_i^2$ como

$$\frac{1}{N} \sum_i \left\{ \right\}_i^2 = q_+ + q_0 + q_- - 2q_{+0} - 2q_{0-} + 2q_{+-}$$

onde

$$q_{\sigma\sigma'} = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh[\beta(m_1 + \sqrt{\alpha r}z + \sigma\theta)] \tanh[\beta(m_1 + \sqrt{\alpha r}z + \sigma'\theta)]. \quad (3.9)$$

Assim, a equação (3.8) pode ser reescrita como,

$$m_\nu^2 = \frac{1}{N} \left\{ \frac{q_+ + q_0 + q_- - 2q_{+0} - 2q_{0-} + 2q_{+-}}{[1 + \beta(1 - q_+ + q_0 - q_-)]^2} \right\} \quad (3.10)$$

com q_σ reescrito sem a variável aleatória ξ_i^1 no termo $\sigma\theta$

$$q_\sigma = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh^2[\beta(m_1 + \sqrt{\alpha r}z + \sigma\theta)] . \quad (3.11)$$

É possível usar a definição de uma média quadrada do *overlap* da configuração do sistema com os padrões não recuperados que havia sido feita no caso monótono. Substitui-se então a equação (3.10) em (2.45),

$$r = \left\{ \frac{q_+ + q_0 + q_- - 2q_{+0} - 2q_{0-} + 2q_{+-}}{[1 + \beta(1 - q_+ + q_0 - q_-)]^2} \right\} . \quad (3.12)$$

Pode-se definir uma variável C ,

$$C \equiv 1 + \beta(1 - q_+ + q_0 - q_-) \quad (3.13)$$

que ao ser reescrita pode-se notar que é possível integrar esta variável analiticamente

$$C = -\beta(1 - q_+) + \beta(1 - q_0) - \beta(1 - q_-) . \quad (3.14)$$

sendo assim é possível reescrever r ,

$$r = \left(\frac{1}{1 - C} \right)^2 \left(q_+ + q_0 + q_- - 2q_{+0} - 2q_{0-} - 2q_{+-} \right) . \quad (3.15)$$

Para encontrar a expressão para m_1 , é escolhido $\nu = 1$ em (3.2).

$$\begin{aligned} m_1 = \frac{1}{N} \sum_i \xi_i^1 & \left[-\tanh(\beta(\sum_\mu \xi_i^\mu m_\mu + \theta)) \right. \\ & \left. + \tanh(\beta \sum_\mu \xi_i^\mu m_\mu) - \tanh(\beta(\sum_\mu \xi_i^\mu m_\mu - \theta)) \right] \end{aligned} \quad (3.16)$$

Os cálculos para m_1 prosseguem de maneira análoga a anterior. Separa-se os termos com $\mu = \nu$,

$$\begin{aligned} m_1 = \frac{1}{N} \sum_i & \left[-\tanh(\beta(m_1 + \sum_{\mu \neq 1} \xi_i^\mu \xi_i^1 m_\mu + \theta)) \right. \\ & \left. + \tanh(\beta(m_1 + \sum_{\mu \neq 1} \xi_i^\mu \xi_i^1 m_\mu)) - \tanh(\beta(m_1 + \sum_{\mu \neq 1} \xi_i^\mu \xi_i^1 m_\mu - \theta)) \right] \end{aligned} \quad (3.17)$$

e usando novamente o teorema do limite central,

$$\begin{aligned} m_1 = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} & \left[-\tanh(\beta(m_1 + \sqrt{\alpha r}z + \theta)) \right. \\ & \left. + \tanh(\beta(m_1 + \sqrt{\alpha r}z)) - \tanh(\beta(m_1 + \sqrt{\alpha r}z - \theta)) \right] \end{aligned} \quad (3.18)$$

3.1.1 Capacidade da rede no limite de temperatura zero

A capacidade da rede pode ser encontrada resolvendo as equações para q em (3.9) e (3.11), r em (3.12) e m_1 em (3.18). Pode-se analisar estas equações no limite em que $T \rightarrow 0$, ou seja, em que $\beta \rightarrow \infty$. Neste limite em que a temperatura tende a zero é possível simplificar um pouco a forma das equações e integrá-las analiticamente, conforme será feito a seguir.

Neste limite em que $\beta \rightarrow \infty$ as seguintes relações são válidas

$$\lim_{\beta \rightarrow \infty} \tanh^2 [\beta(m_1 + \sqrt{\alpha r}z + \sigma\theta)] \rightarrow 1 \quad (3.19)$$

e

$$\lim_{\beta \rightarrow \infty} \tanh [\beta(m_1 + \sqrt{\alpha r}z + \sigma\theta)] = \operatorname{sgn} (m_1 + \sqrt{\alpha r}z\sigma\theta) . \quad (3.20)$$

Assim, pode-se reescrever as equações para q_σ e $q_{\sigma\sigma'}$

$$q_\sigma \rightarrow \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) = 1 \quad (3.21)$$

e

$$q_{\sigma\sigma'} \rightarrow \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z + \sigma\theta) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z + \sigma'\theta) . \quad (3.22)$$

Pode-se separar $q_{\sigma\sigma'}$ para os casos em que $\sigma = 0, +, -$

$$q_{+0} = \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z + \theta) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z) \quad (3.23)$$

$$q_{0-} = \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z - \theta) \quad (3.24)$$

$$q_{+-} = \int \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z + \theta) \operatorname{sgn} (m_1 + \sqrt{\alpha r}z - \theta) \quad (3.25)$$

Estas integrais possuem a mesma forma e podem ser calculadas de maneira similar. Aqui, será feito em detalhes o caso q_{+0} . Separando em q_{+0} os intervalos em que a função degrau é 1 ou -1

$$\begin{aligned} q_{+0} &= \int_{-\infty}^{-\frac{m_1+\theta}{\sqrt{\alpha r}}} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) - \int_{-\frac{m_1+\theta}{\sqrt{\alpha r}}}^{-\frac{m_1}{\sqrt{\alpha r}}} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \\ &\quad + \int_{-\frac{m_1}{\sqrt{\alpha r}}}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) . \end{aligned} \quad (3.26)$$

Estas integrais em q_{+0} podem ser integradas analiticamente. Pode-se calcular em uma forma geral

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sgn} (az + b) &= - \int_{-\infty}^{-b/a} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) + \int_{-b/a}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \\ &= -\frac{1}{2} \left[\operatorname{erf} \left(\frac{-b}{\sqrt{2a}} + 1 \right) \right] + \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{b}{\sqrt{2a}} \right) \right] \\ &= \operatorname{erf} \left(\frac{b}{\sqrt{2a}} \right) . \end{aligned} \quad (3.27)$$

então o resultado de q_{+0} é

$$q_{+0} = 1 - \operatorname{erf} \left(\frac{m_1 + \theta}{\sqrt{2\alpha r}} \right) + \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right). \quad (3.28)$$

De forma análoga ao que foi feito para q_{+0} , as integrais em q_{0-} e q_{+-} podem ser calculadas. O resultado destas integrais é

$$q_{0-} = 1 - \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right) - \operatorname{erf} \left(\frac{\theta - m_1}{\sqrt{2\alpha r}} \right) \quad (3.29)$$

$$q_{+-} = 1 - \operatorname{erf} \left(\frac{m_1 + \theta}{\sqrt{2\alpha r}} \right) - \operatorname{erf} \left(\frac{\theta - m_1}{\sqrt{2\alpha r}} \right). \quad (3.30)$$

A equação para m_1 também pode ser reescrita no limite em que $\beta \rightarrow \infty$

$$m_1 \rightarrow \int \frac{dz}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) \left[\operatorname{sgn} (m_1 + \sqrt{\alpha r} z + \theta) + \operatorname{sgn} (m_1 + \sqrt{\alpha r} z) - \operatorname{sgn} (m_1 + \sqrt{\alpha r} z - \theta) \right]. \quad (3.31)$$

Esta integral pode ser resolvida como em (3.27)

$$m_1 = -\operatorname{erf} \left(\frac{m_1 + \theta}{\sqrt{2\alpha r}} \right) + \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right) - \operatorname{erf} \left(\frac{m_1 - \theta}{\sqrt{2\alpha r}} \right). \quad (3.32)$$

Ao integrar a equação em (??) tem-se

$$C = -\sqrt{\frac{2}{\alpha r \pi}} \left[\exp \left(-\frac{(m_1 + \theta)^2}{2\alpha r} \right) - \exp \left(-\frac{m_1^2}{2\alpha r} \right) + \exp \left(-\frac{(m_1 - \theta)^2}{2\alpha r} \right) \right], \quad (3.33)$$

e é substituído em r os valores de q_σ e $q_{\sigma'}$

$$\begin{aligned} r &= \left(\frac{1}{1-C} \right)^2 \left\{ 3 - 2 \left[1 - \operatorname{erf} \left(\frac{m_1 + \theta}{\sqrt{2\alpha r}} \right) + \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right) \right] \right. \\ &\quad \left. - 2 \left[1 - \operatorname{erf} \left(\frac{m_1}{\sqrt{2\alpha r}} \right) - \operatorname{erf} \left(\frac{\theta - m_1}{\sqrt{2\alpha r}} \right) \right] + 2 \left[1 - \operatorname{erf} \left(\frac{m_1 + \theta}{\sqrt{2\alpha r}} \right) - \operatorname{erf} \left(\frac{\theta - m_1}{\sqrt{2\alpha r}} \right) \right] \right\} \\ &= \left(\frac{1}{1-C} \right)^2 \end{aligned} \quad (3.34)$$

Agora, tem-se somente equações para m_1 em (3.32), r em (3.35) e C em (3.33). Estas equações podem ser integradas numericamente e é possível analisar se o uso de uma rede não-monótona resultou em um ganho na capacidade da rede.

4 Resultados

Neste capítulo serão apresentados e discutidos os resultados obtidos ao integrar numericamente as equações para m , r e C . O método escolhido para integração numérica foi integração pelo método de Newton-Raphson, que pode ser encontrado no livro *Numerical Recipes* (12). Fornecendo condições iniciais de m e C e para diferentes valores de θ , que é um parâmetro de não-monotonicidade da rede, foi variado os valores de α nas integrações, que correspondem a razão do número de padrões armazenados pelo número de neurônios.

Na figura (4) é possível ver um gráfico de m em função de α para diferentes valores de θ . Neste gráfico, é possível observar que para valores grandes de θ os resultados da rede monótona são recuperados, o que já era esperado, uma vez que a função sinal para θ muito grande se aproximará da função monótona original. Para valores menores de θ nota-se um aumento na capacidade de armazenamento da rede, ou seja, há um aumento no número de padrões que a rede consegue armazenar.

Para $\theta = 4.0$, a capacidade crítica α_c da rede é 0.138, como na rede de Hopfield monótona. Para $\theta = 2.0$, o α_c é 0.173, enquanto que para $\theta = 1.5$, o α_c é 0.294 e quando $\theta = 1.1$, o α_c é 0.448. .

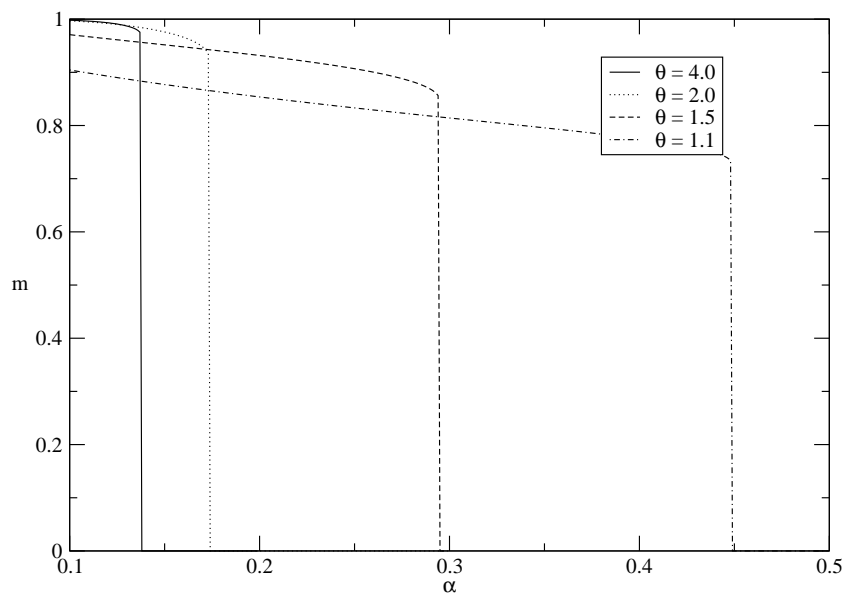


Figura 4 – Gráfico de α em função de m , com diferentes valores de θ

Na figura 5 que também apresenta um gráfico de m em função de α , mas com valores menores de θ é possível observar que conforme o valor de θ diminui, nota-se o aumento de α_c com uma queda na magnetização m . O esperado seria que em um determinado valor de θ seria encontrado um valor máximo de α_c e então ele voltaria a decrescer conforme θ é diminuído. Para valores de θ de até 0.55 foi observado este aumento em α_c , porém para valores de θ menores, há muito ruído na magnetização. Isto sugere que provavelmente há caos nesta região com θ menor que 0.55.

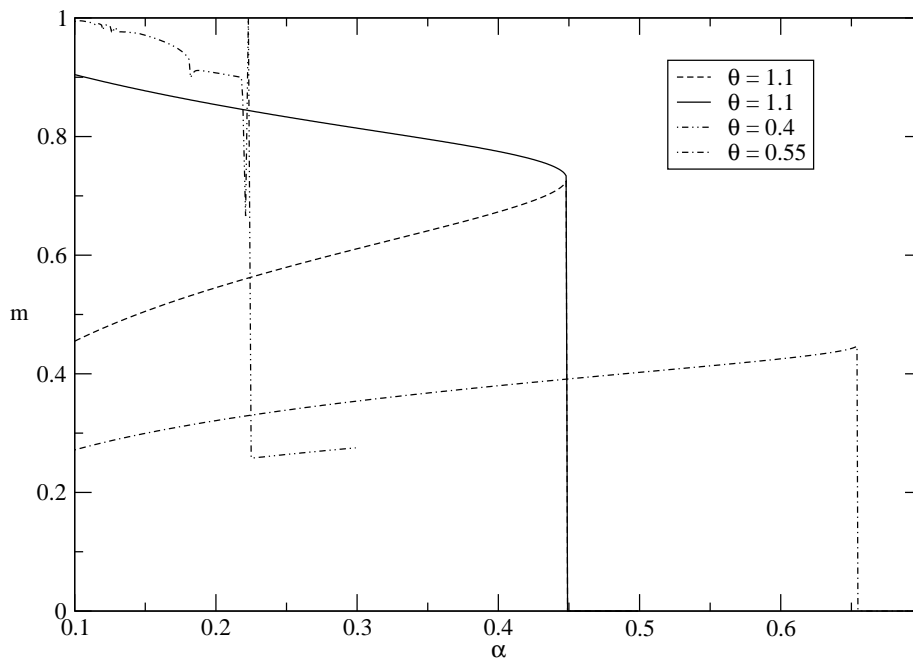


Figura 5 – Gráfico de α em função de m , com diferentes valores de θ

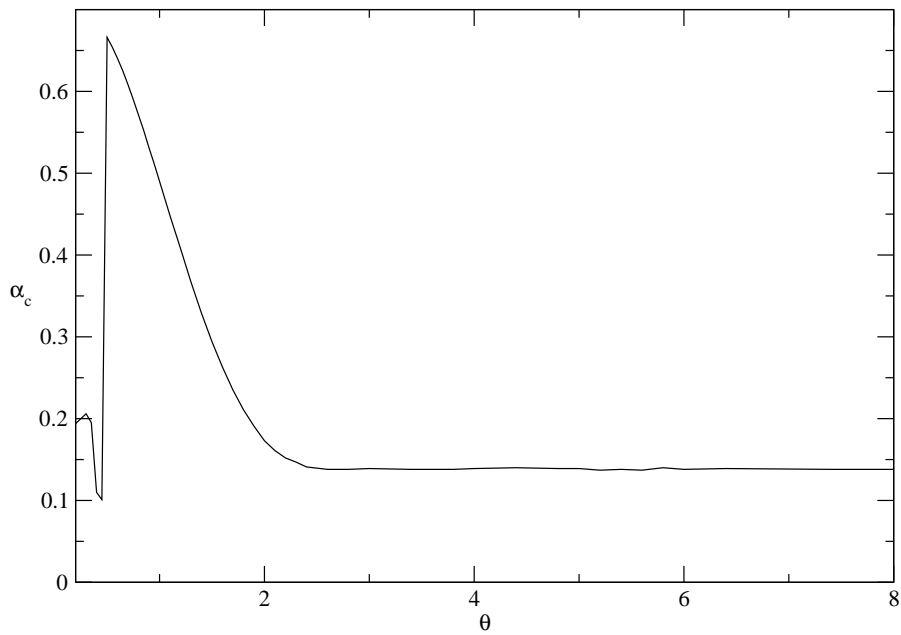


Figura 6 – Gráfico de α_c em função de θ

A figura 6 mostra um gráfico de α_c em função de θ . Para valores grandes de θ o valor de $\alpha_c = 0.138$ é recuperado e conforme a não-monotonicidade da rede aumenta, a capacidade da rede também aumenta, conforme já apresentado na figura 4. Entretanto, na região próxima a $\theta = 0.5$ a magnetização tem muito ruído e este resultado não é muito confiável para esta região, onde provavelmente a rede esteja sob um regime caótico, conforme já apresentado na figura 5 .

5 Conclusões

Neste trabalho foi feita uma breve revisão sobre redes neurais, uma revisão do modelo de Hopfield e foi proposta uma alteração neste modelo, através de uma modificação na função de ativação com a introdução do parâmetro de não-monotonicidade θ . Esta alteração foi estudada em uma rede estocástica recorrente totalmente conexa com uma função de ativação não-monótona.

A escolha da função de ativação não-monótona foi arbitrária, ao contrário da função de ativação para a rede monótona que tem justificativa na mecânica estatística de equilíbrio. A motivação desta escolha não teve base biológica, mas sim por simplicidade matemática. Entretanto, qualquer função não-monótona, derivável, seria apropriada.

Para valores do parâmetro de θ grande, foi recuperado o resultado da rede com a função monótona. Para valores de θ menores que 2,5 observa-se um aumento na capacidade de armazenamento da rede. Observa-se solução estável com $\alpha_c = 0,294$ para $\theta = 1,5$. Para valores de θ menores que 1,5 a solução não é estável. Para valores de θ menores que 0,5, aparentemente, existe a ocorrência de solução caótica. Entretanto, o estudo desta solução não fazia parte do objetivo deste trabalho inicial.

Uma continuidade para este trabalho seria tentar analisar soluções para esta região através de métodos mais adequados. Há vários artigos sobre comportamento caótico em redes neurais em camadas e em redes neurais diluídas (ver (11),(13),(14)), entretanto não encontrou-se literatura versando sobre caoticidade em redes totalmente conexas com uma função de ativação não-monótona, sendo este um campo em aberto para pesquisa.

Referências

- 1 MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115–133, 1943.
- 2 HEBB, D. *The Organization of Behaviour*. [S.l.]: Wiley, 1949.
- 3 HERTZ, J.; PALMER, R.; KROGH, A. *Introduction to the theory of neural computation*. [S.l.]: Addison Wesley, 1991.
- 4 HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, v. 79, p. 2554–2558, 1982.
- 5 HAYKIN, S. *Neural Networks: A comprehensive foundation*. [S.l.]: Prentice Hall, 1998.
- 6 AMIT, D.; GUTFREUND, H.; SOMPOLINSKY, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, v. 55, p. 1530–1533, 1985.
- 7 AMIT, D.; GUTFREUND, H.; SOMPOLINSKY, H. Information storage in neural networks with low levels of activity. *Physical Review A*, v. 35, p. 2293–2303, 1987.
- 8 RICE, J. *Mathematical Statistics and Data Analysis*. [S.l.]: Thomson Brooks/Cole, 1998.
- 9 NEVES, F. S. *Redes de neurônios não monótonos em camadas*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Instituto de Física, Programa de Pós-Graduação em Física, 2006.
- 10 KATAYAMA, K.; SAKATA, Y.; HORIGUCHI, T. Layered neural networks with non-monotonic transfer functions. *Physica A*, v. 317, n. 1–2, p. 270 – 298, 2003.
- 11 DOMINGUEZ, D. R. Interference and chaos by a network of nonmonotonic neurons. *Phys Rev. E*, v. 54, p. 4066–4070, 1982.
- 12 PRESS, W. et al. *Numerical Recipes in C: The Art of Scientific Computing*. [S.l.]: Cambridge University Press, 1992.
- 13 DOMINGUEZ, D. R. C.; THEUMANN, W. K. Generalization and chaos in a layered neural network. *Journal of Physics A: Mathematical and General*, v. 30, n. 5, p. 1403, 1997.
- 14 CAROPPO, D. et al. Chaos in neural networks with a nonmonotonic transfer function. *Physical Review E*, v. 60, n. 2, p. 2186, 1999.