

CAPÍTULO III

O PROBLEMA DO CAIXEIRO VIAJANTE E REDES NEURAIS

III.1 - Introdução:

O problema do Caixeiro Viajante é um dos problemas clássicos em Otimização Combinatória. Dados n cidades, queremos determinar o menor caminho que o caixeiro pode percorrer passando por todas as cidades uma única vez e retornando ao ponto de partida. Apesar do problema ter um enunciado bastante simples é de difícil resolução. A medida que buscamos uma solução de melhor qualidade o tempo de processamento cresce exponencialmente. É um problema representativo de uma importante classe de problemas combinatórios denominados NP-Completo. Em função da elevada complexidade computacional dos algoritmos que buscam uma solução exata, métodos heurísticos mais eficientes são utilizados buscando-se uma solução aproximada para o problema.

Trataremos neste capítulo de duas abordagens heurísticas utilizando métodos analógicos, estaremos interessados na obtenção de soluções aproximadas em um tempo satisfatório.

Estudaremos inicialmente o modelo de Hopfield em Redes Neurais. Como veremos, nenhum processo de aprendizagem será exigido (vide [13]). Por tratar-se de um modelo analógico ele poderá ser implementado em um hardware especialmente dedicado para isso. Buscamos então minimizar duas características conflitantes: obter um circuito de comprimento satisfatório (não necessariamente o menor) e obter uma solução em tempo também satisfatório. Como veremos, a grande quantidade de parâmetros envolvidos no processo criam dificuldades em relação a convergência do método para uma entrada qualquer do problema.

O outro modelo analógico abordado, conhecido como método elástico foi especialmente desenvolvido para a solução do problema do Caixeiro Viajante (embora possa ser aplicado também a problemas distintos em otimização combinatória). A abordagem é bastante geométrica. É dado um anel (círculo) inicial que é deformado gradativamente até que todas as cidades estejam suficientemente próximas de um ponto do anel. O anel obtido dessa forma terá um comprimento correspondente ao "menor" trajeto do caixeiro viajante. O método pode trabalhar com pontos distribuídos em \mathbb{R}^n , entretanto, pela simplicidade e pela fácil visualização geométrica trabalharemos com pontos (cidades) distribuídas no plano.

III.2- Modelo de Hopfield

Se uma ordenação arbitrária de n cidades é dada inicialmente, uma solução do problema do caixeiro viajante poderá ser representada por uma matriz de permutações $n \times n$ (com elementos 0 ou 1), onde cada uma de suas linhas corresponde a uma cidade e cada coluna corresponde a uma posição no circuito (figura III.1).

Seja $x = (x_1, x_2, \dots, x_n)$ um circuito hamiltoniano qualquer. A notação $x_i = k$ nos diz que a cidade i é a k -ésima cidade a ser visitada pelo caixeiro. O exemplo da figura III.1 representa um circuito hamiltoniano com 7 cidades. Note que $x = (1, 7, 2, 3, 4, 6, 5)$ expressa equivalentemente a matriz de permutações. Ambas representam o circuito 1-3-4-5-7-6-2 cujo comprimento total será dado por: $d = d_{13} + d_{34} + d_{45} + d_{57} + d_{76} + d_{62}$, onde d_{ij} representa a distância entre as cidades i e j .

Se o par (i, j) representa a posição de um dado neurônio na matriz de permutações então a função de ativação e a regra de propagação podem ser definidas respectivamente (como discutido em [30]) pelas seguintes equações (vide capítulo I):

$$x_{ij}(t) = \frac{1}{1 + e^{-g(u_{ij} - q_{ij})}} \quad C_{ij} \frac{du_{ij}(t)}{dt} = \sum_{p,q} w_{(ij,pq)} x_{pq}(t) - \frac{u_{ij}(t)}{R_{ij}}$$

onde $w_{(ij,pq)}$ acima representa a sinápsse existente entre os neurônios nas posições (i,j) e (p,q) respectivamente.

		n posições						
		1	2	3	4	5	6	7
n cidades	1	1	0	0	0	0	0	0
	2	0	0	0	0	0	0	1
	3	0	1	0	0	0	0	0
	4	0	0	1	0	0	0	0
	5	0	0	0	1	0	0	0
	6	0	0	0	0	0	1	0
	7	0	0	0	0	1	0	0

Figura III.1: Matriz de Permutações

No modelo de Hopfield cada uma das $n \times n$ posições da matriz de permutações irá corresponder a um neurônio com estado de ativação x_{ij} . Se $x_{ij} \geq 0,5$ o neurônio estará excitado (representado na matriz de permutações pela unidade), caso contrário ele estará inibido (representado por 0). Note também que apenas um neurônio deverá estar excitado em cada linha e cada coluna para que tenhamos um circuito hamiltoniano bem representado.

Esperamos que nossa rede evolua até que um ponto de equilíbrio seja obtido. Neste caso, não teremos mais variação na energia potencial de cada neurônio e os estados de ativação se manterão inalterados. Vejamos então a seguinte definição:

Definição III.1: Chamaremos de *padrão estável* de uma rede neural ao estado de ativação final dos neurônios dessa rede. Em outras palavras, um padrão estável é obtido sempre que a rede neural chega a um ponto de equilíbrio (neste caso a variação da energia potencial de cada neurônio será nula). •

Entretanto, nem sempre um padrão estável corresponde a uma solução desejada para o problema do caixeiro viajante (como expressa na matriz de permutações). Chamaremos de *padrão solução* à solução esperada. Mais formalmente:

Definição III.2: Um *padrão solução* será a configuração final desejada em nossa rede neural. No caso do problema do caixeiro viajante esperamos que apenas um neurônio esteja excitado em cada linha e coluna da matriz de neurônios. •

Precisamos estabelecer portanto, valores para os parâmetros do problema de forma que um padrão estável corresponda sempre a um padrão solução.

Para garantir que não exista nenhuma linha ou coluna com todos os neurônios inibidos, escolheremos inicialmente um valor altamente negativo para θ na função sigmóide (Figura III.2). Note que se fazemos a energia potencial de cada neurônio igual a zero já teremos uma excitação futura de todos os neurônios da rede (vide função sigmóide).

Observe entretanto que para termos um circuito hamiltoniano bem representado, apenas um neurônio deve estar excitado em cada linha e coluna de nossa rede. Neste caso, criamos sinápses altamente inibitórias (w negativo) entre todos os neurônios de uma mesma linha e de uma mesma coluna. Assim, a excitação simultânea entre estes neurônios acarretará um decréscimo na energia

potencial de forma a inibi-los. Como apenas um neurônio de cada linha ou coluna deve permanecer excitado, precisamos estabelecer mais claramente para quais valores de q_j e $w_{(ij,pq)}$ uma solução desejada (circuito hamiltoniano) pode ser obtida.

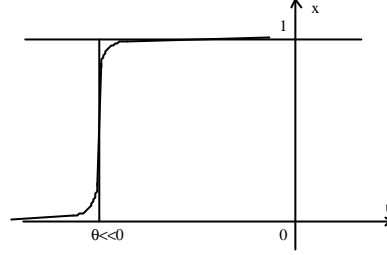


Figura III.2: Criação de sinápsis inibitórias

Note ainda que as sinápses devem ser definidas de forma que o circuito tenha um comprimento "mínimo". Assim, conectamos os neurônios de cada linha k aos neurônios de cada linha l desde que estes neurônios ocupem colunas consecutivas na matriz de neurônios. Neste caso a primeira coluna virá após a última coluna.

Como veremos a seguir, neurônios de colunas consecutivas p e $((p+1) \bmod n)$ (não na mesma linha) terão sinápses inibitórias $-d(k,l)$, onde $d(k,l)$ representa a distância entre as cidades k e l . Desta forma, quando 2 neurônios de colunas consecutivas e linhas k e l estão excitados, eles representam cidades vizinhas na sequência de visitas do caixeiro viajante. Sua contribuição $d(k,l)$ à energia é exatamente a distância entre as cidades k e l .

Na figura III.3 a seguir, são apresentadas as sinápses que contribuem para a energia computacional de uma solução desejada envolvendo 4 cidades:

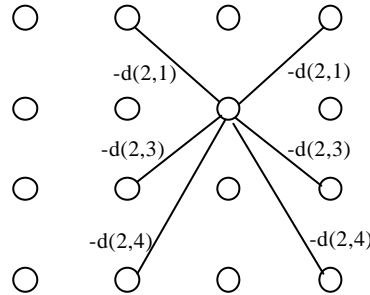


Figura III.3: Construção das Sinápses

Resumindo, para um problema de n cidades, teremos uma rede com n^2 neurônios e com as seguintes sinápses:

(a) Sinápses entre neurônios de uma mesma linha:

$$w_{(ij,iq)} = w; \quad 1 \leq i \leq n, \quad 1 \leq j \leq n, \quad 1 \leq q \leq n \quad e \quad j \neq q \quad (\text{onde } w < 0)$$

(b) Sinápses entre neurônios de uma mesma coluna:

$$w_{(ij,pj)} = w; \quad 1 \leq i \leq n, \quad 1 \leq j \leq n, \quad 1 \leq p \leq n \quad e \quad i \neq p \quad (\text{onde } w < 0)$$

(c) Sinápses entre neurônios de colunas consecutivas (mas não na mesma linha):

$$w_{(ij,k(j+1) \bmod n)} = -d(i,k); \quad 1 \leq i \leq n, \quad 1 \leq k \leq n, \quad 1 \leq j \leq n \quad \text{onde } i \neq k$$

A proposição seguinte sugere valores para \mathbf{q} , R e w de forma que um padrão solução seja sempre obtido quando a rede atinge um ponto de equilíbrio.

Lema IV.1: Em uma rede neural de soluções viáveis cujos neurônios possuem potencial limiar $\mathbf{q} < \min_i \left(-2 \sum_{j \neq i} d(i,j) \right) R$ e sinapses $w < \theta/R$ todos os padrões estáveis são padrões solução.

Demonstração: Demonstraremos equivalentemente que se um padrão de atividade $x(t) = (x_{11}(t), x_{12}(t), \dots, x_{nn}(t))$ em um dado instante t não é padrão solução, esse padrão não será estável. Para isso devemos mostrar que:

- a) Todo padrão de atividade com mais de um neurônio excitado em uma linha ou em uma coluna da matriz não é estável.
- b) Todo padrão de atividade que possui uma linha ou coluna da matriz de neurônios com todos neurônios inibidos não é estável.

Note que os dois itens acima representam as duas únicas situações onde um padrão de atividade não corresponde a um padrão solução.

Seja:

$$A_{kl}(t) = \sum_{\{j: j \neq k\}} (w_{kl,j(l-1)} x_{j(l-1)}(t) + w_{kl,j(l+1)} x_{j(l+1)}(t))$$

a energia potencial em um neurônio N_{kl} considerando-se a contribuição de todos os neurônios vizinhos de N_{kl} a exceção dos demais neurônios na coluna l e linha k respectivamente.

Note que o valor mínimo $-2 \sum_{\{j: j \neq k\}} d(k,j)$ será atingido quando $x_{j(l-1)}$ e $x_{j(l+1)}$ forem todos unitários (para $1 \leq j \leq n$).

a) Seja um padrão de atividade com mais de um neurônio excitado na linha k ou coluna l no instante de tempo t e seja N_{kl} uma destas células. A regra de propagação para este neurônio será:

$$C \frac{du_{kl}(t)}{dt} = A_{kl}(t) + wq - \frac{u_{kl}(t)}{R}$$

onde \mathbf{q} é o número de neurônios excitados na linha k e coluna l excetuando-se N_{kl} . Como N_{kl} está excitado temos $u_{kl}(t) > \mathbf{q}$, assim:

$$C \frac{du_{kl}(t)}{dt} < A_{kl}(t) + wq - \frac{\mathbf{q}}{R}$$

Como todas as sinapses são negativas e $w < \theta/R$ (hipótese) então $C \frac{du_{kl}(t)}{dt} < 0$. Portanto, temos um decréscimo na energia potencial do neurônio N_{kl} no instante t . Logo, para algum t' o impulso $u_{kl}(t)$ se tornará menor que θ , resultando na inibição futura de N_{kl} . Consequentemente, não teremos chegado a uma situação estável em nossa rede.

b) A existência de uma linha k (com neurônios inibidos) resulta também na existência de uma coluna l com todos os neurônios inibidos. Devido ao item (a), não há estabilidade com mais de um neurônio excitado em cada linha ou cada coluna da matriz de neurônios, a existência da linha k resulta na existência de uma coluna l também com todos os seus neurônios inibidos. Assim, a regra de propagação de N_{kl} será:

$$C \frac{du_{kl}(t)}{dt} = A_{kl}(t) - \frac{q}{R}$$

Como o neurônio N_{kl} está inibido temos que $u_{kl}(t) < q$, logo:

$$C \frac{du_{kl}(t)}{dt} > A_{kl}(t) - \frac{q}{R} = \frac{A_{kl}(t)R - q}{R}$$

Temos da hipótese que $q < A_{kl}(t)$, então $C \frac{du_{kl}(t)}{dt} > 0$ o que implica na excitação futura de N_{kl} . c.q.d

Lema IV.2: Em um padrão solução associado a uma solução viável do problema do caixeiro viajante, a energia computacional da rede neuronal de soluções viáveis é igual, a menos de uma constante aditiva, ao comprimento total da trajetória percorrida pelo caixeiro viajante naquela solução viável.

Demonstração: Considere a energia computacional de uma rede neuronal de Hopfield com m neurônios semi-lineares:

$$E(t) = -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m w_{ij} a_i(t) a_j(t) + \sum_{i=1}^m \frac{1}{R_i} \int_0^{a_i(t)} q_i - \frac{1}{g} \ln(1/a - 1) da$$

onde $g \in \mathbb{R}^+$.

Como as resistências elétricas R_i e potenciais limiares q_i são iguais para todos os neurônios excitados, o segundo termo do lado direito da equação acima é constante.

Sabemos também que em um padrão solução os n neurônios excitados se encontram distribuídos de modo que apenas existe um deles em cada linha e coluna da matriz de neurônios, donde concluímos que as sinapses inibitórias w em nada contribuem na energia computacional da rede neuronal. Sejam agora, os n neurônios excitados em um padrão solução numerados por $N_{l_1 1}, N_{l_2 2}, \dots, N_{l_n n}$ e, então, o primeiro termo do lado direito da equação de energia toma a forma:

$$-(a_{l_1 1} w_{(l_1 1, l_2 2)} + a_{l_2 2} w_{(l_2 2, l_3 3)} + \dots + a_{l_n n} w_{(l_n n, l_1 1)})$$

pois na rede neuronal de soluções viáveis só existem sinapses entre neurônios de colunas consecutivas.

Da síntese da rede neuronal de soluções viáveis vemos vemos que $w_{(l_1 k, l_2 (k+1))} = -d(l_1, l_2)$ e, lembrando que o estado de ativação destes neurônios é unitário no padrão estável (o impulso total de entrada é muito maior do que o potencial limiar), a substituição na expressão acima nos fornece:

$$d(l_1, l_2) + d(l_2, l_3) + \dots + d(l_n, l_1)$$

que é exatamente a distância total percorrida pelo caixeiro viajante na solução viável associada ao padrão solução. **c.q.d.**

Vejamos então o seguinte algoritmo, onde fazemos uma simulação dos passos presentes no modelo de Hopfield:

ALGORITMO:Caixeiro Viajante (Hopfield); {Constantes: $C, e, \Delta t, R$ }

Início

- leia $(d(i,j))$; {LêDistâncias}
- $q \leftarrow \min_i \{(-2 \sum_{j \neq i} d(i,j))R - k\}$;
- $w \leftarrow \frac{q}{R} - k$; {sinápses inibitórias}
- Construir sinápses;
- $u(i,j) \leftarrow \emptyset$; $\frac{du(i,j)}{dt} \leftarrow \emptyset$;

- **Repita**

- $u(i,j) \leftarrow u(i,j) + \Delta t * \frac{du(i,j)}{dt}$;
- $a(i,j) \leftarrow [1 + e^{-g[u(i,j) - q(i,j)]}]^{-1}$;
- $\frac{du(i,j)}{dt} = \frac{1}{C(i,j)} \left[\sum_i^n \sum_j^n \sum_p^n \sum_q^n w(i,j;p,q) * a(i,j) - \frac{u(i,j)}{R(i,j)} \right]$
- Calcula $\left\| \frac{du(i,j)}{dt} \right\|$;

Até que $\left\| \frac{du(i,j)}{dt} \right\| < e$

- Variar i e j tal que:
Se $a(i,j) > 0.5$ **então** $x(i) \leftarrow j$;

fim {algoritmo}

Note que o algoritmo para, quando a variação da energia potencial nos $n \times n$ neurônios da rede for arbitrariamente pequena. Teremos consequentemente, uma pequena variação nos estados de ativação desses neurônios chegando portanto a um estado de equilíbrio.

III.3 - Modelo Original de Hopfield-Tank:

No capítulo I, vimos que a função de energia no modelo analógico de Hopfield era obtida através da integração da regra de propagação, associada a cada neurônio, e que procurava modelar de maneira mais fiel o funcionamento de um neurônio biológico.

Uma regra de propagação mais geral (semelhante à primeira) pode ser obtida de maneira a permitir alterações na função de energia que reflitam com maior fidelidade as características do problema em questão. Chamaremos esta nova regra de propagação de equação de movimento, e será definida por:

$$\frac{du_i}{dt} = \sum_j w_{ij} a_j + I_j - u_i; \quad i = 1, 2, \dots, n \times n \quad (I)$$

onde I_i (a nova parcela), representa uma corrente externa sendo adicionada à energia potencial do neurônio i . Note neste caso que não trabalhamos com a capacitância C e a resistência R associada a cada neurônio.

Definiremos também uma "nova" função sigmóide da seguinte forma:

$$a_{ij} = g(u_{ij}) = \frac{1}{2} \left(1 + \tanh \left(\frac{u_{ij}}{u_0} \right) \right) = \frac{1}{1 + \exp(-2u_i / u_0)} \quad (II)$$

Geometricamente, teremos a seguinte situação:

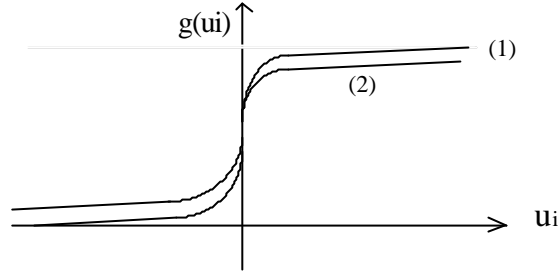


Figura III.4: Função Sigmóide

A constante u_0 é utilizada para alterar a inclinação da função sigmóide. Na figura visualizamos a inclinação na função sigmóide para dois valores distintos atribuídos a u_0 .

Integrando a equação (I) acima para todo neurônio da rede neuronal teremos uma "nova" função de energia dada por:

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} a_i a_j - \sum_i a_i I_i + \int_0^{a_i(t)} g^{-1}(a) da \quad (III)$$

onde $g^{-1}(x)$ é a inversa da função sigmóide.

Note em particular que $\frac{du_i}{dt} = -\frac{dE}{dt}$, assim, podemos afirmar que o decréscimo da energia computacional na rede neural é realizado segundo "passos do tipo gradiente".

O processo termina quando obtemos energia potencial nula para todos os neurônios da rede. Teremos chegado então a um mínimo local da função de energia.

A função de energia como apresentada na equação (III), permite maior versatilidade na representação das características e restrições associadas ao caixeiro viajante. Note que a presença da parcela linear $-\sum_i a_i I_i$ não era presente na função de energia utilizada na seção anterior.

Consideremos então a seguinte função (equação IV):

$$E = \frac{A}{2} \left(\sum_x \sum_i \sum_{j \neq i} a_{xi} a_{xj} \right) + \frac{B}{2} \left(\sum_i \sum_x \sum_{y \neq x} a_{xi} a_{yi} \right) + \frac{C}{2} \left(\sum_x \sum_i a_{xi} - N \right)^2 + \frac{D}{2} \left(\sum \sum \sum d_{xy} a_{xi} (a_{y(i+1)} + a_{y(i-1)}) \right)$$

onde A,B,C e D são os parâmetros (sinápses) utilizadas para representar os vários componentes da função de energia.

O primeiro termo será nulo, se cada linha (na matriz de permutações) correspondendo a uma cidade, possuir um único neurônio ativado (os demais neurônios devem estar desativados).

Analogamente, o segundo termo será nulo se cada coluna, correspondendo a uma posição no circuito, contém apenas um neurônio ativado. Obviamente, ambas as parcelas poderão se anular se tivermos todos os neurônios correspondendo a uma linha ou coluna respectivamente, desativados, entretanto, o terceiro termo presente na função de energia, previne tal situação. Note que a terceira parcela será nula apenas quando n neurônios estiverem ativados em toda a rede neural!

Observe portanto, que as tres parcelas acima penalizam soluções que não correspondem a padrões solução.

A quarta parcela na função de energia é utilizada para assegurar circuitos de menor comprimento. A distância d_{xy} é adicionada a energia sempre para cidades consecutivas na sequência de visitas do caixeiro viajante.

Desenvolvendo a equação (IV), e comparando-a com a equação geral apresentada em (III) identificamos (verificação trivial) parcelas quadráticas e lineares respectivamente:

$$E(t) = -\frac{1}{2} \sum_{xi}^m \sum_{yj}^m w_{(xi,yj)} a_{xi} a_{yj} - \sum_{xi} a_{xi} I_{xi} \quad (V)$$

Note que o último termo foi removido já que não aparece na equação (IV) desenvolvida acima.

Apesar da maior flexibilidade presente na equação de energia, existem ainda sérios problemas relativos à convergência para uma solução viável satisfatória (vide Wilson e Pawley [15]). Abaixo são enumeradas algumas dessas deficiências:

- (a) Em um problema com n cidades necessitamos de $O(n^2)$ neurônios e $O(n^4)$ conexões entre esses neurônios.
- (b) O conjunto de todas as possíveis soluções do caixeiro viajante é de $O(n!)$ permutações, entretanto, na matriz de permutações teremos $O(2^{n^2})$ configurações possíveis!
- (c) Uma solução viável pode ser representada $2n$ vezes na matriz de permutações. Note que qualquer uma das n cidades pode ser escolhida como ponto de partida, além disso, podemos tomar 2 orientações possíveis em um circuito viável (matriz simétrica).
- (d) O método busca sempre a minimização da função de energia caminhando sempre na direção oposta ao gradiente. Entretanto, esta abordagem permite apenas a determinação de mínimos locais e fica sujeita portanto, às condições iniciais do problema.
- (e) O modelo não garante viabilidade. Em outras palavras, muitos pontos de mínimo local da função de energia correspondem a soluções inviáveis.
- (f) A atribuição de valores aos parâmetros A, B, C e D pode ser considerada mais uma arte que uma ciência já que requer um longo processo de tentativa e erro. Atribuindo pequenos valores para os parâmetros A, B e C teremos frequentemente caminhos de pequeno comprimento mas correspondendo a circuitos inviáveis. Da mesma forma, se atribuímos altos valores para esses parâmetros podemos obter soluções viáveis de baixa qualidade.
- (g) A rede não é adaptativa. Os pesos da rede são fixos e derivado dos dados do problema.

- (h) Muitos circuitos inviáveis gerados pela rede possuem apenas um subconjunto de cidades. Isto se deve ao fato da função de energia possuir apenas um termo (termo C) responsável pela penalização de circuitos com menos de n cidades. Observe que os demais termos favorecem a presença de tours inviáveis de pequeno tamanho.
- (i) Em várias situações, podemos ter a convergência da rede para soluções distintas de um dos extremos do hipercubo (vide função sigmóide). Neste caso torna-se imprecisa a solução apresentada pelos estados de ativação dos neurônios da rede.

Apesar das dificuldades apresentadas acima o modelo de Hopfield pode ser implementado facilmente em *hardware*, além disso, pode ser aplicado a problemas não-geométricos, como, por exemplo, o problema do caixeiro viajante não-euclidiano.

Na verdade, os resultados apresentados por Wilson e Pawley[15] não desencorajaram a comunidade de pesquisadores mas sim, estimulou a busca de novos caminhos no sentido de aperfeiçoar o trabalho de Hopfield-Tank. As modificações no modelo original podem ser classificadas basicamente em seis categorias: modificação na função de energia, análise de parâmetros, adição de novas restrições, técnica para "escapar" de mínimos locais, nova representação do problema, alterações na configuração inicial. Essas alterações podem ser estudadas com mais detalhes em [14].

III.4 - Método Elástico:

Sejam $\{x_i\}$ um conjunto de n cidades definidas no plano e $\{y_j\}$ um conjunto de m pontos que definem um anel (círculo). Devemos ter sempre $m \geq n$. Estabelecendo uma ordenação das n cidades e dos m pontos do anel nosso objetivo será, dado um anel $Y^k = (Y_1^k, \dots, Y_m^k)$, fazer uma atualização de cada um destes pontos obtendo um novo anel $Y^{k+1} = (Y_1^{k+1}, \dots, Y_m^{k+1})$.

A atualização de cada ponto é dada por:

$$\Delta Y_j = \alpha \sum_i w_{ij} (X_i - Y_j) + K \beta (Y_{j+1} - 2Y_j + Y_{j-1})$$

onde:

$$w_{ij} = \frac{\exp\left(-\frac{|X_i - Y_j|^2}{2K^2}\right)}{\sum_k \exp\left(-\frac{|X_i - Y_k|^2}{2K^2}\right)}; \quad \text{para } i=1, \dots, n \text{ e } j=1, \dots, m$$

sendo α, β e K parâmetros do problema. Desta forma, para obtermos um novo anel basta fazer $Y_j^{k+1} := Y_j^k + \Delta Y_j^k$, para $1 \leq j \leq m$.

Informalmente falando, o parâmetro α é responsável pela força de atração entre as cidades e os pontos do anel, enquanto que β será responsável pela força de atração entre os pontos do anel. Observe também que w_{ij} representa o peso entre uma cidade X_i e um ponto Y_j do anel, a medida que Y_j se aproxima de X_i ocorre um acréscimo no valor do peso w_{ij} . Portanto, dados α e β adequados ao problema buscamos a obtenção de um circuito viável (passando por todas as cidades) mantendo, simultaneamente, nosso "elástico esticado".

Seja F_1 a força existente entre uma cidade e os demais pontos do anel e F_2 a força existente entre os pontos do anel. Na figura IV.4 abaixo podemos observar melhor a interação existente entre as forças F_1 e F_2 em função dos valores atribuídos aos parâmetros α e $K\beta$.

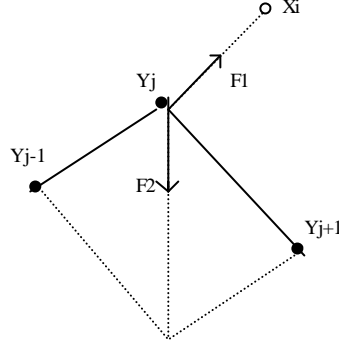


Figura III.5: Iteração das forças F_1 e F_2

Note que $Y_{j-1} - 2Y_j + Y_{j+1} = (Y_{j+1} - Y_j) + (Y_{j-1} - Y_j)$. Assim quanto maior for o produto $K\beta$ (em relação a α) maior será F_2 e consequentemente maior será a força de atração entre os pontos do anel. A força F_1 representa a resultante entre todas as forças existentes entre um ponto qualquer do anel e a cidade X_i .

Podemos interpretar o método elástico equivalentemente ao método gradiente (vide cap.II). Temos que $\Delta Y = (\Delta Y_1, \dots, \Delta Y_m)$ pode ser visto como o gradiente de uma função de energia definida por:

$$E(\{Y_j\}, K) = -\alpha K \sum_{i=1}^n \log \sum_{j=1}^m e^{-\frac{|X_i - Y_j|^2}{2K^2}} + \beta \sum_{j=1}^m \{Y_j - Y_{j+1}\}^2$$

ou seja:

$$\nabla E = \left(\frac{\partial E}{\partial Y_1}, \dots, \frac{\partial E}{\partial Y_m} \right) = -\frac{1}{K} (\Delta Y_1, \dots, \Delta Y_m)$$

Devemos minimizar a função $E(\cdot)$ (fixado um parâmetro K) até que não seja mais possível reduzi-la. Como veremos, a cada parâmetro K e a cada anel inicial Y^0 teremos um ponto de mínimo $Y^*(K, Y^0)$ associado (os parâmetros α e β são definidos a priori). Dessa forma, damos alguns passos na direção do gradiente aproximando-se de $Y^*(K, Y^0)$ (sem necessariamente atingi-lo), em seguida atualizamos K e repetimos o processo até que K seja suficientemente pequeno. Esperamos que ao final do algoritmo os pontos do anel estejam suficientemente próximos das cidades. Teremos chegado então a uma possível solução do caixeiro viajante. Note que, em uma simulação, K representa o tamanho do passo gradiente. Quando $K \rightarrow 0$, as atualizações do anel serão cada vez menores.

Como demonstrado em Durbin, Szeliski e Yuille (DSY-[10]), para K suficientemente grande teremos um único ponto de mínimo associado à função de energia (basta mostrar que os autovalores da matriz hessiana energia são todos positivos). A medida que K decresce é descrita uma trajetória até um valor K_c (K crítico) aonde a trajetória se bifurca (teremos neste caso um autovalor nulo). Em seu trabalho, DSY[10] sugerem o valor K_c como parâmetro inicial. Dependendo do Y^0 inicial utilizado, vários pontos de mínimo local podem ser obtidos, correspondendo cada um a uma possível tour do caixeiro viajante.

Abaixo fazemos uma descrição mais detalhada de nosso algoritmo:

Algoritmo: Elástico;

{**Dados:** posições das n cidades}

Início

- Escolhe \mathbf{a}, \mathbf{b} e K ; {parâmetros do problema}
 - Escolhe ρ ; {taxa de redução de K }
 - Escolhe (Y_1^0, \dots, Y_m^0) inicial;
 - Escolhe G ; {n. de passos do mét. gradiente antes da atualização de K }
 - **Repita**
 - **para** $l:=1$ **até** G **faça**
 - **para** $j:=1$ **até** m **faça** {atualiza anel}
 - $\Delta Y_j = \mathbf{a} \sum_i w_{ij} (X_i - Y_j) + K \mathbf{b} (Y_{j+1} - 2Y_j + Y_{j-1})$;
 - $Y_j = Y_j + \Delta Y_j$;
 - **fim**;
 - **fim**;
 - $K \leftarrow \rho K$;
 - **Até que** $K \leq 0$;
 - Imprime solução;
- fim.** {algoritmo}

Note que se fazemos uma grande redução na atualização de K (ρ pequeno) teremos uma queda brusca na força de atração entre os pontos do anel. Assim, nos aproximaremos com maior rapidez das cidades obtendo uma solução de menor qualidade. O ideal neste caso, é fazer pequenas reduções de K a cada passo mantendo o anel sempre "esticado".

O número de passos G a ser dado pelo método gradiente (antes da atualização de K), é obtido empiricamente, não temos um critério (teórico) que diga qual o melhor momento de atualizarmos o parâmetro K . Da mesma forma os parâmetros \mathbf{a}, \mathbf{b} e \mathbf{r} (taxa de atualização de K) são valores obtidos experimentalmente. Como descrito em [09] e [10] existem boas estimativas para esses parâmetros apesar de não termos uma garantia de convergência (para circuitos viáveis) do algoritmo para uma entrada qualquer.

Mesmo considerando os problemas teóricos encontrados na definição dos parâmetros o método elástico tem mostrado bons resultados computacionais apesar de ser um pouco inferior quando comparado a heurísticas clássicas utilizadas na solução do problema do caixeiro viajante (vide [08]). Devemos lembrar entretanto que se trata de uma tecnologia bastante recente e promissora devido a seu alto grau de paralelização. Podemos ainda implementar este modelo, por exemplo, em um hardware especialmente dedicado ao problema. Em geral, a análise comparativa do desempenho computacional é realizada em máquinas digitais não dando portanto uma idéia exata da performance dos métodos testados.

Como discutido em DSY[10], para α e β qualquer teremos pelo menos um ponto do anel Y_j próximo de cada cidade X_i a medida que K decresce. Assim, o mínimo da função de energia para K suficientemente pequeno corresponde a uma *possível* tour do caixeiro viajante. Para verificar esse fato, basta notar que para cada cidade X_i temos:

$$B(K) = \min_j |X_i - Y_j|$$

Segue então que $B(K) \leq |X_i - Y_j|$ onde Y_j é um ponto qualquer do anel. Elevando ao quadrado, dividindo por $2K^2$ e calculando a exponencial de ambos os termos temos:

$$\exp\left(\frac{B(K)^2}{2K^2}\right) \leq \exp\left(\frac{|X_i - Y_j|^2}{2K^2}\right)$$

Logo, para os m pontos do anel temos:

$$\sum_j \exp\left(\frac{|X_i - Y_j|^2}{2K^2}\right) \geq M \cdot \exp\left(\frac{-B(K)^2}{2K^2}\right)$$

Ou ainda:

$$-aK \log \sum_j \exp\left(\frac{|X_i - Y_j|^2}{2K^2}\right) \geq -aK \log M + \frac{B(K)^2}{2K}$$

A segunda parcela presente na função de energia é cancelada já que aparece em ambos os lados da desigualdade.

Segue então que a energia associada a cada cidade X_i será limitada por uma constante C se:

$$\lim_{K \rightarrow 0} \frac{B(K)^2 - 2K^2 \log M}{2K} = C$$

Note portanto que basta fazer neste caso: $B(K) = O(\sqrt{K})$. Assim, quando $K \rightarrow 0$ teremos $B(K) \rightarrow 0$, ou seja, teremos pelo menos um ponto do anel suficientemente próximo de cada cidade X_i .

Durbin mostra que se $\frac{a}{b} < A$, apenas um único ponto do anel é atraído para cada cidade (onde A é o menor caminho entre um ponto do anel sendo atraído por uma cidade e outro ponto do anel não atraído pela mesma cidade). Apesar disso, pode ocorrer a presença de Spikes (Figura IV.5) como discutido a seguir.

Como demonstrado por Simmen[09], um ponto de equilíbrio da função de energia pode não corresponder a um circuito viável para o caixeiro viajante. Podemos ter as seguintes situações, como representada na figura abaixo. Os pequenos quadrados representam as cidades, e os círculos, os pontos do anel.

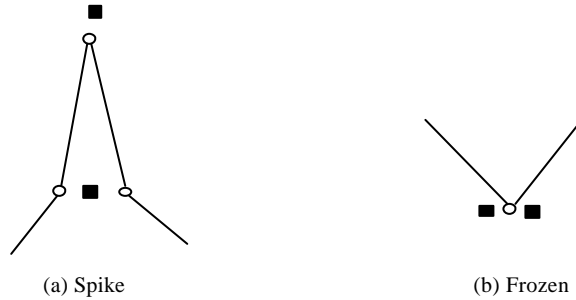


Figura III.6: Circuitos Não-Viáveis

No primeiro caso (Spike), repetimos a visita em uma cidade (circuito não-viável). Como descrito em Simmen [09] a razão α/β deve ser arbitrariamente pequena para a eliminação de spikes. Esta estratégia entretanto, pode ocasionar outros problemas como o representado no item (b) acima (Frozen). No limite, quando K tende a zero, nenhuma das duas cidades consegue atrair definitivamente o ponto do anel. Este fato, demonstrado por Simmen [09] contradiz o resultado obtido anteriormente por DSY[10], já que para $K \rightarrow 0$ não temos necessariamente $\min_j |X_i - Y_j| \rightarrow 0$. Isto ocorre por existir pontos de mínimo local com alta energia, implicando que algumas cidades permaneçam não-visitadas quando $K \rightarrow 0$. Para minimizar a ocorrência de circuitos não-viáveis Simmen [09] sugere $b/a \approx 1/2m$ onde $m \in A/2$.