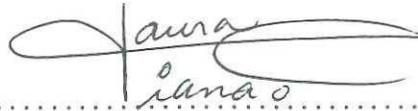




RESUMEN de la Tesis de CITLALI MARTÍNEZ SISNIEGA presentada como requisito para la obtención del grado de Maestro en Ciencias en Física Aplicada con opción en Física de Materiales. Ensenada, Baja California, México, junio de 1994..

## TAMAÑO DE LAS CUENCAS DE ATRACCIÓN DE UNA RED NEURONAL TIPO HOPFIELD Y DINÁMICA DE DICHA RED



Resumen aprobado por: .....

Dra. Laura Viana Castrillón  
Director de Tesis

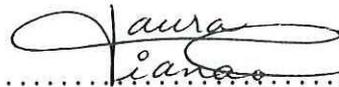
Con el propósito de recuperar la información almacenada en una red neuronal, se evalúa el tamaño de las cuencas de atracción en dos diferentes redes tipo Hopfield. (a) La primera de ellas consiste de una red compuesta de  $N$  neuronas con dos estados posibles modelados por los espines de Ising. En dicha red se almacenan  $p$  patrones de acuerdo con la regla de Hebb modificada y se estudia el tamaño relativo de las cuencas de atracción como función del peso asignado a cada una de las memorias y como función del parámetro de ruido  $T$ . Este estudio se lleva a cabo iterando numéricamente las ecuaciones de flujo o traslapes entre un estado almacenado y el estado dinámico del sistema. Se encuentra que un cierto tipo de memorias (memorias espurias) aparecen abruptamente y que dependen de los valores específicos de los pesos asignados a cada memoria. De igual manera se encuentra que el porcentaje de recuperación de las memorias espurias disminuye considerablemente con la temperatura. (b) La segunda red consiste de una red neuronal compuesta por dos subredes tipo Hopfield, las cuales están conectadas unidireccionalmente, es decir, cada neurona de la primera subred envía información exclusivamente a una neurona de la segunda subred. En este sistema se almacena información compuesta que consta de una parte para cada una de las subredes. Se estudia la recuperabilidad de la información de la segunda subred, como función del estado inicial de la primera, desde los puntos de vista micro y macroscópico.

## BASINS OF ATTRACTION OF A HOPFIELD NEURAL NETWORK AND IT DYNAMICS.

The basins of attraction associated to the stored information, in a Hopfield neural network, are studied for the case of a network composed of long range Ising spins. A finite number of  $p$  patterns are stored according to the modified Hebb rule, each one having a different weight,  $w_\mu$  ( $\mu = 1, \dots, p$ ), and the size of the basins of attraction is studied by varying, in a systematical way,  $w_\mu$  and the noise parameter  $T$ . This is done by iterating numerically the flux equation for the overlaps between the stored patterns and the dynamical state of the system. We find that, in  $w_\mu$  space, there are two qualitatively different kinds of behaviour, divided by a line of the form  $w_2 + w_3 = c(T)$ , with  $c(0) = 1$ .

The dynamical behaviour of a neural network composed of two Hopfield subnetworks interconnected unidirectionally is also studied by means of numerical simulations. The connections are such that each one of the neurons on the first subnet sends information to, exclusively, one of the neurons on the second. A pattern, consisting of two subpatterns, is stored in the subnets using a Hebb-like rule, and the recovery of a particular subpattern in the second subnet is studied as a function of the initial state of the first one. Our microscopical approach was complemented by a macroscopic study, in which the flux equations for the overlaps were derived.

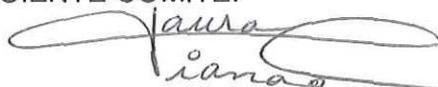
Abstract aproved by: .....



Dra. Laura Viana Castrillón  
Thesis Director

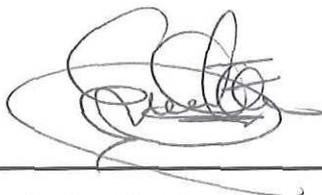
TESIS DEFENDIDA POR: **CITLALI MARTINEZ SISNIEGA**

Y APROBADA POR EL SIGUIENTE COMITE:



---

**DRA. LAURA CECILIA VIANA CASTRILLON.-** Director del Comité



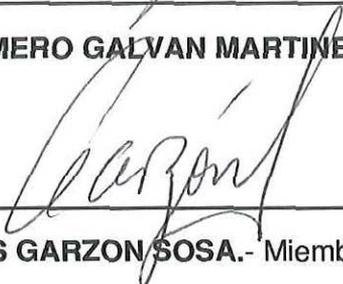
---

**DR. ERNESTO COTA ARAIZA.-** Miembro del Comité



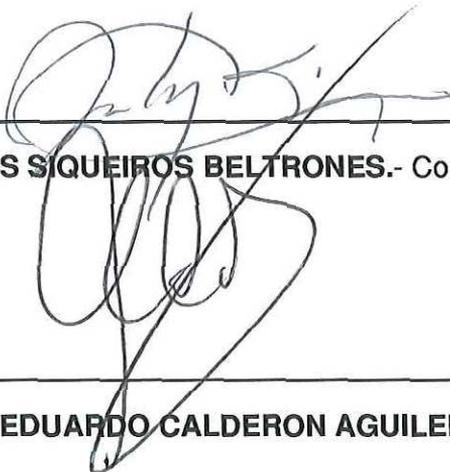
---

**DR. DONALD HOMERO GALVAN MARTINEZ.-** Miembro del Comité



---

**DR. IGNACIO LUIS GARZON SOSA.-** Miembro del Comité



---

**DR. JESUS SIQUEIROS BELTRONES.-** Coordinador del Progrado en Física de Materiales

**DR. LUIS EDUARDO CALDERON AGUILERA.-** Director de Estudios de Posgrado

21 DE JUNIO DE 1994

CENTRO DE INVESTIGACION CIENTIFICA Y EDUCACION  
SUPERIOR DE ENSENADA



DIVISIÓN DE FÍSICA APLICADA  
FÍSICA DE MATERIALES

**TAMAÑO DE LAS CUENCAS DE ATRACCIÓN DE UNA  
RED NEURONAL TIPO HOPFIELD Y DINÁMICA DE  
DICHA RED**

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el  
grado de **Maestro en Ciencias** presenta:

**CITLALI MARTÍNEZ SISNIEGA**

Ensenada, Baja California, México, junio de 1994.

A JUAN

## AGRADECIMIENTOS

Al presentarse el producto final de una etapa de formación académica, es imposible agradecer por su nombre a todas las personas que de una manera u otra contribuyeron en ella. Quisiera agradecer de manera muy especial la participación de algunos de ellos.

A la Dra. Laura Viana por la dirección de esta tesis.

A los Drs. Ernesto Cota, Ignacio Garzón, Héctor Echavarría por su participación en el comité de tesis y de manera muy especial al Dr. Homero Galván quien además de haber formado parte de dicho comité me brindó su confianza y amistad.

Al Consejo Nacional de Ciencia y Tecnología por su apoyo económico y al centro de cómputo de la Universidad Nacional Autónoma de México por permitirme hacer uso de la supercomputadora CRAY.

A Armando Reyes, Alvaro Posada, Sergio Carnalla, Arturo Gamietea, Noboru Takeuchi, Arnulfo Castellanos, Francisco Villa y Felipe Castellón por su amistad, compañerismo y múltiples comentarios.

A mi amigo Alberto Duarte, por habernos aguantado el uno al otro durante todo este período. Siempre recibí su apoyo. A Leonardo Morales.

A mi familia Juan, Lucía y Gerardo por el amor y apoyo que me han brindado siempre.

Y finalmente, de manera que no admite descripción, a Eugenio, a Andrés e Itzel, nuestros hijos, por enseñarme lo más valioso que existe.

# CONTENIDO

	Página
<b>I. INTRODUCCION</b>	<b>1</b>
I. 1. Redes neuronales . . . . .	3
I. 2. Modelo de Hopfield . . . . .	7
I. 2. 1. Memoria asociativa . . . . .	10
I. 2. 2. Regla de Hebb . . . . .	11
I. 3. Estructura del trabajo . . . . .	14
<b>II. RED NEURONAL TIPO HOPFIELD</b>	<b>15</b>
II. 1. Introducción . . . . .	15
II. 2. Introducción al problema . . . . .	19
II. 2. 1. Antecedente analítico . . . . .	20
II. 3. Red tipo Hopfield: Dependencia en $T$ . . . . .	24
II. 3. 1. Antecedentes y motivación . . . . .	25
II. 3. 2. Método . . . . .	27
II. 3. 3. Resultados . . . . .	28
II. 3. 4. Conclusiones . . . . .	39
<b>III. REDES ACOPLADAS</b>	<b>40</b>
III. 1. Introducción . . . . .	40

III. 2. Modelo . . . . .	42
III. 2. 1. Estudio macroscópico de las redes acopladas. . . . .	44
III. 2. 1. 1. Solución de las ecuaciones de flujo (método iterativo). . . . .	52
III. 2. 1. 2. Conclusiones . . . . .	53
III. 2. 2. Estudio microscópico de las redes acopladas. . . . .	54
III. 2. 2. 1. Simulación numérica. . . . .	56
III. 2. 2. 2. Resultados y discusión. . . . .	58
 <b>LITERATURA CITADA</b>	 <b>66</b>
 <b>A. ECUACIONES DE FLUJO</b>	 <b>68</b>
 <b>B. AUTO PROMEDIO</b>	 <b>74</b>

## LISTA DE FIGURAS

Figura	Página
1 Dibujo esquemático de una neurona común. . . . .	4
2 Correspondencia uno a uno entre los espines de Ising y la red de Mc Culloch-Pitts. Un espín apuntando hacia arriba (abajo) es identificado por una neurona activa (inactiva). . . . .	7
3 Representación esquemática del espacio de configuraciones con tres estados memorizados. . . . .	9
4 Porcentaje de recuperación de las memorias espurias como función de $\alpha$ para valores de $p = 3$ hasta $p = 7$ , tanto para la regla de Hebb como para la regla de Hebb modificada. Ambos casos están indicados por (pA) y (pB) respectivamente. . . . .	25
5 Representación del espacio de las $w$ 's para $p = 3$ a $T = 0$ . El comportamiento de esta red en el espacio de las $w$ 's esta separado en dos regiones por la línea $w_2 + w_3 = w_1 = 1$ . Las líneas de contorno en la parte clara indican el porcentaje de microestados que conducen hacia las cuencas de atracción de las memorias puras. . . . .	26
6 Porcentaje de recuperación de las memorias espurias en el espacio de las $w$ 's para $p = 3$ , $\sigma = 1 \times 10^{-5}$ y desde $T = 0$ hasta $T = .9$ . En esta gráfica se pueden visualizar dos comportamientos divididos por la línea $w_2 + w_3 = c(T)$ . A medida que la temperatura aumenta la línea $w_2 + w_3 = c(T)$ se desliza y el porcentaje de las memorias espurias disminuye. . . . .	30
7 Relación entre $C(T)$ y $T$ para distintos valores de $\sigma$ . . . . .	33
8 Porcentaje de recuperación de las memorias espurias como función de $\ln \sigma$ para el caso de la regla de Hebb $w_\mu = 1$ y para $T = 0$ hasta $T = .3$ . . . . .	33

- 9 Porcentaje de recuperación de las memorias espurias en el espacio de las  $w$ 's para  $p = 3$ ,  $\sigma = 1 \times 10^{-1}$  y desde  $T = 0$  hasta  $T = .9$ . En esta gráfica se pueden visualizar dos comportamientos divididos por la línea  $w_2 + w_3 = c(T)$ . A medida que la temperatura aumenta la línea  $w_2 + w_3 = c(T)$  se desplaza y el porcentaje de las memorias espurias disminuye. . . . . 34
- 10 Porcentaje de recuperación de las memorias espurias en el espacio de las  $w$ 's para  $p = 3$ ,  $\sigma = 1 \times 10^{-5}$  y  $T = 0$  (negro),  $T = .2$  (azul),  $T = .3$  (rosa),  $T = .4$  (verde) y  $T = .5$  (rojo). . . . . 37
- 11 Porcentaje de recuperación de las memorias espurias como función de  $T$  para la regla de Hebb modificada con  $p = 3$  y  $w_\mu = .7$ . La figura (a) considera el caso de  $\sigma = 1 \times 10^{-5}$  mientras que la figura (b) el caso con  $\sigma = 1 \times 10^{-1}$ . En ambas, en el extremo superior derecho se muestra una amplificación de la zona comprendida entre  $T = 0$  hasta  $T = 0.1$ . . . . . 38
- 12 Representación esquemática de una red neuronal compuesta por dos subredes. . . . . 42
- 13 Porcentaje de recuperación de la segunda subred como función del traslape final  $m_f$  sobre un conjunto de 1000 redes con  $N = 100$  y distintos valores de  $p$ . En el extremo superior derecho se indica el porcentaje total de veces que se recuperó un estado con traslape macroscópico  $m_f$  con valores distribuidos de acuerdo con las figuras. . . . . 61
- 14 Porcentaje de recuperación como función del número de patrones almacenados  $p$  para la segunda subred con  $N = 100$  elementos. La línea sólida representa valores impares de  $p$  y la línea punteada valores pares. Desviación estandar del 10%. . . . . 64
- 15 Valor promedio sobre 1000 muestras del traslape final  $m_f$  como función de  $p$  para  $N = 100$ . La línea sólida representa valores impares de memorias almacenadas ( $p = 3, 5, 7, 9, 11, 13, 15$ ) y la línea punteada representa valores pares de  $p$  ( $p = 2, 4, 6, 8, 10, 12, 14$ ). . . . . 64
- 16 Relación entre el traslape inicial  $m_0$  y el traslape final  $m_f$  para la segunda subred ( $N = 64$  y  $p = 4$ ) cuando la primera subred se encuentra en un estado estable. La gráfica (a) muestra casos típicos, sin embargo existen pocos casos como los mostrados en las gráficas (b), (c) y (d) en los que el traslape final depende del traslape inicial. Este comportamiento sólo tiene dependencia en  $p$  par. . . . . 65

# TAMAÑO DE LAS CUENCAS DE ATRACCIÓN DE UNA RED NEURONAL TIPO HOPFIELD Y DINÁMICA DE DICHA RED

## I. INTRODUCCION

Recientemente hemos sido testigos de la creación de una nueva rama de la ciencia, la cual quizá sea centro de atracción para un gran número de científicos del siglo XXI. A ésta se le conoce como la física de los sistemas complejos (SC). No se puede decir que el estudio de los sistemas complejos forme parte de alguna de las disciplinas ya conocidas. Esta nueva rama surgió como resultado de la convergencia de varias ramas tradicionales de la ciencia, tales como la biología, las ciencias computacionales, la física y las matemáticas. También, como consecuencia de nuevas ideas, nuevas preguntas y técnicas. Para su entendimiento sí se requiere de conocimientos y técnicas provenientes de dichas áreas.

Es difícil llegar a una definición satisfactoria, y de aceptación general, de sistema complejo basándonos solamente en su composición. Para el propósito de nuestro trabajo definiremos sistema complejo en términos del tipo de comportamiento que presenta, pues existen algunos sistemas muy simples que exhiben un comportamiento sumamente complicado. En nuestro caso, consideremos a un SC como aquel que está compuesto por un número muy grande de elementos y que muestra diversos comportamientos colectivos teniendo un número considerable o infinito de soluciones o evoluciones posibles.

Ahora bien, cuando el propósito principal de estudio de estos sistemas es encontrar la mejor solución dentro de todas las posibles, se dice que nos hemos topado con un problema de optimización [Martínez,1989]. El problema clásico de optimización es

el problema del “agente viajero” (traveling salesman problem), que presentaremos a continuación: supongamos que un agente viajero debe visitar  $N$  ciudades sin pasar por ninguna de ellas en más de una ocasión, hasta llegar finalmente al lugar de partida. ¿Cuál es la secuencia que debe seguir el viajero de tal manera que su recorrido sea el más corto posible? De la misma manera, se pueden mencionar una gran cantidad de problemas de optimización que tienen aplicaciones tanto tecnológicas como económicas. Sin embargo, quizás uno de los problemas más fascinantes es el de las redes neuronales, en el que un sistema compuesto por muchos elementos simples, interconectados, es capaz de funcionar como una memoria asociativa (a la que nos referiremos más adelante) al almacenar, reconocer y recuperar información.

Las redes neuronales (RN) son sistemas que están formados de un número muy grande de elementos simples operando en paralelo; a estos elementos se les llama “neuronas” por analogía con las neuronas biológicas. El funcionamiento de estas redes es determinado por la topología de las conexiones entre las neuronas, por la “fuerza” de dichas conexiones y por el tipo de proceso que desarrollan los elementos. Aunque el proceso llevado a cabo por cada uno de los elementos es muy sencillo, el sistema completo es capaz de realizar funciones muy complicadas tales como: almacenaje de información por contenido, clasificación, optimización, generalización, etcetera. Todas estas características constituyen lo que llamamos memoria asociativa. También a las redes neuronales se les puede ver desde diferentes puntos de vista. Por un lado, constituyen modelos idealizados para esclarecer los principios por los cuales el cerebro de un sistema biológico funciona, y por otro, estos modelos pueden ser implementados como circuitos electrónicos con el fin de ejecutar algún procesamiento específico de datos sin tomar en cuenta que, originalmente, se tomó como referencia a los sistemas biológicos, i.e., en principio se podrían construir máquinas capaces de resolver problemas que para una máquina convencional no son tarea sencilla.

## I. 1 Redes neuronales

Como se mencionó anteriormente, las RN constituyen modelos idealizados y útiles para esclarecer los principios por los cuales un sistema biológico funciona. Bajo este aspecto, el cerebro ha sido objeto de estudio de diversas disciplinas como son, entre otras, la anatomía, que se refiere a cómo están distribuidos los diferentes elementos del cerebro; la fisiología, que estudia cómo funcionan las partes que involucra el cerebro y la manera en la que éstas interactúan; la psiquiatría, que estudia y trata las desviaciones de la conducta de individuos con respecto a la conducta promedio, así como la psicología que analiza el modo de razonar y pensar. Sin embargo, a pesar de que se tiene gran conocimiento del cerebro, no se sabe a ciencia cierta el proceso por el cual se almacena y recupera la información. Es aquí donde recientemente se ha visto que los modelos utilizados por la Física Estadística pueden jugar un papel muy significativo, ya que estos ofrecen un procedimiento para estudiar, desde un punto de vista estadístico, el comportamiento colectivo de las neuronas para almacenar y procesar la información. Como comportamiento colectivo entendemos que debe haber un mecanismo en el que no es relevante la contribución individual específica de cada una de las neuronas; en esta medida, es posible hacer un tratamiento estadístico de la actividad neuronal de un conjunto muy grande de neuronas.

El cerebro está compuesto de células nerviosas o neuronas altamente especializadas y cuyo número es del orden de  $10^{10}$ . Una neurona típica consiste de un cuerpo, del cual emanan el axón y las dendritas. Las señales de entrada de información, provenientes de un número muy grande de otras neuronas, son recibidas a través de las dendritas; estas señales son sumadas o promediadas en el cuerpo de la célula de tal manera que si la intensidad del estímulo recibido es mayor que el umbral característico de dicha neurona, ésta generará y emitirá una señal de respuesta; dicha respuesta, a su vez, será enviada a

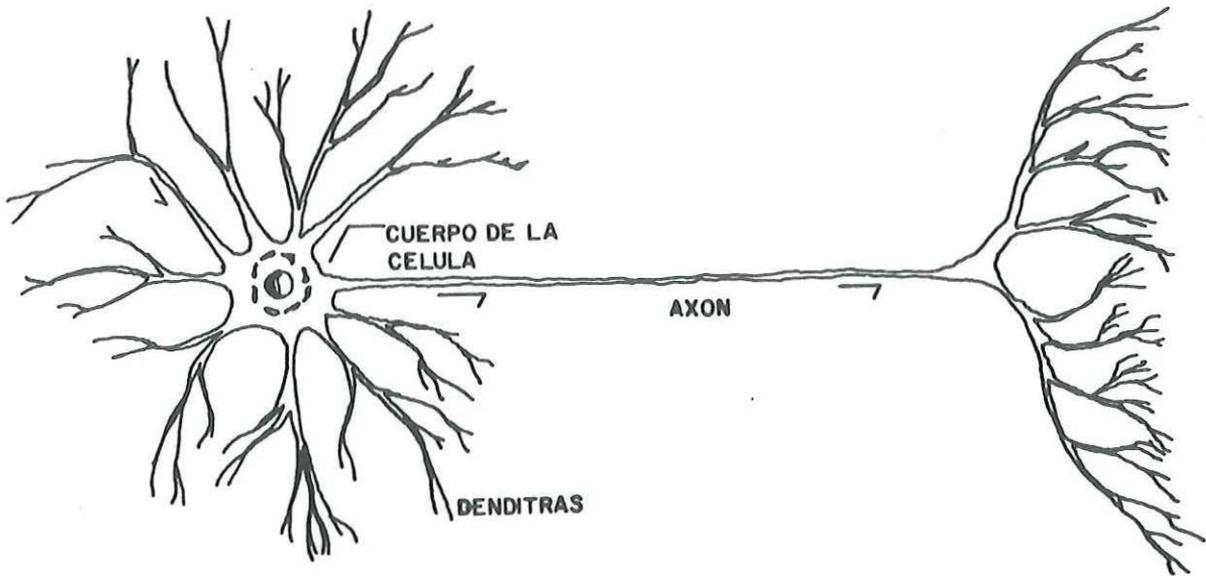


Figura 1: Dibujo esquemático de una neurona común.

través de una fibra larga llamada axón, que eventualmente se ramifica terminando en las juntas sinápticas de otras neuronas (figura 1).

En el cerebro existen dos tipos de señales: eléctricas y químicas. La señal transportada por el axón es un impulso eléctrico, pero la señal es transmitida de célula a célula por moléculas de sustancias neurotransmisoras que fluyen a través de las uniones sinápticas. Una sinápsis será inhibitoria o excitadora, dependiendo de la presencia de diferentes tipos de receptores y transmisores en la membrana celular. Por otro lado, se ha comprobado experimentalmente en organismos simples conocidos como *aplysia* [Kandel, 1979], que tanto el aprendizaje como la información genética son determinantes de la eficiencia sináptica.

El interés por modelar el comportamiento colectivo de un conjunto de neuronas interconectadas data de la primera mitad del siglo XX. En el artículo publicado en 1943 por Warren McCulloch y Walter Pitts, se establecen las bases matemáticas para modelar una red de neuronas. En particular, ellos propusieron una teoría general (lejos de la realidad biológica) para el procesamiento de información, basados en la suposición de que el código de las respuestas de las neuronas es binario. Como respuestas binarias

entendemos: que cada elemento de la red  $i = 1, \dots, N$  puede tomar los valores de salida  $\sigma_i = 0, 1$ , donde  $\sigma_i = 0$  representa el estado de reposo y  $\sigma_i = 1$  representa el estado de actividad de la  $i$ -ésima neurona. Ahora bien, esta respuesta sucede cuando cada neurona ajusta su estado de acuerdo con la suma de señales que ésta recibe del resto de las neuronas, es decir, que la suma sea mayor o menor que un cierto umbral  $\mu_i$ . Así, la evolución de la red está determinada por:

$$\sigma_i(t+1) = \theta(h_i(t) - \mu_i) \quad (1)$$

donde  $\theta(x)$  es la función escalón o la función Heaviside dada por:

$$\theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otro caso,} \end{cases} \quad (2)$$

y  $h_i$  representa la suma de señales del resto de las neuronas sobre la neurona  $i$ , y está dada por:

$$h_i = \sum_j J_{ij} \sigma_j(t). \quad (3)$$

En esta expresión  $J_{ij}$  representa la sinapsis que conecta a la neurona  $i$  con la neurona  $j$ , y ésta puede ser de tipo excitadora ( $J_{ij} > 0$ ), inhibidora ( $J_{ij} < 0$ ) o es cero si no hay sinápsis entre las neuronas. El parámetro  $\mu_i$  en la ecuación (1) representa el umbral de la neurona  $i$ ; esto es, la suma de las señales de entrada debe alcanzar o exceder dicho valor para que la neurona se active. Finalmente, el tiempo  $t$  se considera discreto.

Otro avance muy significativo fue cuando William Little (1974), mostró la similitud entre la red neuronal propuesta por McCulloch y Pitts y un sistema físico constituido por elementos con momentos magnéticos diferentes de cero denominados espines (los cuales se caracterizan por su magnitud y la dirección en la que están orientados) (figura 2). La interacción entre los espines es generada por diversos mecanismos, por lo cual sus efectos pueden ser muy variados, sin embargo, uno de los efectos más importantes que originan dichas interacciones es que la orientación final de los espines será aquella en donde la energía libre del sistema sea mínima. Esto significa que cada par de espines tratará localmente de orientarse de la forma más natural posible, es decir, de tal manera que la energía de ese par sea mínima. Como consecuencia de este resultado tenemos que hay una gran diversidad de comportamientos en los diferentes materiales magnéticos que existen en la naturaleza dependiendo de la interacción entre cada par de espines (ferromagnéticos, diamagnéticos, paramagnéticos, vidrios de espín, etcetera) [Viana, 1988]. Existen varios modelos para describir los espines magnéticos, uno de ellos es el llamado *modelo de Ising*, según el cual el espín  $S_i$  puede tomar dos orientaciones diferentes, hacia arriba o hacia abajo, denotadas por  $S_i = 1$  ó  $S_i = -1$  respectivamente. La analogía con la red neuronal se lleva a cabo cuando se identifica cada espín con una neurona, asociándole el valor positivo  $S_i = 1$  si la neurona  $\sigma_i = 1$  está en estado activo, y el valor negativo  $S_i = -1$  si la neurona  $\sigma_i = 0$  está en estado de reposo.

Posteriormente, Hopfield (1982) estudió cómo la red neuronal o sistema de espines, propuesta por Little, puede almacenar y recuperar la información. El modelo de Little y el de Hopfield difieren en la forma en que el estado del sistema cambia. Es decir, en el modelo de Little todas las neuronas (espines) cambian sincronizadamente de acuerdo con la ecuación (3), mientras que en el modelo de Hopfield las neuronas se actualizan secuencialmente (asincrónicamente), ya sea en un orden fijo o al azar.

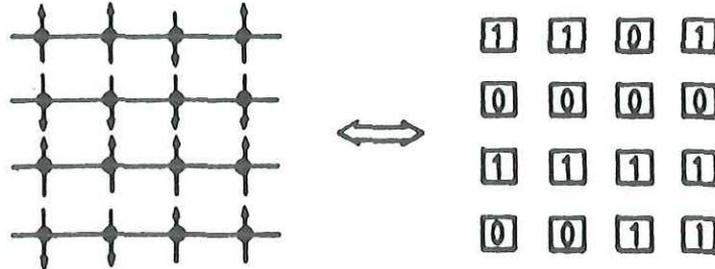


Figura 2: Correspondencia uno a uno entre los espines de Ising y la red de McCulloch-Pitts. Un espín apuntando hacia arriba (abajo) es identificado por una neurona activa (inactiva).

## I. 2 Modelo de Hopfield

Una vez hecha la analogía entre los sistemas magnéticos y una red de neuronas, sin olvidar que la interacción (sinápsis) entre las neuronas puede ser de tipo excitadora ( $J_{ij} > 0$ ) o inhibidora ( $J_{ij} < 0$ ), Hopfield mostró que es posible establecer una analogía matemática, con la cual se expresarían los estados del sistema. Dicha analogía consistió en proponer para la red neuronal el hamiltoniano de interacción de Heisenberg:

$$\mathcal{H} = -\frac{1}{2} \sum_{ij} J_{ij} S_i S_j - \sum_i h_i S_i, \quad (4)$$

en donde el primer término incluye la suma sobre los  $N$  elementos de interacción  $J_{ij}$  entre las neuronas  $S_i$  y  $S_j$  modeladas por los espines de Ising ( $S_i = \pm 1$ ), y el segundo término,

el acoplamiento de las neuronas con una señal de entrada  $h_i$  que varía de lugar a lugar, como sería por ejemplo, el sistema sensorial. Tal función de energía existe únicamente si las conexiones sinápticas son simétricas, i.e.,  $J_{ij} = J_{ji}$ . En una red de neuronas reales esto no tiene sentido.

Ahora bien, para entender un poco más acerca del papel que juega el ya mencionado hamiltoniano, conviene dar un enfoque mecánico-estadístico del problema, esto es así ya que nuestro sistema de interés posee un número muy grande de elementos y las propiedades que se calculan provienen de promedios sobre ensambles.

La mecánica estadística data de principios de siglo con las contribuciones independientes de Einstein y Gibbs (Einstein 1902, 1903, 1904, Gibbs 1902). Ambos desarrollos llegaron a los mismos resultados, pues son generalizaciones de la teoría cinética de Boltzmann, pero por caminos diferentes (Boltzmann 1896, 1898). En particular el planteamiento de Gibbs introduce el concepto de ensemble de sistemas. Este consiste de una colección de un número muy grande de sistemas que son una réplica macroscópica del sistema en estudio, sin embargo, a nivel microscópico no son necesariamente iguales entre ellos. Por otro lado, el ensemble debe cumplir con el principio de *igual probabilidad a priori*, esto es que todos y cada uno de los estados microscópicos estén igualmente representados en el ensemble.

Consideremos un sistema aislado de  $N$  componentes contenido en un volumen  $V$  y con energía constante. El estado de este sistema quedara definido por  $x = (\vec{q}_1, \dots, \vec{q}_N, \vec{p}_1, \dots, \vec{p}_N)$  donde  $(\vec{q}_i, \dots, \vec{p}_i)$  son las coordenadas y momentos generalizados de la  $i$ -ésima componente. Ahora bien, la solución de las ecuaciones de movimiento, que consideran las coordenadas generalizadas, determinan la trayectoria del sistema. Dicha trayectoria se estudia en el espacio fase, el cual es un espacio de  $2N$  dimensiones con ejes ortogonales, uno para cada coordenada  $q_i$  y otro para cada momento  $p_i$ . De aquí que cada estado microscópico de un ensemble está representado por un punto en el espacio fase. La mecánica estadística para un sistema como el antes mencionado considera la hipótesis

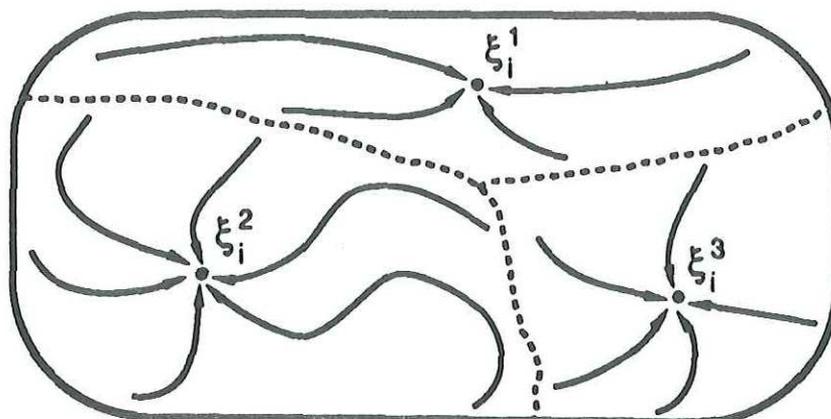


Figura 3: Representación esquemática del espacio de configuraciones con tres estados memorizados.

ergódica, que nos dice que es equivalente dejar evolucionar a un sistema en el tiempo y promediar sobre esta evolución, a hacer  $n$  copias macroscópicamente iguales que reflejan todos los microestados posibles del sistema y promediar sobre éstas.

En el caso de los sistemas magnéticos, se sabe que existen ciertas configuraciones o estados que reducen la energía del sistema. Así, suponemos que el sistema en contacto térmico con el exterior va a evolucionar o cambiar su estado hasta que llegue a un estado de energía libre mínima. Más específicamente, en el caso de los vidrios de espín, en donde existe una “competencia” entre las interacciones  $J_{ij}$  tenemos que no hay una, sino muchas configuraciones de energía libre mínima del sistema. Lo mismo sucede en las redes neuronales, donde dichos mínimos actúan como atractores dinámicos del sistema, esto es, dado un estado inicial del sistema, éste se relaja llevando una trayectoria “cuesta abajo” hasta alcanzar un estado estacionario del espacio de las energías. Como consecuencia de esto, los mínimos pueden ser considerados como estados memorizados de la red (figura 3), ya que actúan como atractores de la dinámica de la red. Por esto entendemos que es posible almacenar información de manera “asociativa”.

## I. 2. 1 Memoria asociativa

Como memoria asociativa entendemos la posibilidad de almacenar y recuperar la información a partir de una parte de esta misma información almacenada previamente.

En una computadora tradicional, la recuperación de la información almacenada requiere de la dirección o localización precisa del lugar en que ésta se encuentra. Es claro que el cerebro no está organizado de esta manera; si nosotros tratamos de recordar el nombre de cierta persona, no sirve de nada saber que esta persona fue la número 797 que conocimos durante nuestra vida, sin embargo, sí es útil recordar varios hechos como: color de pelo, altura, trabajo, familia, etcetera. Estas características de alguna manera son combinadas en nuestra mente para formar la memoria del individuo. Por otro lado, a veces se da el caso de que estos datos no son suficientes para relacionar a quien le pertenecen, entonces sería necesario otro tipo de fuente, como por ejemplo, escuchar la voz, la cual quizás es una característica más fuerte para recordar a la persona.

De esta manera, podemos definir la memoria asociativa como sigue: Si almacenamos en la red información codificada en términos de neuronas activas e inactivas a las cuales llamaremos patrones (atractores dinámicos de la red) y los denotamos por  $\{\xi_i^\mu\}$  con  $\mu = 1, \dots, p$  y  $i = 1, \dots, N$ , de tal forma que cuando le presentamos a la red un estado inicial  $\vec{S}_0 = (S_1^0, \dots, S_N^0)$ , ésta va a responder produciendo el patrón almacenado que más se asemeje a él. El traslape nos proporciona una medida de qué tan cerca o lejos nos encontramos de un estado y se define por:

$$q_\mu = \frac{1}{N} \sum_i^N \xi_i^\mu S_i^0. \quad (5)$$

siendo  $\xi_i^\mu$  el patrón almacenado en la red y  $S_i^0$  el estado dinámico de la misma. También, este traslape es equivalente al parámetro de orden que representa el espacio de configuraciones de un sistema magnético.

Ahora bien, es necesario cuantificar de alguna forma la cantidad de ruido permitida para que sea aceptable la información recuperada. Una manera de “medir” el parecido que existe entre el patrón almacenado  $\alpha$  (mínimo local de la energía y el estado inicial  $\beta$  de la red (de entrada) es por medio de la distancia entre el estado inicial  $\vec{S}_\beta$  y el patrón modelo  $\vec{\xi}_\alpha$ . Dicha distancia, que cuantifica la fracción de errores, está dada por:

$$d_{\alpha\beta} = \frac{(1 - q_{\alpha\beta})}{2} \quad (6)$$

en donde  $q_{\alpha\beta}$  representa el traslape entre los estados  $\alpha$  y  $\beta$ .

## I. 2. 2 Regla de Hebb

Hasta el momento no hemos dicho nada acerca de la forma que toman las interacciones  $J_{ij}$ . La manera explícita de escoger las interacciones es una de las características que distingue a cada modelo, pues esto equivale a especificar la posición de los mínimos en el espacio de las energías.

El modelo de Hopfield está basado en la regla de Hebb (1949), quien propuso una hipótesis de cómo las conexiones sinápticas en el cerebro cambian en respuesta al aprendizaje. Mediante el uso de esta regla es posible almacenar información en la red y está dada por:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu}. \quad (7)$$

Aquí,  $p$  es el número de patrones  $\vec{\xi}_i^{\mu}$  (conjunto  $\xi_i^{\mu} = \pm 1$ ) por almacenar.

Del hecho de que los mínimos del sistema están definidos por las interacciones, y que éstos a su vez actúan como atractores dinámicos o patrones almacenados, es posible considerar como puntos de interés los siguientes:

- Proponer y evaluar diferentes reglas de aprendizaje para  $\{J_{ij}\}$ .
- Dada una regla de aprendizaje, evaluar la capacidad máxima y óptima de almacenamiento.

Hasta el momento hemos mencionado que es posible formar atractores en el sistema al almacenar la información utilizando la regla de Hebb. A dichos atractores se les conoce como estados almacenados, memorias puras o estados nominales y corresponden a mínimos locales de la función de energía. Sin embargo, dichos mínimos no son los únicos. Se sabe que en este modelo [Amit *et al.*, 1985] además de los estados almacenados a propósito, aparecen estados mixtos estables  $\xi_i^{mix}$  llamados estados espurios. Estos estados no nominales resultan de la combinación de la información proveniente de un número impar de memorias puras y, deterioran el funcionamiento de la red, ya que entorpecen la recuperación de la información al actuar como atractores adicionales.

Otro parámetro que juega un papel importante dentro de la teoría de las redes neuronales, es el que se asocia a la temperatura  $T$ . Es importante tener muy claro que este parámetro no necesariamente representa la temperatura real de una red neuronal,

es más bien un parámetro que introduce estocasticidad a la red y que se denomina temperatura por la analogía que juega ésta en los sistemas físicos.

Analícemos ahora el funcionamiento de una red neuronal en el régimen de  $T = 0$ . En este caso la evolución del sistema sólo acepta descensos en la energía, por lo que el sistema evoluciona hacia el mínimo local más cercano. Por otro lado, a  $T \neq 0$ , se aceptan incrementos en la energía del sistema con una probabilidad dependiente de  $T$  de acuerdo con el factor de Boltzmann de la mecánica estadística. Esto quiere decir que al evolucionar el sistema puede salirse de un mínimo local pequeño para caer en uno más profundo. En este sentido, la temperatura puede ser vista como agente útil para eliminar el efecto de los estados espurios, lográndose la recuperación de la información almacenada en la red con mayor facilidad.

También se tiene que un parámetro muy importante en el funcionamiento de una memoria asociativa es la razón que existe entre el número de memorias almacenadas en la red  $p$  y el número total de neuronas  $N$ , i.e.,

$$\alpha = \frac{p}{N} \quad (8)$$

a esta cantidad se le conoce como parámetro de capacidad de almacenamiento de la red. Este parámetro puede interpretarse como una medida de qué tanta información es posible almacenar en la red de tal manera que ésta pueda ser extraída sin errores, o al menos, sin un número apreciable de éstos. Simulaciones numéricas, así como análisis estadísticos, han mostrado que dicho parámetro juega un papel similar a la temperatura. Con esto se quiere decir que, mientras más patrones se almacenan en la red la calidad de reconocimiento se deteriora. Esto es, cuando  $\alpha$  se aproxima a un valor crítico  $\alpha_c$  ( $\alpha_c \approx 0.138$  en el límite termodinámico), la red súbita y catastróficamente dejará de

reconocer los patrones almacenados. En otras palabras, la red saltará de manera abrupta, de un reconocimiento casi perfecto (con menos del 3% de errores), a estados de confusión completa. Para el caso de  $p$  finita y  $N \rightarrow \infty$  ( $\alpha = 0$ ), tenemos que los mínimos coinciden con los patrones almacenados ( $\xi_i^\mu = S_i$ , para toda  $i$ ), cosa que para valores finitos de  $\alpha$ ,  $0 < \alpha < \alpha_c$  no sucede, ya que los mínimos tienden a desplazarse con respecto a su valor deseado, lográndose de esta manera que patrones se recuperen con algunos bits equivocados.

### I. 3 Estructura del trabajo

La estructura de este trabajo es la siguiente: en el capítulo II se presenta un problema específico de una red neuronal tipo Hopfield. El propósito principal en dicho problema, es calcular el tamaño de las cuencas de atracción desde el punto de vista macroscópico, es decir, por medio de la iteración de las ecuaciones de flujo o traslapes. Estos traslapes dan una medida de el parecido que existe entre el estado dinámico presente en el sistema y uno de los patrones almacenados en la red.

Un problema novedoso de redes neuronales es el que concierne al capítulo III. Éste consiste en una red neuronal compuesta por dos subredes conectadas unidireccionalmente, de tal manera que cada una de las neuronas de la primera subred manda información solamente a una de las neuronas de la segunda subred. En este sistema la primera subred funciona como una red neuronal tipo Hopfield por lo tanto, su evolución es independiente de la segunda subred. Igual que en el capítulo II, se estudia el tamaño de las cuencas de atracción, pero esta vez, también desde el punto de vista microscópico. El aspecto microscópico concierne al valor específico que toma cada uno de los espines  $S_i$  del sistema.

## II. RED NEURONAL TIPO HOPFIELD

### II. 1 Introducción

En el capítulo anterior vimos que se puede visualizar el espacio de configuraciones de una red neuronal (RN) como un espacio compuesto de valles y montañas, donde los valles actúan como atractores de la dinámica del sistema. Este hecho ha causado gran interés dentro de la comunidad científica, ya que este tipo de sistemas pueden funcionar como memorias asociativas. Es decir, permiten el almacenamiento de configuraciones o patrones que corresponden a mínimos en la energía del sistema. De esta manera es posible recuperar la información contenida en la red a partir de señales con ruido o fallas.

Por analogía a la forma en que la mecánica estadística trata a los sistemas físicos, las redes neuronales también pueden ser tratadas desde dos puntos de vista distinto, el microscópico y el macroscópico. En el primer caso se considera una RN compuesta de  $N$  elementos en donde  $p$  patrones han sido almacenados, y exploramos el espacio de configuraciones con simulaciones tipo Monte Carlo a partir de un estado inicial, en el espacio microscópico  $S_i$ . En el segundo caso, que se refiere al tratamiento macroscópico, se estudia al sistema en términos del traslape  $\vec{q}$ , cuyas componentes  $q_\mu$  constituyen una medida de la semejanza entre el estado dinámico  $S_i$  del sistema y el patrón  $\mu$ -ésimo  $\xi_i^\mu$  almacenado en la red. El procedimiento consiste en encontrar las ecuaciones de flujo o

evolución para los traslapes a partir de un estado inicial. A continuación presentamos una revisión bibliográfica de algunos de los trabajos en el área más relevantes para nuestro estudio.

En general, se han realizado estudios analíticos y numéricos que han comprobado la existencia de estados espurios como configuraciones estables de la red (mínimos que entorpecen la recuperación de los patrones aprendidos en el sistema). Amit *et al.* (1985) encontraron que para una red grande ( $N \rightarrow \infty$ ), en el caso de la regla de Hebb (mencionada en el capítulo anterior), todos los estados espurios corresponden a la mezcla de varios patrones almacenados a propósito. Este resultado lo obtuvieron estudiando el espacio de configuraciones de una RN en forma sistemática, haciendo uso de las herramientas de la mecánica estadística y de la simulación computacional. En sus estudios, ellos encontraron que para  $0 \leq T \leq 1$  las energías de los estados base estaban relacionadas con memorias puras; donde  $T = 1$  es el nivel crítico de ruido abajo del cual aparecen los estados ordenados. Para  $0.461 \leq T < 1$  los estados puros son los únicos mínimos; sin embargo, en tanto que el nivel de ruido decrece, los estados espurios aparecen. Estos mínimos espurios corresponden a la mezcla de un número impar de memorias puras y pueden ser caracterizados como estados simétricos o asimétricos. Los estados simétricos (asimétricos) son mínimos equidistantes (no equidistantes) con respecto a los patrones puros. Para  $T < 0.461$  algunos de los estados simétricos son estables y así mientras  $T \rightarrow 0$  todos los mínimos se vuelven mínimos locales. Los estados asimétricos aparecen por debajo de  $T \sim 0.57$  y ninguno de ellos se encontró que fuera estable, al menos para  $0.461 \leq T \leq 0.57$ .

Los resultados antes mencionados han sido utilizados como base para estudiar otros modelos tipo Hopfield. Este estudio nos da información sobre la existencia y la estabilidad de los atractores, sin embargo no nos permite evaluar su importancia relativa, es decir, no es posible cuantificar el tamaño de las cuencas de atracción. Para esto, ha sido necesario hacer otro tipo de estudios.

Vale la pena señalar que un punto fijo puede ser un atractor o un repulsor de la dinámica del sistema, los mínimos son atractores y los máximos son repulsores. Para estudiar la importancia de las memorias puras (atractores) en el espacio de configuraciones, se han usado diferentes aproximaciones. Esto con el fin de evaluar el tamaño de las cuencas de atracción en diversos modelos. Como ejemplo, podemos mencionar el trabajo realizado por Gardner *et al.* (1989). En éste se encontró que es posible recuperar un patrón en casi un 100% si utilizamos una RN diluída con interacciones óptimas (construida de acuerdo con el llamado algoritmo de Edinburgo), siempre y cuando el traslape (sobre éste hablaremos más adelante) entre el estado inicial del sistema y el patrón a recordar sea macroscópico y  $\alpha = \frac{p}{N} < \alpha_c = 0.42$ . Por otro lado, Kepler y Abbott (1988) calcularon el tamaño de las cuencas de atracción de una red entrenada también con el algoritmo de Edinburgo pero conectada completamente, suponiendo dos puntos: a) que el efecto dinámico del traslape microscópico inicial con patrones no definidos puede ser despreciado; b) si el estado “viaja” al menos la mitad de la distancia a un punto fijo en su primer actualización, entonces eventualmente alcanzará el punto fijo. Encontraron que la dinámica del sistema tiene el mismo comportamiento que el modelo de un único patrón. Sin embargo, posteriormente se encontró que el efecto acumulativo de estos traslapes microscópicos no puede ser despreciado, ya que cuando  $\alpha$  aumenta, éstos juegan un papel importante en la dinámica de una red neuronal. También en 1988, Cottrell aproximó el problema calculando el radio máximo de una hiperesfera alrededor de un estado estable en el espacio de configuraciones. Esta esfera se define de manera que el patrón correspondiente al estado estable va a ser recuperado si el estado inicial de la red está dentro de la esfera. Este análisis no tiene sólo aplicaciones en el modelo de Hopfield sino también en un gran número de modelos de RN totalmente conectadas. Krauth, *et. al* (1988), propusieron un modelo para un único patrón y estudiaron la simetría de las interacciones y la estabilidad de los patrones.

Es importante tener en mente que el pretender determinar el tamaño de las cuencas de atracción de una red neuronal involucra también un problema dinámico. Forrest (1988)

realizó una aproximación numérica para el modelo de Hopfield (para el caso  $\alpha = 0$ ) que consistió en generar estados con ruido, consiguiendo de esta manera un traslape inicial  $m_0$  con los estados nominales (almacenados). Estudió su evolución de manera sincrónica y consideró que un patrón había sido recuperado con éxito si el estado final de la red no difería del patrón almacenado en más de  $N/16$  neuronas.

Un método diferente y más directo, en el que este problema ha sido aproximado, consiste en calcular la fracción promedio  $f_p(\mu)$  de microestados  $S_i$  que evolucionan al  $\mu$ -ésimo patrón almacenado en la red. Esta evaluación se ha hecho de diferentes maneras, una de ellas consiste en simular la dinámica de una RN compuesta de  $N$  elementos  $S_i$  ( $i = 1, \dots, N$ ) donde  $p$  patrones  $\xi_i^\mu$  ( $\mu = 1, \dots, p$ ) han sido almacenados, empezando de un estado inicial generado al azar [Viana et al, 1993]; sin embargo se ha visto que hay efectos de tamaño finito, por lo que no es tan sencillo extrapolar al límite termodinámico.

Otra manera de evaluar el tamaño de las cuencas de atracción consiste en estudiar la evolución dinámica del sistema a nivel “macroscópico”, esto es, considerando las ecuaciones de flujo, que describen la evolución del valor de los traslapes en el tiempo. Estos parámetros de orden o traslapes dan una medida entre el parecido que existe entre el estado dinámico presente en el sistema y uno de los patrones almacenados en la red. Un sistema con  $p$  patrones almacenados en la red va a estar caracterizado por  $p$  parámetros de orden o traslapes. En particular si el estado de la red no está correlacionado con alguno de los patrones, el parámetro correspondiente va a ser de orden cero, mientras que si está correlacionado con algún patrón el traslape en particular tendrá un valor cercano a uno.

## II. 2 Introducción al problema

Anteriormente se mencionó un poco sobre el trabajo que se ha realizado, tanto analítico como numérico, para medir el tamaño de las cuencas de atracción en redes neuronales tipo Hopfield. En esta sección nos enfocaremos a estudiar los tamaños de las cuencas de atracción en una red neuronal en particular, también tipo Hopfield. Nos referimos a una red en donde cada patrón ha sido almacenado de acuerdo a la regla de Hebb modificada (la que veremos posteriormente), de manera que nos es posible aumentar o disminuir el tamaño de los dominios de atracción aumentando o disminuyendo el peso  $w_\mu$  (con  $\mu = 1, \dots, p$ ) asociado a cada una de las memorias.

El análisis que llevaremos a cabo es en función del número  $p$  de patrones almacenados en la red, del peso  $w_\mu$  asignado a cada memoria y de la temperatura  $T$ , ya que ésta es un parámetro que mide el nivel de ruido. Se tendrá mayor énfasis en el espacio ocupado por las memorias espurias, ya que como se ha mencionado, éstas entorpecen las capacidades de recuperación de la red en cuestión.

Algunos de los puntos más relevantes en los que se basa este modelo, de tal manera que se simplifique, son los siguientes:

- Las neuronas son discretas, i.e., toman sólo dos estados posibles  $S_i = \pm 1$ .
- La red está totalmente conectada, i.e., cualesquiera dos neuronas de la red están conectadas.
- La red está compuesta de un número muy grande de neuronas ( $N \rightarrow \infty$ ).
- La conexión sináptica es simétrica ( $J_{ij} = J_{ji}$ ), i.e., la influencia de la neurona  $j$  sobre la neurona  $i$  es igual a la influencia de la  $i$  sobre la  $j$ .

## II. 2. 1 Antecedente analítico

Consideremos el hamiltoniano dado por:

$$H = -\frac{1}{2} \sum_{ij}^N J_{ij} S_i S_j, \quad (9)$$

donde éste describe a un sistema de  $N$  neuronas interactuantes, que pueden ser vistas como espines de Ising  $S_i \pm 1$ . Suponemos que en esta red ha sido almacenado un número finito  $p$  de patrones  $\{\xi_i^\mu\}$  ( $\mu = 1, \dots, p$ ) a través de la interacción  $J_{ij}$ , simétrica y de largo alcance, entre los espines  $S_i$  y  $S_j$ . La interacción  $J_{ij}$  está dada por la expresión:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p w_\mu \xi_i^\mu \xi_j^\mu. \quad (10)$$

Los patrones  $\{\xi_i^\mu\} = \pm 1$  están distribuidos de acuerdo con la probabilidad:

$$P(\xi_i^\mu) = \frac{1}{2} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)] \quad (11)$$

y  $\{w_\mu\}$  es un factor de peso que se introduce con el fin de simular entrenamiento ya que como se mencionó anteriormente es posible aumentar o disminuir el tamaño de las cuencas de atracción. Sin pérdida de generalidad, esto se logra considerando un conjunto  $\{w_\mu\}$

de  $p$  diferentes  $w$ 's para cada estado almacenado tales que  $1 = w_1 > w_2 > \dots > w_p > 0$ . En el caso en que  $w_\mu = 1$  para toda  $\mu$  se reduce a la regla de Hebb.

Para este sistema la función de partición está dada por:

$$Z = Tr \exp \{-\beta H\}, \quad (12)$$

en donde  $H$  está dado por la ecuación (9),  $Tr$  es la traza o suma sobre los valores posibles de los estados y  $\beta = 1/T$  está relacionado con el nivel de ruido  $T$  de la red neuronal, donde  $T$  pequeña corresponde a bajos niveles de ruido. Sustituyendo la ecuación (9) en la expresión (12) llegamos a que la función de partición queda expresada de la forma siguiente (ver apéndice **A**):

$$Z = Tr \exp\{\beta/2N \sum_{\mu} w_{\mu} (\sum_i \xi_i^{\mu} S_i)^2\}, \quad (13)$$

Como podemos apreciar, en esta expresión, es imposible tomar la traza sobre los posibles valores de  $S_i$  debido a que estos aparecen en un término cuadrático.

Con el fin de linealizar la expresión (13) hacemos uso de la transformación de Hubbard-Stratonovich [Hubbard, 1959], y obtenemos:

$$Z = \int \prod_{\mu} dq_{\mu} \exp[-Nf\{q_{\mu}\}], \quad (14)$$

donde  $f\{q_\mu\}$  actúa como una función de energía efectiva dada por:

$$f\{q_\mu\} = (\beta/2N) \sum_{\mu} w_{\mu} q_{\mu}^2 - N^{-1} \sum_i \ln Tr \exp\{\beta \sum_{\mu} w_{\mu} \xi_i^{\mu} S_i q_{\mu}\}, \quad (15)$$

en la que  $\{q_\mu\}$  es un conjunto de  $p$  variables auxiliares introducidas con el fin de linealizar el término  $(\sum_i \xi_i^{\mu} S_i)^2$ , y que actúan como parámetros macroscópicos del sistema.

Las propiedades de equilibrio de este sistema están caracterizadas a nivel macroscópico por la existencia de  $p$  parámetros de orden  $q_{\mu}$ , que miden el parecido entre un estado estable microscópico del sistema y el patrón  $\mu$ -ésimo. En el límite termodinámico ( $N \rightarrow \infty$ ), la integral de la ecuación (14) está dominada por el mínimo de  $f\{q_\mu\}$  por lo que es posible hacer un desarrollo en serie de Taylor alrededor de estos puntos. De esta manera el valor de los términos auxiliares o parámetros de orden están determinados por las condiciones  $\frac{\delta f}{\delta q_{\mu}} = 0$  la cual, después de tomar la traza, da lugar a un conjunto de  $p$  ecuaciones acopladas dadas por:

$$q_{\mu} = \langle \xi_i^{\mu} \tanh\{\beta \sum_{\nu} w_{\nu} q_{\nu} \xi_i^{\nu}\} \rangle_{\xi} \quad \text{con} \quad \mu, \nu = 1, 2, \dots, p \quad (16)$$

en donde ahora la función de energía efectiva está dada por:

$$f\{q_{\mu}\} = (\beta/2N) \sum_{\mu} w_{\mu} q_{\mu}^2 - \langle \ln[2 \cosh \beta \sum_{\mu} w_{\mu} \xi_i^{\mu} q_{\mu}] \rangle_{\xi} \quad (17)$$

En estas expresiones hemos usado la notación  $\langle \dots \rangle_\xi$  para indicar el promedio sobre las variables al azar  $\{\xi_i^\mu\}$  dadas por la ecuación (11). En este límite ( $N \rightarrow \infty$ ) se puede aplicar un autopromedio [Van Hemmen, 1986], que consiste en no considerar un promedio sobre las variables aleatorias  $\{\xi_i^\mu\}$  sino un promedio sobre variables macroscópicas correspondientes a  $\eta_\kappa = \pm 1$  con  $\kappa = 1, \dots, 2^{p-1}$  sobre  $2^p$  esquinas de un hipercubo (para mayor entendimiento ver apéndice **B**). De aquí que esta ecuación queda escrita como:

$$q_\mu = \langle \langle \eta_\mu \tanh \left\{ \sum_\gamma w_\gamma q_\gamma \eta_\gamma \right\} \rangle \rangle_{\vec{\eta}} \quad (18)$$

donde el doble paréntesis  $\langle \langle \dots \rangle \rangle_{\vec{\eta}}$  indica que se está promediando sobre las  $2^p$  esquinas del hipercubo alrededor de  $\vec{\eta} = 0$ . Esta descripción estática tiene una contraparte dinámica; para un sistema con dinámica paralela sincronizada, la evolución en el tiempo de los traslapes está dada por la expresión [Coolen, 1989].

$$\vec{q}(n+1) = F(\vec{q}(n)) \quad (19)$$

donde  $F(\vec{q})$  está representada por la expresión:

$$F(\vec{q}) = \langle \langle \vec{\eta} \tanh \left\{ \sum_\gamma w_\gamma q_\gamma \eta_\gamma \right\} \rangle \rangle_{\vec{\eta}} \quad (20)$$

De esta manera los puntos fijos (atractores) de la dinámica, i.e., los puntos que

satisfacen  $F^n(\vec{q}) = F^{n-1}(\vec{q})$ , para algún valor de  $n$ , corresponden a puntos estables dados por la ecuación (18).

Las ecuaciones (19) y (20) tienen puntos fijos, de los cuales algunos actúan como atractores dinámicos del sistema. Para el modelo de Hebb, se ha visto que para  $N \rightarrow \infty$  y un número finito de patrones  $p$  almacenados en la red, existen puntos fijos relacionados con las  $p$  memorias puras, y algunos que resultan de la combinación de dichas memorias (espurias). En lo subsecuente, se tendrá mayor énfasis en el espacio ocupado por las memorias espurias.

### II. 3 Red tipo Hopfield: Dependencia en $T$

El tamaño de las cuencas de atracción puede verse como la fracción de estados que evolucionan al  $\mu$ -ésimo patrón. Una de las formas en que puede llevarse a cabo este estudio es a través del mapeo del espacio de configuraciones a nivel microscópico, realizando simulaciones tipo Monte Carlo. En particular se han efectuado estudios para el modelo de la regla de Hebb modificada antes descrita [Martínez, 1989, Martínez *et al.*, 1990] a  $T = 0$  como función de  $\alpha = p/N$ . En la figura (4) se muestra el porcentaje de memorias espurias recuperadas como función de  $\alpha$ . Comparando los resultados obtenidos para la regla de Hebb (pA) con los de la regla de Hebb modificada (pB), se observa que el porcentaje ocupado por las memorias espurias es menor en el segundo caso. También observamos que, el tamaño de las cuencas de los patrones nominales depende del peso asignado a cada memoria, y por ende podemos suponer que existe una tendencia muy fuerte al entrenamiento. De igual manera, en la figura (4) nótese que para ambos casos, si el valor de  $N$  aumenta el porcentaje ocupado por las memorias espurias en el espacio de configuraciones aparentemente tiende a un límite.

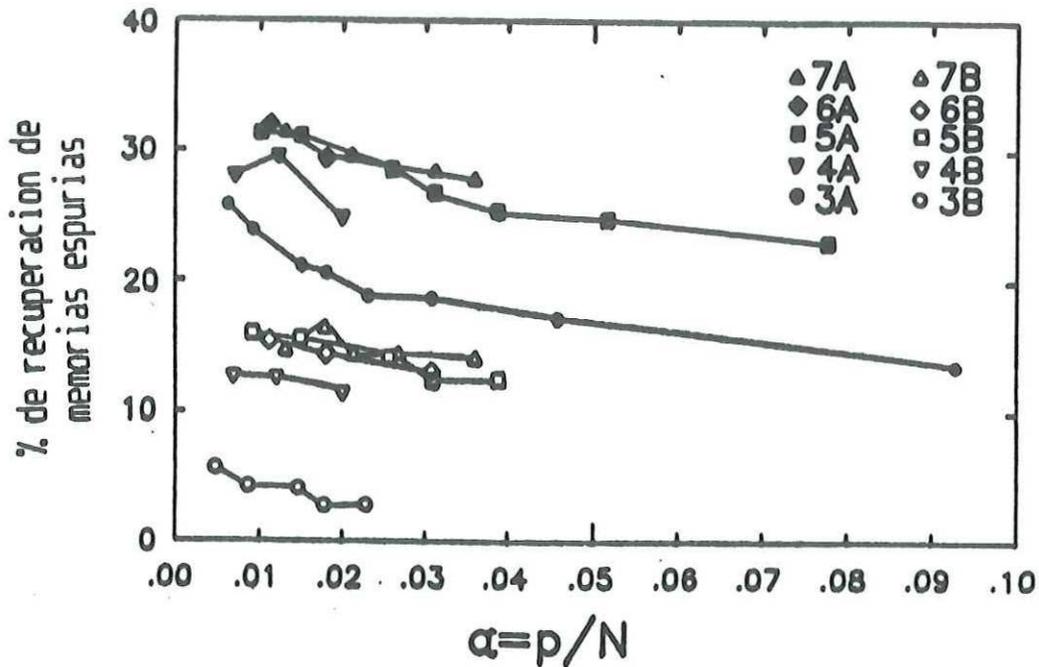


Figura 4: Porcentaje de recuperación de las memorias espurias como función de  $\alpha$  para valores de  $p = 3$  hasta  $p = 7$ , tanto para la regla de Hebb como para la regla de Hebb modificada. Ambos casos están indicados por (pA) y (pB) respectivamente.

Otra manera de evaluar el tamaño de las cuencas es mediante un estudio de carácter macroscópico, que involucra a los traslapes  $\vec{q}$ , cuyas componentes  $q_\mu$  constituyen una medida de la semejanza entre el estado dinámico  $\{S_i\}$  y el patrón almacenado  $\{\xi_i^\mu\}$  en la red. En seguida nos enfocaremos a un análisis de este tipo como función de los pesos  $\{w_\mu\}$  asignados a cada una de las memorias, del número  $p$  de patrones almacenados en la red y de la temperatura  $T$ .

### II. 3. 1 Antecedentes y motivación

En 1993, Viana y Coolen realizaron un estudio sistemático a nivel macroscópico a  $T = 0$ , con el fin de medir el tamaño de las cuencas de atracción de una red neuronal en la que la información se almacenó de acuerdo a la regla de Hebb modificada. Algunos resultados de este estudio se pueden ver en la figura (5), en la que se muestra que por debajo de la línea  $w_2 + w_3 < w_1 = 1$  todos los atractores corresponden a memorias

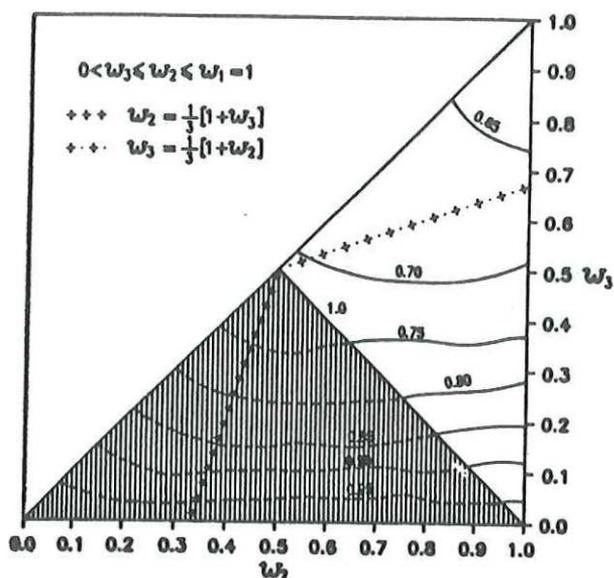


Figura 5: Representación del espacio de las  $w$ 's para  $p = 3$  a  $T = 0$ . El comportamiento de esta red en el espacio de las  $w$ 's está separado en dos regiones por la línea  $w_2 + w_3 = w_1 = 1$ . Las líneas de contorno en la parte clara indican el porcentaje de microestados que conducen hacia las cuencas de atracción de las memorias puras.

puras, es decir, no existen memorias espurias en esta región. Dentro de esta zona, las ecuaciones de flujo convergen a un punto estable después de una o dos iteraciones de las ecuaciones (20) mostrándose con las líneas de contorno punteadas, el porcentaje de veces que esto sucede. Si la línea  $w_2 + w_3 = w_1 = 1$  es cruzada, entrando en el régimen de  $w_2 + w_3 > w_1 = 1$  ocurre una transición abrupta hacia una región en la que empiezan a aparecer memorias espurias, es decir, cuando las ecuaciones de flujo convergen a la primera iteración el punto de convergencia siempre está relacionado con memorias puras en cambio, si la convergencia tarda más de una iteración se refiere a memorias espurias. Las líneas de contorno sólidas muestran el porcentaje de recuperación de las memorias puras.

Cuando un sistema se encuentra a  $T = 0$ , se espera que al evolucionar en el espacio de las energías (compuesto por valles y montañas), quede atrapado dentro de la cuenca en que se encuentra, ya que las barreras de ésta son tan altas que le es imposible pasar

de una a otra. Sin embargo, si en el sistema se permite un poco de ruido ( $T \neq 0$ ), puede disminuir la altura de dichas barreras haciendo que el sistema explore otras regiones de mínima energía. En este sentido el ruido puede desempeñar un papel favorable en la recuperación de la información; tal efecto tiene limitaciones para valores grandes de ruido ( $T > T_c$ ).

Para extender los resultados de Viana y Coolen, se considera el caso de temperaturas finitas y su efecto sobre la recuperación de la información. Se pondrá mayor énfasis en las memorias espurias que como ya se ha mencionado entorpecen la recuperación de la información.

### II. 3. 2 Método

El procedimiento utilizado para evaluar el tamaño de las cuencas de atracción en el sistema mencionado, es por medio de la iteración de las ecuaciones de flujo, o traslapes, dadas por la ecuación (16), a partir de un estado inicial al azar que obedece una distribución gaussiana con dispersión  $\sigma \sim \frac{1}{N^{1/2}}$ . En este método se considera que el número de patrones almacenados en la red es pequeño comparado con el número de neuronas, y que el traslape entre patrones es de orden cero con el fin de eliminar algunos efectos de tamaño finito.

Se realizaron dos estudios diferentes para medir el tamaño de las cuencas de atracción: **a)** el primero de ellos consistió en un estudio sistemático de las cuencas de atracción de las memorias puras y las memorias espurias como función de los pesos asignados a cada memoria  $\{w_\mu\}$  para  $p = 3$ , y como función de  $T$ . **b)** El segundo estudio tuvo la finalidad de analizar la evolución de un punto en dicho espacio como función de  $T$ ; para esto, se consideró un conjunto de pesos  $\{w_\mu\}$  fijo .

En los estudios anteriores, cada red memoriza  $p$  estados en base a la regla de Hebb modificada, ecuación (10). En el caso **a**), esto se hizo como función de los pesos asignados  $\{w_\mu\}$  a cada memoria con la convención de que  $w_1 = 1$  y  $0 < w_i < w_j, \forall i > j$ , y considerando incrementos en los pesos de  $\Delta w = 0.04$ . En el caso **b**), se fijaron los valores para las  $w$ 's tales que  $\langle w_\mu \rangle = 0.7$ . En ambos casos, se varió la temperatura en un rango de  $0 \leq T < 1$ .

### II. 3. 3 Resultados

Como se mencionó con anterioridad en este trabajo se estudió el espacio de configuraciones. Más específicamente, se midió el tamaño de los dominios de atracción en términos del porcentaje de microestados recuperados de una red neuronal cuyos patrones  $p$  fueron almacenados de acuerdo a la regla de Hebb modificada. Los resultados de ambos estudios se representan en las gráficas de las figuras 6 y 9 respectivamente.

La figura 6 corresponde al primer caso, en el cual se fijó el valor de  $p = 3$  y se estudiaron las cuencas de atracción como función de los parámetros  $T$  y  $w_\mu$  para  $\sigma = 1 \times 10^{-5}$ , lo cual corresponde a un valor de  $N = 1 \times 10^{10}$ . En esta figura pueden observarse dos comportamientos diferentes en la red neuronal, que están divididos por una línea que se desplaza conforme aumenta la temperatura  $T$  en el rango  $0 \leq T < 1$  y está dada por  $w_2 + w_3 = c(T)$ ; debajo de esta línea no existen puntos fijos asociados a memorias espurias. Para distintos valores de  $\sigma$  se encontró que el parámetro  $C(T)$  varía como se muestra en la figura (7) por lo que nos permitimos concluir que el valor de  $C(T)$  no es un efecto de tamaño finito. Es interesante notar que para el caso de la regla de Hebb se ha visto que el porcentaje de recuperación de las memorias espurias es independiente de  $\sigma$  (figura 8).

Por otro lado, se encontró que un pequeño porcentaje de memorias espurias que se desplaza y crece, en el espacio de los pesos  $\{w_\mu\}$ , conforme se incrementa el valor de  $T$ . Esto se muestra en la figura (9). para  $\sigma = 1 \times 10^{-1}$  ( $N = 100$ ). Con el objeto de entender a qué se debe este comportamiento variamos el valor de  $\sigma$  entre  $1 \times 10^{-7}$  y  $1 \times 10^{-1}$  y se observó que mientras más pequeño es el valor de  $\sigma$  más pequeño es dicho porcentaje de memorias espurias. Aunque esto no nos permite dar una explicación clara y precisa de este fenómeno, nos permite concluir que se debe a efectos de tamaño finito pues en el tratamiento de las ecuaciones de flujo el tamaño de la red se refleja en el tamaño de la  $\sigma = \frac{1}{\sqrt{N}}$ . Finalmente nótese que, independientemente del valor de  $\sigma$  (figura.10) el porcentaje de memorias espurias disminuye conforme aumenta la temperatura.

En el segundo estudio se consideran tres memorias almacenadas ( $p = 3$ ) así como  $\sigma = 1 \times 10^{-5}$  ( $\alpha \sim 10^{-10}$ ) y  $\sigma = 1 \times 10^{-1}$  ( $\alpha = .03$ ). Se estudió al sistema en la región  $w_2 + w_3 = c(T)$  consiguiendo siempre valores para los pesos  $\{w_\mu\}$  tales que  $\langle w_\mu \rangle = 0.7$ , es decir,  $w_1 = 1$ ,  $w_2 = 0.7$  y  $w_3 = 0.4$ .

Los resultados están representados en la figura (11) que muestra el porcentaje de memorias espurias en función de  $T$ . Se observa que al aumentar la temperatura hay un punto en el que ocurre una caída abrupta en el porcentaje de las memorias espurias. Arriba de este valor crítico de  $T$  el porcentaje de dichas memorias es cero. Con esto puede decirse que el 100% de la información almacenada en ese punto puede ser recuperada. Para valores de  $T$  arriba de .9 el porcentaje de memorias espurias está relacionado con ruido.

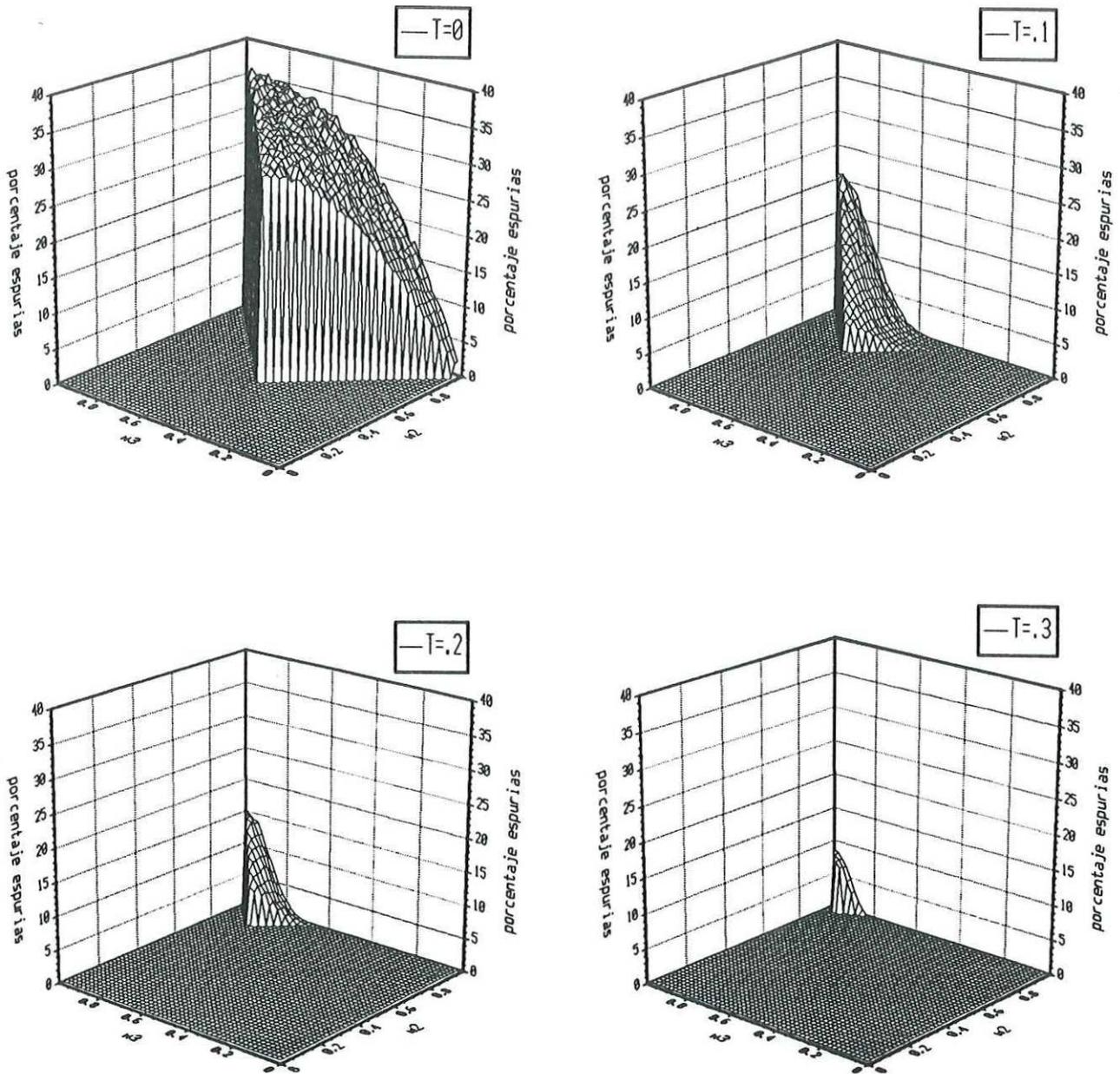


Figura 6: Porcentaje de recuperación de las memorias espurias en el espacio de las  $w$ 's para  $p = 3$ ,  $\sigma = 1 \times 10^{-5}$  y desde  $T = 0$  hasta  $T = .9$ . En esta gráfica se pueden visualizar dos comportamientos divididos por la línea  $w_2 + w_3 = c(T)$ . A medida que la temperatura aumenta la línea  $w_2 + w_3 = c(T)$  se desplaza y el porcentaje de las memorias espurias disminuye.

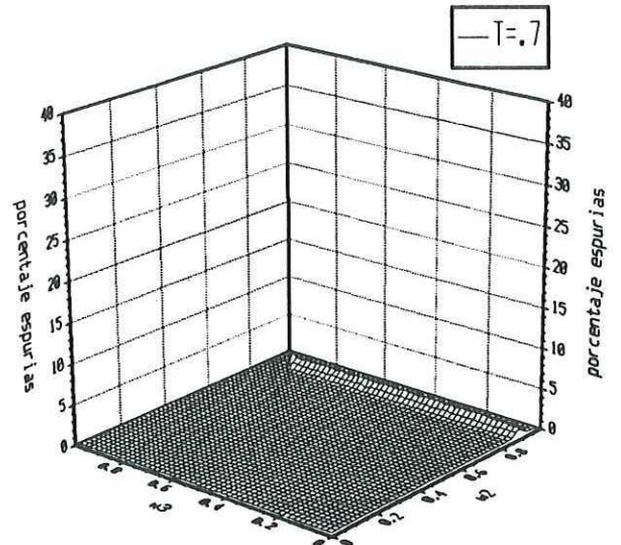
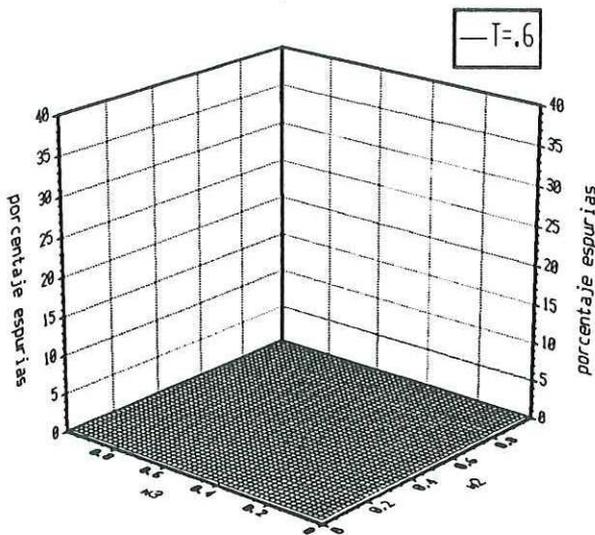
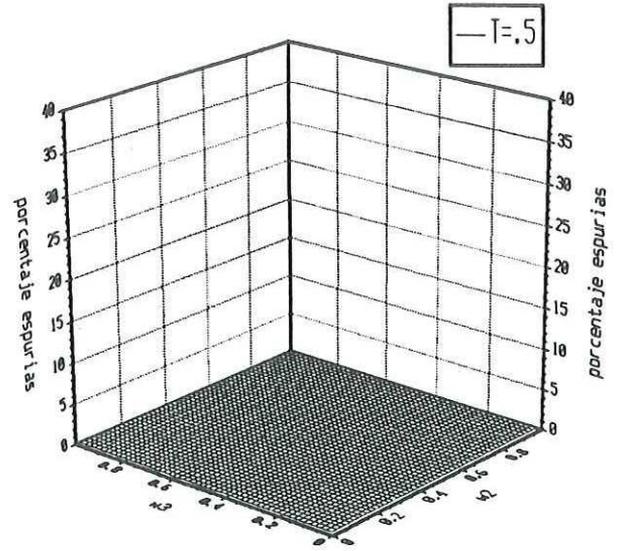
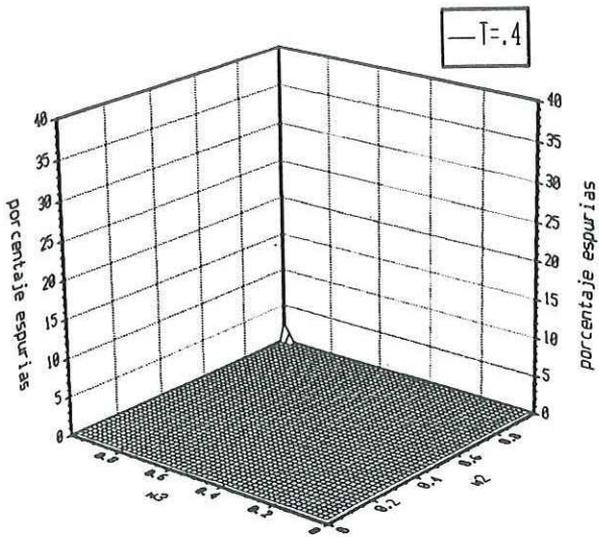


Figura 6: ... (Continuación)

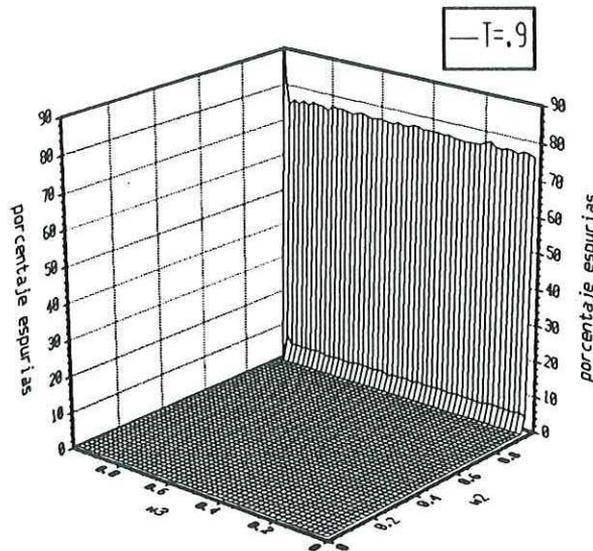
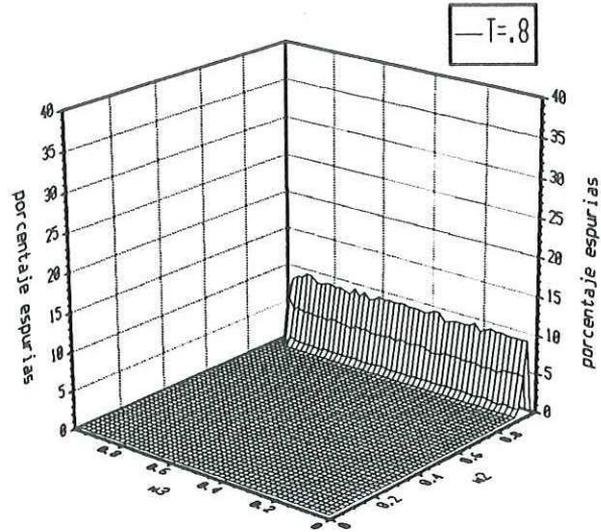


Figura 6: ... (Continuación)

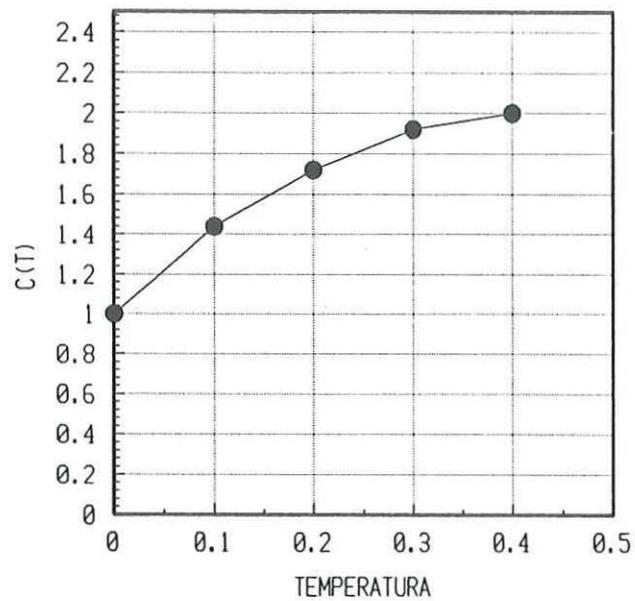


Figura 7: Relación entre  $C(T)$  y  $T$  para distintos valores de  $\sigma$ .

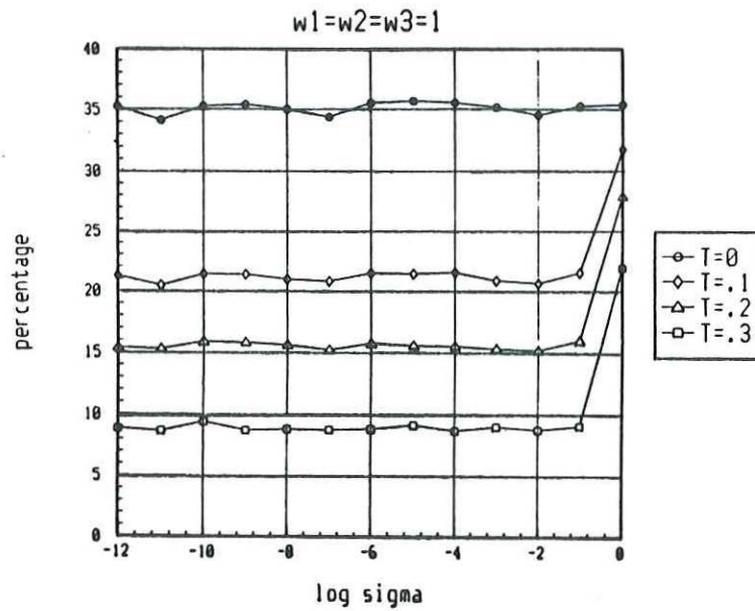


Figura 8: Porcentaje de recuperación de las memorias espurias como función de  $\ln \sigma$  para el caso de la regla de Hebb  $w_\mu = 1$  y para  $T = 0$  hasta  $T = .3$ .

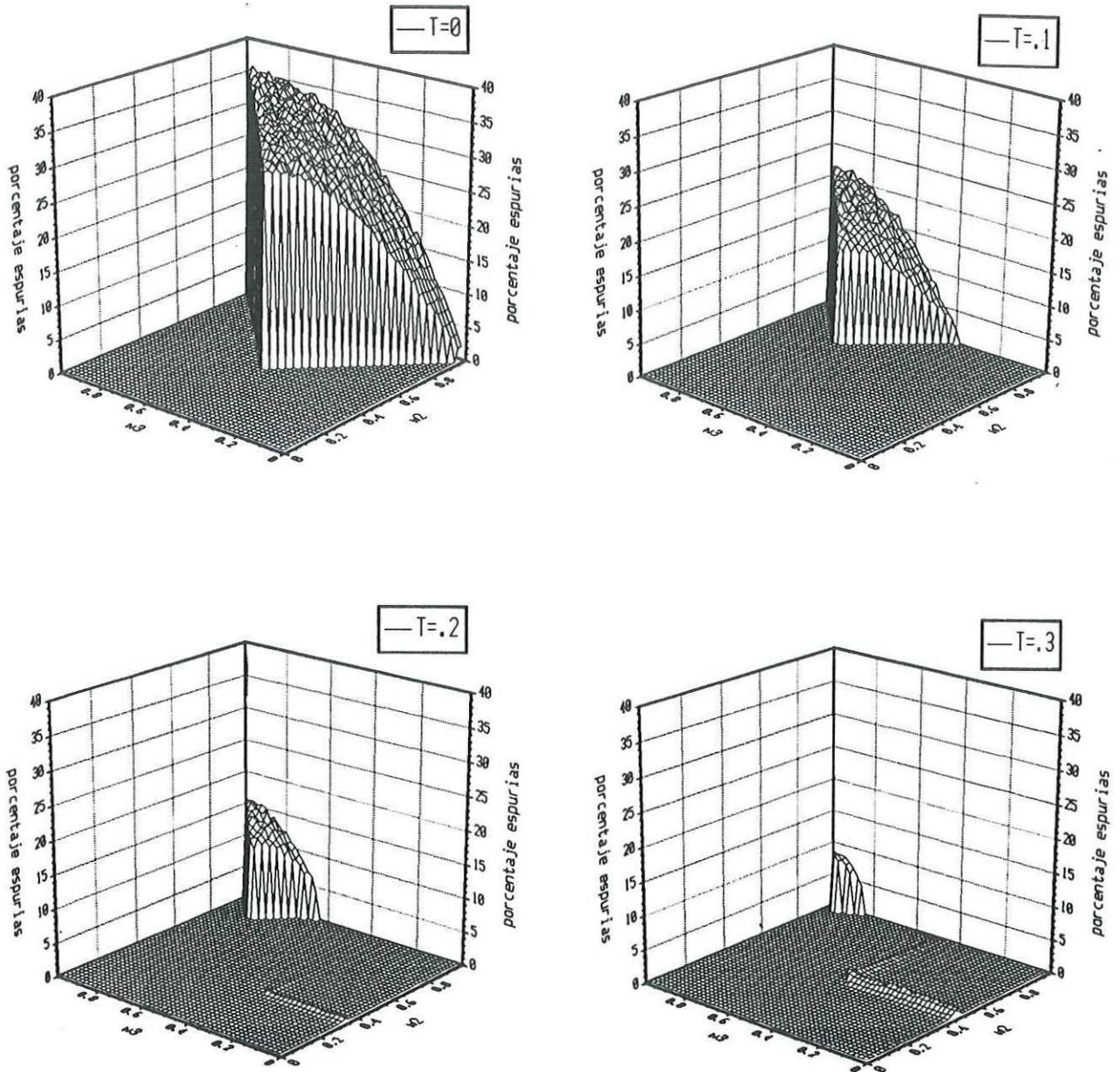


Figura 9: Porcentaje de recuperación de las memorias espurias en el espacio de las  $w$ 's para  $p = 3$ ,  $\sigma = 1 \times 10^{-1}$  y desde  $T = 0$  hasta  $T = .9$ . En esta gráfica se pueden visualizar dos comportamientos divididos por la línea  $w_2 + w_3 = c(T)$ . A medida que la temperatura aumenta la línea  $w_2 + w_3 = c(T)$  se desplaza y el porcentaje de las memorias espurias disminuye.

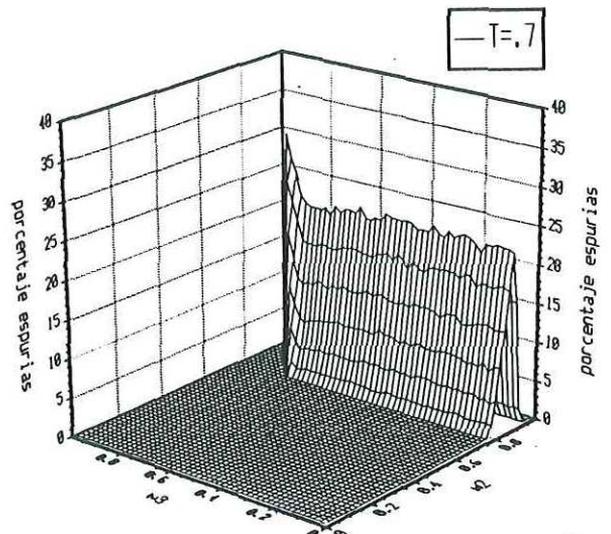
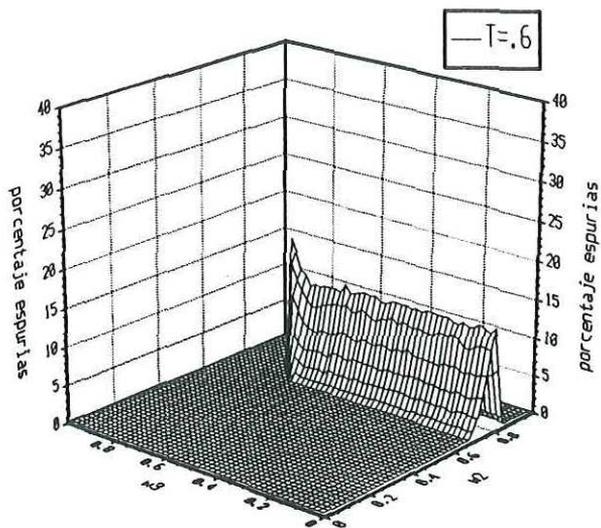
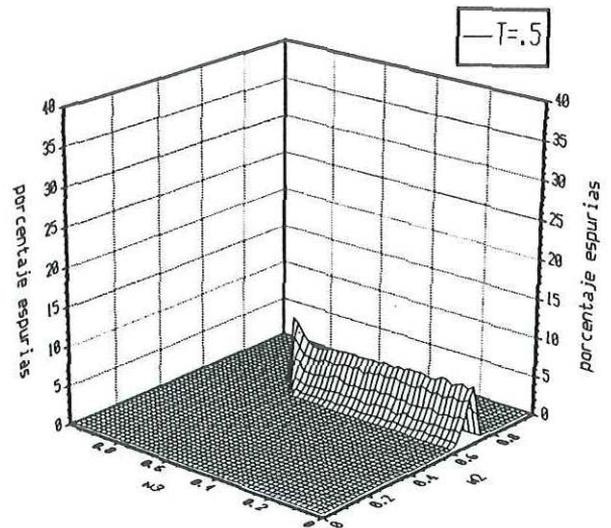
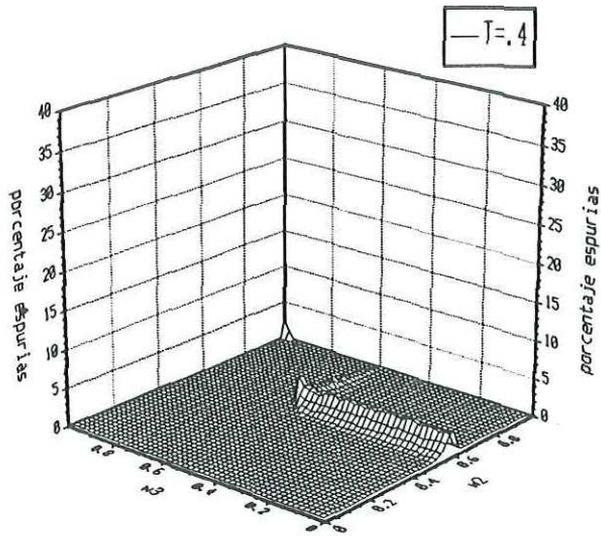


Figura 9: ... (Continuación)

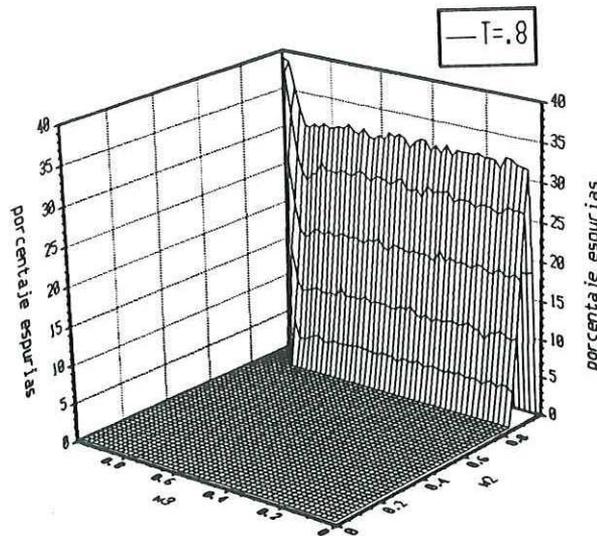
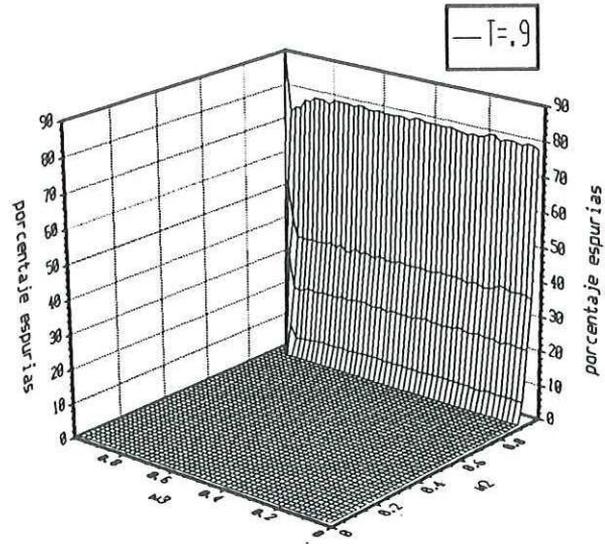


Figura 9: ... (Continuación)

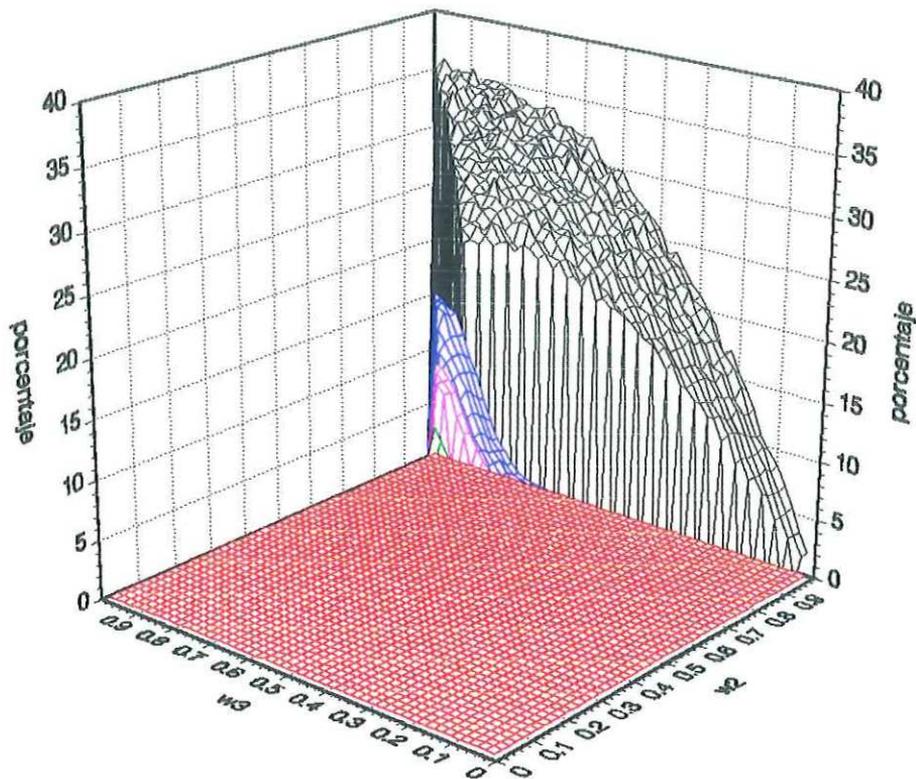
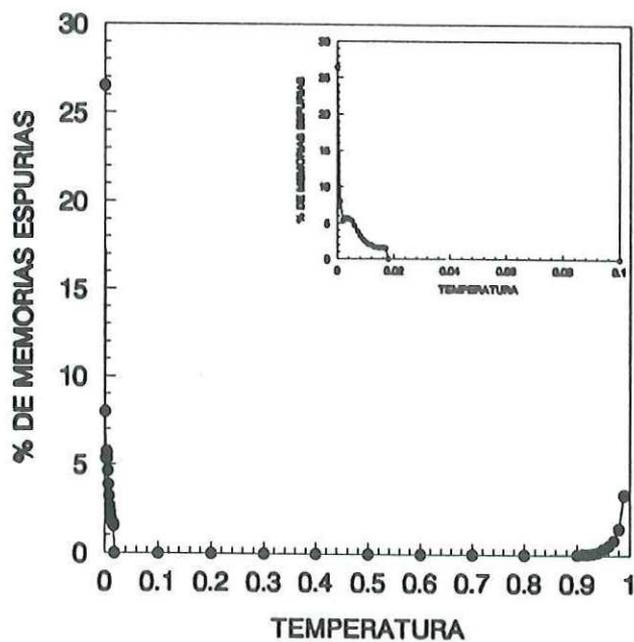
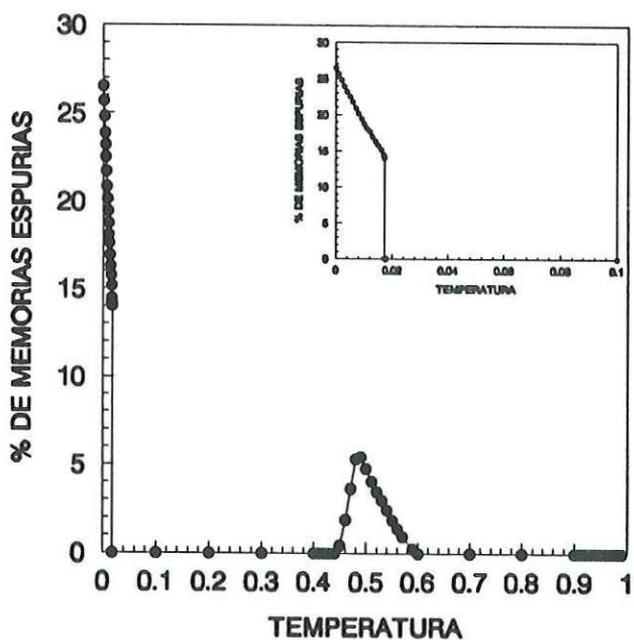


Figura 10: Porcentaje de recuperación de las memorias espurias en el espacio de las  $w$ 's para  $p = 3$ ,  $\sigma = 1 \times 10^{-5}$  y  $T = 0$  (negro),  $T = .2$  (azul),  $T = .3$  (rosa),  $T = .4$  (verde) y  $T = .5$  (rojo).



(a)



(b)

Figura 11: Porcentaje de recuperación de las memorias espurias como función de  $T$  para la regla de Hebb modificada con  $p = 3$  y  $w_\mu = .7$ . La figura (a) considera el caso de  $\sigma = 1 \times 10^{-5}$  mientras que la figura (b) el caso con  $\sigma = 1 \times 10^{-1}$ . En ambas, en el extremo superior derecho se muestra una ampliación de la zona comprendida entre  $T = 0$  hasta  $T = 0.1$ .

## II. 3. 4 Conclusiones

Antes del presente trabajo, y del realizado por Viana y Coolen [1993], se consideraba que el tener tres ( $p = 3$ ) o más memorias almacenadas en una red neuronal implicaba la existencia de memorias espurias. Esto es cierto en el caso de la regla de Hebb, sin embargo para el caso de la regla de Hebb modificada, se comprobó que la aparición de las memorias espurias depende de los valores asignados a las  $w$ 's y que esto es válido para cualquier temperatura por debajo de la temperatura crítica.

Con los resultados obtenidos, se comprobó que para el modelo considerado el porcentaje de recuperación de las memorias espurias disminuye conforme aumenta la temperatura, de lo que podemos afirmar que al permitirle un poco de ruido externo a la red, se favorece la recuperación de los estados nominales (puros). Sin embargo, el hecho de que no aparezcan memorias espurias para ciertos valores de  $w$ 's no significa que se haya eliminado totalmente el ruido en la red y se haya recuperado la información perfecta de los patrones almacenados, ya que el valor final del traslape  $q$  no siempre es uno. Del hecho de que el valor de  $q$  sea cercano a la unidad se infiere que a  $T = 0$  la posición del mínimo se encuentra desplazada con respecto al mínimo almacenado, o sea, que la información se ha recuperado con una pequeña fracción de errores dada por  $d = \frac{1}{2}(1 - q)$ . Por otro lado, para  $T \neq 0$  tenemos que dentro de la cuenca existe una cierta dinámica del sistema entre estados microscópicos compatibles con un valor macroscópico  $q$ . En particular el valor de  $q \neq 1$  lo que quiere decir que el sistema no está en un estado fijo microscópico, como ocurre a  $T = 0$ , sino que está fluctuando entre estados con ese valor  $q$ . De aquí que, un punto de estudio para recuperar en su totalidad la información almacenada, es por medio del enfriado lento (simulated annealing) que consiste en bajar la temperatura lentamente para ir alcanzando estados de equilibrio dentro de la misma cuenca.

## III. REDES ACOPLADAS

### III. 1 Introducción

El reconocimiento de patrones por medio de memoria asociativa es un problema que ha recibido mucha atención y que está relativamente bien estudiado. En particular, en el capítulo anterior estudiamos el tamaño de las cuencas de atracción de una red neuronal tipo Hopfield en la que la información fue almacenada asignándole un peso estrictamente diferente a cada una de las memorias, esto con el fin de simular diferentes niveles de reforzamiento para diferentes memorias. El estudio se realizó como función del peso asignado a cada memoria  $\{w_\mu\}$ , del número  $p$  de patrones y de la temperatura  $T$ .

En este capítulo estudiaremos una red neuronal (RN) compuesta por dos subredes tipo Hopfield interconectadas. Uno de los objetivos a alcanzar ahora es encontrar una forma en la cual éstas trabajen acopladamente. Con el propósito de estudiar la recuperabilidad (tamaño de las cuencas) de la información en la segunda subred como función del estado inicial de la primera, el estudio se hace desde dos puntos de vista diferentes, microscópico y macroscópico. En particular, este tipo de sistema permite estudiar otra clase de problemas, como es el caso de los patrones compuestos, es decir, aquellos patrones que se pueden descomponer en segmentos con significado propio, por ejemplo, una palabra en letras. De esta manera, es posible reconocer a una palabra por segmentos, es decir, el

reconocimiento asociativo se lleva a cabo letra por letra, en pares tales que cada subred almacena a una de ellas.

Con el fin de ejemplificar lo anterior supongamos que tenemos una red neuronal compuesta por dos subredes tipo Hopfield (figura 12). En dicha red se almacenan patrones que son palabras tales como: **si**, **no**, **el**, etcetera, de tal manera que la primera letra de la palabra queda almacenada en la primera subred y la segunda letra en la segunda subred, en ambos casos de acuerdo con la regla de Hebb. Ahora bien, si consideramos entre las dos subredes conexiones tipo Hebb, pero unidireccionales, de tal manera que cada una de las  $N$  neuronas de la primera subred mande información solamente a una de las  $N$  neuronas de la segunda subred. En estas circunstancias, esperaríamos que si se presenta a la primera subred un estado inicial tal que coincida con una de las letras almacenadas, ésta a su vez mande información a la segunda subred de tal forma que esta última responda con el patrón correspondiente a la letra con la que está asociada. Es decir, si el estado inicial coincide con la letra **s** en la primera subred, ésta va a influir en el proceso dinámico de la segunda subred de manera que ésta recupere la letra **i**. Sin embargo, en principio esto no quita la posibilidad de recuperar otro patrón que corresponda a otra letra.

En este sistema la primera subred es una RN tipo Hopfield pues su evolución, a partir de un estado inicial, es independiente de la segunda subred. Se han realizado varios estudios referentes a la precisión con la cual es posible recuperar la información en una red de este tipo. En particular, Forrest [1988] calculó numéricamente la fracción de estados  $f(q_0)$  que pueden ser recuperados con un número de espines equivocados menor que  $N/16$  para  $\alpha = p/N$  para varios traslapes iniciales  $q_0$ . Este criterio es equivalente a considerar un traslape final  $q_f > \frac{7}{8}$ . En general, se ha visto que para una red totalmente conectada es posible recuperar la información almacenada a partir de cualquier estado inicial  $q_0$  tal que  $q_0 \geq q_c$  (el traslape crítico), consiguiéndose de esta manera un traslape

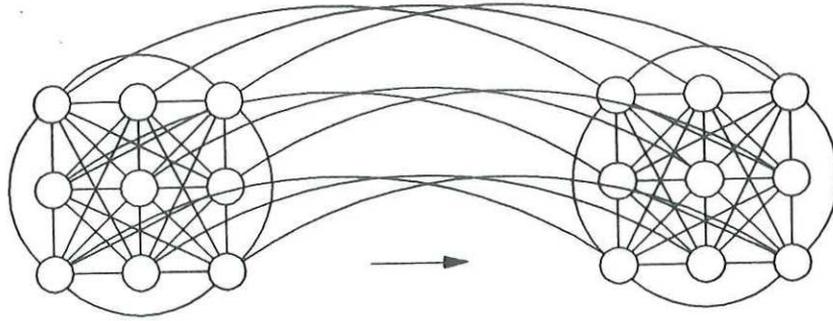


Figura 12: Representación esquemática de una red neuronal compuesta por dos subredes.

final  $q_f$  de orden uno entre el estado inicial y el estado almacenado. Por otro lado, el estudio realizado en el capítulo anterior, junto con el trabajo de Viana y Coolen [1993] es una manera alternativa de estudiar este tipo de redes.

## III. 2 Modelo

Consideremos una red neuronal compuesta de dos subredes, tal que cada una de ellas consta de  $N$  neuronas con estados  $S_i = \pm 1$  ( $i = 1, 2, \dots, N$ ). La interacción  $J_{ij}$  entre las neuronas  $S_i$  y  $S_j$ , en cada una de las subredes, es a través de conexiones simétricas tipo Hebb y está dada por:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu}, \quad (21)$$

donde hemos almacenado  $p$  patrones  $\{\xi_i^\mu\}$  con  $\mu = 1, \dots, p$ , y  $\xi_i^\mu = \pm 1$  con la misma probabilidad.

Consideremos que las dos subredes constan de  $N$  elementos, la primera con  $i = 1, \dots, N$ , y la segunda con  $i = N + 1, \dots, 2N$ . Supongamos que existen interacciones unidireccionales que conectan a ambas subredes de manera que la primera subred envía información a la segunda por medio de conexiones dadas por:

$$w_{i+N,j} = \sum_{\mu=1}^p \xi_{i+N}^\mu \xi_i^\mu \delta_{i+N,i}, \quad (22)$$

con  $i = 1, \dots, N$ . Así tenemos que cada neurona de la primera subred envía información exclusivamente a una neurona de la segunda, siendo entonces que las conexiones entre ambas subredes **no** sean simétricas ( $w_{j,i+N} = 0$ ). Es importante mencionar que cada subred tiene  $N$  neuronas, por lo que el número total de conexiones internas en cada una de las subredes es  $N^2$ . Por otro lado, hemos escogido conectar cada una de las  $N$  neuronas de la primera subred con una de las  $N$  neuronas de la segunda subred de manera que, el número de interacciones entre las subredes es de orden  $N$ . Por lo anterior, nos vemos obligados a normalizar el peso a la regla de aprendizaje indicada en la ecuación (22), de lo contrario, en el límite termodinámico ( $N \rightarrow \infty$ ) las subredes se desconectarían.

Para medir el tamaño de las cuencas de atracción de la RN compuesta por las dos subredes, es posible hacer el estudio deseado desde dos puntos de vista diferentes: a nivel microscópico y a nivel macroscópico. El primero de ellos parte de considerar la evolución temporal del estado  $\vec{S}(t)$  de la red neuronal, que consiste en un vector de  $N$  dimensiones, tal que la  $i$ -ésima componente  $S_i$  representa el estado de la  $i$ -ésima neurona y tiene una probabilidad de transición asociada. En cambio, el nivel macroscópico consiste

en estudiar la evolución temporal del traslape  $q_\mu(t)$  (o parámetro de orden) del estado dinámico microscópico de la red, con el  $\mu$ -ésimo patrón almacenado.

En el caso de la descripción microscópica lo apropiado es llevar a cabo una simulación tipo Monte Carlo para analizar el tamaño de las cuencas de atracción, mientras que para el nivel macroscópico se requiere obtener primeramente  $p$  ecuaciones de evolución temporal para los traslapes  $q_\mu(t)$ , las cuales resultan acopladas y no lineales, de modo que su solución no es posible de manera simple. En consecuencia se recurre a un método de iteración computacional. Para el caso que es nuestro objeto de estudio en el presente capítulo, se describen ambos enfoques a continuación.

### III. 2. 1 Estudio macroscópico de las redes acopladas.

En esta sección haremos un estudio para medir el tamaño de las cuencas de atracción de la RN antes mencionada. Dicho estudio es desde el punto de vista macroscópico y consiste en obtener una expresión analítica para los parámetros de orden o traslapes, que miden el parecido entre un estado dinámico microscópico del sistema y el  $\mu$ -ésimo patrón de la red. Es importante recordar que esta versión del tratamiento sólo es válida para el caso  $\alpha = 0$ .

Ahora bien, nos interesa dar una descripción analítica del sistema compuesto por dos subredes conectadas unidireccionalmente. Para esto consideramos que cada una de las subredes que constituyen a la RN están compuestas por  $N$  elementos interconectados entre ellos, con conexiones simétricas de tipo Hebb. Con el objeto de cargar el menor número de índices durante el cálculo consideraremos la convención de que la primera subred consta de  $i = N + 1, \dots, 2N$  elementos mientras que la segunda de  $i = 1, \dots, N$  elementos. Por otro lado, para que estas subredes trabajen acopladamente se propone que

la conectividad entre ellas no sea simétrica, es decir, que sea unidireccional de tal manera que cada una de las  $N$  neuronas de la primera subred mande información exclusivamente a una de las  $N$  neuronas de la segunda subred.

Dado que la conexión entre las subredes es unidireccional, enviando información de la primera hacia la segunda, podemos decir que el comportamiento de la primera subred no se ve afectado por la presencia de la segunda. Esto implica que podemos ver a la primera subred como una red neuronal tipo Hopfield. Por esto entendemos que, dado un estado inicial en dicha subred, ésta va a evolucionar hacia a uno de los atractores de la misma independientemente del proceso que se lleve a cabo en la segunda subred [Martínez *et al.*, 1990, Viana *et al.*, 1993]. Sin embargo, la segunda subred sí se verá afectada por la primera. Una vez que la primera subred recupere el estado almacenado mandará una señal persistente a las  $N$  neuronas de la segunda subred. Como veremos más adelante, estas señales actúan como umbral para las neuronas de la segunda subred logrando, colectivamente, que el proceso dinámico que se lleva a cabo en dicha subred recupere la información relacionada con la primera subred.

Del hecho que la primera subred se comporte como una típica RN tipo Hopfield, la cual está relativamente bien estudiada, nos permite centrar nuestra atención en la segunda subred. Por otro lado, consideramos que la primera subred se encuentra en un estado estable  $\{S_{i+N} = \xi_{i+N}\}$  en el espacio de las energías, de tal manera que la información que ésta envía a través de la sinapsis a la segunda subred, es constante.

A una red con conexiones asimétricas no se le puede asignar una función de energía ya que éstas no permiten que el sistema alcance un estado de equilibrio. Para nuestro caso, es importante tener en mente que se considera para la segunda subred un hamiltoniano condicionado a que la primera red llegue a estado de equilibrio y que la señal que ésta envía (asimétrica), funja como un campo externo para la segunda subred.

Sea el hamiltoniano en la segunda subred, dado por:

$$H_2 = -\frac{1}{2} \sum_{ij} J_{ij} S_i S_j - \sum_i^N w_{i+N,i} S_{i+N} S_i. \quad (23)$$

Aquí el primer término de la expresión incluye la interacción entre cada par de neuronas de la segunda subred, de tal manera que están almacenados un número  $p$  finito de patrones fijos  $\{\xi_i^\mu\}$  al azar de acuerdo con la regla de Hebb dada por la ecuación (21). El segundo término incluye la interacción que existe entre las dos subredes. Esta interacción dada por la ecuación (22), como ya se mencionó con anterioridad, es unidireccional, es decir, cada neurona de la primera subred envía información exclusivamente a una neurona de la segunda subred, siendo esta una señal persistente, esto es, que no cambia en el tiempo.

Sustituyendo las expresiones de ambas interacciones en la ecuación (23) y suponiendo que el estado dinámico de la primera subred se encuentra en la memoria uno  $S_{i+N} = \xi_{i+N}^1$ , se tiene que el Hamiltoniano que describe a la segunda subred queda expresado por:

$$H_2 = -\frac{1}{2} \sum_{ij} J_{ij} S_i S_j - \sum_i^N h_i S_i, \quad (24)$$

en donde  $h_i$  es la señal persistente que actúa como umbral para las neuronas de la segunda subred y está dada por:

$$\begin{aligned} h_i &= \sum_{\mu=1}^p \xi_i^\mu \xi_{i+N}^\mu \xi_{i+N}^1 \\ &= \xi_i^1 + \xi_{i+N}^1 \sum_{\mu=2}^p \xi_i^\mu \xi_{i+N}^\mu. \end{aligned} \quad (25)$$

Ahora bien, haciendo un análisis semejante al del capítulo anterior con el fin de obtener una expresión analítica para los parámetros de orden o traslapes, consideramos la función de partición dada por:

$$Z = \text{Tr} \exp\{-\beta H_2\}, \quad (26)$$

en donde ahora  $H_2$  está dado por la expresión (24) y  $\beta = 1/T$  está relacionado con el nivel de ruido  $T$  de la red neuronal.

Procediendo en la forma usual, con el fin de linealizar, hacemos uso de la identidad de Hubbard-Stratonovich [Hubbard, 1959], dada por:

$$\exp\{a^2/b\} = (b/4\pi)^{1/2} \int_{-\infty}^{\infty} \exp\{-\frac{b}{4}x^2 + ax\} dx, \quad (27)$$

en donde, para nuestro caso,  $a$  y  $b$  están representadas por:

$$\begin{aligned} a &= \beta(\sum_i \xi_i^\mu S_i) \\ b &= 2N\beta \end{aligned}, \quad (28)$$

de esta manera, antes de calcular la traza, la función de partición queda expresada por:

$$Z = \left(\frac{N\beta}{2\pi}\right)^{p/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{N\beta}{2} \sum_{\mu=1}^p \vec{q}^2\right\} \times \text{Tr}_{S_i} \prod_i \exp\{\beta \sum_{\mu=1}^p \xi_i^\mu S_i \vec{q} + \beta h_i S_i\} d\vec{q}, \quad (29)$$

en donde  $\{q_\mu\}$  es una variable auxiliar introducida con el fin de linealizar el término  $(\sum_i \xi_i^\mu S_i)^2$  (apéndice A).

Calculando la traza de la ecuación (29), podemos expresar la función de partición como:

$$Z = C \int_{-\infty}^{\infty} \exp[-N\beta f_{ef}] d\vec{q}, \quad (30)$$

en donde  $f_{ef}$  es una función efectiva dada por:

$$f_{ef} = \frac{1}{2} \sum_{\mu=1}^p \vec{q}^2 - \frac{1}{N\beta} \sum_i \ln 2 \cosh\{\beta \sum_{\mu} \xi_i^\mu \vec{q} + \beta h_i\}. \quad (31)$$

Recordemos que para un sistema de este tipo, las propiedades de equilibrio a nivel macroscópico están caracterizadas por los parámetros de orden  $q_\mu$ , que miden el parecido entre el estado dinámico del sistema y el  $\mu$ -ésimo patrón. En el límite termodinámico ( $N \rightarrow \infty$ ), la integral de la ecuación (30) está dominada por el mínimo de  $f_{ef}$ , por lo que es posible hacer un desarrollo en serie de Taylor alrededor de estos puntos. Así los parámetros  $q_\mu$  están determinados por las condiciones  $\frac{\delta f}{\delta q_\mu} = 0$ . Considerando esto obtenemos:

$$\begin{aligned}
q_\mu &= \frac{1}{N} \sum_i \xi_i^\mu \tanh \left\{ \beta \left( \sum_\mu \xi_i^\mu q_\mu + h_i \right) \right\} \\
&= \left\langle \xi_i^\mu \tanh \left\{ \beta \left( \sum_\mu \xi_i^\mu q_\mu + h_i \right) \right\} \right\rangle
\end{aligned} \tag{32}$$

en donde  $h_i$  está dada por (25). La notación  $\langle \dots \rangle_\xi$  se refiere al promedio sobre las variables al azar  $\xi_i^\mu$  con probabilidad:

$$P(\xi_i^\mu) = \frac{1}{2} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)]. \tag{33}$$

Si recordamos que nos encontramos en un espacio  $p$  dimensional entonces, como se consideró en el capítulo anterior, es posible aplicar un autopromedio [Van Hemmen, 1986] (apéndice **B**) a la ecuación (32). Este consiste en no considerar un promedio sobre las variables aleatorias  $\{\xi_i^\mu\}$  sino un promedio de variables macroscópicas  $\eta_\kappa = \pm 1$  con  $\kappa = 1, \dots, 2^{p-1}$  sobre  $2^p$  esquinas de un hipercubo que incluye todas las posibles realizaciones de los  $p$  patrones. Con esto pondremos atención en las variables  $\xi_i^\mu$  y  $\xi_{i+N}^\mu$ .

Definamos:

$$x_\mu = \xi_{i+N}^\mu \xi_{i+N}^1. \tag{34}$$

De esta manera la expresión (32) queda expresada por:

$$q_\mu = \left\langle \eta^\mu \tanh \left\{ \beta \left[ (q_1 + 1) \eta^1 + \sum_{\mu=2}^p (q_\mu + x_\mu) \eta^\mu \right] \right\} \right\rangle_{\eta^\mu x^\mu}. \tag{35}$$

Es importante tener claro que las variables  $\eta^\mu$  y  $x^\mu$  también están distribuidas de acuerdo con la expresión (33). Nótese, que en tal caso el problema tiene simetría  $q_\mu \leftrightarrow -q_\mu$  para toda  $\mu$ , . Por lo que de aquí en adelante nos restringiremos al caso  $q_\mu \geq 0$ .  $q_\mu$  es la coordenada  $\mu$ -ésima del vector  $\vec{q}$ .

La expresión (35) es una ecuación que tiene un gran número de soluciones, siendo algunas de éstas puntos fijos de la dinámica del sistema. Para un número finito de patrones  $p$  almacenados en en el límite termodnámico ( $N \rightarrow \infty$ ), existen puntos fijos relacionados con cada una de las  $p$  memorias puras almacenadas en la red, sin embargo, como se mencionó en el capítulo anterior, también existen los puntos fijos relacionados con cualquier combinación de  $s$  memorias puras, donde  $3 \leq s \leq p$ .

Procediendo de la manera usual (apéndiceA), para el caso de dos y tres memorias almacenadas en cada subred se obtuvieron las siguientes expresiones para los parámetros de orden:

- Dos memorias almacenadas:

$$q_1 = \frac{1}{4}[\tanh\{\beta(q_1 + q_2)\} + \tanh\{\beta(q_1 + q_2 + 2)\} + \tanh\{\beta(q_1 - q_2)\} + \tanh\{\beta(q_1 + q_2 + 2)\}]$$

$$q_2 = \frac{1}{4}[\tanh\{\beta(q_1 + q_2)\} + \tanh\{\beta(q_1 + q_2 + 2)\} - \tanh\{\beta(q_1 - q_2)\} - \tanh\{\beta(q_1 - q_2 + 2)\}]$$

- Tres memorias almacenadas:

$$q_1 = \frac{1}{8}[\tanh A + \tanh B + \tanh C + \tanh D] + \frac{1}{16}[\tanh E + \tanh F + \tanh G + \tanh H + \tanh I + \tanh J + \tanh K + \tanh L]$$

$$q_2 = \frac{1}{8}[\tanh A + \tanh B - \tanh C - \tanh D] + \frac{1}{16}[\tanh E + \tanh F - \tanh G - \tanh H + \tanh I + \tanh J - \tanh K - \tanh L]$$

$$q_3 = \frac{1}{8}[\tanh A - \tanh B - \tanh C + \tanh D] + \frac{1}{16}[\tanh E - \tanh F - \tanh G + \tanh H + \tanh I - \tanh J - \tanh K + \tanh L]$$

En donde  $A, B, C, D, E, F, G, H, I, J, K$  y  $L$  están dadas por:

$$\begin{aligned} A &= \beta(q_1 + q_2 + q_3 + 1) \\ B &= \beta(q_1 + q_2 - q_3 + 1) \\ C &= \beta(q_1 - q_2 - q_3 + 1) \\ D &= \beta(q_1 - q_2 + q_3 + 1) \\ E &= \beta(q_1 + q_2 + q_3 - 1) \\ F &= \beta(q_1 + q_2 - q_3 - 1) \\ G &= \beta(q_1 - q_2 - q_3 - 1) \\ H &= \beta(q_1 - q_2 + q_3 - 1) \\ I &= \beta(q_1 + q_2 + q_3 + 3) \\ J &= \beta(q_1 + q_2 - q_3 + 3) \\ K &= \beta(q_1 - q_2 - q_3 + 3) \\ L &= \beta(q_1 - q_2 + q_3 + 3) \end{aligned} \tag{36}$$

Hasta aquí, los parámetros de orden obtenidos nos dan una idea de la fracción de espacio ocupada por los patrones. Sin embargo, si se desea medir el tamaño de las cuencas de atracción es necesario hacer un estudio dinámico del problema. Para esto se considera la evolución en el tiempo de los parámetros  $q$ , es decir:

$$\mathbf{q}(n+1) = F(\mathbf{q}(n)) \quad (37)$$

en donde

$$F(\mathbf{q}) = \left\langle \eta \tanh \left\{ \beta \left[ (q_1 + 1) \eta^1 + \sum_{\mu=2}^p (q_\mu + x_\mu) \eta^\mu \right] \right\} \right\rangle_{\eta \mathbf{x}}. \quad (38)$$

De esta manera los puntos fijos (atractores) de la dinámica, i.e., los puntos que satisfacen  $F^n(\mathbf{q}) = \mathbf{F}^{n-1}(\mathbf{q})$ , para algún valor de  $n$ , corresponden a puntos estables dados por la ecuación (35).

### III. 2. 1. 1 Solución de las ecuaciones de flujo (método iterativo).

En la sección anterior encontramos las expresiones de los parámetros de orden o traslapes para los casos de dos y tres memorias almacenadas en cada una de las subredes respectivamente. Como es de notarse éstas involucran la solución de  $p$  ecuaciones trascendentales, correspondientes a cada una de las coordenadas del vector  $\vec{q}$ , por lo que

es necesario recurrir a un método alternativo tal que nos permita tener una idea más clara de lo que sucede en el espacio de configuraciones. Un ejemplo sería localizar un punto fijo en dicho espacio a partir de un estado inicial.

Debido a lo anterior, se consideró iterar las ecuaciones con los siguientes procedimientos:

- Enfriado súbito; de acuerdo con una distribución de probabilidad uniforme entre  $+1$  y  $-1$  a temperatura  $T = 0$ . Se generaron valores al azar para las  $q$ 's iniciales. Se iteraron las ecuaciones de flujo, tanto para dos como para tres memorias, hasta que el valor de  $q$  no cambiara. Se observó que independiente del valor inicial de  $q$  el traslape final era el mismo para cualquiera de los casos, sin embargo el traslape final relacionado con la memoria uno era mucho mayor que los traslapes finales relacionados con el resto de la memorias.
- Enfriado lento; éste consistió en empezar la iteración a partir de un estado inicial para las  $q$ 's entre  $+1$  y  $-1$  a temperatura relativamente alta  $T = 0.3$ . Una vez que se llegó a un punto fijo, es decir, que el traslape final a la temperatura inicial se mantuviera constante, se bajaba la temperatura y se consideraba como  $q$  inicial el traslape final anterior, esto se hizo hasta llegar a  $T = 0$ . En 100 redes diferentes se observó, por un lado, que al descender la temperatura la calidad de recuperación del patrón uno en la segunda subred era mayor con respecto a los otros, y por el otro, que independientemente del estado inicial el traslape final era el mismo.

### III. 2. 1. 2 Conclusiones

En esta sección se encontraron, principalmente, dos resultados de suma importan-

cia. El primero de ellos es que, al menos para los casos considerados, el traslape final  $q_f$  no depende del valor inicial de  $q_i$  para cualquier valor de  $T$ , y el segundo (como se esperaba) es que invariablemente la segunda subred recupera el patrón uno. Sin embargo, éste se encuentra desplazado obteniendo, para el caso de  $p = 3$ , un traslape final del orden de  $q_f = 0.5$ , es decir, con el 25 por ciento de los espines volteados, mientras que el traslape con otras memorias es típicamente menor que .1.

Hasta aquí hemos realizado un estudio que concierne al caso de  $p$  finita  $N \rightarrow \infty$ . En general, nos hemos topado con limitaciones de tipo analítico para encontrar las ecuaciones de flujo para  $p > 3$ . Con el fin de extender el número de memorias almacenadas y poder comparar con los resultados obtenidos en esta sección procederemos a hacer un estudio microscópico de este problema.

### III. 2. 2 Estudio microscópico de las redes acopladas.

El objetivo principal en este capítulo es estudiar el tamaño de las cuencas de atracción para el caso de las redes acopladas, pues esto nos permite visualizar el comportamiento conjunto de ambas redes para almacenar y recuperar la información. Con este propósito se hace un estudio a  $T = 0$  del comportamiento dinámico de una red neuronal compuesta de dos subredes tipo Hopfield con conexiones unidireccionales que llevan la información de la primera subred a la segunda. Retomando la convención original, se considera que en cada una de las subredes hay  $N$  elementos cuyos estados dinámicos están descritos por  $S_i = \pm 1$ , de manera que para la primera subred tenemos  $i = 1, \dots, N$  mientras que para la segunda  $i = N + 1, \dots, 2N$ . El estado de cada neurona al tiempo  $t$  está gobernado por la dinámica siguiente:

$$S_i(t+1) = \text{sgn} \{h_i^k(t)\}, \quad (39)$$

en donde  $h_i$  es el campo local que siente cada neurona  $i$  causado por el resto de las neuronas que la rodean en la  $k$ -ésima subred. Para la primera subred este campo está expresado de la siguiente manera:

$$h_i^1 = \sum_{j=1}^N J_{ij} S_j, \quad (40)$$

mientras que para la segunda subred se escribe como sigue:

$$h_i^2 = \sum_{j=1+N}^{2N} J_{ij} S_j + \sum_{j=1}^N w_{i,j} S_j \delta_{j+N,i}, \quad (41)$$

en donde la interacción  $J_{ij}$  es simétrica y está dada por la regla de Hebb (ecuación 21) y  $w_{ij}$  es la interacción existente entre las dos subredes (ecuación 22) siendo ésta unidireccional mandando exclusivamente información de un elemento de la primera subred a un elemento de la segunda. Cabe recordar que son las interacciones las que determinan los mínimos en el espacio de las energías.

Una manera de caracterizar la recuperabilidad de la información en cualquiera de las subredes, es estudiando al sistema en términos de los traslapes microscópicos, los cuales miden el parecido o la diferencia entre el estado dinámico de cada una de las subredes y el  $\mu$ -ésimo estado nominal, y están dados por:

$$m_{1(2)}^\mu = \frac{1}{N} \sum_{i=1+l}^{N+l} S_i \xi_i^\mu, \quad (42)$$

aquí  $S_i$  es el estado dinámico y  $\xi_i^\mu$  es el patrón almacenado en la red. Para  $l = 0$  nos referimos a la primera subred, mientras que para  $l = N$  nos referimos a la segunda. En esta expresión el traslape  $m_{1(2)}^\mu$  se refiere al patrón  $\mu$ -ésimo de la primera o la segunda subred además, éste ha sido normalizado a la unidad para cada subred por lo que la fracción de unidades (bits), que coinciden con el subpatrón  $\mu$ , está dada por  $\frac{1}{2}(1 + m_{1(2)}^\mu)$ . En este trabajo nos referiremos a la recuperabilidad de la memoria uno, por lo que nos ahorraremos el índice que indica  $\mu = 1$ . Cuando consideremos otros traslapes, lo mencionaremos explícitamente. De igual manera para referirnos a los traslapes iniciales o finales usaremos como subíndices “0” ó “ $f$ ” respectivamente.

### III. 2. 2. 1 Simulación numérica.

Es importante tener en mente que el comportamiento de la primera subred no se ve afectado por la presencia de la segunda, por lo que escogemos que el estado inicial en la primera subred sea  $S_i = \xi_i^1$  para  $i = 1, \dots, N$ . Una vez que esta subred haya alcanzado dicho estado mandará una señal persistente hacia la segunda subred, lográndose que el estado final alcanzado por la misma esté relacionado con el estado recuperado en la primera.

Dentro de la simulación se efectuaron dos diferentes estudios. (a) El primero de ellos consistió en almacenar  $p$  patrones  $\vec{\xi}_i^\mu$  en la red de acuerdo con las reglas dadas por las expresiones (21) y (22). Ahora bien, con el fin de medir el tamaño de las cuencas de atracción en la segunda subred se varió el traslape inicial  $m_{02}$  y se estudió la recuperabili-

dad como función de dicho traslape. Se vió que casi para toda  $p$  y  $N$  el estado final ( $m_{f2}$ ) de la memoria uno es independiente del estado inicial. Más adelante explicaremos a qué nos referimos con “casi para toda  $p$  y  $N$ ”. (b) El segundo estudio llevó como propósito determinar la existencia del desplazamiento de la “posición” de los estados de mínima energía, ya que se observó tanto en el estudio macroscópico como el estudio arriba mencionado, que dicha señal persistente tiende a mover en la segunda subred la posición de los mínimos con respecto al estado almacenado  $\xi_1^1$ . Se encontró que efectivamente ocurre este fenómeno cuando varían los parámetros  $\alpha$  y  $p$ . Para mostrar esto, se realizó la simulación que se describe a continuación:

Se almacenaron  $p$  patrones  $\vec{\xi}_i^\mu$  de acuerdo a la regla de Hebb (21), para que de esta manera estos fueran puntos fijos de la dinámica (39):

$$\vec{\xi}_\mu = (\xi_i^\mu, \dots, \xi_N^\mu, \xi_{N+1}^\mu, \dots, \xi_{2N}^\mu), \quad (43)$$

aquí los primeros  $N$  elementos se refieren a la primera subred y los restantes a la segunda. Posteriormente se generó un estado inicial tal que tuviera un traslape  $m_{01}^1 = 1$  con la memoria uno de la primera subred, y con traslape arbitrario  $m_{02} > 0$  con la segunda subred. De esta forma, cuando la primera subred se encuentra en estado de equilibrio, una vez que ha recuperado la memoria uno, mandará una señal persistente hacia la segunda subred. En este punto se midió el traslape inicial de la segunda subred con la memoria uno definido por:

$$m_{02} = \frac{1}{N} \sum_{i=1}^N S_{i+N} \xi_{i+N}^1, \quad (44)$$

para después permitir que el sistema completo evolucione sincrónicamente de acuerdo con la ecuación (39) hasta que la segunda subred alcance un estado de equilibrio. Por último se midieron los traslapes finales, con cada una de las memorias, para ver si la memoria uno fue recuperada.

### III. 2. 2. 2 Resultados y discusión.

Con el fin de evaluar el tamaño de las cuencas de atracción en la segunda subred se tomó como criterio de recuperabilidad que el traslape final  $m_f$  del patrón uno en la segunda subred fuese considerablemente más grande que el valor del traslape con respecto a los otros patrones de la misma. El número de patrones almacenados de acuerdo con las expresiones (21) y (22) varió desde  $p = 3$  hasta  $p = 120$  para diferentes RN tomando valores para  $N = 32$  hasta  $N = 1024$ .

El desplazamiento del patrón uno en la segunda subred está relacionado con la señal persistente proveniente de la primera subred, cuando ésta ha alcanzado su estado de equilibrio es decir,  $S_i = \xi_i^1$  para  $i = 1, \dots, N$ . El efecto que ejerce dicha señal sobre la segunda subred, es el de conducir al estado dinámico de la misma lo más cerca posible de la cuenca que está relacionada con el patrón uno (patrón nominal) de la primera subred. Dicho desplazamiento de los patrones almacenados en la segunda subred con respecto a los patrones nominales de la primera se muestra en la figura (13), que es una muestra representativa para el caso  $N = 100$  y distintos valores de  $p$ . En estas gráficas se observa el porcentaje de veces para el cual se obtiene un traslape final  $m_f$  en 1000 redes diferentes con mismo traslape inicial  $m_0$ . También vemos que se presenta un comportamiento que depende de la paridad de  $p$ , es decir, cuando  $p$  es par se presenta un corrimiento hacia la derecha mientras que para valores no pares de  $p$  el corrimiento es hacia la izquierda. Por otro lado, es interesante notar que tanto el porcentaje de recuperación como la calidad

de la misma es mayor para valores pares de  $p$  ( $p = 4, 6, \dots$ ) que para valores impares de  $p$  ( $p = 3, 5, \dots$ ) figuras (14) y (15) respectivamente.

Los valores del traslape final para casi la totalidad de las redes eran los mismos independientemente del estado inicial en la segunda subred. Esto lo podemos ver en la figura (16) en la que se muestra la relación entre los traslapes iniciales y finales de la segunda subred considerando 20 redes distintas para valores de  $p = 4$  y  $N = 64$ . Nótese (figura 16(a)) que para valores del traslape inicial  $m_o = -1$  el sistema evoluciona a un valor final con el mismo traslape  $m_f$  que los estados con traslape inicial  $m_o = 1$ . También existen otros casos en los que la información es recuperada para cualquier estado inicial, sólo que la calidad de la información recuperada depende del traslape inicial figura (16(b),(c),(d)) en estos casos al parecer la cuenca de atracción relacionada con el patrón uno esta compuesta a su vez de pequeñas cuencas que no permiten al sistema recuperar en su totalidad la información. Este comportamiento se encontró para cualquier valor de  $N$  pero únicamente para valores pares del número de patrones almacenados.

Los resultados obtenidos en esta sección, en particular para  $p = 3$  en las figura (15) nos da un punto de concordancia con respecto al estudio macroscópico. Esto es, la calidad de recuperación, para ambos estudios, es de  $\langle m_f \rangle = 0.5$  independientemente del estado inicial. Por otro lado, para  $p = 3$  en la figura (13), se puede observar el desplazamiento de la memoria uno en la segunda subred con respecto a la memoria uno ( $m_{f1} = 1$ ) de la primera subred.

Finalmente la cantidad de información necesaria para recuperar la información almacenada en una red neuronal como la descrita a lo largo del capítulo, es menor que para una típica RN tipo Hopfield. Esto es debido a la existencia de la señal persistente que va de una de las subredes a la otra, actuando como umbral para las neuronas de la subred que le sigue.

Con esto y con los resultados obtenidos en este capítulo se sugiere la posibilidad

de formar una cadena compuesta de varias subredes, de tal manera que éstas trabajen acopladamente al almacenar y recuperar la información. Un estudio alternativo a seguir es medir la recuperabilidad de la información en la segunda subred una vez que la señal persistente proveniente de la primera subred sea desconectada, es decir, cuando dicha señal haya colocado a la segunda subred en un estado cercano a la cuenca en cuestión.

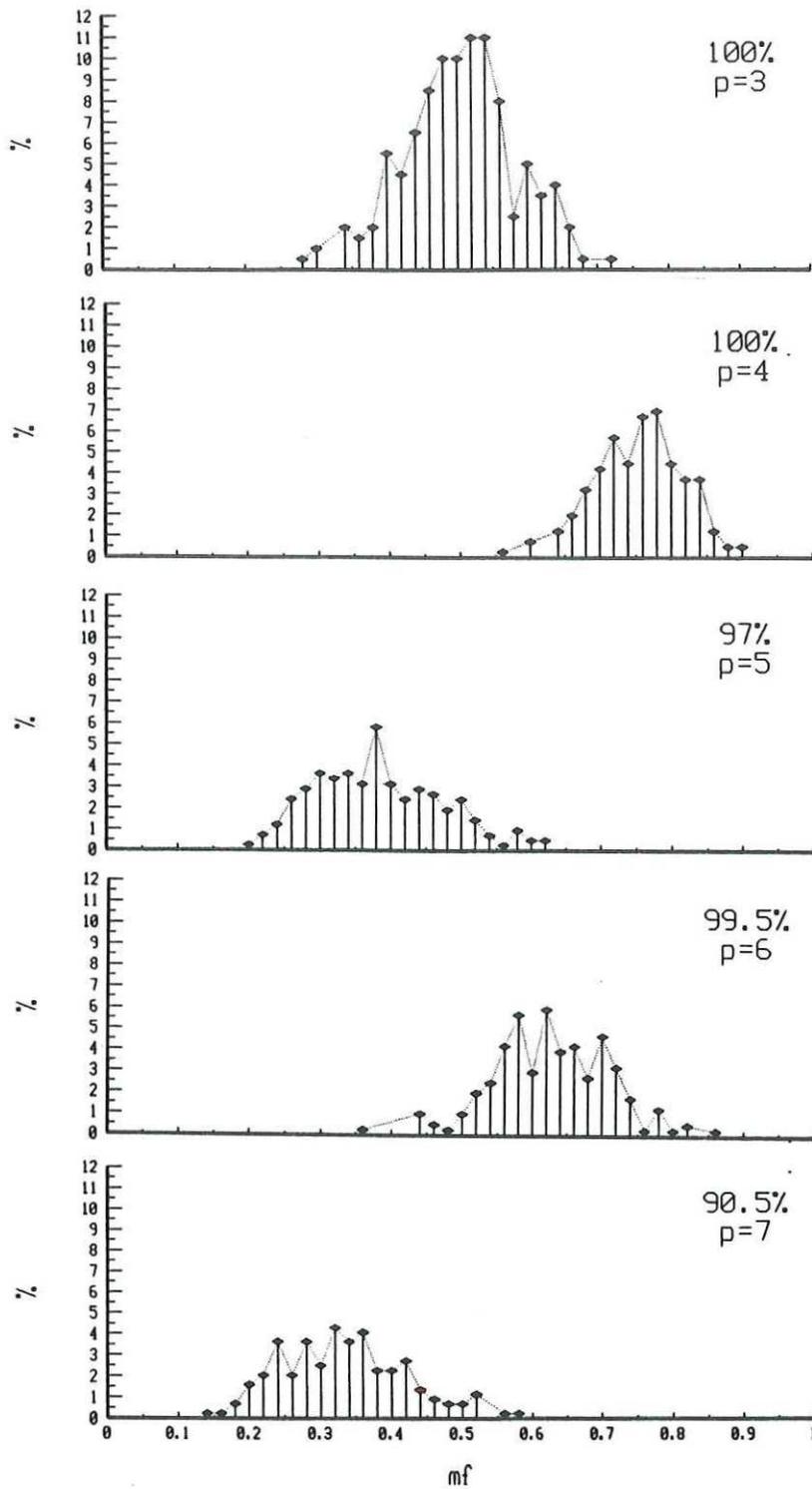


Figura 13: Porcentaje de recuperación de la segunda subred como función del traslape final  $m_f$  sobre un conjunto de 1000 redes con  $N = 100$  y distintos valores de  $p$ . En el extremo superior derecho se indica el porcentaje total de veces que se recuperó un estado con traslape macroscópico  $m_f$  con valores distribuidos de acuerdo con las figuras.

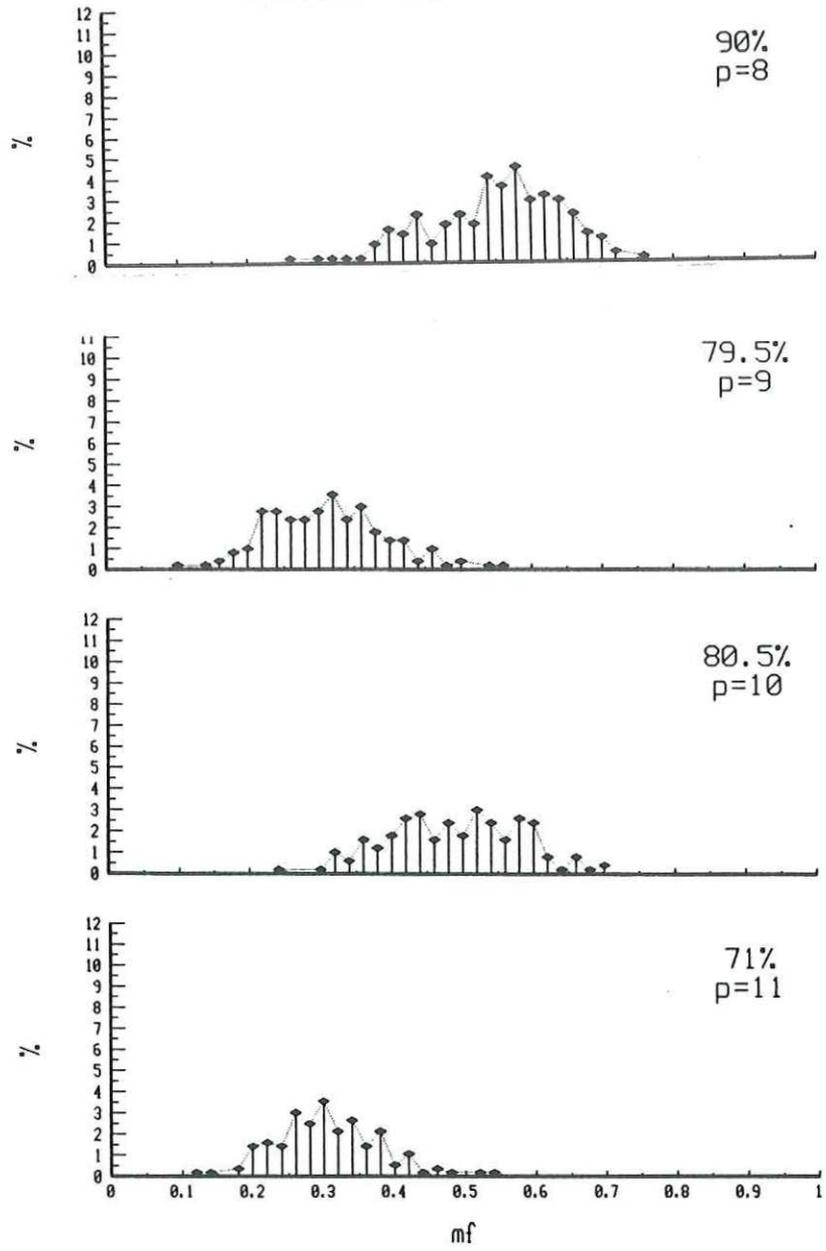


Figura 13: (Continuación)

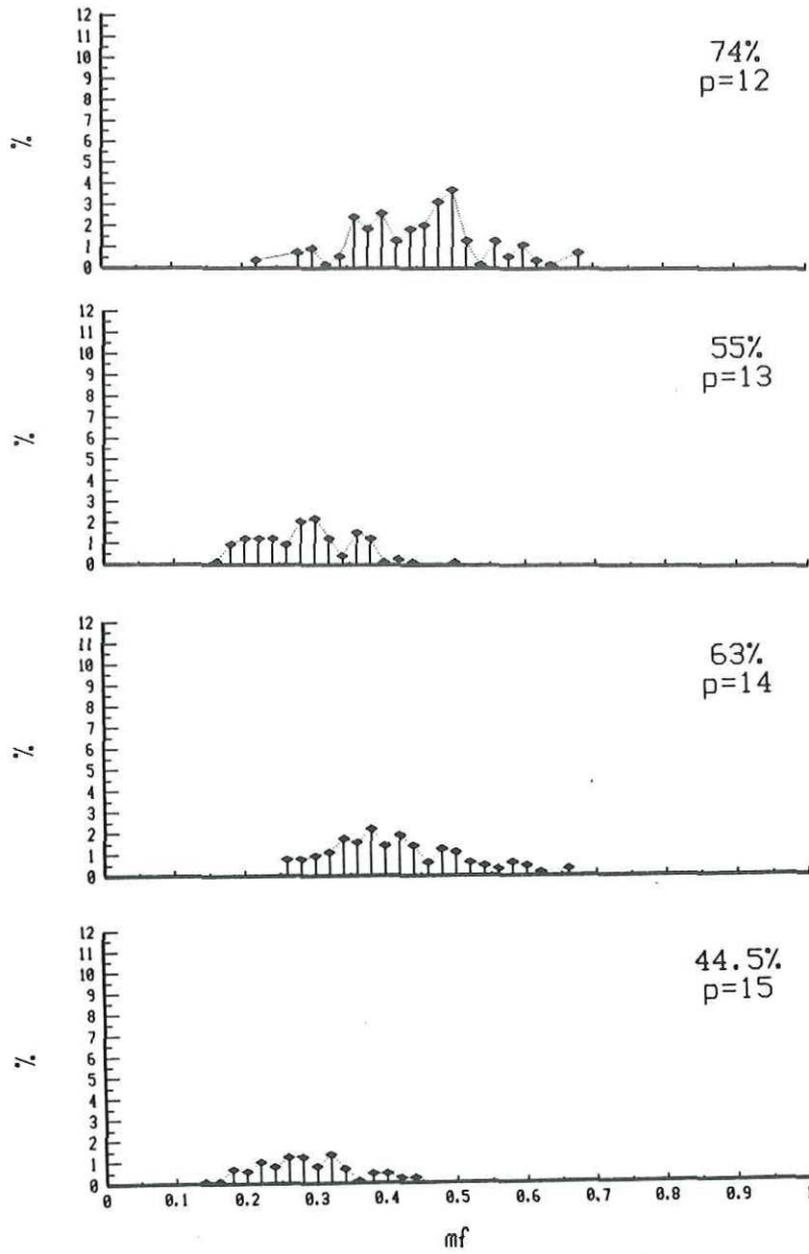


Figura 13: (Continuación)

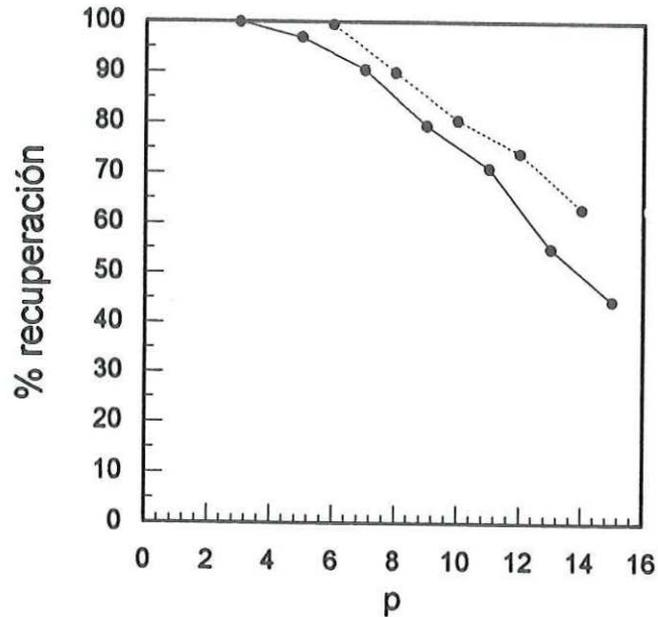


Figura 14: Porcentaje de recuperación como función del número de patrones almacenados  $p$  para la segunda subred con  $N = 100$  elementos. La línea sólida representa valores impares de  $p$  y la línea punteada valores pares. Desviación estandar del 10%.

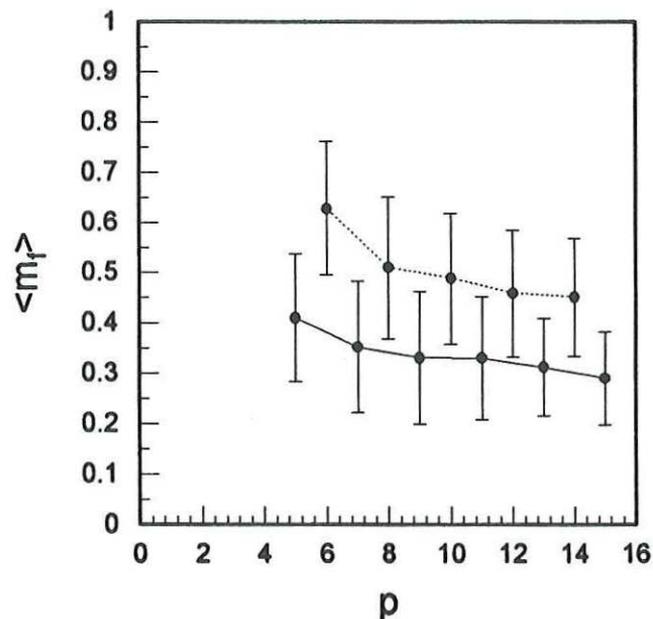


Figura 15: Valor promedio sobre 1000 muestras del traslape final  $m_f$  como función de  $p$  para  $N = 100$ . La línea sólida representa valores impares de memorias almacenadas ( $p = 3, 5, 7, 9, 11, 13, 15$ ) y la línea punteada representa valores pares de  $p$  ( $p = 2, 4, 6, 8, 10, 12, 14$ ).

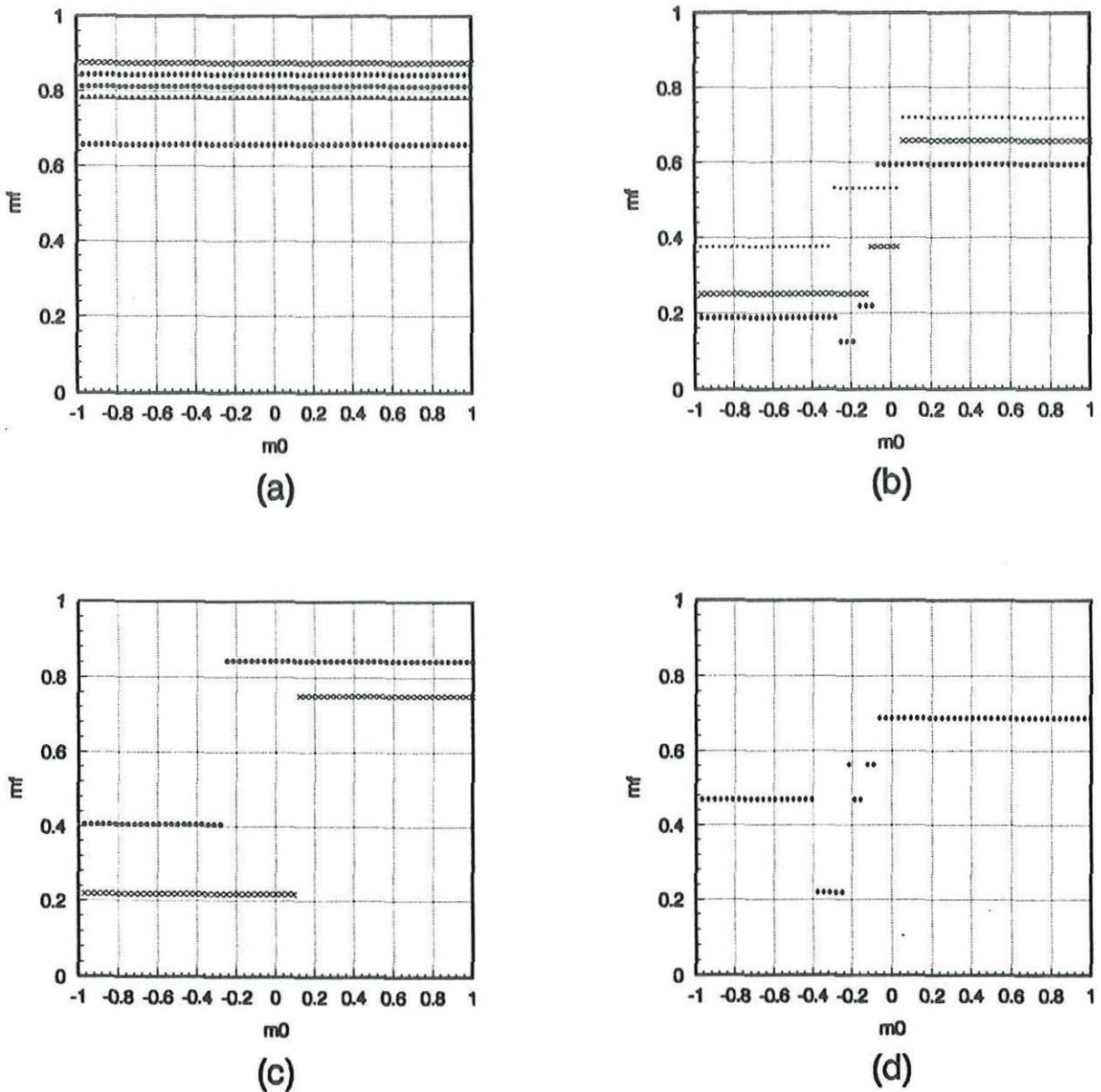


Figura 16: Relación entre el traslape inicial  $m_0$  y el traslape final  $m_f$  para la segunda subred ( $N = 64$  y  $p = 4$ ) cuando la primera subred se encuentra en un estado estable. La gráfica (a) muestra casos típicos, sin embargo existen pocos casos como los mostrados en las gráficas (b), (c) y (d) en los que el traslape final depende del traslape inicial. Este comportamiento sólo tiene dependencia en  $p$  par.

## LITERATURA CITADA

- Amit, D., H. Gutfreund y H. Sopolinsky, 1985. "Spin-Glass Models of Neural Networks". *Physical Review A* **32**. pp. 1007-1018.
- Coolen, A. C. C., H. J. J. Jonker y Th. W. Ruijgrok, 1989. "Size of the Domains of Attraction in the Hopfield Model". *Phys. Rev.* **40**. pp. 5295.
- Cottrell. 1988. *Biol. Cybern.* **58**. pp. 129.
- Einstein, A., 1902. "Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik". *Ann. Physik.* **9**. pp. 417.
- Einstein, A., 1902. "Eine Theorie der Grundlagen der Thermodynamik". *Ann. Physik.* **11**. pp. 170.
- Einstein, A., 1902. "Zur allgemeinen molekularen Theorie der Wärme". *Ann. Physik.* **14**. pp. 354.
- Forrest, B. M., 1988. "Content-Addressability and Learning in Neural Networks". *J. Phys. A: Math. Gen.* **21**. pp. 245-255.
- Gardner, E., H. Gutfreund, y I. Yekutieli, 1989. "The Phase Space of Interactions in Neural Networks with Definite Symmetry". *Journal of Physics A* **22**. pp. 1995-2008.
- Gibbs, J. W., 1902 *Elementary Principles in Statistical Mechanics*, Yale University Press. Reimpreso por Dover, 1963.
- Hebb, D. O., 1949. *The Organization of Behavior*. New York: Wiley. Reimpreso parcialmente por Anderson y Rosenfeld (1988).
- Hopfield, J. J., 1982. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities". *Proceedings of the National Academy of Sciences. USA* **79**, Reimpreso por Anderson y Rosenfeld (1988).
- Hubbard, J., 1959. "Calculation of Partition Functions". *Phys. Rev. Lett.* **3**. pp. 77.
- Kandel, E. R., 1979. "Small Systems of Neurons" *Scientific America.* **241**. pp. 66.
- Kepler, T. B. y L. F. Abbott, 1988. "Domains of Attraction in Neural Networks". *J. Phys. France* **49**. pp. 1656.
- Krauth, J. P. Nadal y M. Mezard, 1988. *J. of Phys. A* **21**. pp. 2995.
- Little, W. A., 1974. "The Existence of Persistent States in the Brain". *Mathematical Biosciences.* **19**. pp. 101-120.

- Martínez, C., 1989. *Física Estadística de Sistemas Complejos*. Tesis de Licenciatura, Facultad de Ciencias UNAM.
- Martínez, C., E. Cota y L. Viana, 1990. "Spin Glass Model for a Neural Network: Configuration Space Studies". *Revista Mexicana de Física*, **36**. pp. 415-423.
- McCulloch, W. S. y W. Pitts, 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bull. Math. Biophys.* **5**. pp. 115.
- Van Hemmen, J. L., D. Greising, A. Huber y R. Kuhn, 1986. *Z. Phys.* **B65**. pp.53.
- Viana, L. 1988., "Spin Glass Model for a Neural Network: Associative Memories Stored with Unequal Weights". *J. Phys. France* **49**. pp. 167-174.
- Viana, L. 1991., *Memoria Natural y Artificial*. La ciencia desde México, 88. Fondo de Cultura Económica.
- Viana, L. y A. C. C. Coolen, 1993. "Attraction Domains in Neural Networks". *J. Phys. I France* **3**. pp. 777-786.

## A. ECUACIONES DE FLUJO

El propósito de este apéndice es exponer en detalle la manera de obtener la expresión analítica para las ecuaciones de flujo o traslapes, que miden el parecido entre un patrón almacenado en la red y el estado dinámico de la misma.

El estudio sistemático consiste, primeramente, en calcular la función de partición, ya que ésta nos permite obtener las cantidades deseadas después de ciertas consideraciones como son: el suponer que el sistema se encuentra a volumen  $V$ , número de elementos  $N$  y energía fijos.

$$Z = Tr \exp \{-\beta H\}, \quad (45)$$

en donde la traza,  $Tr$ , se refiere a la suma sobre todos los estados posibles  $\{S_i = \pm 1\}$  y  $H$  la función de energía:

$$H = \frac{1}{2} \sum_{i \neq j}^N J_{ij} S_i S_j. \quad (46)$$

En esta expresión  $J_{ij}$  está dada por la regla de Hebb modificada,

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p w_{\mu} \xi_i^{\mu} \xi_j^{\mu}. \quad (47)$$

donde  $\xi_i^{\mu} = \pm 1$  son patrones escogidos al azar con la misma probabilidad.

Sustituyendo (47) en (46) y analizando esta expresión obtenemos:

$$H = -\frac{1}{2N} \sum_{\mu=1}^p w_{\mu} \sum_{ij}^N \xi_i^{\mu} \xi_j^{\mu} S_i S_j.$$

Manipulando ésta de la siguiente manera:

$$\begin{aligned} H &= -\frac{1}{2N} \sum_{\mu=1}^p w_{\mu} \left[ \sum_{i \neq j}^N \xi_i^{\mu} \xi_j^{\mu} S_i S_j - \sum_i (\xi_i^{\mu})^2 (S_i)^2 \right] \\ &= -\frac{1}{2N} \sum_{\mu=1}^p w_{\mu} \left[ \sum_i \xi_i^{\mu} S_i \sum_j \xi_j^{\mu} S_j + \frac{p}{2} \right] \end{aligned}$$

obtenemos el Hamiltoniano:

$$H = -\frac{1}{2N} \sum_{\mu=1}^p w_{\mu} \left[ \sum_i \xi_i^{\mu} S_i \right]^2, \quad (48)$$

en donde la constante  $\frac{p}{2}$  ha sido eliminada.

Dado que nuestro punto de partida es calcular la función de partición, nos permitimos sustituir la expresión (48) en (45) obteniendo:

$$Z = Tr \exp \left\{ \frac{\beta}{2N} \sum_{\mu=1}^p w_{\mu} \left( \sum_i \xi_i^{\mu} S_i \right)^2 \right\}, \quad (49)$$

esta expresión sería fácil de evaluar si el término de la exponencial fuera lineal en  $S_i$ , así, la traza sería el producto de  $N$  términos independientes, uno para cada  $i$ . Desafortunadamente esto no es así, ya que como vemos el término de la exponencial es cuadrático, lo que nos lleva a hacer uso de la identidad de Hubbard-Stratonovich [Hubbard, 1957] para linealizar dicho término:

$$e^{\frac{b^2}{4a}} = \left(\frac{a}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} \exp \left\{ -ax_{\mu}^2 + bx_{\mu} \right\} dx_{\mu}, \quad (50)$$

en donde  $a$  y  $b$ , para nuestro caso, estan dadas por:

$$\begin{aligned} a &= \frac{1}{2}N\beta w_\mu \\ b &= \beta w_\mu \sum_i \xi_i^\mu S_i \end{aligned}$$

Al hacer uso de la identidad (50) es necesario introducir las variables auxiliares  $x_\mu$ . En nuestro caso vamos a tener  $p$  términos de este tipo, uno para cada  $\mu$ :

$$Z = Tr \Pi_{\mu=1}^p \left( \frac{N\beta w_\mu}{2\pi} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2}N\beta w_\mu x_\mu^2 + \beta w_\mu x_\mu \sum_i \xi_i^\mu S_i \right\} dx_\mu. \quad (51)$$

Definamos la constante como:

$$C_1^\mu = \left( \frac{N\beta w_\mu}{2\pi} \right)^{\frac{1}{2}} \quad (52)$$

Separando cada término de la expresión (51) en sus respectivas componentes:

$$\begin{aligned} Z &= Tr C_1^\mu \left( \int e^{\{-\frac{1}{2}N\beta w_1 x_1^2 + \beta w_1 x_1 \sum_i \xi_i^1 S_i\}} dx_1 \times \dots \times \int e^{\{-\frac{1}{2}N\beta w_p x_p^2 + \beta w_p x_p \sum_i \xi_i^p S_i\}} dx_p \right) \\ &= Tr C_1^\mu \int \dots \int dx_1 dx_2 dx_3 \dots e^{\{-\frac{1}{2}N\beta w_1 x_1^2 + \beta w_1 x_1 \sum_i \xi_i^1 S_i - \frac{1}{2}N\beta w_2 x_2^2 + \beta w_2 x_2 \sum_i \xi_i^2 S_i - \dots\}} \\ &= Tr C_1^\mu \int \dots \int dx_1 dx_2 \dots e^{\{-\frac{1}{2}N\beta \{w_1 x_1^2 + w_2 x_2^2 + \dots\} + \beta \{w_1 x_1 \sum_i \xi_i^1 S_i + w_2 x_2 \sum_i \xi_i^2 S_i + \dots\}\}} \\ &= Tr C_1^\mu \int d\vec{x}_\mu e^{\left\{ -\frac{1}{2}N\beta \sum_\mu w_\mu \vec{x}_\mu^2 + \beta \sum_\mu w_\mu \vec{x}_\mu \sum_i \xi_i^\mu S_i \right\}} \\ &= C_1^\mu Tr \int d\vec{x}_\mu e^{\left\{ -\frac{1}{2}N\beta \sum_\mu w_\mu \vec{x}_\mu^2 \right\}} e^{\left\{ \beta \sum_\mu w_\mu \vec{x}_\mu \sum_i \xi_i^\mu S_i \right\}} \\ &= C_1^\mu Tr \int d\vec{x}_\mu e^{\left\{ -\frac{1}{2}N\beta \sum_\mu w_\mu \vec{x}_\mu^2 \right\}} \prod_i e^{\left\{ \beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu S_i \right\}} \end{aligned}$$

para finalmente obtener:

$$Z = C_1^\mu \int d\vec{x}_\mu e^{\left\{-\frac{1}{2}N\beta \sum_\mu w_\mu \vec{x}_\mu^2\right\}} Tr \Pi_i e^{\left\{\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu S_i\right\}}, \quad (53)$$

en donde  $\vec{x}_\mu$  es un vector de  $p$  componentes. Si observamos la expresión (53) vemos que el término cuadrático ha desaparecido por lo que ahora sí es posible calcular la traza sobre  $S_i$ , es decir:

$$\begin{aligned} Tr \Pi_i e^{\left\{\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu S_i\right\}} &= \Pi_i \left( e^{\left\{\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu\right\}} + e^{\left\{-\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu\right\}} \right) \\ &= \Pi_i 2 \cosh \left\{\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu\right\} \end{aligned}$$

en donde hemos utilizado la identidad:  $2 \cosh x = e^x + e^{-x}$ .

Sustituyendo la expresión obtenida después de aplicar la traza en (53) tenemos:

$$Z = C_1^\mu \int d\vec{x}_\mu \exp \{-Nf(x_\mu)\}, \quad (54)$$

en donde:

$$f(x_\mu) = \frac{1}{2}\beta \sum_\mu w_\mu \vec{x}_\mu^2 - \frac{1}{N} \sum_i \ln 2 \cosh \left\{\beta \sum_\mu w_\mu \vec{x}_\mu \xi_i^\mu\right\}. \quad (55)$$

La integral de la ecuación (54) puede resolverse por el método de *descenso por el camino más abrupto*. Este método es posible utilizarlo ya que por un lado, tenemos el caso de  $N \rightarrow \infty$  y por el otro, que la integral queda dominada por el valor mínimo que toma el integrando. Dado esto, es posible desarrollar en serie de Taylor alrededor del punto que minimiza la expresión, de esta manera el valor de los parámetros auxiliares queda determinado por la condición  $\frac{\delta f}{\delta x_\mu} = 0$ , es decir:

$$\begin{aligned}\frac{\partial f(x_\mu)}{\partial x_\mu} &= \beta w_\mu \bar{x}_\mu - \frac{1}{N} \sum_i \beta w_\mu \xi_i^\mu \tanh \left\{ \beta \sum_\mu w_\mu \bar{x}_\mu \xi_i^\mu \right\} \\ &= 0\end{aligned}$$

Obteniendo finalmente un conjunto de  $p$  ecuaciones acopladas dadas por:

$$\bar{x}_\mu = \frac{1}{N} \sum_i \beta w_\mu \xi_i^\mu \tanh \left\{ \beta \sum_\mu w_\mu \bar{x}_\mu \xi_i^\mu \right\},$$

o lo que es lo mismo:

$$\bar{x}_\mu = \left\langle \beta w_\mu \xi_i^\mu \tanh \left\{ \beta \sum_\mu w_\mu \bar{x}_\mu \xi_i^\mu \right\} \right\rangle_\xi. \quad (56)$$

En esta expresión hemos usado la notación  $\langle \dots \rangle_\xi$  para indicar el promedio sobre las variables al azar  $\{\xi_i^\mu\}$  dadas por la ecuación:

$$P(\xi_i^\mu) = \frac{1}{2} [\delta(\xi_i^\mu - 1) + \delta(\xi_i^\mu + 1)] \quad (57)$$

Ahora bien, una vez que hemos obtenido la expresión analítica para los traslapes (56) proseguimos a calcular explícitamente el valor del traslape para cada memoria. Para esto tengamos en mente que el promedio de una función está dado por:

$$\langle f \rangle = \int_{-\infty}^{\infty} P(x) f(x) dx, \quad (58)$$

en donde para nuestro caso la  $P(x)$  está dada por la expresión (57) y  $f(x)$  por la ecuación (56).

Finalmente después de un tedioso cálculo llegamos a las expresiones siguientes:

Para 2  $q$ 's diferentes de cero tenemos:

$$q_1 = \frac{1}{2}(\tanh \beta \{q_1 w_\mu + q_2 w_\mu\} + \tanh \beta \{q_1 w_\mu - q_2 w_\mu\})$$

$$q_2 = \frac{1}{2}(\tanh \beta \{q_1 w_\mu + q_2 w_\mu\} - \tanh \beta \{q_1 w_\mu - q_2 w_\mu\})$$

De igual manera para tres  $q$ 's diferentes de cero se obtiene:

$$q_1 = \frac{1}{4}(\tanh A + \tanh B + \tanh C + \tanh D)$$

$$q_2 = \frac{1}{4}(\tanh A + \tanh B - \tanh C - \tanh D)$$

$$q_3 = \frac{1}{4}(\tanh A - \tanh B + \tanh C - \tanh D)$$

en donde  $A$ ,  $B$ ,  $C$  y  $D$  están dadas por las expresiones siguientes:

$$A = \beta \{q_1 w_1 + q_2 w_2 + q_3 w_3\}$$

$$B = \beta \{q_1 w_1 + q_2 w_2 - q_3 w_3\}$$

$$C = \beta \{q_1 w_1 - q_2 w_2 + q_3 w_3\}$$

$$D = \beta \{q_1 w_1 - q_2 w_2 - q_3 w_3\}$$

## B. AUTO PROMEDIO

Un punto importante a tratar en la teoría de las redes neuronales es el del autopromedio, ya que este nos permite pasar de una representación en el espacio  $N$ -dimensional a una en el espacio de  $2^p$  dimensiones.

Consideraremos una red neuronal compuesta de  $N$  neuronas con  $p$  memorias almacenadas de acuerdo a la regla de Hebb:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} \quad (59)$$

en donde  $p$  es el número de patrones por grabar, y  $\{\xi_i^{\mu}\}$  es un patrón escogido al azar donde cada uno de los términos tiene probabilidad  $1/2$  igual a  $\pm 1$ .

Sea un arreglo de  $p$  vectores (que representan los estados almacenados de la red neuronal) dados por:

$$\vec{\xi}^1 = (\xi_1^1, \xi_2^1, \dots, \xi_i^1, \dots, \xi_N^1),$$

.

.

.

$$\vec{\xi}^\mu = (\xi_1^\mu, \xi_2^\mu, \dots, \xi_i^\mu, \dots, \xi_N^\mu),$$

.

.

.

$$\vec{\xi}^p = (\xi_1^p, \xi_2^p, \dots, \xi_i^p, \dots, \xi_N^p).$$

Ahora bien, si consideramos la  $i$ -ésima componente de cada uno de los estados, tendremos  $N$  vectores de la forma siguiente:

$$\vec{\xi}_i = (\xi_i^1, \xi_i^2, \dots, \xi_i^\mu, \dots, \xi_i^p), \quad (60)$$

con  $i = 1, \dots, N$ . Cada uno de estos vectores (de  $p$  dimensiones) corresponden a puntos de un hipercubo con  $2^p$  esquinas, de tal forma que cada una de ellos, en el límite termodinámico, tiene una probabilidad  $P(\xi)$  de alcanzar una de dichas esquinas. En el caso

en que las memorias estén distribuidas al azar, con  $P(\xi_i^\mu) = P(-\xi_i^\mu)$  para toda  $i, \mu$ , cada esquina ocurrirá con una probabilidad de  $1/2^p$ .

Si se considera el límite cuando  $N \rightarrow \infty$  y  $p$  como el número de patrones fijos entonces, el promedio sobre todas las neuronas puede ser remplazado por un promedio sobre la distribución de las variables  $\xi_i^\mu$ . En este caso la frecuencia con la que  $\xi_i^\mu$  toma los valores  $\pm 1$  está determinada por la distribución de probabilidad  $P(\xi)$ . Todo esto se puede expresar de la siguiente manera:

Sea  $G(f)$  una función de variables aleatorias entonces,

$$\frac{1}{N} \sum_{i=1}^N G(\xi_i^\mu) \rightarrow \sum_{\xi^\mu} P(\xi^\mu) G(\xi^\mu) = \frac{1}{2^p} \sum_{\xi^\mu} G(\xi^\mu) = \langle\langle G(\xi) \rangle\rangle \quad (61)$$

en donde  $\langle\langle \dots \rangle\rangle$  indica el promedio sobre la distribución al azar de los  $p$  patrones  $\{\xi_i^\mu\}$  patrones.

En la expresión (61) la suma sobre todas las neuronas de una realización dada de la variable aleatoria, contiene la mismas propiedades estadísticas que ofrece un promedio sobre todos los diferentes sistemas, cada uno con una realización diferente de las  $\xi_i^\mu$ .

Se debe tener en cuenta que si el número de variables aleatorias aumenta, en este caso los patrones  $\vec{\xi}^\mu$ , no es posible considerar un autopromedio pues el número  $p$  de  $\vec{\xi}^\mu$  ya no permite que cada uno de ellos aparezca una cantidad suficientemente grande de veces en una realización dada.