

Métodos de Monte Carlo

Apresentação 10

Física Computacional

Departamento de Física
Universidade de Aveiro

28 de maio de 2020

Métodos de Monte Carlo — Integração numérica

A denominação "Métodos de Monte Carlo" aplica-se a todos os algoritmos que usam sequências de números aleatórios¹ para obter soluções numéricas aproximadas de diversos tipos de problemas.

Os métodos de Monte Carlo são utilizados em várias áreas da física, química, matemática, finanças, etc. Uma das principais classes de aplicações é a simulação molecular em física estatística, sendo uma das duas abordagens mais comuns a esses problemas (a outra é a dinâmica molecular).

Aqui vamos usar métodos de Monte Carlo para calcular o valor de integrais definidos e para estimar a magnetização de um sistema descrito pelo modelo de Ising, em equilíbrio a uma dada temperatura.

¹Na realidade, pseudo-aleatórios.

Métodos tradicionais de integração numérica

Os métodos usados habitualmente para calcular numericamente integrais a uma dimensão baseiam-se na divisão do domínio de integração (entre $x = a$ e $x = b$) em N intervalos iguais. A função $f(x)$ a integrar é depois calculada em todos os pontos

$$x_k = a + kh$$

com

$$h = \frac{b - a}{N}, \quad k = 0, 1, 2, \dots, N$$

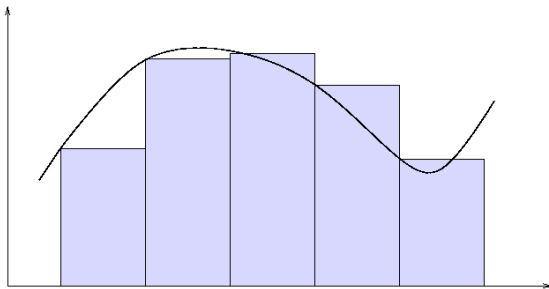
Diferentes métodos usam diferentes maneiras de calcular, a partir dos valores $f(x_k)$, uma estimativa do integral

$$I = \int_a^b f(x)dx$$

Métodos tradicionais de integração numérica

Regra dos retângulos

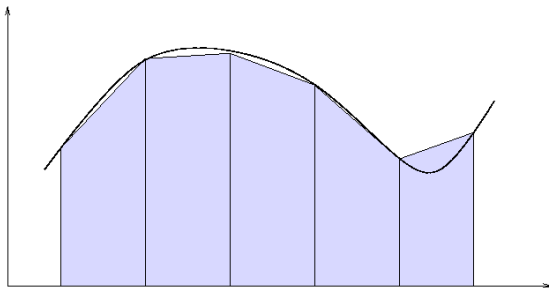
$$I = \sum_{k=0}^{N-1} f(x_k)h$$



Métodos tradicionais de integração numérica

Regra dos trapézios

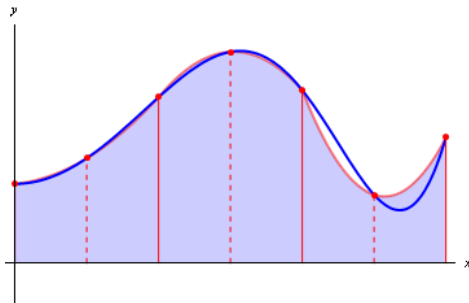
$$I = \left[\frac{1}{2}f(x_0) + \sum_{k=1}^{N-1} f(x_k) + \frac{1}{2}f(x_N) \right] h$$



Métodos tradicionais de integração numérica

Regra de Simpson

$$I = \frac{1}{3} \left[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + 2f(x_{N-2}) + 4f(x_{N-1}) + f(x_N) \right] h$$



Métodos tradicionais de integração numérica

Expandindo a função em série de Taylor, verifica-se que, para integrais a uma dimensão, os erros totais dos 3 métodos apresentados acima são, respetivamente, de ordens $\mathcal{O}(h)$, $\mathcal{O}(h^2)$ e $\mathcal{O}(h^4)$, ou seja, $\mathcal{O}(N^{-1})$, $\mathcal{O}(N^{-2})$ e $\mathcal{O}(N^{-4})$.

A ordem do erro das estimativas de integrais d -dimensionais tem a mesma dependência com h . No entanto, o número de pontos cresce com o decréscimo de h da forma $N \propto h^{-d}$. Assim, os métodos de ordem $\mathcal{O}(h)$, $\mathcal{O}(h^2)$ e $\mathcal{O}(h^4)$ são de ordem $\mathcal{O}(N^{-1/d})$, $\mathcal{O}(N^{-2/d})$ e $\mathcal{O}(N^{-4/d})$.

Para garantir a precisão das estimativas de integrais multidimensionais, é necessário um aumento significativo do número de pontos onde se calcula a função.

Iremos ver mais tarde que os métodos de Monte Carlo para cálculo de integrais definidos se tornam melhores que os métodos convencionais para integrais multi-dimensionais.

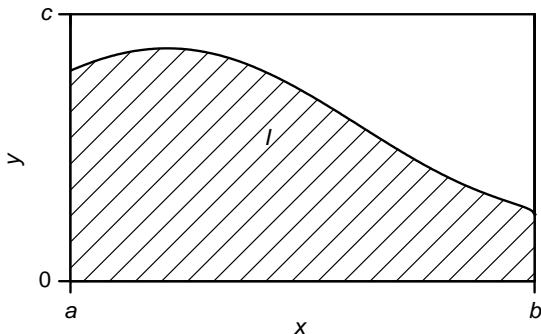
Vamos mostrar duas maneiras de usar números aleatórios para determinar o integral

$$I = \int_a^b y(x)dx$$

Métodos de Monte Carlo — Integração numérica I

Vamos assumir que, no domínio de integração $[a, b]$, $y(x)$ é sempre positivo e inferior a um dado valor c .

- A área do retângulo da figura é $A_0 = c \cdot (b - a)$.
- O integral I é uma fração f da área do retângulo, i.e., $I = fA_0$.



O algoritmo do primeiro método é:

- 1 Gera-se um conjunto $\{x_i\}$ de N números (pseudo-)aleatórios com uma distribuição de probabilidades uniforme entre a e b e um conjunto $\{y_i\}$ de N números (pseudo-)aleatórios com uma distribuição de probabilidades uniforme entre 0 e c . Obtemos um conjunto de N pontos com coordenadas (x_i, y_i) .
- 2 Conta-se o número de pontos N_I desse conjunto que se encontram na zona tracejada.
- 3 Como N_I/N é uma estimativa para f , a estimativa numérica do integral é:

$$I = \frac{N_I}{N} \cdot c \cdot (b - a)$$

Métodos de Monte Carlo — Integração numérica II

O segundo método é baseado no seguinte: se soubermos o valor médio $\langle y \rangle$ de $y(x)$ no domínio de integração $[a, b]$, é imediato obter o valor do integral:

$$I = \int_a^b y(x) dx = (b - a) \langle y \rangle$$

Esta relação resulta diretamente da definição de valor médio de uma função contínua.

O algoritmo é:

- 1 Gera-se um conjunto $\{x_i\}$ de N números (pseudo-)aleatórios com uma distribuição de probabilidades uniforme entre a e b .
- 2 Calcula-se os valores $\{y(x_i)\}$.
- 3 Calcula-se a estimativa numérica do integral:

$$I = (b - a) \langle y \rangle = (b - a) \cdot \frac{1}{N} \sum_{i=1}^N y(x_i)$$

Erro do método de Monte Carlo

Para um dado valor de N , o erro do método de integração numérica de Monte Carlo é diferente para duas repetições do cálculo, com números aleatórios diferentes.

Contudo, se fizermos várias simulações independentes para cada N , verificamos que, **em média**, o valor absoluto do erro é proporcional a $N^{-1/2}$.

Podemos levantar duas questões:

- Porquê essa dependência?
- Podemos saber a constante de proporcionalidade?

Dado que a estimativa do integral resulta de uma média, vamos analisar alguns conceitos estatísticos, em particular, o erro médio quando este é estimado usando uma amostra da população.

Variância — Conceito estatístico

Variância da amostra

$$\tilde{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N \left[y_i - \langle y \rangle \right]^2$$

onde

$$\langle y \rangle = \frac{1}{N} \sum_{i=1}^N y_i$$

Para $N \gg 1$, a variância da amostra é aproximadamente igual à variância da população.

Variância da população

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N \left[y_i - \langle y \rangle \right]^2 = \langle y^2 \rangle - \langle y \rangle^2$$

Desvio padrão e erro da média

O desvio padrão da amostra ou da população é dado pela raiz quadrada da variância

$$\sigma = \sqrt{\sigma^2}$$

A média da população é estimada pela média da amostra. A esta estimativa está associado um desvio padrão, usualmente designado de erro da média, dado por:

Erro da média

$$\sigma_m = \frac{\sigma}{\sqrt{N-1}} \approx \frac{\sigma}{\sqrt{N}}$$

Erro do Método de Monte Carlo

No método de Monte Carlo, I é calculado como uma média (multiplicada pelo tamanho do domínio), assim, o erro que lhe está associado é proporcional ao erro da média. O erro do integral I do slide 11 é:

$$\text{erro}(I) = |b - a| \frac{\sigma_y}{\sqrt{N}}$$

onde σ_y é o desvio padrão das amostras de $y(x)$ e N é o tamanho das amostras.

Para integrais multidimensionais temos

$$\text{erro}(I) = V_D \frac{\sigma_y}{\sqrt{N}},$$

onde V_D é o volume multidimensional do domínio de integração.

Erro do método de M. C. — explicação qualitativa

Para tentar perceber melhor porque o erro é proporcional a σ_y , vamos considerar dois casos limite.

- Vamos supor que a função a integrar é uma constante C em todo o intervalo $x \in [a, b]$. Como queremos determinar o valor médio da função dentro do intervalo, qualquer cálculo da função, para qualquer ponto, vai dar o valor exato: o erro é nulo para qualquer valor de N .
- Suponhamos agora uma segunda função com o mesmo valor médio que a primeira, mas que tem um valor muito pequeno na maior parte do intervalo e um pico muito alto e estreito centrado num valor x_0 . É fácil de perceber que o erro do método de Monte Carlo vai ser muito elevado.
- A diferença significativa entre os dois casos é que dado um conjunto de N números aleatórios $\{x_k\}$, com uma distribuição de probabilidade uniforme entre a e b , a amostra de valores $\{y_k\}$ vai ter um desvio padrão nulo no primeiro caso e um desvio padrão muito grande no segundo.

Comparação dos erros dos métodos tradicionais e de Monte Carlo

A discussão dos últimos slides não depende de todo do número d de dimensões do integral. Quer dizer que, ao contrário dos métodos tradicionais, o aumento do d não altera a forma da dependência com N . Assim, os métodos de Monte Carlo para cálculo de integrais definidos tornam-se melhores que os métodos convencionais para integrais multi-dimensionais (o que acontece, na prática, logo a partir de valores de d iguais a 5 ou 6).

Se quisermos diminuir o erro de uma estimativa, temos que aumentar o número de pontos ou diminuir a variância. Para atingir o segundo objetivo, usa-se uma distribuição não uniforme de números aleatórios — amostragem por importância.

Amostragem por importância

Vamos voltar ao integral a determinar:

$$I = \int_a^b y(x) dx = (b - a) \langle y \rangle$$

A estimativa numérica do integral é calculada através de

$$I = (b - a) \langle y \rangle = (b - a) \cdot \frac{1}{N} \sum_{i=1}^N y(x_i)$$

Onde os números aleatórios x_i foram criados a partir de uma função de distribuição de probabilidade uniforme

$$\rho(x) = \frac{1}{b - a}$$

Amostragem por importância

Se escrevermos o integral como

$$I = \int_a^b \frac{y(x)}{\rho(x)} \rho(x) dx,$$

vemos que a estimativa numérica do integral é dada por

$$I = \frac{1}{N} \sum_{i=1}^N \frac{y(x_i)}{\rho(x_i)},$$

onde, repete-se, os números aleatórios x_i foram criados a partir da função de distribuição de probabilidade $\rho(x)$. Pode mostrar-se que isto é válido para outras distribuições de probabilidade.

Amostragem por importância

Se conseguíssemos escolher uma distribuição de probabilidade $\rho(x)$ proporcional a $y(x)$, a variância e o erro da média seriam nulos.

No entanto, se soubéssemos escolher exatamente essa distribuição de probabilidade, já teríamos o integral calculado.

Sem entrar em muitos pormenores, é fácil de entender que se escolhermos uma distribuição de probabilidade que é em geral maior para valores de x para os quais $y(x)$ é maior, vamos ter um menor erro.

Modelo de Ising

Consideremos um sistema bidimensional de N spins que se encontram nos locais de uma rede quadrada. Cada spin é descrito por um número quântico S que pode tomar apenas os valores -1 e $+1$. A interação entre os spins é descrita pelo modelo de Ising: a energia de uma dada configuração α do sistema é dada por uma soma sobre todos os pares de vizinhos próximos:

$$E_{\alpha} = - \sum_{\langle i,j \rangle} J S_i S_j ,$$

onde $J > 0$. No trabalho prático vamos usar unidades reduzidas de energia, de tal maneira que $J = 1$:

$$E_{\alpha}^* = - \sum_{\langle i,j \rangle} S_i S_j ,$$

Modelo de Ising

Se o sistema está em contacto térmico com um reservatório de calor a uma temperatura T , a probabilidade de uma dada configuração α é dada por:

$$P_{\alpha} = \frac{e^{-\beta E_{\alpha}}}{Z},$$

onde $\beta = 1/(k_B T)$ e Z é a função de partição (soma dos fatores de Boltzmann de todas as 2^N possíveis configurações).

Usando unidades reduzidas de temperatura, vem

$$P_{\alpha} = \frac{e^{-E_{\alpha}^*/T^*}}{Z}.$$

Cálculo da magnetização média

Uma medida macroscópica da magnetização corresponde à média ponderada das magnetizações de todas as configurações que podem ocorrer durante o intervalo de medida. Todas as configurações são possíveis, mas as suas probabilidades são diferentes. Podemos então determinar a magnetização média a partir

$$\langle M \rangle = \sum_{\alpha} P_{\alpha} M_{\alpha}$$

onde P_{α} é a probabilidade da configuração α e M_{α} a respetiva magnetização, dada por

$$M_{\alpha} = \sum_j S_j$$

Sabe-se da física estatística que esta média também pode ser calculada conceptualmente a partir de uma média de ensemble.

Cálculo da magnetização média

Em equilíbrio térmico com o reservatório de calor, a magnetização média é dada pela média de ensemble da coletividade canónica:

$$\langle M \rangle = \frac{\sum_{\alpha} M_{\alpha} e^{-E_{\alpha}^*/T^*}}{\sum_{\alpha} e^{-E_{\alpha}^*/T^*}}$$

Se quiséssemos calcular numericamente o valor exato, os somatórios teriam que ser feitos, em princípio, sobre todas as 2^N configurações possíveis, o que seria impraticável, mesmo para sistemas relativamente pequenos.

Podemos fazer uma média usando apenas L configurações. Se essas configurações forem selecionadas com igual probabilidade, a estimativa de Monte Carlo da magnetização média será

$$\langle M \rangle_{\text{MC}} = \frac{\sum_{l=1}^L M_l e^{-E_l^*/T^*}}{\sum_{l=1}^L e^{-E_l^*/T^*}},$$

onde M_l e E_l^* são a magnetização e a energia da configuração l .

Cálculo da magnetização média — amostragem por importância

Para cada valor da temperatura, apenas uma fração das configurações contribui significativamente para a média de ensemble. A estratégia de selecionar configurações com igual probabilidade faz com que se gaste muito tempo a analisar configurações com probabilidades baixas, levando a que não se possa realizar uma amostragem boa das configurações que realmente contam.

Quando se aplica uma amostragem por importância, o algoritmo de Monte Carlo seleciona a configuração l com uma dada probabilidade ρ_l . No cálculo da média de ensemble, é preciso corrigir o resultado:

$$\langle M \rangle_{\text{MC}} = \frac{\sum_{l=1}^L M_l e^{-E_l^*/T^*} \rho_l^{-1}}{\sum_{l=1}^L e^{-E_l^*/T^*} \rho_l^{-1}}$$

Cálculo da magnetização média — amostragem por importância

Para podermos ter estimativas com significado, temos que escolher apropriadamente a distribuição de probabilidade ρ_l . Se conseguirmos que essa distribuição seja proporcional ao fator de Boltzmann, teremos a situação desejável em que cada configuração é visitada com uma probabilidade proporcional ao peso que tem no cálculo das médias de ensemble. Neste caso,

$$\langle M \rangle_{\text{MC}} = \frac{\sum_{l=1}^L M_l}{\sum_{l=1}^L 1} = \frac{\sum_{l=1}^L M_l}{L}$$

e a estimativa será simplesmente a média aritmética das magnetizações das configurações utilizadas.

Cálculo de $\langle M \rangle_{MC}$ – Algoritmo de Metropolis

A seleção de configurações com uma probabilidade proporcional ao fator de Boltzmann é conseguida da maneira que se descreve a seguir:

- 1 Escolhe-se uma configuração inicial.
- 2 Propõe-se uma nova configuração baseada numa alteração aleatória da configuração anterior². Calcula-se a variação de energia ΔE^* que resultaria dessa alteração.
- 3 Gera-se um número aleatório com uma distribuição uniforme entre 0 e 1. Se e só se esse número for menor ou igual que $\exp(-\Delta E^*/T^*)$, o sistema passa para a nova configuração. Note que novas configurações que diminuem a energia são sempre aceites.
- 4 Repetem-se os dois pontos anteriores um número suficiente de vezes.

²No programa que vai ser feito na aula prática, as novas configurações são propostas a partir da inversão de um dos spins, escolhido aleatoriamente.

Cálculo de $\langle M \rangle_{MC}$ – Algoritmo de Metropolis

As medidas dos valores de grandezas termodinâmicas não são registadas para todas as configurações assim geradas, pois há uma grande correlação estatística entre configurações consecutivas.

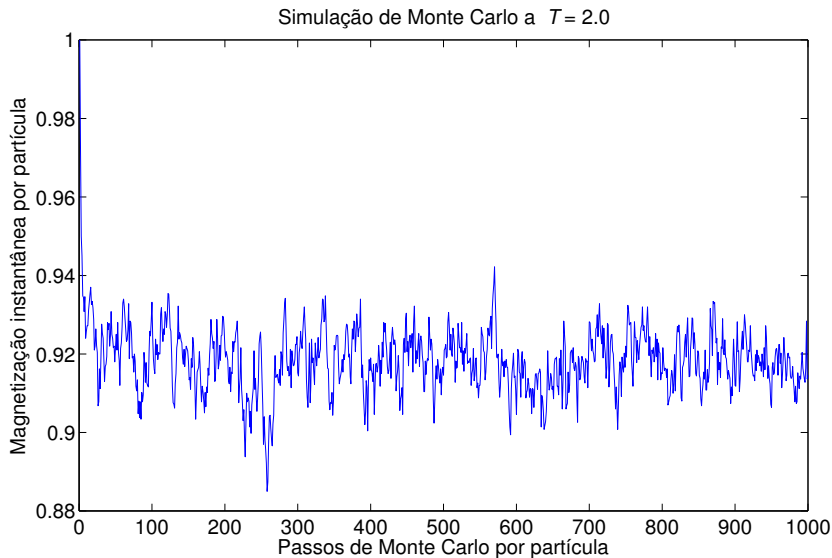
Tipicamente, as configurações M_l que são selecionadas para o cálculo das médias estão separadas por alguns passos de Monte Carlo³ por partícula.

Além disso, são necessários vários passos de Monte Carlo por partícula para que o sistema atinja o equilíbrio termodinâmico à temperatura T^* , pelo que só se deve iniciar as medidas das grandezas termodinâmicas quando se tem a certeza que esse equilíbrio foi atingido.

Vamos começar todas as simulações com o valor máximo da magnetização. Note que estamos a estudar o sistema sem a presença de um campo magnético exterior.

³Tentativas de alteração da configuração, quer sejam aceites ou não.

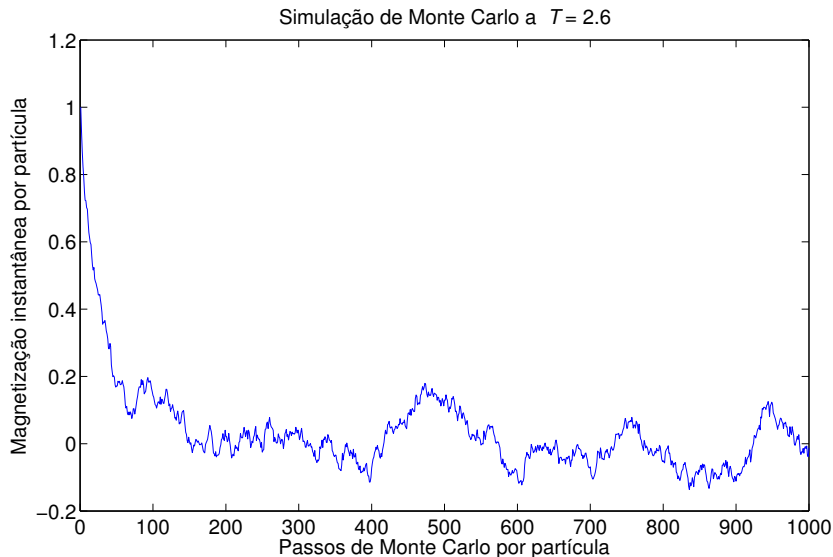
Modelo de Ising — Ferromagnetismo



Modelo de Ising — Ferromagnetismo

Na figura do slide anterior, vemos que para $T^* = 2.0$, a magnetização média em equilíbrio não é nula, mesmo sem campo magnético exterior. O sistema é ferromagnético.

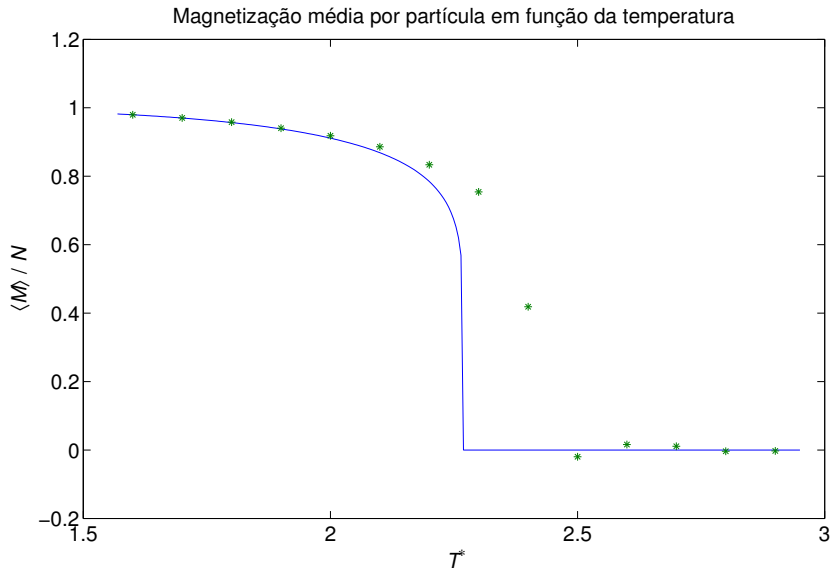
Modelo de Ising — Paramagnetismo



Modelo de Ising — Paramagnetismo

Na figura do slide anterior, vemos que para $T^* = 2.6$, a magnetização média em equilíbrio tende para 0. O sistema é paramagnético.

Modelo de Ising — Transição de fase



Modelo de Ising — Transição de fase

Lars Onsager obteve a solução exata do modelo de Ising a duas dimensões quando $N \rightarrow \infty$ e verificou que, para um valor crítico da temperatura, há uma transição de fase de segunda ordem de um regime ferromagnético para um regime paramagnético. A expressão exata da magnetização média por partícula em função da temperatura reduzida está representada por uma linha no gráfico do slide anterior.

Os pontos obtidos por simulação de Monte Carlo estão representados por asteriscos. Os cálculos necessários para produzir o gráfico demoraram menos que 3 minutos num processador a 1.86 GHz.

A má qualidade dos resultados das simulações próximas do ponto crítico era inevitável. Para estudar o sistema nessa vizinhança, seria necessário usar outras técnicas.

Condições fronteira periódicas

As simulações que podemos fazer têm necessariamente um número de spins muito pequeno. Como queremos obter estimativas da magnetização no limite termodinâmico, $N \rightarrow \infty$, devemos tentar reduzir o efeito dos limites do nosso sistema: quanto menor ele for, maior será a fração de spins que se encontram junto de uma das quatro superfícies que delimitam a rede quadrada.

Para evitar que existam spins que não interagem com 4 vizinhos próximos, usam-se condições fronteira periódicas: assume-se que a mesma rede finita se repete, ocupando todo o espaço. Isso significa que, por exemplo, um spin na coluna mais à direita e numa dada linha central vai ter como quarto vizinho próximo o spin da mesma linha que se encontra na coluna mais à esquerda.