

## Capítulo 5

### Teoría de procesos estocásticos

#### 5.1 Generalidades

Los campos aleatorios nos proporcionan medidas de probabilidad sobre un dominio de definición que tenga relaciones de tipo espacial o temporal. El conjunto de posiciones donde se define el campo va a denominarse malla o rejilla y se va a denotar por el conjunto  $S$ .

Hablaremos de textura cuando el conjunto  $S$  represente al conjunto de píxeles en una estructura matricial (dos dimensiones) [8]

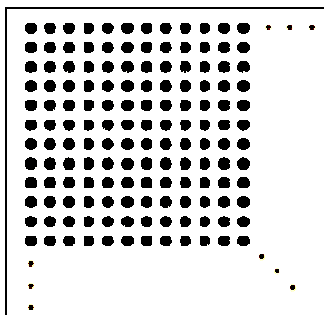


Figura 5.1.1 Textura [8]

Y hablaremos de contorno cuando el conjunto  $S$  represente a cada uno de los radios trazados desde el centro de una estructura lineal (una dimensión)

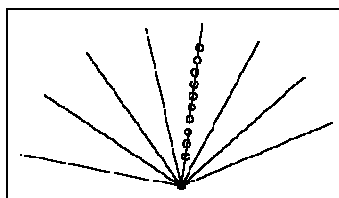


Figura 5.1.2 Contorno [8]

Para cada posición  $s \in S$  se va a definir un espacio de estados  $\Lambda_s$ . Para la textura,  $s$  es un píxel y el espacio de estados  $\Lambda_s$  para cada píxel corresponde a los niveles de gris. Para el contorno,  $s$  es un radio y el espacio de estados  $\Lambda_s$  para cada radio corresponde a cada uno de los puntos del radio por donde puede pasar el contorno. En la figura del contorno, estos corresponden a los círculos blancos a lo largo del radio. Con  $x_s \in \Lambda_s$  vamos a denotar un valor de gris del píxel o una posición para el radio correspondiente.  $\Omega = (\Lambda_s)_{s \in S}$  es el espacio de todas las configuraciones. Por lo tanto, va a ser el conjunto de todas las texturas o contornos posibles que se pueden definir en  $S$ . Por otro lado, con  $x = (x_s)_{s \in S}$  vamos a denotar una configuración en particular, es decir, una textura o un contorno. Se va a suponer que todos los espacios y configuraciones son finitos. Si todos los píxeles tienen el mismo rango de niveles de gris para el caso de la textura y el mismo rango de

posiciones para todos los radios en el caso del contorno, se dice que el espacio  $\Omega$  es homogéneo.

Una vez que dado el dominio de definición del campo, nos falta definir una medida de probabilidad o distribución de probabilidad en el espacio de configuraciones  $\Omega$ . A cada configuración (textura o contorno)  $x \in \Omega$  se le asigna una probabilidad  $\Pi(x) \geq 0$ , tal que  $\sum_{x \in \Omega} \Pi(x) = 1$

Un suceso  $E \subset \Omega$  corresponderá a un conjunto de configuraciones. Será un conjunto de texturas o un conjunto de contornos para los que se define su probabilidad simplemente como  $\Pi(E) = \sum_{x \in E} \Pi(x)$

Igualmente se pueden definir subconjuntos  $A \subset S$  del espacio de definición  $S$ . En el caso de la textura, corresponderá a una subimagen y, en el caso del contorno, a una porción es éste.  $\Omega_A = (\Lambda_s)_{s \in A}$  corresponde al espacio de configuraciones de la subimagen o porción del contorno.  $x_A = (x_s)_{s \in A}$  corresponde a una subimagen o trozo de contorno en particular. La probabilidad en este subespacio se define utilizando el concepto clásico de probabilidad marginal. Se va a denotar con  $X$ ,  $X_A$  ó  $X_s$  las variables aleatorias correspondientes a los espacios  $\Omega$ ,  $\Omega_A$  ó  $\Lambda_s$  respectivamente.  $X$  es la variable aleatoria que representa a la textura o el contorno,  $X_A$  la que representa a una subimagen de la textura o una porción del contorno y  $X_s$  la que representa a un píxel de la textura o a un radio del contorno.

Se dice que un campo definido en la rejilla  $S$  con espacio de configuraciones  $\Omega$  y media de probabilidad asociada  $\Pi$  es un campo aleatorio a estocástico si para todo  $x \in \Omega$  se cumple que  $\Pi(x) > 0$ , es decir, si la distribución o medida de probabilidad cumple la condición de positividad. Por lo tanto, si el modelo de textura es un campo aleatorio, significa que todas las imágenes o configuraciones son posibles (con mayor o menor probabilidad). En el caso del contorno, de la misma forma, significa que todos los contornos van a ser posibles.

## 5.2 Topologías

Para un campo aleatorio se puede definir un tipo de probabilidad condicionada denominada característica local del campo, definida para  $A \subset S$  como  $\Pi(X_A = x_A \mid X_{S \setminus A} = x_{S \setminus A})$

Las características locales siempre están definidas gracias a la propiedad de positividad de los campos aleatorios. En el caso de la textura, sería la probabilidad de que una subimagen de la textura tome un cierto valor, condicionada al resto de la textura. En el caso del contorno, sería la probabilidad de que una porción del contorno tome un cierto valor, condicionada al resto del mismo.

### 5.2.1 Vecindades

Las dependencias en  $S$  van a ser, en general, locales. Esto quiere decir que en una textura un píxel va a depender de los píxeles cercanos y que en un contorno el valor de éste en un radio va a depender del valor del contorno en los radios cercanos. Por esta razón vamos a definir para cada posición  $s \in S$  un conjunto  $\partial(s) \subset S$ , que corresponde a las posiciones de  $S$  de las que  $s$  depende. Los elementos de  $\partial(s)$  se

denominan vecinos de  $s$ . La colección de conjuntos  $\partial = \{\partial(s) : s \in S\}$  se denomina sistema de vecindario o vecindario de la rejilla  $S$ . Un sistema de vecindario debe cumplir dos propiedades:

- Que  $s$  no sea vecino de sí mismo  $s \notin \partial(s)$
- Que si  $s$  es vecino de  $t$ , este último lo sea del primero  $s \in \partial(t) \Leftrightarrow t \in \partial(s)$

En general, los sistemas de vecindario son homogéneos. Esto quiere decir que sabiendo los vecinos de una posición  $s$  se puede saber cuáles son los vecinos de otra posición  $t$  sin más que desplazar a  $t$  el sistema de vecinos de  $s$ . Se puede decir que son invariantes en el espacio. Los sistemas de vecinos pueden ser también isótropos. Esto quiere decir que se comportan de la misma manera en todas las direcciones (son invariantes a rotaciones en las direcciones principales de la rejilla).

En el caso del modelo de textura, sea  $s=(i, j)$  y  $t=(k, l)$  dos pixeles de la imagen tal que  $t \in \partial(s)$ ; se define el orden  $c$  del vecindario (para el caso homogéneo e isótropo) como el menor entero que cumpla  $c > (k-i)^2 + (l-j)^2$  para todos los vecinos  $t=(k, l)$  de  $s=(i, j)$ .

A continuación podemos ver vecindarios de los órdenes más comunes: 1, 2, 4, 5 y 8, para el caso de la textura.

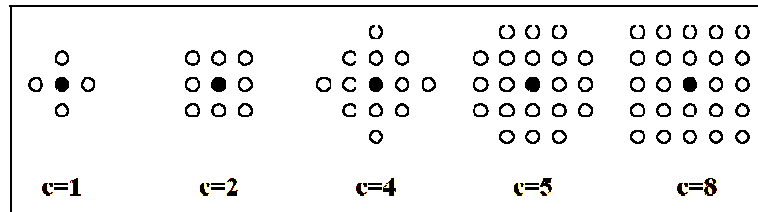


Figura 5.2.1.1 Vecindarios [8]

El punto negro representa el píxel  $s$  y el punto blanco cada uno de los pixeles  $t$  vecinos de  $s$ . Puesto que la imagen definida en el rejilla  $S$  tiene dimensiones finitas, los vecindarios de los pixeles cerca del borde de la imagen no pueden ser iguales que los vecindarios de los pixeles interiores. El concepto de homogeneidad no se puede cumplir para los pixeles de borde; sin embargo, esto siempre va a ocurrir y se va a seguir diciendo que el vecindario es homogéneo, suponiendo implícitamente el efecto de los bordes. Hay que recordar que  $\partial(s) \subset S$  para todo  $s \in S$ , como ya hemos dicho.

En el caso del modelo de contorno, si  $s$  y  $t$  son radios tales que  $t \in \partial(s)$ , se define el orden de vecindario (para el caso homogéneo e isótropo) como el menor entero que cumpla  $c \geq |s-t|$ , para todos los vecinos  $t$  de  $s$ .

A continuación podemos ver vecindarios de los órdenes más comunes: 1, 2 y 3, para el caso del contorno.

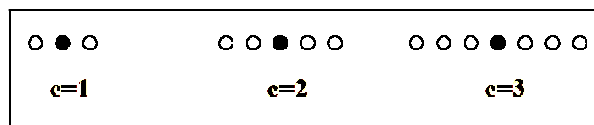


Figura 5.2.1.2 Vecindarios contorno [8]

Se ha representado cada radio mediante un punto. Puesto que el contorno es cerrado, no tenemos problemas con los bordes y podemos tener un vecindario homogéneo en sentido estricto que cumpla que  $\partial(s) \subset S$  para todo  $s \in S$ .

### 5.2.2 Cliques

Dado un sistema de vecindario  $\partial$  en  $S$ , se dice que un subconjunto  $C \subset S$  es un clique si dos elementos cualesquiera de  $C$  (diferentes entre sí) son vecinos. El conjunto de todos los cliques de  $S$  lo vamos a denotar con  $\zeta$ . En el caso de vecindarios homogéneos los cliques se van a poder clasificar en tipos, según la relación espacial entre los elementos que lo forman.

A continuación se representan los tipos de cliques definibles a partir de vecindarios de órdenes uno y dos, respectivamente, a partir del modelo de textura.

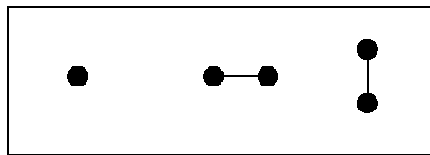


Figura 5.2.2.1 Cliques de contorno de orden 1 [8]

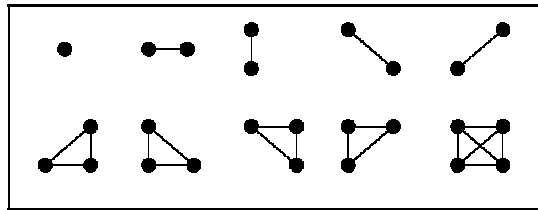


Figura 5.2.2.2 Cliques de textura de orden 2 [8]

Los cliques de dos elementos para orden cinco se pueden ver en la siguiente figura. En esta figura el círculo negro representa un elemento que pertenece al clique y el círculo blanco un elemento que no pertenece al clique. Los cliques de dos elementos se denominan cliques de pares.

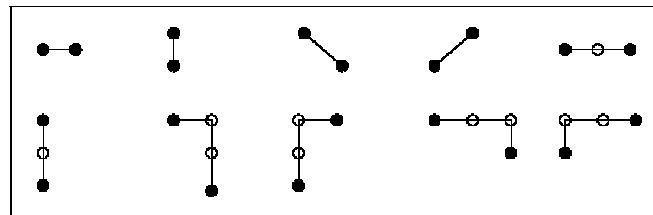


Figura 5.2.2.3 Cliques de textura de orden 5 [8]

Con respecto al modelo de contorno, en las siguientes figuras podemos observar los tipos de cliques para el vecindario de orden uno y dos respectivamente.



Figura 5.2.2.4 Cliques de contorno de orden 1 [8]

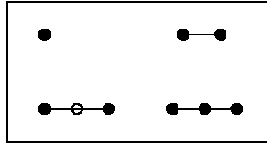


Figura 5.2.2.5 Cliques de contorno de orden 2 [8]

### 5.3 Propiedades de Markov

Se dice que un campo aleatorio es de Markov o MRF con respecto al vecindario  $\partial$  si para todo  $x \in \Omega$

$$\Pi(X_s = x_s | X_r = x_r, r \neq s) = \Pi(X_s = x_s | X_r = x_r, r \in \partial(s))$$

Se dice que un campo es de Gibbs o GRF si la medida o distribución de probabilidad se puede poner de la forma  $\Pi(x) = \frac{e^{-H(x)}}{Z}$  donde H es la función de energía que induce al campo y Z es la función de partición dada por  $Z = \sum_{z \in \Omega} e^{-H(z)}$ .

La función energética H está definida sobre el espacio de configuraciones  $\Omega$ . En la práctica, debido al tamaño de este espacio, se trabaja con el dominio de las características locales. Para poder separar las diferentes interacciones locales presentes globalmente en la función de energía, se va a definir una familia de funciones con dependencias locales  $\{U_A : A \subset S\}$  en  $\Omega$ . A esta familia se le denomina potencial U y a cada función elemento función potencial. Tenemos una función potencial  $U_A$  para cada subconjunto  $A \subset S$ . El número de funciones potenciales es igual al número de subconjuntos posibles de S, que es  $2^{|S|}$ .

Para que un potencial U induzca un GRF, ésta debe cumplir dos condiciones:  $U_\emptyset = 0$  y  $U_A(x) = U_A(y) \Leftrightarrow x_A = y_A$

La primera de estas condiciones significa que la función potencial para el conjunto vacío es la función nula y la segunda que la función potencial  $U_A$  sólo va a depender de la configuración en  $A \subset S$ , es decir, podríamos poner  $U_A = U_A(x_A)$ . En base a ellas, la energía del GRF se calcula  $H_U = \sum_{A \subset S} U_A$

Se dice que el GRF viene inducido por el potencial U, siempre que la energía del campo  $H_U$  venga dada por la ecuación anterior. La forma de estas funciones potenciales puede ser todo lo arbitraria que se desee.

En el caso de que las dependencias sean locales y dadas por un sistema de vecindario  $\partial$ , el número de funciones potenciales no nulas se reduce considerablemente. De hecho de todos los subconjuntos  $A \subset S$  sólo es necesario considerar los dados por el conjunto de cliques  $\zeta$ . Esto simplifica notablemente la definición de un GRF utilizando funciones potenciales. Dado un sistema de vecindario  $\partial$ , un potencial U es un potencial de vecindario con respecto a  $\partial$  si  $U_A = 0$ , siempre que A no sea un clique, es decir, siempre que  $A \notin \zeta$ . Por lo tanto, si C es un clique, las funciones potenciales no nulas se van a denotar por  $U_C$ . La energía del campo se calcula ahora  $H_U(x) = \sum_{C \in \zeta} U_C(x) = \sum_{C \in \zeta} U_C(x_C)$ .

Esta energía induce un GRF denominado de vecindario para el potencial U con respecto a  $\partial$ . Para vecindarios homogéneos, en general, el potencial va a ser

homogéneo. Esto, en la práctica, significa que tendremos en lugar de una función potencial diferente para cada cliqué  $C \in \zeta$ , una función potencial para cada tipo de cliqué. En general, aunque el sistema de vecindario sea isótropo, las funciones potenciales no lo van a ser.

En el caso de las funciones potenciales  $U_A$  sean nulas para  $|A| > 2$ , es decir, para conjuntos de más de dos elementos, el potencial  $U$  se denomina potencial de pares. Si éste viene referido a cierto vecindario  $\partial$  entonces además  $A$  debe ser un cliqué. En este caso, se dice que el potencial  $U$  es un potencial de pares referido a  $\partial$ .

Dado un sistema de vecindario  $\partial$  vamos a ver la relación que existe entre GRFs y MRFs. Esta relación se conoce como el teorema de Hammersley y Clifford y establece que un campo aleatorio es un MRF con respecto al vecindario  $\partial$  si y sólo si es un GRF para  $\partial$ . Es decir, son totalmente equivalentes. Dado un campo de Markov para un vecindario  $\partial$ , debe existir un potencial de vecindario  $U$  que induzca dicho campo. Un GRF para el vecindario  $\partial$  inducido por el potencial de vecindario  $U$ , es un MRF para ese vecindario.

En general, la relación entre un GRF o un MRF y un potencial no es unívoca. De hecho dos potenciales diferentes pueden inducir el mismo MRF. Para evitar esta ambigüedad, el potencial se normaliza con respecto a una configuración fijada por convenio, de forma que todas las funciones potenciales para esa configuración sean nulas. Ahora sí se puede decir que la relación entre un MRF y un potencial normalizado es unívoca. Esta configuración que anula todas las funciones potenciales se suele denominar configuración de referencia o de vacío. Puesto que la energía se determina como la suma global de funciones potenciales, la energía de configuración de referencia también es cero. Esta función energética inducida se dice que está normalizada. La relación entre un MRF y la energía normalizada también es unívoca.

En base a la ecuación que definía un MRF a partir de las características locales para cada  $s \in S$  con respecto al sistema de vecindario  $\partial$  se puede determinar a partir del potencial  $U$  como

$$\begin{aligned} \Pi(X_s = x_s \mid X_r = x_r, r \neq s) &= \Pi(X_s = x_s \mid X_r = x_r, r \in \partial(s)) \\ &= \frac{e^{-\sum_{C \in \zeta, C \ni s} U_C(x_C)}}{\sum_{z_s \in \Lambda_s} e^{-\sum_{C \in \zeta, C \ni s} U_C(z_s, x_C \setminus \{s\})}} \end{aligned}$$

Además, gracias a la definición de las funciones potenciales, se puede extender la expresión anterior para las características locales de cualquier subconjunto  $A \subset S$

$$\begin{aligned} \Pi(X_s = x_s \mid X_r = x_r, r \neq s) &= \Pi(X_s = x_s \mid X_r = x_r, r \in \partial(s)) \\ &= \frac{e^{-\sum_{C \in \zeta, C \cap A \neq \emptyset} U_C(x)}}{\sum_{z_A \in \Omega_A} e^{-\sum_{C \in \zeta, C \cap A \neq \emptyset} U_C(z_A, x_s \setminus A)}} \end{aligned}$$

Donde se define el vecindario para cada subconjunto  $A \subset S$  como  $\partial(A) = \{U \partial(s) \mid A\}$

Finalmente podemos escribir la distribución de probabilidad usando el potencial de vecindario  $U$  como  $\Pi(x) = \frac{e^{-\sum_{c \in \mathcal{C}} U_c(x)}}{Z}$  donde  $Z$  es la función de partición que en función del potencial  $U$  se puede determinar como  $Z = \sum_{z \in \Omega} e^{-\sum_{c \in \mathcal{C}} U_c(z)}$

### 5.4 Modelos de Markov

En los años 20, basado en el modelo de Ising, un nuevo tipo de proceso estocástico apareció en la teoría de la probabilidad bajo el nombre de Campos Aleatorios de Markov (MRF por sus siglas en inglés). Los MRFs se convirtieron rápidamente en una herramienta muy usada para la resolución de múltiples problemas, no solamente estadísticos, si no también de procesamiento de imágenes. [12]

#### 5.4.1 Modelo de Ising

Si consideramos una secuencia  $0, 1, 2 \dots, n$  en línea. Cada punto hay una pequeña vuelta ya sea para abajo o para arriba en cualquier momento dado.

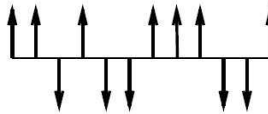


Figura 5.5.1.1 Modelo de Ising de una dimensión [12]

Ahora definimos una medida de probabilidad para el conjunto  $\Omega$  de todas las posibles combinaciones  $w = (w_0, w_1, \dots, w_n)$ . Es este contexto, cada vuelta es una

función: 
$$\delta_i(w) = \begin{cases} 1 & \text{si } w_i \text{ es arriba} \\ -1 & \text{si } w_i \text{ es abajo} \end{cases}$$

Una energía  $U(w)$  es asignada a cada configuración: 
$$U(w) = -J \sum_{i,j} \delta_i(w) \delta_j(w) - mH \sum_i \delta_i(w)$$

En la primera suma, Ising hace una simplificación asumiendo que únicamente las interacciones de puntos con una unidad de separación tienen que ser tomados en cuenta. Este término representa la energía causada por la interacción de las vueltas. La constante  $J$  es una propiedad del material. Si  $J > 0$ , las interacciones tienden a mantener las vueltas vecinas en la misma dirección (caso de atracción). Si  $J < 0$ , vueltas vecinas con orientación opuesta son favorecidas (caso de repulsión). El segundo término representa la influencia de un campo magnético externo de intensidad  $H$  y  $m > 0$  es una propiedad del material. La medida de probabilidad en  $\Omega$

está dada por  $P(w) = \frac{e^{-\frac{U(w)}{kT}}}{Z}$ , donde  $T$  es la temperatura y  $k$  es la constante universal.

La constante normalizada (o función de partición)  $Z$  se define como  $Z = \sum_{w \in \Omega} e^{-\frac{U(w)}{kT}}$ .

La función de probabilidad definida anteriormente se le conoce como distribución de Gibbs. Uno puede extender el modelo a dos dimensiones de forma natural. Las vueltas están alineadas a la cuadrícula, se representan como dos coordenadas y un punto tiene cuatro vecinos, a menos que se encuentre en el borde. En un caso de dos dimensiones, la probabilidad  $P$  es inestable, hay una transición de fase. Considerando un caso de atracción y un campo externo  $h$ , la probabilidad  $P_h$  converge a  $P$  si  $h$  llega a cero desde valores negativos, pero converge a  $P^+ \neq P^-$  si  $h$  va a cero desde valores positivos. Siempre existe una temperatura crítica  $T_c$  que debajo de esta temperatura siempre tenemos una fase de transición. La temperatura depende de los parámetros de interacción vertical  $J_1$  y horizontal  $J_2$ .

Como un ejemplo especial, tenemos el modelo de árbol de Cayley, originalmente propuesto por Bethe como una aproximación al modelo de Ising. En este caso, los puntos están en un árbol (ver figura 5.2.1.2). La raíz es llamada nivel 0. Desde la raíz, tenemos  $q$  ramas. El caso donde  $q = 1$  se simplifica a una cadena de Markov de una dimensión. La configuración en un árbol de  $n$  niveles es una asignación de etiquetas arriba y abajo para cada punto y podemos definir una función de energía para cada punto como en el modelo de Ising.

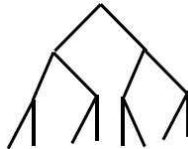


Figura 5.2.1.2 Modelo de árbol de Cayley [12]

### 5.4.2 Modelo de Potts

El modelo de Potts es una extensión del modelo de Ising para más de dos estados por punto. El problema recae en que el modelo de Ising considera un sistema de vueltas interaccionando que pueden ser paralelas o antiparalelas. Más genéricamente, podemos considerar un sistema de vueltas, cada vuelta apuntando una de las direcciones igualmente espaciadas  $q$ . Estos vectores son combinaciones lineales de  $q$  vectores unitarios apuntando a las  $q$  direcciones simétricas de un hipertetraedro en  $q-1$  dimensiones. Para ejemplos de  $q = 2, 3, 4$  vea la figura 5.2.2.1. La función de energía para el modelo de Potts se puede escribir como  $U(w) = \sum_{i,j} J(\Theta_{ij})$  donde  $J(\Theta)$  es una función de periodo  $2\pi$  y  $\Theta_{ij}$  es el ángulo entre dos vueltas vecinas en  $i$  y  $j$ . El caso de  $q = 2$  es equivalente al modelo de Ising.



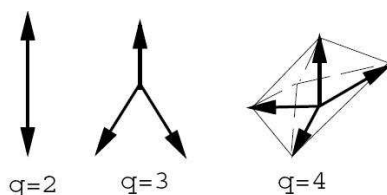


Figura 5.2.2.1 Modelo de Potts [12]

## 5.5 Optimización

### 5.5.1 Recocido simulado

En el espacio de configuraciones  $\Omega$  se tiene definido un GRF  $\Pi$  con función de energía  $H$ . Esta distribución se denomina a priori. Además, se conoce la función densidad de probabilidad  $f(y | x)$  del espacio de observaciones dada la configuración  $x$ . A esta función se le denomina verosimilitud. Utilizando el teorema de Bayes se puede determinar la distribución de las configuraciones  $x$  dada la observación  $y$ , también denominada distribución a posteriori, que viene dada por

$$P(x|y) = \frac{\Pi(x)f(y|x)}{f(y)} = \frac{\Pi(x)f(y|x)}{\sum_{z \in \Omega} \Pi(z)f(y|z)}$$

En la mayoría de los casos la distribución a posteriori de  $x$  dada la observación  $y$  se encuentra en cierto subespacio  $\hat{\Omega} \subset \Omega$  y la distribución a posteriori es otra distribución de Gibbs, si lo es la distribución a priori, definida en  $\hat{\Omega}$  y por lo tanto se podrá poner como  $P(x|y) = \frac{e^{-H(x|y)}}{Z(y)}$ , para todo  $x \in \hat{\Omega}$ , donde la función de partición depende de la observación  $y$ , y viene dada por  $Z(y) = \sum_{z \in \hat{\Omega}} e^{-H(z|y)}$ .

La función energética  $H(x|y)$  de la distribución de Gibbs a posteriori viene dada por  $H(x|y) = H(x) - \ln f(y|x) + c(y)$ , donde el primer término es la energía a priori, el segundo término es la función de verosimilitud (logarítmica) y el último término es una constante que depende de la observación  $y$ .

Una vez que hemos determinado la distribución de Gibbs a posteriori, el objetivo es, dada una observación  $y$ , determinar, de todas las configuraciones  $x \in \hat{\Omega}$ , cuál es la más probable. Es un problema de estimación MAP. Dado que el tamaño del espacio  $\hat{\Omega}$  es muy grande, determinar de forma directa la configuración que maximice globalmente la distribución de Gibbs a posteriori, o lo es que lo mismo, que minimice su función energética  $H(x|y)$ , es computacionalmente inviable. Para resolver este problema, Geman y Geman idearon un algoritmo que nos permite obtener la maximización de las funciones de Gibbs, o la minimización de las funciones energéticas, a partir de las características locales del campo; este algoritmo se conoce como Recocido Simulado o SA por sus siglas en inglés (Simulated Annealing) y es una modificación al algoritmo GS.

Vamos a suponer que tenemos un GRF  $\Pi$  con función de energía  $H$  definido en el espacio de configuraciones  $\Omega$ . Los mínimos de  $H$  corresponden con los

máximos de  $\Pi$ . Si ahora la función energética se multiplica por una constante  $\beta$ , tendríamos otro GRF con función energética  $\beta H$ . La posición de los mínimos de la energía y de los máximos del campo sigue siendo la misma. Si  $\beta$  es pequeño, los máximos del campo serán mucho más suaves, mientras que si es grande, los máximos del campo serán mucho más abruptos. Al parámetro  $\beta$ , el cual controla la suavidad de la forma del campo sin modificar la posición de sus máximos, se le denomina temperatura inversa (en muchos casos se utiliza el parámetro temperatura  $T = \frac{1}{\beta}$ ).

Si obtuviéramos una realización de nuestro campo usando el algoritmo GS para un valor de  $\beta$  elevado, puesto que el campo tiene una forma muy abrupta, la probabilidad de no estar en un máximo sería muy pequeña y por lo tanto el algoritmo GS nos proporcionaría una configuración que maximizase el GRF. El problema fundamental que deseamos evitar es que dado que la energía tiene muchos mínimos locales o el campo muchos máximos locales, nos quedemos atrapados en uno de ellos y no se evolucione hacia el máximo global. Además, cuanto mayor sea la temperatura inversa  $\beta$ , la convergencia del algoritmo GS es menor, por lo que esta solución no sería válida. Si por el otro lado, el valor de  $\beta$  toma un valor pequeño, el campo sería suave y el algoritmo GS converge rápidamente, pero tendríamos configuraciones que no maximizan el campo. Para solucionar este problema se utiliza el algoritmo GS para  $\beta$  pequeño y para cada barrido se va aumentando  $\beta$ , o lo que es lo mismo, se va disminuyendo la temperatura, hasta hacer que el campo sea muy abrupto. Así vamos a obtener una configuración que maximiza globalmente el campo.

Dado el GRF  $\Pi$  con función energética  $H$ , vamos a denotar como  $\Pi^\beta$  al GRF con función energética  $\beta H$ . Y esta viene dada por  $\Pi^\beta = \frac{e^{-\beta H(x)}}{Z^\beta}$ , donde

$$Z^\beta = \sum_{z \in \Omega} e^{-\beta H(z)}$$

Si denotamos como  $M \subset \Omega$  el conjunto de configuraciones que minimizan globalmente la función de energía  $H$ . Las configuraciones de  $M$  van a maximizar globalmente tanto el GRF original  $\Pi$  como el GRF  $\Pi^\beta$  para la temperatura inversa  $\beta$ . Según la temperatura inversa tienda a infinito el GRF  $\Pi^\beta$  converge a la distribución uniforme en los minimizadores de la energía  $M$ , es decir

$$\lim_{\beta \rightarrow \infty} \Pi^\beta(x) = \begin{cases} \frac{1}{|M|} & \text{si } x \in M \\ 0 & \text{en otro caso} \end{cases}$$

Por otro lado, si  $\beta$  tiende a cero, el GRF se hará cada vez más suave hasta degenerar en la distribución uniforme en  $\Omega$ , es decir  $\lim_{\beta \rightarrow 0} \Pi^\beta(x) = \frac{1}{|\Omega|}$ . En este caso

las posiciones de  $S$  serían independientes.

Partiendo de una configuración arbitraria  $v$  y fijando  $\beta$  en el GRF, el algoritmo GS nos proporciona una realización de  $\Pi^\beta$ . Además, según aumentamos  $\beta$  el GRF tiende a la distribución uniforme en los minimizadores  $M$  de la energía. Para poder hacer ambas cosas manteniendo la convergencia, la condición fundamental está en la

forma en la cuál aumentamos el valor de  $\beta$ , o lo que es lo mismo en la forma en la que enfriamos el campo.

### 5.5.2 Esquemas de decremento de temperatura

Llamemos  $n$  al índice del número de barrido, dado que en cada barrido vamos aumentando a  $\beta$ , disminuyendo la temperatura, denotaremos como  $\beta(n)$  para  $n \geq 0$  a la secuencia de temperaturas inversas del algoritmo SA. Esta secuencia recibe el nombre de plan de enfriado y es la clave para que el algoritmo SA converja a la distribución uniforme  $M$ . [9]

Para que el método converja, es necesario que se satisfaga la siguiente ecuación del plan de enfriado  $\beta(n) \leq \frac{1}{\Delta|S|} \ln(n)$ , donde  $\Delta$  es la oscilación local máxima de la función energética  $H$  del campo  $\Pi$  dada por  $\Delta = \max_{s \in S} \{\delta_s\}$  donde  $\delta_s$  es la oscilación local de la energía  $H$  en la posición  $s$  dada por  $\delta_s = \sup_{x, y \in \Omega} \{H(x) - H(y) : x_{s\{s\}} = y_{s\{s\}}\}$

En general el plan de enfriado que se sigue es  $\beta = c \ln n$  y la constante  $c$  se determina de tal forma que  $\beta$  tenga un valor razonable en un número viable de barridos.

El problema del plan de enfriado anterior es que resulta ser muy lento, así como en general todos los planes logarítmicos; para obtener un plan de enfriamiento más rápido se utilizan planes como  $\beta(n) = n$ ,  $\beta(n) = \alpha^n$  e incluso  $\beta(n) = \infty$ . Estos planes nos ofrecen rapidez, con el defecto de que dan resultados semióptimos, debido a que es posible que la configuración obtenida sea solamente un mínimo local de energía.