

Capítulo 4

Simulaciones numéricas en Mecánica Estadística

4.1 Introducción: Simulaciones numéricas

Uno de los métodos mas poderosos en Física Teórica en la actualidad son las simulaciones numéricas. Podemos pensar que una simulación numérica consiste esencialmente en la resolución numérica de las ecuaciones que determinan el estado de un dado sistema, ya sea un conjunto de ecuaciones diferenciales, el cálculo de una función partición, etc. No obstante, ya sea por limitaciones en las capacidades de las computadoras disponibles o bien por las características del problema en sí, en general resulta necesario introducir nuevos elementos en los modelos a fin de elaborar un algoritmo que permita resolver numéricamente las ecuaciones correspondientes. Así, por ejemplo, en el caso de ecuaciones diferenciales se hace necesario introducir una discretización en las variables dinámicas del problema. Las ecuaciones resultantes constituyen ahora un *mapa discreto*, cuya solución se espera constituya una buena aproximación a las trayectorias continuas que resuelven las ecuaciones originales.

Las ecuaciones que resultan de introducir las modificaciones antes mencionadas, pueden entonces interpretarse como las ecuaciones exactas de un *nuevo modelo* simplificado, el cual **simula** el comportamiento del sistema original y del cual podemos

extraer (excepto posibles errores estadísticos, en el caso de utilizar variables aleatorias) *información exacta* a través de la computadora. En este sentido, las simulaciones numéricas pueden ser consideradas como auténticos *experimentos numéricos*, en los cuales se estudia el comportamiento exacto del modelo simplificado, con la ventaja frente a un experimento real de que en este caso poseemos un completo control sobre las condiciones microscópicas del problema. Así, por ejemplo, en tanto que un experimental puede controlar los parámetros macroscópicos (tales como temperatura o presión) sobre su muestra, en principio le resulta sumamente difícil determinar los efectos de una variación en el potencial interatómico. Esto, obviamente resulta trivial en el caso de una simulación numérica.

Por otra parte, dichas simplificaciones introducen una dificultad adicional ya que es posible que el modelo simplificado a analizar presente comportamientos nuevos, los cuales no estaban presentes en el modelo original.

Tomemos como ejemplo el tema que nos interesa: el estudio de los fenómenos críticos.

Repasemos entonces el programa de la mecánica estadística: dadas las energías del sistema microscópico (Hamiltoniano) $\{E_\alpha\}$, donde α numera los posibles estados microscópicos del sistema, nuestro objetivo es el cálculo de valores medios de observables X :

$$\langle X \rangle_p = \sum_{\alpha} X_{\alpha} p_{\alpha} \quad (4.1)$$

donde

$$p_{\alpha} = \frac{e^{-\beta E_{\alpha}}}{Z} \quad (4.2)$$

$$Z = \sum_{\alpha} e^{-\beta E_{\alpha}} \quad (4.3)$$

y donde $\beta = 1/k_B T$. Veamos entonces la primera dificultad que se presenta. Sabemos que un modelo puede exhibir una transición de fase (singularidad en las

ecuaciones termodinámicas) *solamente en el límite termodinámico* $N \rightarrow \text{inf.}$ Obviamente, en una computadora solo podemos simular sistemas *finitos*. Ahora bien, los sistemas reales que presentan transiciones de fase son finitos, pero con N muy grande ($N \sim 10^{23}$). Se espera entonces que para N suficientemente grande el comportamiento del sistema finito en torno de un punto crítico resulte indistinguible, dentro de cierta precisión, del correspondiente al sistema infinito.

A partir de estas últimas consideraciones surge la suposición básica en toda simulación numérica de un fenómeno crítico: es posible extraer información sobre el comportamiento de un sistema infinito *extrapolando* el comportamiento del correspondiente sistema *finito* para valores crecientes de N .

Bajo esta suposición, la manera mas simple de obtener los valores medios termodinámicos consiste en la evaluación directa de la función partición Z . Tomemos por ejemplo el modelo de Ising bidimensional, definido en una red cuadrada de L sitios de lado. En la Fig.(4.1) se muestra un cálculo del calor específico para $L = 4, 6$, en comparación con la solución exacta ($L = \infty$). Vemos que los resultados numéricos dan una aproximación razonable de la solución exacta a altas y bajas temperaturas pero, si bien se evidencia la formación de un pico para valores de L crecientes, los resultados difieren mucho en el entorno del punto crítico. Esto resulta facil de interpretar, si recordamos que las divergencias en las funciones respuesta están asociadas a una divergencia en la longitud de correlación $\xi(T)$. Para un sistema finito tendremos que $\xi(T) \leq L$. Para valores de L pequeños, al aproximarnos a la región crítica, ξ rápidamente alcanza su valor máximo, con lo cual el comportamiento de las funciones respuesta difiere mucho del crítico. La solución entonces consiste en tomar valores suficientemente grandes de L . No obstante, recordemos que el número de configuraciones es $W = 2^N = 2^{L^2}$. Para los casos $L = 4, 6$ tenemos que $W = 65536$ y 6.8719×10^{10} . Para el modesto caso de una red $L = 10$ tendremos que $W \sim 10^{30}$. Tomando en cuenta el tiempo de procesamiento por operación elemental de la mejor supercomputadora existente hasta el presente, este cálculo llevaría del orden 10^{15} seg, tiempo que es alrededor de 30 millones de años!

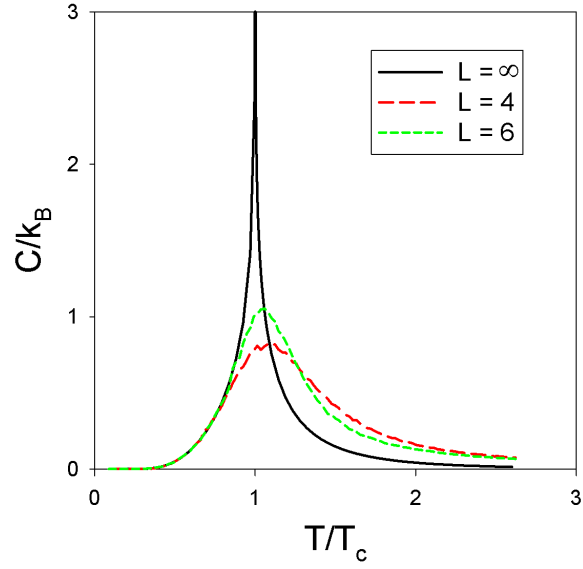


Figura 4.1: Calor específico en función de T para una red $L = 4$, $L = 4$ y $L = \infty$ (solución exacta).

El crecimiento exponencial en el número de configuraciones con el tamaño del sistema (N) es completamente general e independiente del modelo, ya que siendo la entropía una magnitud extensiva tendremos que, a bajas temperaturas $S \propto \ln W \propto N$. La evaluación directa de Z por enumeración exhaustiva de todas las configuraciones resulta por lo tanto un método inviable.

4.2 Métodos de Monte Carlo

De la sección anterior concluimos que para evaluar Z resulta necesario encontrar algún método que nos evite la enumeración exhaustiva de todas las configuraciones del sistema, ya que necesariamente tenemos que trabajar con N razonablemente grande. Esto en principio es factible, ya que, como veremos, en general no todas las configuraciones contribuyen de manera significativa en el cálculo de Z . La selección apropiada de las configuraciones importantes puede llevarse a cabo mediante los llamados **Métodos de Monte Carlo**. En términos generales, se denomina así

a todo algoritmo numérico que involucre de alguna manera el uso de **números aleatorios** (ver Apendice 4.2.1). Discutiremos en esta sección dos métodos generales [?, ?] (si bien ejemplificaremos con el problema que nos interesa) y en la próxima sección veremos la aplicación concreta de uno de ellos al problema de fenómenos críticos.

4.2.1 Números aleatorios

Estrictamente hablando, se define como números aleatorios a una secuencia de valores que adopta una variable aleatoria asociada a algun proceso físico. Asi, por ejemplo, el conjunto de los intervalos de tiempo en los cuales un contador Geiger colocado frente a una muestra radioactiva emite una señal, constituye una secuencia de números aleatorios. En términos prácticos, resulta inviable utilizar algun proceso de este tipo en una simulación numérica. Se recurre entonces a algoritmos que generan una secuencia de números, los cuales *emulan* números aleatorios. Dichos algoritmos se denominan *generadores de números aleatorios*. Siendo estos algoritmos necesariamente deterministas, resulta un poco extraña su utilización para obtener valores que en principio se suponen impredecibles en algun grado. Pensemos entonces de que manera se determina la “aleatoriedad” de un conjunto de valores en un proceso físico, esto es, mediante tests estadísticos, tales como la determinación de la distribución de probabilidad de la variable correspondiente (histograma), cálculo de valores medios, correlaciones y otros mas sofisticados. Si un conjunto de valores, producido por un dado generador, resulta indistinguible de un conjunto de “auténticos” números aleatorios (con una dada distribución de probabilidad) respecto de un cierto conjunto de test estadísticos, diremos que tenemos un buen generador respecto de ese conjunto de tests.

En la práctica, uno de los aspectos mas importantes a tener en cuenta, ademas de la distribución de probabilidad, es la *independencia estadística* de los números de una secuencia, la cual se comprueba mediante tests que involucran el cálculo de correlaciones.

Existe en la literatura una enorme variedad de generadores disponibles, cada uno de ellos con sus ventajas y desventajas[?, ?]. No obstante, este es una tema de investigación abierto y permanentemente se inventan nuevos generadores los cuales supuestamente superan a todos los anteriores. En la práctica, la elección de un dado generador depende del problema a calcular, siendo que en muchos casos los resultados de una simulación no se alteran substancialmente al cambiar un generador “razonable” (y posiblemente rápido) por un generador altamente sofisticado, los cuales en general resultan mucho mas lentos.

La mayor parte de los generadores standard producen números con distribución uniforme, ya sea números reales $r \in [0, 1]$ o bien enteros entre 0 y m , aunque este último caso puede transformarse simplemente en el primero dividiendo por el mayor entero m . Como veremos mas adelante, a partir de un generador con distribución uniforme pueden obtenerse números con distribución (en principio) arbitraria.

Veamos como ejemplo uno de los generadores mas populares por su rapidez, siendo de hecho el que se usa en la mayor parte de los lenguajes de alto nivel, tales como FORTRAN o C, el cual se conoce como *congruencial multiplicativo lineal* [?]. Sean los números enteros r_0 (semilla), m (modulo) a y c . La secuencia de números alatorios enteros $\{r_i\}$, con distribución uniforme entre 0 y m se contruye a partir de r_0 mediante el mapa iterativo:

$$r_i = (a r_{i-1} + c) \bmod m$$

Este generador, obviamente presentará un *ciclo*, cuyo máximo valor posible será m (el cual se toma en general como el mayor entero que admite el lenguaje), ya que una vez que se repita un valor en la secuencia, la misma se repetirá de manera completa. La elección de los enteros a y c en este caso resulta crucial, ya que una mala elección produce ciclos mucho menores que m . Este generador es en general bastante malo (excepto por su rapidez), ya que ademas de presentar ciclos relativamente pequeños se verifica que presenta una alta *correlación* entre distintos valores de las secuencias.

El problema de los ciclos se encuentra presente en todo generador, por la naturaleza determinista de los mismos. No obstante, existen en la actualidad generadores bastante rápidos con ciclos mucho mayores, los cuales en general se obtienen como combinación de varios congruenciales multiplicativos. De esta manera, se necesita la repetición de un cierto conjunto de números sucesivos para que se produzca el ciclo.

Un test simple para comprobar la existencia de correlaciones es el siguiente [?]: supongamos un generador con $r \in [0, 1]$. Se construyen vectores en d dimensiones tomando conjuntos de d sucesivos valores en la secuencia (r_1, r_2, \dots, r_d) . Un conjunto grande de estos vectores debe llenar de manera uniforme el hipercubo de lado 1 en d dimensiones. La existencia de correlaciones fuertes en general se evidencia mediante la aparición de regiones en el hipercubo que nunca son visitadas.

4.2.2 Sorteo de una variable aleatoria

Supongamos que disponemos de un generador con distribución uniforme $r \in [0, 1]$, y que deseamos a partir del mismo obtener los valores de una variable aleatoria x , la cual puede tomar los valores discretos $\{x_1, x_2, \dots, x_N\}$, con probabilidad $\{p_1, p_2, \dots, p_N\}$, tal que:

$$\sum_{i=1}^N p_i = 1$$

Comencemos con el caso mas simple $N = 2$. x tomará el valor x_1 con probabilidad p_1 y el valor x_2 con $1 - p_1$. Si dividimos el intervalo $[0, 1]$ en dos segmentos $[0, p_1]$ y $[p_1, 1]$ (ver Fig. 4.2a), la probabilidad de que r caiga en el primer segmento es precisamente p_1 y de que caiga en el segundo segmento es $1 - p_1$. De esta manera, si $0 \leq r \leq p_1$ asignamos $x \rightarrow x_1$ y $x \rightarrow x_2$ en caso contrario.

Tomemos ahora el caso $N = 3$. Dividimos ahora el intervalo $[0, 1]$ en los tres segmentos: $[0, p_1]$, $(p_1, p_1 + p_2]$ y $(p_1 + p_2, 1]$ (ver Fig. 4.2b). En este caso la probabilidad de que r caiga en el primer segmento es p_1 , en el segundo es p_2 y

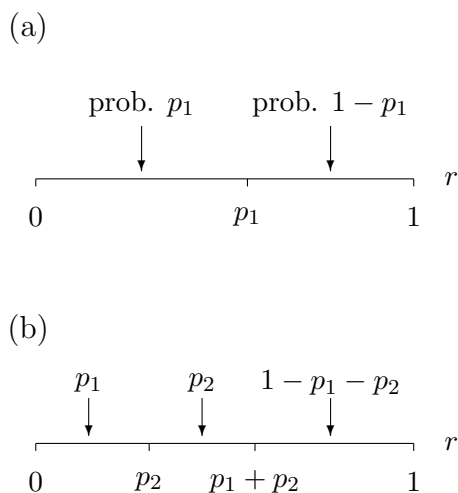


Figura 4.2: Sorteo de una variable aleatoria discreta.

en el tercero $1 - p_1 - p_2$. De esta manera, si $0 \leq r \leq p_1$ asignamos $x \rightarrow x_1$, si $p_1 < r \leq p_1 + p_2$ asignamos $x \rightarrow x_2$ y si $p_1 + p_2 < r \leq 1$ entonces $x \rightarrow x_3$.

El mismo proceso se repite para N arbitrario.

Supongamos ahora que disponemos de un generador para una variable x continua, con

$$p(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases} \quad (4.4)$$

y deseamos a partir del mismo obtener los valores de una variable y mediante una transformación $y = f(x)$, tal que y presente una distribución continua $g(y)$, no-uniforme. Cambiemos por un momento la pregunta y supongamos que conocemos $f(x)$ y deseamos averiguar $g(y)$, y supongamos por simplicidad que $f(x)$ es invertible y no decreciente. Tenemos entonces que $p(y)$ debe satisfacer la relación:

$$p(x)dx = g(y)dy$$

de donde

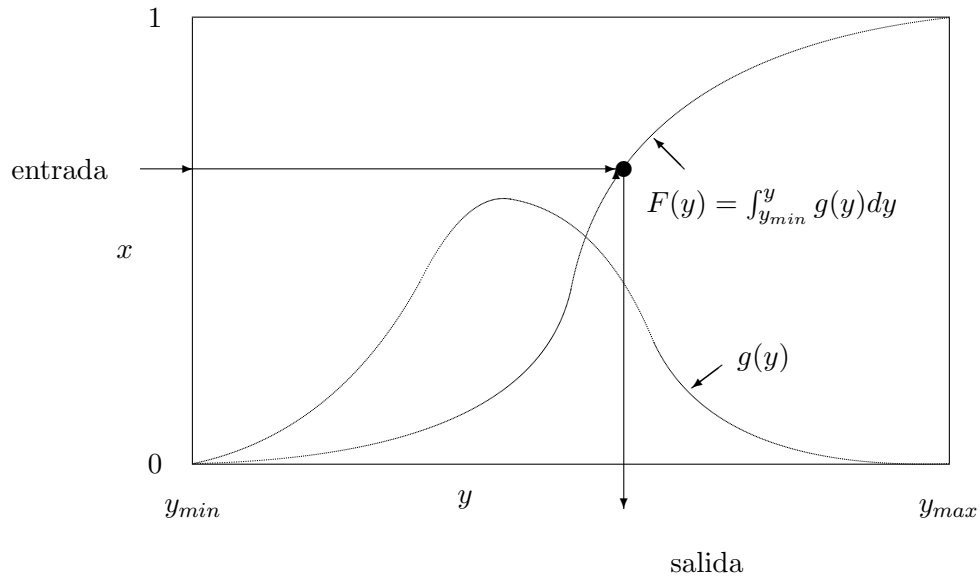


Figura 4.3: Sorteo de una variable aleatoria continua.

$$g(y) = p(x) \frac{dx}{dy} = p(F(y)) F'(y) \quad (4.5)$$

donde $x = F(y) \equiv f^{-1}(y)$. Integrando la Ec.(4.5) desde el mínimo de la distribución y_{min} y usando la Ec.(4.4) tenemos que

$$F(y) = \int_{y_{min}}^y g(t) dt \quad (4.6)$$

Si somos capaces de integrar analíticamente la Ec.(4.6) e invertir la función resultante tendremos el resultado $f(x)$ buscado. De otro modo, siempre es posible obtener estos resultados numéricamente, mediante algún procedimiento de interpolación adecuado (ver figura 4.3).

4.2.3 Muestreo simple (“simple sampling”)

Una posible estrategia para la selección de un conjunto reducido de configuraciones consiste en tomar un subconjunto de M configuraciones completamente al azar, de tal manera que la probabilidad de tomar una dada configuración es la misma para

todas ellas. En el caso del modelo de Ising esto puede realizarse sorteando cada vez la dirección de todos los spines de manera independiente, con igual probabilidad de que tomen la dirección hacia arriba o hacia abajo.

De esta manera podemos aproximar los valores medios $\langle X \rangle_p$ mediante los *estimadores*:

$$\bar{X} = \frac{\sum_k^M e^{-\beta E_{\alpha_k}} X_{\alpha_k}}{\sum_k^M e^{-\beta E_{\alpha_k}}}$$

Este método, si bien resulta de bastante utilidad cierta clase de problemas, presenta graves dificultades en el caso de sistemas que ordenan a bajas temperaturas, tales como el modelo de Ising. A altas temperaturas tenemos que $e^{-\beta E_{\alpha_k}} \sim 1$, $\forall \alpha$. Así, todas las configuraciones tienen aproximadamente la misma probabilidad y el método dará buenos resultados. A bajas temperaturas, por el contrario, la mayor parte de las configuraciones poseen probabilidad casi nula, excepto el estado fundamental y los primeros estados excitados (en la red $L = 3$, 7 de los 512 estados dan el 99.9 % de la contribución a Z para $T = T_c/2$). Así, los resultados varían mucho dependiendo de que dichos estados sean incluidos en el muestreo o no.

4.2.4 Muestreo pesado (“importance sampling”)

Los problemas discutidos anteriormente pueden resolverse si conseguimos un algoritmo que nos permita elegir las configuraciones del muestreo preferencialmente entre aquellas que contribuyen más a la función partición a una dada temperatura. Este tipo de algoritmos constituyen lo que se denomina *muestreo pesado*. Supongamos que elegimos las configuraciones con una distribución no-uniforme

$$\rho_\alpha \geq 0 \quad \forall \alpha$$

donde

$$\sum_\alpha^W \rho_\alpha = 1$$

Introduzcamos las variables

$$\xi_\alpha \equiv \frac{p_\alpha}{\rho_\alpha} X_\alpha$$

donde p_α viene dado por la Ec.(4.2). Es decir, tomamos una distribución arbitraria ρ y elegimos las variables ξ de tal manera que la contribución de $\xi_\alpha \rho_\alpha$ sea igual a la del término $X_\alpha p_\alpha$ en la suma (4.1). Introducimos entonces el estimador $\bar{\xi}$:

$$\bar{\xi} = \frac{1}{M} \sum_{l=1}^M \xi_{\alpha_l}. \quad (4.7)$$

Calculemos entonces el valor de expectación de $\bar{\xi}$ de acuerdo con la distribución ρ :

$$\langle \bar{\xi} \rangle_\rho = \langle \xi \rangle_\rho = \sum_{\alpha=1}^W \xi_\alpha \rho_\alpha = \sum_{\alpha=1}^W p_\alpha X_\alpha = \langle X \rangle_p \quad (4.8)$$

Es decir, la cantidad $\bar{\xi}$ promediada sobre todas las posibles elecciones de M configuraciones distribuidas según ρ resulta igual al valor medio térmico $\langle X \rangle_p$. No obstante, esto no nos garantiza que el método funcione correctamente, ya que como vimos anteriormente las fluctuaciones pueden ser muy grandes (de hecho este análisis incluye como caso particular el de muestreo simple para $\rho = 1/W$). Analicemos entonces la varianza del estimador $\bar{\xi}$:

$$\begin{aligned} \text{var}(\bar{\xi}) &= \left\langle \left(\bar{\xi} - \langle \bar{\xi} \rangle_\rho \right)^2 \right\rangle_\rho \\ &= \frac{1}{M} \text{var}(\xi) \\ &= \frac{1}{M} \sum_{\alpha=1}^W \left(\xi_\alpha - \langle X \rangle_p \right)^2 \rho_\alpha \end{aligned} \quad (4.9)$$

De esta manera podemos reducir $\text{var}(\bar{\xi})$ eligiendo ρ de tal manera que $\xi_\alpha \approx \langle X \rangle_p$ para todo α . La manera mas simple de realizar esto es conseguir que $\rho_\alpha = p_\alpha$ para todo α . En este caso tendremos que $\xi_\alpha = X_\alpha$ y el estimador (4.7) se transforma en un simple promedio aritmético de X_α sobre las M configuraciones.

4.3 Cadenas de Markov

Los métodos de muestreo pesado que nos interesan se basan en el concepto de *cadena de Markov*. Antes de definir una cadena de Markov repasemos algunos conceptos de procesos estocásticos [?]. Un proceso estocástico consiste en una variable aleatoria x (supongamos por simplicidad que x toma valores discretos x_1, x_2, \dots) cuya distribución de probabilidad $P(x, t)$ evoluciona en el tiempo con:

$$\sum_{\{x_i\}} P(x_i, t) = 1$$

A un conjunto particular de valores $x(t)$ se lo denomina una *realización* del proceso. El valor medio de x viene dado por

$$\langle x \rangle(t) = \sum_{\{x_i\}} x_i P(x_i, t)$$

Un proceso se dice *estacionario* si $P(x, t) = P(x)$.

Para definir completamente un proceso estocástico es preciso conocer además las distribuciones *conjuntas* de probabilidad $P(x_1, t_1; \dots, x_n, t_n)$, que dan la probabilidad de que la variable aleatoria tome el valor x_1 al tiempo t_1 , x_2 al tiempo t_2 , etc. Se introduce también la probabilidad *condicional* $P(x_1, t_1; x_2, t_2; \dots; x_{n-1}, t_{n-1} \| x_n, t_n)$, donde $t_1 < t_2 < \dots < t_n$, como la probabilidad de que x tome el valor x_n al tiempo t_n , *dado* que tomo los valores x_1 al tiempo t_1 , x_2 al tiempo t_2 , etc. La probabilidad condicional $P(x_1, t_1 \| x_2, t_2)$ se denomina *probabilidad de transición*, y se define mediante la identidad:

$$P(x_1, t_1)P(x_1, t_1 \| x_2, t_2) = P(x_1, t_1; x_2, t_2)$$

Un proceso **markoviano** se define como aquel que solo posee memoria de su pasado inmediato, es decir:

$$P(x_1, t_1; x_2, t_2; \dots; x_{n-1}, t_{n-1} \| x_n, t_n) = P(x_{n-1}, t_{n-1} \| x_n, t_n)$$

En este caso la jerarquía completa de distribuciones conjuntas puede describirse en términos de $P(x, t)$ y de la probabilidad de transición. Por ejemplo:

$$\begin{aligned}
P(x_1, t_1; x_2, t_2; x_3, t_3) &= P(x_1, t_1; x_2, t_2) P(x_1, t_1; x_2, t_2 \| x_3, t_3) \\
&= P(x_1, t_1) P(x_1, t_1 \| x_2, t_2) P(x_2, t_2 \| x_3, t_3)
\end{aligned}$$

De esta manera, un proceso markoviano puede pensarse como una secuencia de transiciones, las cuales son *estadísticamente independientes* entre si.

Una **cadena de Markov** es un proceso de Markov, donde la variable estocástica x asume valores discretos en tiempos *discretos*. La probabilidad $P(x_1, t \| x_2, t+1)$ se denomina en este caso probabilidad de transición *por unidad de tiempo* y la denotaremos por $P_t(x_1 \rightarrow x_2)$. En los casos que nos interesan esta última probabilidad no depende de t y por lo tanto omitiremos el subíndice t .

4.4 Métodos de Monte Carlo en Mecánica Estadística

Estos métodos consisten en algoritmos que generan una cadena de Markov (ver Apéndice 4.3) de estados $\{\alpha_1, \alpha_2, \dots\}$, con distribución de probabilidad estacionaria $P(\alpha, t) = p_\alpha = e^{-\beta E_\alpha} / Z$, los cuales se utilizan para el muestreo pesado de configuraciones. En otras palabras, dada una configuración inicial α se genera una nueva configuración α' , de acuerdo a una probabilidad de transición $P(\alpha \rightarrow \alpha')$. El proceso se repite con un nuevo sorteo, independiente del anterior, para $\alpha' \rightarrow \alpha''$, etc. La probabilidad de transición se elige de tal manera que, al cabo de un cierto tiempo los sucesivos estados obedecen una distribución estacionaria igual a p_α . Para ello es suficiente con que $P(\alpha \rightarrow \alpha')$ satisfaga las siguientes hipótesis:

1. **Accesibilidad:** Dados dos estados cualesquiera α y α' , existe una sucesión (*finita*) de estados $\alpha_1, \alpha_2, \dots, \alpha_M$, tal que

$$P(\alpha \rightarrow \alpha_1) P(\alpha_1 \rightarrow \alpha_2) \dots P(\alpha_M \rightarrow \alpha) \neq 0$$

2. **Balance detallado (o microreversibilidad):** $P(\alpha \rightarrow \alpha')$ satisface la relación:

$$p_\alpha P(\alpha \rightarrow \alpha') = p_{\alpha'} P(\alpha' \rightarrow \alpha) \quad (4.10)$$

Teorema: Si se cumplen (1) y (2) entonces:

$$\lim_{t \rightarrow \infty} P(\alpha, t) = p_\alpha$$

El siguiente es un argumento de plausibilidad. Sabemos que las probabilidades de transición deben satisfacer la normalización:

$$\sum_{\alpha'} P(\alpha \rightarrow \alpha') = 1 \quad (4.11)$$

Mostraremos primero que p_α es efectivamente un estado estacionario de la cadena de Markov bajo la condición 2 (Ec.(4.10)). Esto es, si para $t = n$ se cumple que:

$$P(\alpha, n) = p_\alpha \quad \forall \alpha$$

tenemos que:

$$\begin{aligned} P(\alpha, n+1) &= \sum_{\alpha'} P(\alpha', n) P(\alpha' \rightarrow \alpha) \\ &= \sum_{\alpha'} p_{\alpha'} P(\alpha' \rightarrow \alpha) \\ &= p_\alpha \sum_{\alpha'} P(\alpha \rightarrow \alpha') \\ &= p_\alpha \end{aligned} \quad (4.12)$$

donde en el segundo paso hemos utilizado la hipótesis 2 (Ec.(4.10)) y en el tercero la Ec.(4.11).

Supongamos ahora que $P(\alpha, n) \neq p_\alpha$. Definimos la desviación entre $P(\alpha, n)$ y p_α al tiempo $t = n$ como:

$$D_n \equiv \sum_{\alpha} |P(\alpha, n) - p_\alpha| \quad (4.13)$$

Tenemos entonces:

$$\begin{aligned}
 D_{n+1} &= \sum_{\alpha} |P(\alpha, n+1) - p_{\alpha}| \\
 &= \sum_{\alpha} \left| \sum_{\alpha'} P(\alpha', n) P(\alpha' \rightarrow \alpha) - p_{\alpha} \right| \\
 &= \sum_{\alpha} \left| \sum_{\alpha'} \{P(\alpha', n) P(\alpha' \rightarrow \alpha) - p_{\alpha} P(\alpha \rightarrow \alpha')\} \right| \\
 &= \sum_{\alpha} \left| \sum_{\alpha'} \{P(\alpha', n) - p_{\alpha'}\} P(\alpha' \rightarrow \alpha) \right|
 \end{aligned}$$

donde en el tercer paso hemos usado la Ec.(4.11) y en el cuarto el balance detallado.

Usando la desigualdad triangular tenemos que:

$$\begin{aligned}
 D_{n+1} &\leq \sum_{\alpha, \alpha'} |P(\alpha', n) - p_{\alpha'}| P(\alpha' \rightarrow \alpha) \\
 &\leq \sum_{\alpha'} |P(\alpha', n) - p_{\alpha'}| \\
 &\leq D_n
 \end{aligned}$$

es decir, la desviación D_n decrece a lo largo de la cadena.

Finalmente, notemos que para p_{α} dado por la Ec.(4.2), la condición de balance detallado puede reescribirse como:

$$\frac{P(\alpha \rightarrow \alpha')}{P(\alpha' \rightarrow \alpha)} = e^{-\beta \Delta E} \quad (4.14)$$

donde $\Delta E \equiv E_{\alpha} - E_{\alpha'}$.

4.4.1 Algoritmo de Metropolis

Esquema general del algoritmo [?]:

- Se elige una configuración arbitraria α .
- Se elige una nueva configuración α' por algún método, y se calcula ΔE .

- Si $\Delta E < 0$ se acepta la nueva configuración α' como la siguiente en la cadena.
- Si $\Delta E > 0$ la nueva configuración se acepta con probabilidad $e^{-\beta\Delta E}$ y se rechaza con probabilidad $1 - e^{-\beta\Delta E}$.
- Se repite el proceso con la nueva configuración (α o α').

La nueva configuración puede elegirse tanto secuencial como aleatoriamente, de acuerdo con las características del problema.

Tomemos como ejemplo el modelo de Ising en una red de N spines. Una manera simple de elegir la nueva configuración es elegir un spin S_i e invertirlo (“flippear”): $S_i \rightarrow -S_i$. Los spines S_i pueden elegirse en las sucesivas configuraciones de las siguientes maneras:

- Secuencialmente, recorriendo la red de alguna manera hasta completar los N spines.
- Aleatoriamente, sorteando en cada paso un spin con igual probabilidad ($1/N$) entre todos los spines de la red.

Las probabilidades de transición serán en este caso:

$$P(\alpha \rightarrow \alpha') = \begin{cases} \frac{1}{A} & \text{si } \Delta E \leq 0 \\ \frac{1}{A} e^{-\beta\Delta E} & \text{si } \Delta E > 0 \end{cases} \quad (4.15)$$

para las configuraciones α' que difieran de α en un spin y cero para las restantes. A es una constante de normalización que vale 1 para el caso secuencial y N para el caso aleatorio. Es fácil verificar que las probabilidades de transición (4.15) satisfacen balance detallado.

4.4.2 Algoritmo de Glauber para el modelo de Ising

Tomemos, por ejemplo, el Hamiltoniano de Ising con interacción primeros vecinos a campo nulo, definido sobre una red de N spines:

$$H = -J \sum_{\langle i,j \rangle} S_i S_j \quad S_i = \pm 1 \quad (4.16)$$

Para cada sitio se define el *campo local*:

$$h_i = J \sum_{\delta} S_{i+\delta} \quad (4.17)$$

donde el subíndice δ corre sobre los primeros vecinos del sitio i . Notemos que H puede escribirse entonces como:

$$H = -\frac{1}{2} \sum_i S_i h_i$$

El algoritmo (o *dinámica*) de Glauber es el siguiente:

- Para una dada configuración α se sortea un sitio i con equiprobabilidad entre los N de la red.
- Independientemente del valor de S_i en α , el valor de S_i en la nueva configuración α' se sortea de acuerdo con

$$P(S_i = 1) = f(h_i) \quad y \quad P(S_i = -1) = 1 - f(h_i)$$

donde $f(h)$ es la función de Fermi:

$$f(h) = \frac{1}{1 + e^{2\beta h}} = \frac{1}{2} [1 + \tanh(\beta h)] \quad (4.18)$$

- Se repite el proceso.

Notemos que $f(h)$ satisface:

$$f(h) + f(-h) = 1$$

de donde $P(S_i = -1) = f(-h_i)$. Así, podemos escribir en forma compacta:

$$P(S_i) = f(S_i h_i) \quad (4.19)$$

Analicemos las probabilidades de transición. Las únicas transiciones posibles son: aquella en la cual se invierte el spin sorteado $S_i \rightarrow -S_i$ y aquella en la cual S_i no cambia. Dado que el estado final de S_i no depende de su estado inicial, tenemos que

$$\begin{aligned} P(S_i \rightarrow S_i) &= P(S_i) = f(S_i h_i) \\ P(S_i \rightarrow -S_i) &= P(-S_i) = f(-S_i h_i) \end{aligned}$$

de donde

$$\frac{P(S_i \rightarrow -S_i)}{P(-S_i \rightarrow S_i)} = \frac{f(-S_i h_i)}{f(S_i h_i)} = e^{2\beta h_i S_i} \quad (4.20)$$

Pero notemos que $\Delta H = -2\beta h_i S_i$. Reemplazando en la Ec.(4.20) obtenemos así la condición de balance detallado.

Antes de continuar haremos algunos comentarios generales acerca de los algoritmos vistos hasta aquí.

- Luego de un transitorio (finito) en el cual $P(\alpha, n) \rightarrow p_\alpha$ (al menos dentro de cierta precisión), el promedio *aritmético* de un observable cualquiera X_α tomado sobre M estados de la cadena de Markov resultante

$$\langle X \rangle = \frac{1}{M} \sum_j^M X_{\alpha_j},$$

según la prescripción del muestreo pesado, representa un promedio **en ensemble**.

- Los estados sucesivos de la cadena de Markov no son estadísticamente independientes, ya que existe una fuerte correlación entre ellos. Para obtener un buen muestreo, conviene entonces esperar un cierto número de iteraciones para acumular los valores de las variables a promediar. El número de pasos a esperar entre dos acumulaciones sucesivas es variable, ya que depende de

cuan rápido mueran las correlaciones a lo largo de la cadena, lo cual a su vez depende de la cercanía al punto crítico.

Para modelos de spines, se define de manera standard un *paso de MC (PMC)*, como el tiempo (nro. de iteraciones) necesario para recorrer *en media* los N spines. En general, se promedia como mínimo cada PMC.

- En tanto que el algoritmo de Glauber es específico para modelos tipo Ising (variables de estado binarias), el algoritmo de Metropolis es completamente general para cualquier modelo de tipo Hamiltoniano.
- En el caso del modelo de Ising, es importante recalcar las siguientes diferencias entre ambos algoritmos: en tanto que en Metropolis la transición fundamental consiste en un spin “flip” ($S_i \rightarrow -S_i$), en el caso de Glauber el estado del spin S_i se sortea de manera independiente del estado del spin en el instante anterior. Además, la probabilidad de transición en este último algoritmo no es necesariamente 1 en el caso en que $\Delta E < 0$. Estas características, como veremos a continuación, permiten una interpretación física del algoritmo como proceso dinámico.

4.4.3 Interpretación dinámica del algoritmo de Glauber

Supongamos que interpretamos cada iteración del algoritmo como la discretización de una variable continua t , dentro de una cierta escala tiempo τ asociada con cada iteración. Tendremos así definidas dentro del problema dos escalas de tiempo: una “microscópica” τ y una “macroscópica” dada por el PMC: $N\tau$. A cada n iteraciones asociaremos un tiempo $t' = n\tau$, y llamaremos t al tiempo medido en unidades de PMC, es decir

$$t \equiv \frac{t'}{N\tau} = \frac{n}{N},$$

de manera que $t = 1$ corresponde a 1 PMC. Tenemos entonces que $P(\alpha, t) \equiv P(\alpha, n)$. De la Ec.(4.12) tenemos:

$$P(\alpha, n+1) = P(\alpha, n)P(\alpha \rightarrow \alpha) + \sum_{\alpha' \neq \alpha} P(\alpha', n)P(\alpha' \rightarrow \alpha). \quad (4.21)$$

De la Ec.(4.11) podemos escribir

$$P(\alpha \rightarrow \alpha) = 1 - \sum_{\alpha' \neq \alpha} P(\alpha' \rightarrow \alpha). \quad (4.22)$$

Reemplazando en la Ec.(4.21) obtenemos

$$\begin{aligned} P(\alpha, n+1) - P(\alpha, n) &= P(\alpha, t + \Delta t) - P(\alpha, t) \\ &= \sum_{\alpha'} \{P(\alpha', t)P(\alpha' \rightarrow \alpha) - P(\alpha, t)P(\alpha \rightarrow \alpha')\} \end{aligned} \quad (4.23)$$

donde $\Delta t = 1/N$ corresponde a un paso microscópico (τ) en unidades de PMC. Recordemos que para una actualización aleatoria $P(\alpha \rightarrow \alpha') \propto 1/N$. Definimos entonces las probabilidades de transición *por unidad de tiempo* τ como:

$$\mathcal{W}(\alpha \rightarrow \alpha') \equiv N P(\alpha \rightarrow \alpha') = P(\alpha \rightarrow \alpha')/\Delta t$$

Dividiendo la Ec.(4.23) por Δt tendremos, en el límite de tiempo continuo $N \rightarrow \infty$, que el proceso estocástico esta gobernado por la *ecuación maestra* [?]:

$$\frac{dP(\alpha, t)}{dt} = \sum_{\alpha'} \{P(\alpha', t)\mathcal{W}(\alpha' \rightarrow \alpha) - P(\alpha, t)\mathcal{W}(\alpha \rightarrow \alpha')\}, \quad (4.24)$$

la cual describe un proceso *difusivo*. En otras palabras, el algoritmo simula una dinámica estocástica de los spines, la cual puede ser interpretada físicamente en términos de acoplamientos muy debiles de los mismos a un baño térmico (las vibraciones térmicas del cristal subyacente, por ejemplo); este último induce transiciones aleatorias entre los spines, las cuales dependen de la temperatura y del campo *local* instantaneo observado por el spin. Dado que el tiempo medio entre dos saltos sucesivos de un spin resulta varios ordenes de magnitud mayor que las escalas temporales

características de las vibraciones en un sólido, estas últimas pueden ser razonablemente aproximadas por un *baño térmico*, nombre con el cual se conoce también este algoritmo. De hecho, el algoritmo fue originalmente propuesto por Glauber [?] como un modelo físico de la relajación al equilibrio del modelo de Ising. Al modelo de Ising dotado de esta dinámica se lo suele denominar también **modelo de Ising cinético**.

Bajo esta última interpretación, tenemos que

$$\langle X \rangle = \frac{1}{M - M_0} \sum_{j=M_0+1}^M X_{\alpha_j} = \frac{1}{t_M - t_{M_0}} \int_{t_{M_0}}^{t_M} X(t) dt \quad (4.25)$$

donde $t_{M_0} = M_0/N$ es el tiempo transcurrido luego de generar las M_0 primeras configuraciones (transitorio) en unidades de PMC y $t_M = M/N$. Es decir, los promedios en ensemble a través de los cuales estudiamos las propiedades de equilibrio son obtenidos a través de promedios *temporales* a lo largo de una trayectoria $X(t) \equiv X(\alpha(t))$, donde las variables a promediar relajan al equilibrio gobernadas por la ecuación (4.24). Esto trae a colación el problema de la **ergodicidad**. Es decir, los promedios temporales serán representativos de los promedios en ensemble, en la medida en que el sistema tenga acceso durante su evolución a todas las configuraciones del sistema. Esto se encuentra de hecho garantizado en la simulación, ya que trabajamos con sistemas *finitos*.

4.4.4 Ruptura espontánea de simetría y tamaño finito

Existen, no obstante, situaciones en las que la ergodicidad del sistema finito es un efecto indeseado, ya que el sistema físico que se desea simular *no es ergódico*.

Tomemos por ejemplo un ferromagneto uniaxial (Ising). A campo magnético nulo $B = 0$ el Hamiltoniano de este sistema es simétrico ante una inversión del eje de magnetización. A altas temperaturas $T > T_c$ esta simetría se refleja en las funciones termodinámicas. Para $T < T_c$ el sistema presenta magnetización espontánea, apareciendo así una orientación preferencial de los spines. Esto es, el estado termodinámico del sistema no presenta las mismas simetrías que el Hamiltoniano y el

sistema presenta ruptura espontanea de simetría. Desde el punto de vista dinámico, un sistema infinito cuya evolución este gobernada por una dinámica del tipo (4.24), *no será ergodico* para $T < T_c$. Es decir, supongamos que el sistema se encuentra en un instante dado en equilibrio termodinámico con magnetización hacia arriba. En este caso resultará imposible que transicione al estado con magnetización hacia abajo, ya que esto involucra pasar por estados con *energía infinita*.

Analicemos esto con mas detalle. Llamemos m_α a la magnetización microscópica en la configuración α , esto es:

$$m_\alpha \equiv \frac{1}{N} \sum_{i=1}^N S_i^\alpha$$

donde S_i^α denota el valor del spin S_i en la configuración α .

En el equilibrio, la distribución probabilidad para la magnetización m viene dada por

$$P(m) = \sum_{\alpha} \delta(m - m_\alpha) p_\alpha \quad (4.26)$$

A campo nulo tendremos que $P(m) = P(-m)$, dada la simetría de inversión del Hamiltoniano. Para $T < T_c$, $P(m)$ presentará un máximo doblemente degenerado en torno de un valor $m = |m^*|$, donde $m^* \rightarrow m_s(T)$ cuando $N \rightarrow \infty$, siendo $m_s(T)$ la magnetización espontanea del sistema infinito (ver Fig.4.4). Para N finito $P(0)$ será finita ($m = 0$ corresponde al estado antiferromagnético de energía máxima). Así, en todo sistema finito el sistema podra incursionar, mediante la dinámica estocástica, entre las regiones de magnetización positiva y negativa pasando por estados de equilibrio. Esto es, el sistema siempre es ergódico. El tiempo medio t_e que el sistema permanece en una de las regiones antes de transicionar a la otra se denomina *tiempo de ergodicidad*. t_e depende de T y N . Si el tiempo de observación t_{obs} , durante el cual se lleva a cabo el promedio temporal de m_α , es $t_{obs} \gg t_e$, tendremos que $\langle m \rangle = 0$, contrario a lo que ocurre en un sistema infinito.

Para $N \rightarrow \infty$ y $T < T_c$ tenemos que $P(0) \rightarrow 0$ y $t_e \rightarrow \infty$. Es en este caso que se produce la ruptura espontanea de simetría y el sistema deja de ser ergódico. En

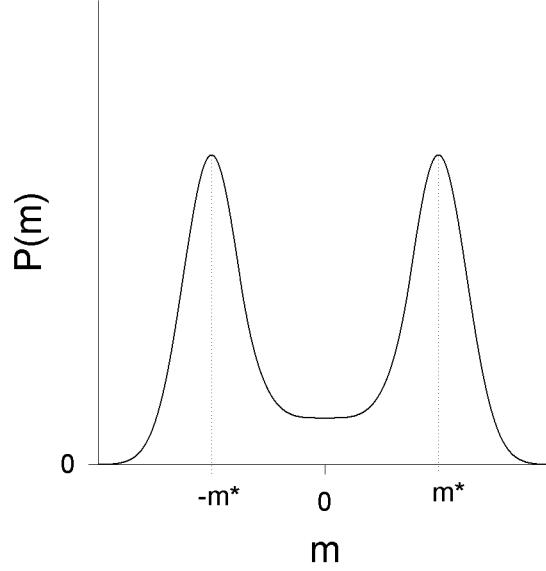


Figura 4.4: Distribución de probabilidad para la magnetización cuando $T < T_c$ en un sistema finito.

otras palabras, si el sistema comienza su evolución en estados con magnetización positiva, jamás alcanzará estados de magnetización negativa y viceversa. En este caso $\langle m \rangle = m_s(T)$. Dado que la mecánica estadística de equilibrio no toma en cuenta la parte dinámica, la magnetización a campo nulo en un sistema infinito promedia a cero, ya que ambas regiones $m = \pm m_s(T)$ son igualmente probables. Así, la definición correcta de $m_s(T)$ es

$$m_s(T) = \lim_{B \rightarrow 0} \lim_{N \rightarrow \infty} \langle m \rangle (T, B). \quad (4.27)$$

La distribución $P(m)$ a campo no nulo es asimétrica y en el límite termodinámico $P(m)$ se anula en la región correspondiente a magnetización anti-alineada con el campo B . De esta manera el límite $B \rightarrow 0$ *a posteriori* del límite termodinámico reproduce el resultado físicamente correcto.

Con respecto a la simulación, debemos tomar en cuenta entonces la existencia de ergodicidad si queremos evaluar correctamente la magnetización espontánea. Dado

que el límite (4.27) resulta sumamente engorroso de calcular numéricamente, las soluciones posibles a este problema son (i) evaluar t_e para cada par de valores N y T y promediar con $t_{obs} < t_e$; (ii) aprovechar la simetría $P(m) = P(-m)$ y promediar $|m|$. Este último método resulta el mas conveniente especialmente cerca de T_c donde, debido a las grandes fluctuaciones de m , t_e resulta muy pequeño, aún para valores grandes de N [?].

4.4.5 Dinámica de Kawasaki

Existe un segundo algoritmo específico para el modelo de Ising conocido como *dinámica de Kawasaki*. En este algoritmo las probabilidades de transición son del tipo Glauber:

$$P(\alpha \rightarrow \alpha') = \frac{1}{2} [1 + \tanh(2\beta\Delta E)],$$

pero la regla para elegir la nueva configuración α' consiste en elegir dos spines al azar S_i, S_j e intercambiarlos.

Notemos que este algoritmo mantiene invariante la magnetización. Esto equivale a promediar en un ensemble de magnetización constante. En términos de un gas de red, esto equivale a mantener constante la densidad de partículas:

$$\rho = \frac{1}{2} (1 - m).$$

El ensemble de equilibrio obtenido a lo largo de la cadena de Markov corresponde entonces al ensemble *canónico* (número de partículas constante), en tanto que el obtenido mediante la dinámica de Glauber corresponde al *gran canónico*.

4.4.6 Observaciones generales

Hasta aquí no hemos mencionado nada acerca de las condiciones de contorno en una simulación. Salvo en problemas particulares, en general conviene utilizar condiciones de contorno periódicas, ya que estas disminuyen los efectos de borde. No obstante, deben observarse ciertas precauciones con respecto a las simetrías del problema. Por

ejemplo, en el caso de un Hamiltoniano antiferromagnético se debe tener cuidado de no introducir frustración artificialmente con el tamaño del sistema. Si la red utilizada es un cuadrado, el lado L debe ser par.

Frecuentemente se desean obtener a partir de simulaciones numéricas las funciones respuesta del sistema (calor específico, susceptibilidad, etc.). En general resulta mas facil y mas preciso calcular las mismas a través de las fluctuaciones de las magnitudes extensivas asociadas. Asi, por ejemplo, para el cálculo del calor específico y la susceptibilidad isotérmica podemos utilizar las expresiones ($k_B = 1$):

$$c = \frac{1}{NT^2} \left(\langle E^2 \rangle - \langle E \rangle^2 \right) \quad (4.28)$$

$$\chi = \frac{1}{NT} \left(\langle M^2 \rangle - \langle M \rangle^2 \right) \quad (4.29)$$

Respecto de optimizaciones, mencionaremos dos trucos usuales:

- Ciertas cantidades no necesitan ser recalculadas completamente a cada iteración, sino solamente su corrección. Tomemos por ejemplo la magnetización m en el modelo de Ising. Llamemos $m^{(j)}$ al valor de m en el paso j . Tenemos que:

$$m^{(j+1)} = m^{(j)} + \Delta m$$

donde, para un “spin flip” $S_i \rightarrow -S_i$, tenemos que $\Delta m = -2 S_i$.

- Si las posibles transiciones en un paso son pocas, conviene calcular las probabilidades de transición una única vez al comienzo del programa. Tomemos como ejemplo el modelo de Ising primeros vecinos en $d = 2$, y supongamos que el algoritmo es Metropolis. Tenemos que:

$$\Delta H = 2 S_i J \sum_{\delta} S_{i+\delta}$$

En la red cuadrada, los posibles valores de $S_i \sum_{\delta} S_{i+\delta}$ son $\{-4, -2, 0, 2, 4\}$. Podemos definir entonces un arreglo $W[k]$, donde el índice k toma los valores $\{-4, -2, 0, 2, 4\}$ y

$$W[k] = \begin{cases} 1 & \text{para } k \geq 0 \\ \exp(-2\beta J k) & \text{para } k < 0 \end{cases} \quad (4.30)$$

A cada iteración podemos generar entonces un número aleatorio r con distribución uniforme en el intervalo $[0, 1]$ y compararlo con $W[S_i \sum_{\delta} S_{i+\delta}]$. Dado que el acceso a memoria es una operación computacional mucho mas rápida que un cálculo en punto flotante, esto produce una optimización considerable en velocidad, ya que esta operación es la que mas se repite en este tipo de simulaciones.