eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología
Euskadi 2006

Trabajo Fin de Grado
Grado en Física

Memoria Asociativa en Redes Tipo Hopfield Balanceadas

Autor:
Ibon Recio
Director
Joaquin J. Torres
Director en la UPV
Jon Urrestilla

Leioa, 4 de Septiembre de 2015

Índice

I Introducción y objetivos **2**

II Desarrollo **5**

1. El cerebro humano **5**
1.1. Comprensión acerca del cerebro 5
1.1.1. McCulloch-Pitts 6
1.1.2. Hopfield 7
1.1.3. Hebb 7
1.2. Modelo estándar 7
1.3. Curva de magnetizacion 12

2. Modelo de red balanceada **15**
2.1. Motivacion Biologica 15
2.2. Modelo 16
2.3. Resultados 22

III Conclusiones **30**

Parte I
Introducción y objetivos

El presente trabajo ha sido desarrollado bajo la tutela del profesor e investigador Joaquín J. Torres integrante del Departamento de Electromagnetismo y Física de la Materia en la Universidad de Granada y miembro del Instituto "Carlos I" de Física Teórica y Computacional.

En la actualidad, las distintas ramas del saber no están divididas en compartimentos cerrados y cada vez más aparecen interfases que las conectan de tal manera que, por ejemplo, el análisis matemático y riguroso inherente a la física, hace que pueda ser aplicadas sus técnicas a innumerables ámbitos del saber que hasta hace poco podrían ser impensables como pueden ser la economía, la biología, la ecología y la sociedad.

La física estadística, en particular, es una de las principales disciplinas de la física que permite esta conexión interdisciplinar, principalmente debido a que permite estudiar de forma rigurosa sistemas con un alto numero de grados de libertad y entender su comportamiento emergente complejo. Un ejemplo de estos sistemas es el cerebro, que está considerado como un ejemplo de complejidad ya que une un alto numero de unidades básicas como son las neuronas con una altísima conectividad entre ellas, y que tiene fenómenos emergentes, no del todo comprendidos, que no se derivan de las dinámicas microscópicas que rigen las neuronas y sinapsis, si no del efecto colectivo de todas ellas. Surge por ello la necesidad de potentes instrumentos físico-estadísticos que resultan imprescindibles para su correcto entendimiento. En este contexto y aparejado al desarrollo de los ordenadores, han surgido nuevas disciplinas como la neurociencia computacional y la neurofísica, que versan sobre el estudio de las funciones cerebrales de alto nivel como la memoria, el aprendizaje, la conciencia, el procesamiento de información espacio-temporal, usando la técnicas fisico-matemáticas y mediante el uso del ordenador, simulando neuronas sencillas con alto grado de precisión y redes neuronales complejas que imitan ciertos medios neuronales. Se trata de ramás del saber altamente interdisciplinares que conectan distintos campos como la neurociencia, la ciencia cognitiva y la fisiología con otros como la ingeniería electrónica, informática, ciencia computacional, matemáticas o física.

Los modelos actuales de redes de neuronas, son por una parte lo más sencillos posibles para que puedan ser tratados mediante las técnicas de la física estadística, y por otra parte lo sufientemente complejos para que puedan reproducir comportamientos emergentes complejos en el cerebro. El hecho de la simplicidad a la hora de formalizarlos implica que muchas veces no incluyan muchos aspectos biológicos observados en experimentos en sistemas reales. Por lo tanto, dentro de la neurofísica computacional el enfoque que se plantea es el siguiente: construir nuevos paradigmas de redes neuronales con inspiración biológica y que sean capaces de reproducir los resultados experimentales observados. El fin último que se persigue es la aplicación de dichos modelos a ciencia, medicina e ingeniería.

En el presente trabajo nos hemos centrado en un tipo de redes neuronales capaces de reproducir la propiedad de memoria asociativa presente en el cerebro. Mediante esta propiedad el cerebro permite que seamos capaces de reconocer a un amigo al que llevamos años sin ver y cuyo rostro ha cambiado de forma considerable. Así la memoria asociativa es capaz tanto de reconocer patrones previamente aprendidos como de recuperarlos a partir de muestras defectuosas o alteradas del patrón original.

Durante el año que ha durado la realización de este Trabajo Fín de Grado, he ido aprendiendo técnicas de modelado, simulación y análisis de la física estadística y de la neurociencia computacional, y en particular he profundizado en el conocimiento, las virtudes y las limitaciones de los modelos de redes de neuronas autoasociativas presentes en la literatura, y en particular en el modelo estándar de este tipo de sistemas, el llamado Modelo de Hopfield. El aprendizaje ha sido de tipo teórico y computacional, ya que una vez comprendidos los modelos matemáticos que describen este tipo de redes se ha procedido a simularlos en ordenador para explorar su comportamiento emergente en función de los parámetros relevantes. También se han explorado variantes del mismo en cuanto a la topología de la red neuronal, el tipo de sinapsis y la excitabilidad neuronal. Una vez adquiridos los conocimientos teóricos y computacionales referentes al modelo estándar de Hopfield, en la segunda parte de la memoria se propone un modelo que extiende el modelo estándar asumiendo consideraciones más biológicas para hacerlo más realista. Concretamente se asume que las sinapsis en la red neuronal están balanceadas entre excitadoras e inhibidoras en la misma forma que se ha observado en la corteza cerebral, un hecho que se ha demostrado que tiene gran importancia e implicaciones en la función cerebral. El balanceado que se propone es en forma de un ruido sináptico balanceado que compite con el aprendizaje Hebbiano característico del modelo de Hopfield, de tal forma que preserva para cierta región de los parámetros la propiedad de memoria asociativa. Una vez diseñado el modelo de red de memoria asociativa con balance E/I, se ha procedido a su análisis teórico y computacional para explorar las implicaciones computacional que presenta. El modelo se ha simulado con el Superordenador Proteus (del Instituto Carlos I de Física Teórica y Computacional de la Universidad de Granada) y los resultados del análisis del mismo son los que se incluyen en este trabajo.

OBJETIVOS Los objetivos planteados con este trabajo han sido principalmente dos:

- Introducción al mundo de la neurofísica y neurociencia computacional mediante la obtención de conocimientos básicos de distintos modelos de redes neuronales, su fundamento biológico y su análisis teórico mediante técnicas de la física estadística y del estudio de los sistemas complejos, así como su correcta programación y simulación en un ordenador para el estudio de su fenomenología emergente.
- Entender el ámbito actual y las limitaciones que tienen los modelos actuales de redes neuronales autoasociativas y proponer un modelo que extienda

los modelos actuales y supere sus limitaciones, incluyendo aspectos biológicos no incluidos hasta la fecha. Así mismo una vez propuesto el modelo, analizar en profundidad la física resultante del mismo y su relación con fenómenos y funciones de alto nivel observados en el cerebro.

Parte II
Desarrollo

1. El cerebro humano

El cerebro es paradigma de lo que se define como sistema complejo *«sistema con dinámicas e interacciones no lineales que muestra sensibilidad a las condiciones iniciales y cuyo comportamiento emergente no se deriva del de sus constituyentes elementales"* [10] donde mediante la interacción de unidades básicas bien conocidas como son las neuronas este, el cerebro, es capaz de desarrollar una fenomenología emergente no predecible a la vez de útil. Desconocemos además muchos detalles de los mecanismos que utiliza el cerebro para coordinar el resto del sistema nervioso, procesar constantemente gran cantidad de complicada información y controlar el comportamiento global del individuo, incluso sus emociones e inteligencia. El cerebro humano es capaz de desarrollar tareas que ni las ordenadores más potentes diseñados hasta la fecha son capaces de emular. Por ejemplo, en tareas tan ordinarias hoy en día como la de reconocer, después de muchísimos años, a un amigo o identificar y escribir imágenes como las que aparecen en la Figura 1.

El cerebro humano tiene alrededor de 10^{11} células nerviosas, las neuronas (las cuales podemos entender como la unidad mas básica de procesamiento de información en el cerebro). Además, estas neuronas exhiben una altísima cooperación entre ellas, ya que se relacionan mediante más de 10^{15} conexiones, llamadas sinapsis, con una media de 10.000 conexiones por neurona. A pesar del altísimo numero de neuronas y conexiones que componen nuestro cerebro, debemos remarcar el hecho de que la materia ordinaria exhibe una cantidad de unidades básicas, en este caso moléculas, notoriamente superior (recordemos que un mol de materia contiene del orden de 10^{24} moléculas). Esta gran diferencia de unidades básicas nos sugiere que tiene que existir algo esencial o quizá sutil en la manera en la que cooperan estas unidades y que diferencia los efectos cooperativos de la materia (condensación, magnetismo...) de los del cerebro (conocimiento, inteligencia...). El cerebro, coordinando relativamente pocas neuronas es capaz de conseguir fenomenología emergente con un orden aún más sorprendente que los cambios de fase observados en la materia ordinaria. Si bien es verdad que estudios recientes abren la puerta a la existencia de cambios de fase dentro del cerebro [16].

1.1. Comprensión acerca del cerebro

Los detalles básicos de la complicada estructura del cerebro ya fueron revelados por Santiago Ramón y Cajal (1852-1934), que recibió el premio Nobel por ello en 1906. Mirando a través de un sencillo microscopio, Cajal acumuló datos y formuló hipótesis que son el fundamento de la neurociencia moderna. Demostró que el sistema nervioso y el cerebro no son medios continuos, sino una

Figura 1: Este tipo de imágenes o patrones suelen ser usadas a menudo para verificar que el usuario de cierto servicio es un ser humano y no una máquina, debido a que a un ordenador, por muy potente que sea, le resultaría, a día de hoy, una tarea muy difícil incluso imposible recuperar las letras y números escondidas en estos patrones.

combinación discreta de neuronas, que describió como "células largas y filiformes...misteriosas mariposas del alma, cuyo batir de alas puede algún día -¿quien sabe?- clarificar el secreto de la vida mental". También notó que la conexión sináptica entre neuronas era mediante uniones discontinuas, y que cada neurona parecía recibir señales de las otras y reaccionar consecuentemente, quizá con el envío de otra señal [4].

Con el transcurrir de los años y el desarrollo de nuevos dispositivos y técnicas de extracción de datos y análisis cada vez menos invasivas, se ha profundizado en el conocimiento de las unidades básicas del cerebro y su comportamiento emergente. Aunque se haya visto que cada una de estas neuronas es un mundo en si mismo, se conocen casi todos los tipos de neuronas y sinapsis, sus constituyentes y los mecanismos básicos que determinan sus funciones. Es decir, conocemos con alto grado de precisión la naturaleza de las unidades del cerebro, pero como paradigma de complejidad, el conocimiento de las unidades básicas no nos permite entender las propiedades emergentes del mismo.

A partir de la segunda mitad del siglo XX se comenzaron a proponer modelos sencillos matemáticos que trataban de imitar aceptablemente algunas funciones cerebrales básicas, y que así demostraban que son consecuencia de la cooperación entre muchas neuronas y sinapsis, y que vamos a describir muy brevemente a continuación.

1.1.1. McCulloch-Pitts

Un primer paso a la hora de desarrollar un modelo estándar para el cerebro fue dado por el neurofisiólogo Warren S. McCulloch (1899-1969) y el matemático Walter Pitts (1923-1969) al condensar el concepto de neurona que, obviando su complejidad, imaginaron como un interruptor elemental [12]. Según ellos, la función de la neurona consiste esencialmente en disparar cuando la suma de señales que le llegan desde otras neuronas supera un determinado umbral de excitación. Construyeron dispositivos con esta idea para evaluar su capacidad en la solución de problemas lógicos, y hoy se sabe que, por lo que respecta a fenómenos debidos a la cooperación, las neuronas se pueden imaginar como variables binarias binarias McCulloch-Pitts.

1.1.2. Hopfield

El paso decisivo hacia un modelo estándar de medio neuronal real fue dado por el físico John J. Hopfield (1933), quien propuso un modelo de cerebro capaz de recrear la memoria [7]. En su modelo considera una red con N neuronas binarias McCulloch-Pitts, $s_i = 0, 1$, donde los dos estados corresponden a disparo y reposo respectivamente. Cada neurona está relacionada con todas las demás mediante sinapsis cuyos "pesos" o "fuerza" respectivos vienen descritos por una variable real ω_{ij} . Aquí los índices i y j, cambian de 1 a N para describir todas las neuronas y sus relaciones sinápticas. En cada instante discreto t , el estado del sistema, llamado *configuración del sistema* o de la *red de neuronas*, se caracteriza mediante el conjunto de valores de las actividades de las neuronas y pesos sinápticos. Suponiendo que los pesos sinápticos son siempre los mismos, el modelo queda determinado al detallar la dinámica de las neuronas mediante una regla para los disparos y un método para ir actualizando las actividades.

1.1.3. Hebb

El psicólogo Donald O. Hebb (1904-1985), observando procesos de aprendizaje y los efectos de la cirugía y lesiones accidentales, consiguió establecer un relación plausible entre fisiología y funciones cerebrales. Hebb conjeturó que la estimulacion repetida de receptores específicos inducia la formación de "agrupaciones de células" con sus actividades correlacionadas que perduraban y servían luego como recuerdo del estímulo [5]. El mismo Hebb enuncio su idea como un punto de vista más funcional. Es la llamada *regla de Hebb*, según la cual el fortalecimiento de algunas sinapsis–y por tanto, el aumento relativo del peso ω correspondiente– es consecuencia de la activación repetida de las neuronas que conectan, mientras que las sinapsis entre neuronas predominantemente inactivas tenderían a debilitarse y, eventualmente, desaparecerían.

1.2. Modelo estándar

Como se ha comentado previamente el modelo estándar (que es el que vamos a usar en el presente trabajo) consta del modelo de Hopfield donde los pesos sinápticos son calculados en base a la regla de aprendizaje de Hebb [19].

Modelo de Hopfield: Como ya hemos introducido antes, en el modelo de Hopfield, el cerebro se estructura como una red, donde los N nodos son ocupados por las neuronas, las cuales serán las unidades básicas de procesamiento de información. Cada una de estas neuronas podrá estar en los estados de reposo y disparo, $s_i = 0, 1$ respectivamente. Se asume una red *fully connected* ya que todas las neuronas están conectadas con el resto de neuronas. Estas conexiones, o sinapsis, vienen definidas por sus pesos ω_{ij}, los cuales indican lo bien o mal que se transportaran los impulsos de una neurona j a una neurona i .

Supongamos, por ejemplo, una neurona i conectada a otras muchas que representamos genéricamente por j , como se muestra en la Figura 2. En este

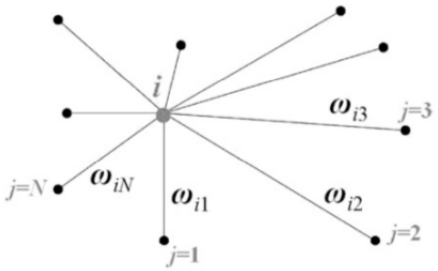

Figura 2: En la imagen se ha dibujado un esquema simplificado de la estructura que tiene una red neuronal de tipo Hopfield. En ella una neurona esta conectada con el resto y sus conexiones se indetifican con la variable ω_{ij}, indicando así la eficacia con la que puede transmitirse información [11].

diagrama, la neurona i recibe una señal de cada una de las otras que se verá modulada por el peso sináptico de cada una de las conexiones que las une, por lo que la señal que llegue a la neurona i desde la neurona j , será $\omega_{ij}s_j$. De esta forma que la señal depende de la actividad de la neurona j (tan solo llegara señal a la neurona i, si la neurona j esta disparando) y de la eficacia de la transmisión ω_{ij}, que puede ser negativa (cuando la sinapsis es inhibidora). Sumando todas estas señales para todo valor de j , se obtiene la señal total o corriente sináptica neta h_i (también llamado campo local). Con esto se quiere decir que la señal en i sera igual a la suma de todos los pesos sinápticos cuando todas las demás neuronas estén activas, $s_j = 1$; en caso de que las neuronas estén en reposo $s_j = 0$, no aportaran nada a la señal total.

La regla básica para introducir una dinámica en el sistema consiste en inducir al disparo neuronas que están en un estado silencioso de reposo. Supongamos que tenemos una neurona en el instante t en estado de reposo $s_i = 0$. Para activar dicha neurona necesitamos que la señal neta que llegue a la neurona desde sus vecinos supere un umbral, especifico de cada neurona, θ_i. Esto se puede modelar con una función de tipo escalón en la que si la señal total supera al umbral ($h_i \geq \theta_i$) la neurona puede disparar, y en caso contrario ($h_i < \theta_i$) se quedará en reposo. Inspirándose en los estudios del magnetismo, esta regla de actualización se hace probabilística, esto es, en lugar de hacer el cambio de s_i siempre que $h_i \geq \theta_i$, se hace con una probabilidad que depende de la diferencia entre h_i y θ_i o mas concretamente de $\beta\,(h_i - \theta_i)$, donde β es un parámetro y cuya $T = 1/\beta$, denominado temperatura, controla la estocasticidad del proceso.

En la práctica se procede de la siguiente manera: partiendo de una configuración inicial dada para todas las actividades s_i, calculamos todas las señales h_i. Esto permite decidir que actividades han de intercambiarse usando la regla de disparo especificada. Una vez actualizados los estados s_i de las neuronas, se repite el calculo de las h_i para la nueva configuración de la red. La actualización de la red puede hacerse de dos maneras, de manera *paralela* donde todas las neuronas son actualizadas en $t + 1$ a la vez usando los estados del instante t, o

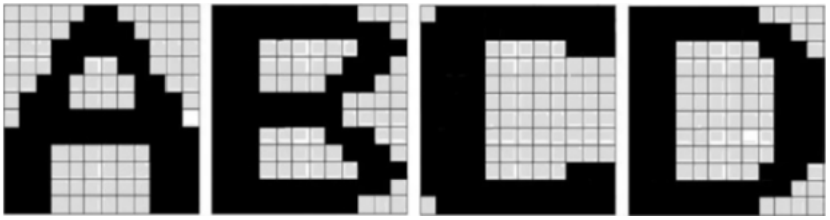

Figura 3: La imagen muestra de izquierda [11] a derecha cuatro patrones, en este caso con forma de letra, que una red es capaz de aprender. Estos patrones tienen 12x12 píxeles que están codificados como celdas binarias que la red recordará reforzando o debilitando los pesos sinápticos que relacionen los distintos píxeles.

de manera *secuencial* visitando las neuronas de una en una de forma aleatoria cambiando solo la actividad del lugar visitado a cada paso.

El modelo de Hopfield ha de completarse con un conjunto de valores para los pesos ω_{ij} ya que éstos son necesarios para calcular las señales h_i, como hemos comentado anteriormente. La manera de elegir estos pesos en el modelo estándar está dada mediante la regla de aprendizaje de Hebb.

Regla de Hebb: La regla de aprendizaje de Hebb, es la pieza que complementa al modelo de Hopfield para constituir juntas el modelo estándar. Esta regla se basa en la experiencia biológica de que cuando ciertas neuronas se activan a la vez ante ciertos estímulos, estas tienden a reforzar sus conexiones, mientras que sus conexiones se debilitan con aquellas que permanecen en reposo.

Imaginemos un estimulo –por ejemplo, visual– o patrón que consta de un retículo cuyas celdas se corresponden con los píxeles de una cierta imagen. Por sencillez del ejemplo, nos limitamos a caso donde cada píxel solo puede ser blanco o negro como los de la Figura 3.

En este caso binario, cada patrón, digamos ξ^μ, con $\mu = 1(\mathbf{A}), 2(\mathbf{B}), 3(\mathbf{C}), 4(\mathbf{D})$, es un conjunto de variables $\xi^\mu = 0, 1$ que representan el color (claro o negro) de cada píxel. La regla de Hebb se implementa entonces de la siguiente manera: el peso ω_{ij} de la conexión entre las neuronas i y j se hace igual a la suma de los productos $(\xi_i^\mu - a)(\xi_j^\mu - a)$ de los píxeles en los lugares i y j de cada patrón μ restados por el valor medio de los pixeles $a = \langle(\xi_i^\mu\rangle$ (y se normaliza diviendiendo por el numero de patrones M, para mantener el peso pequeño). La receta puede escribirse $\omega_{ij} \sim \big((\xi_i^\mu - a)(\xi_j^\mu - a) + (\xi_i^\mu - a)(\xi_j^\mu - a) + (\xi_i^\mu - a)(\xi_j^\mu - a) + (\xi_i^\mu - a)(\xi_j^\mu - a)\big)$ para los cuatro patrones indicados. O escrito de una manera mas general y compacta,

$$\omega_{ij} = \frac{1}{M}\sum_{\mu=1}^{M}(\xi_i^\mu - a)(\xi_j^\mu - a). \tag{1}$$

Cuando las sinapsis se construyen de este modo, esto es, como combinación lineal de una serie de patrones, se dice que estos patrones están almacenados

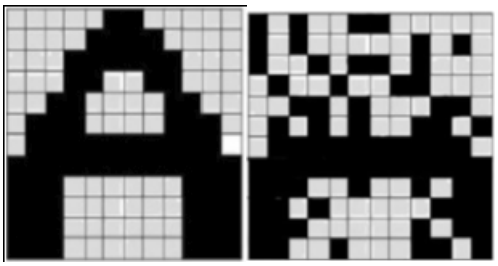

Figura 4: La imagen de la derecha muestra una deconstrucción del patrón **A** (izquierda). A partir de este estado inicial la red sera capaz de recuperar el patrón original que se había aprendido previamente.

en la red de neuronas. Este es el mecanismo del modelo estándar para aprender y acumular experiencias. Es bastante general ya que los patrones pueden representar, además de imágenes, cualquier tipo de información compleja.

Usando este modelo vamos a hacer un pequeño experimento para ver la potencia del mismo. Tomando uno de los patrones almacenados y degradándolo a propósito como en la imagen que aparece en la derecha de la Figura 4, hasta que la letra sea prácticamente irreconocible.

Con este experimento se esta recreando la situación en la que nos encontramos inesperadamente con una persona que no veíamos desde la infancia, esto es, con una copia ciertamente "deteriorada" de una imagen que guardamos entre nuestros recuerdos. En estas condiciones, el cerebro suele reaccionar rápidamente asociando las dos imágenes y volcando los recuerdos –el nombre de la persona y las vivencias comunes– relacionadas con la antigua imagen. Pues bien, el modelo resulta comportarse de manera semejante. Con la **A** falseada como condición inicial, al aplicar constantemente la regla de disparo por todo el sistema, la configuración va cambiando hasta que, generalmente, recobra un estado muy próximo al **A** puro.

La Figura 5 ilustra el comportamiento en una evolución temporal típica de la red para distintas temperaturas, o niveles de ruido en el sistema. La situación no cambia al almacenar más patrones mientras no se alcance un límite alto de saturación a partir del cual la red neuronal no puede memorizar más patrones y recordarlos sin error (la relación entre los patrones almacenados y numero de neuronas, que se conoce como capacidad de almacenamiento $\alpha = M/N$, no puede exceder el valor de $\alpha_c \sim 0.14$ [2]). Para temperaturas bajas y salvo fluctuaciones, la figura muestra que el estado que se alcanza después de un TRANSITORIO se parece mucho a uno de los patrones memorizados, pues el solapamiento m^μ que mide el parecido entre el estado de la red y el patrón almacenado se aproxima rápidamente a 1.0. Esto quiere decir que, en un sentido, el sistema restaura la imagen deteriorada que le hemos dado como condición inicial, mostrando así una notable tolerancia frente a datos incompletos. Se dice entonces que los patrones almacenados son atrayentes de la dinámica del modelo estándar y que éste tiene la propiedad de *memoria asociativa*. De hecho, en condiciones

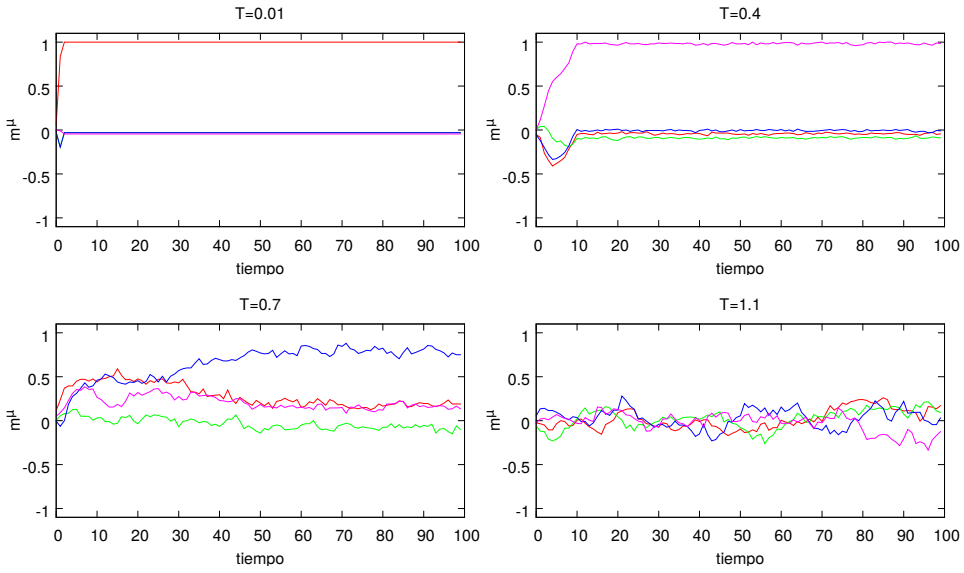

Figura 5: Los cuatro paneles muestran simulaciones de recuerdo en una red neuronas que almacena M=4 memorias en los pesos sinápticos (concretamente las letras $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$) definidas como en la figura 3. Partiendo de una configuración inicial aleatoria se ha dejado correr el sistema y se ha medido el parecido que tiene a lo largo de toda la simulación con cada uno de los patrones almacenados, medido en términos de cierta función solapamiento m^μ. Para temperaturas por debajo de la temperatura critica ($T < T_c = 1.0$) el sistema recupera uno de los patrones con bastante eficacia. Cada una de las simulaciones se ha hecho con distinto valor del parámetro de ruido T, para que se vea el efecto que tiene este a la hora de desarrollar recuerdos de memorias. En la última imagen vemos como la red es incapaz de recuperar ninguno de los patrones, ya que el ruido es excesivamente alto.

adecuadas, el modelo evoluciona generalmente hacia uno de los estados puros de la Figura 3, aquel que se encuentre más próximo de la condición inicial. Sin embargo a medida que la temperatura aumenta, la cuenca de atracción de estos estados atrayentes de memoria disminuye y la temperatura desestabiliza la soluciones alejándolas de estados de memoria (más abajo comentaremos como interpretar esta temperatura, que no es más que un parámetro de ruido que introduce estocasticidad en el sistema).

Para temperaturas bajas y salvo fluctuaciones, el estado estado se parece mucho, al cabo de una evolución relativamente corta, a uno de los patrones, pues el solapamiento m^μ que mide el parecido entre el estado de la red y el patrón almacenado se aproxima rápidamente a 1.0. Esto quiere decir que, en un sentido, el sistema restaura la imagen deteriorada que le hemos dado como condición inicial, mostrando así una notable tolerancia frente a datos incompletos. Se dice

que los patrones almacenados son atractores de la dinámica del modelo estándar y que este tiene la propiedad de *memoria asociativa*. De hecho, en condiciones adecuadas, el modelo evoluciona generalmente hacia uno de los estados puros de la Figura 3, al mas próximo de la condición inicial. Sin embargo a medida que la temperatura aumenta, la cuenca de atracción de estos atractores de memoria mengua y la temperatura desestabiliza la soluciones alejándolas de estados de memoria, mas abajo comentaremos como interpretar esta temperatura, que no es mas que un ruido estocástico.

1.3. Curva de magnetizacion

Como hemos visto el modelo estándar es capaz de "memorizar" o "aprender" una serie de patrones, que mas tarde mediante las reglas de disparo y actualización de la red será capaz de recordar a partir de un patrón de entrada. En las imágenes que aparecen en la Figura 5 se observan distintas evoluciones temporales (la unidad de tiempo es un paso Monte Carlo que consiste en N actualizaciones de neuronas), esto es, como desde una condición inicial el sistema evoluciona hasta parecerse en mayor o menor medida a uno de los patrones almacenados. En el ejemplo anterior, teníamos cuatro patrones, y el sistema era capaz a partir de una muestra defectuosa de uno de los patrones almacenados evolucionar hacia una configuración muy próxima, o idéntica, de ese patrón.

Al cabo de un tiempo determinado, el sistema habrá evolucionado hacia un estado estacionario caracterizado por un valor medio para el solapamiento m^μ y unas fluctuaciones que van a depender del parámetro T. Se aprecian distintas series temporales en las que el sistema evoluciona hacia distintos estados estacionarios. En la imagen izquierda de la parte superior ($T = 0$), el sistema alcanza rápidamente su estado estacionario $m^\mu = 1$, sin apenas fluctuaciones alrededor de este valor. Si observamos el resto de imágenes de la Figura 6, se puede ver como al sistema le cuesta mas alcanzar estos estados estacionarios a medida que la temperatura del sistema T aumenta de valor, y como además las fluctuaciones en el estado estacionario son cada vez de mayor intensidad, ya que al aumentar T se incrementa la agitación térmica en el sistema. Este comportamiento se explica de la siguiente forma. A T=0, el sistema alcanzará el estado asociado a uno de los patrones almacenados, ya que se puede demostrar [14] la dinámica determinista ($T = 0$) lleva antes o después a estos estados (son estados atrayentes de la dinámica). Cuando T aumenta los cambios estocásticos perturban dicha dinámica que hace que nunca se alcance exactamente el estado de memoria sino un estado estacionario fluctuante alrededor del mismo. Cuando las fluctuaciones son muy grandes, e.g., para $T = 1$, la memoria se pierde pues las fluctuaciones son tan grandes que en cada paso pueden llevar el estado fuera de la cuenca de atracción de cada memoria, que por lo tanto deja de ser atrayente (ver por ejemplo imagen inferior derecha de Figura 6).

Una vez visto lo importante que son los efectos de la temperatura en la memoria del modelo estándar resulta comprensible la necesidad de ilustrar el comportamiento de la memoria del modelo con respecto a la temperatura del sistema y las curvas que emergen de este análisis reciben el nombre de *curvas*

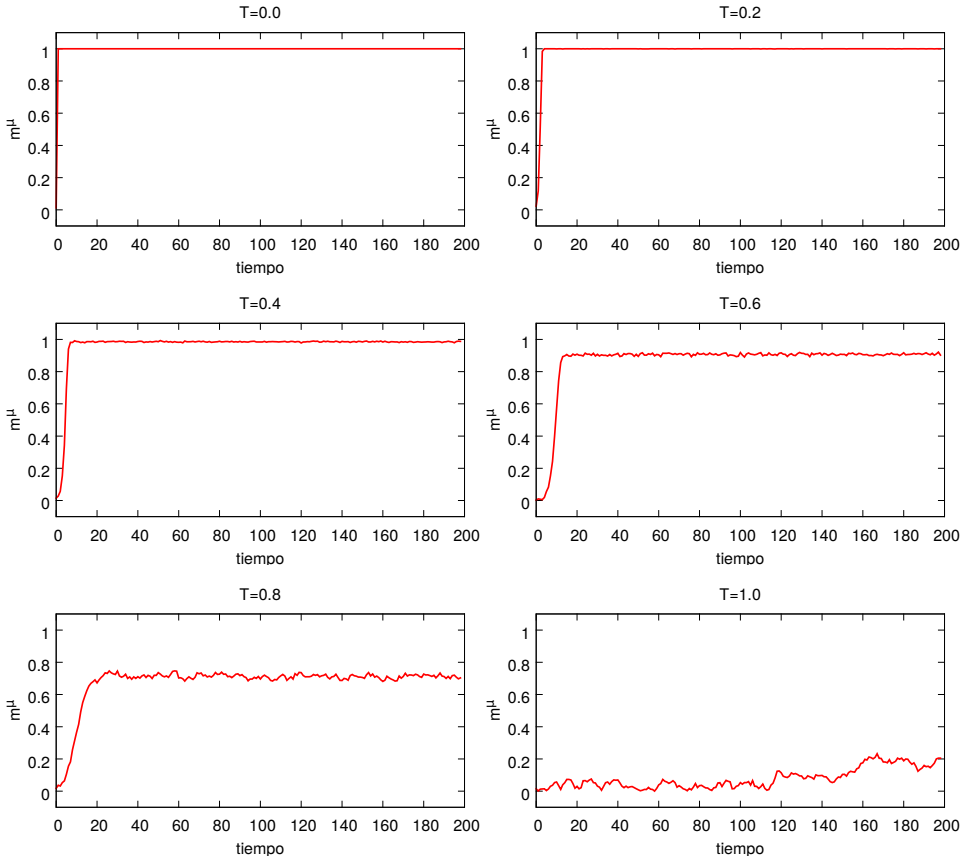

Figura 6: Las distintas imágenes muestran evoluciones temporales de la variable de solapamiento ,m^{μ}, para distintos valores de temperatura. Se observa como a medida que la temperatura es mayor al sistema le cuesta mas recuperar el patrón almacenado y además una vez recuperado fluctúa alrededor de ese valor.

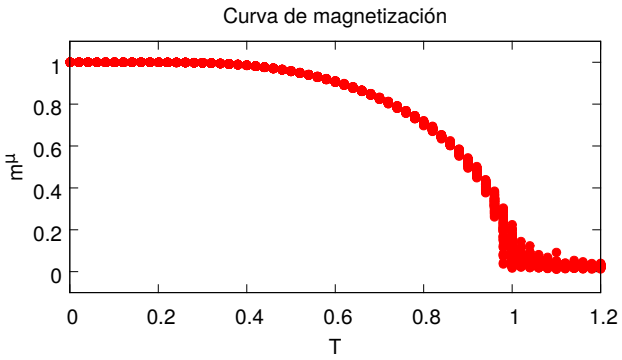

Figura 7: Curva de magnetización para el modelo estándar. Los puntos han sido obtenidos a partir de simulación.

de magnetización o de solapamiento.

Para realizar este tipo de curvas la metodología es la que sigue: deberán realizarse series temporales como las representadas en la Figura 6 lo suficientemente largas como para que el sistema alcance un sistema estacionario y y no esté en un estado transitorio. Por ejemplo, si en la imagen con $T = 0.8$ de la Figura 6 hubieramos cogido el valor de solapamiento en el instante $t = 10$. Dado que el estado estacionario puede estar caracterizado por grandes fluctuaciones, no vale con coger el ultimo valor de la serie temporal ya que este no tiene por que ser representativo. Lo que hay que hacer es determinar el numero de puntos del estado estacionario que permitan mediante estadística darnos un valor medio representativo del mismo, así como una cuantización de la intensidad de las fluctuaciones en dicho estado. Dado que partimos de condiciones iniciales aleatorias, el sistema podría evolucionar hacia distintos estados estacionarios (por ejemplo si tuviéramos biestabilidad), cosa que no ocurre en el modelo estándar, por lo que deberemos repetir todo el proceso numerosas veces y después promediar entre todas ellas. Cuando repetimos cada uno de los pasos anteriores para distintas temperaturas, entonces obtenemos la curva de magnetización.

En la figura 7 aparece la curva de magnetización propia del modelo estándar cuando se almacena una sola memoria en las sinapsis. Esta curva nos muestra que para el límite de bajas temperaturas, allí donde el modelo es muy poco estocástico, el sistema recuerda la memoria aprendida sin apenas error ya que el valor que se alcanza de solapamiento (también conocido como overlap o hamming) es casi la unidad (se alcanzaría exactamente el valor uno en T=0) y éste se define como:

$$m^{\mu}(\mathbf{s}) \equiv \frac{1}{Na(1-a)} \sum_i (\xi_i - a)(s_i - a). \tag{2}$$

que no es más que el producto escalar entre la configuración actual de la red $s = \{s_i = 1, 0; i = 1, ..., N\}$ y la configuración particular de la red asociada al

patron μ $\{\xi_i^\mu = 1, 0; i = 1, ..., N\}$. Por lo tanto en cada instante t ,m^μ mide lo se parecen ambas configuraciones de red. Los patrones memorizados en principio no tienen ninguna limitación a excepción de tener que ser del mismo tamaño que la red y que sus componentes sean binarios. En este trabajo sin embargo vamos a considerar – por simplicidad y por que permite hacer ciertos cálculos analíticos en el modelo estándar – patrones aleatorios tal que $a =< \xi_i >= 0.5$. Es decir, configuraciones aleatorias de la red neuronal que en promedio tengan el mismo número de neuronas disparando ($\xi_i = 1$) y neuronas silenciosas ($\xi_i = 0$).

Si volvemos a la ecuación (2) y sustitimos las variables s_i y ξ_i por los dos valores que pueden obtener cada una pueden darse las siguientes combinaciones

$$\sum_i (\xi_i - a)(s_i - a) = \begin{cases} 0.25 & si \quad s = 0\, y\, \xi_i = 0 \quad ó \quad s_i = 1\, y\, \xi_i = 1 \\ -0.25 & si \quad s = 0\, y\, \xi_i = 1 \quad ó \quad s_i = 1\, y\, \xi_i = 0 \end{cases} \tag{3}$$

esto es, si el elemento i de la configuración de la red coincide con el elemento i del patrón, siendo ambos 0 ó 1 entonces ellos aportaran un término positivo al valor total del solapamiento m^μ. Mientras que si estos dos elementos son diferentes entre ellos de la manera que se indica en la parte inferior de (3), entonces ellos añadirán un termino negativo al valor del solapamiento. Para valores grandes de N y suponiendo una configuración inicial aleatoria de la red, el valor inicial del solapamiento será nulo o casi nulo y se necesitara de un tiempo, para que la red mediante su dinámica vaya pareciéndose cada vez más al patrón previamente almacenado. Es importante notar que debido a la simetría inherente del modelo estándar y que si un patrón ξ es un atrayente de la dinámica del sistema, se puede demostrar que el estado $1 - \xi$ (tambiém llamado antipatrón) es también un atrayante de la dinámica del sistema [1]. De forma trivial se puede demostrar que si el estado del sistema alcanza el antipatrón se tiene $m^\mu = -1$ (ya que m^μ representa el solapamiento con el patrón).

2. Modelo de red balanceada

2.1. Motivacion Biologica

La mayoría de los comportamientos humanos están mediados por el sistema nervioso central, constituido por la médula espinal y el encéfalo, éste último dividido a su vez en cerebro, tallo cerebral y cerebelo. El cerebro está anatómicamente dividido en dos hemisferios y todo el conjunto puede ser subdividido en otras áreas anatómica y funcionalmente distintas que constituyen una red compleja o conectome.

La división del cerebro que a nosotros nos interesan son los hemisferios cerebrales y especialmente la capa más externa de los mismos denominada corteza cerebral. Mientras que muchas funciones del sustento de la vida son llevadas a cabo por la médula espinal, tallo cerebral y el diencéfalo (parte del cerebro justo encima del tallo cerebral), la responsable de muchas de las acciones de planificación y ejecución de acciones diarias es la corteza cerebral.

La corteza cerebral esta dividida en seis capas celulares. El numero de capas y la organización funcional de las mismas esta determinada por la zona de la corteza en la que se encuentren. La estructura más general del neocortex (las áreas más evolucionadas de la corteza cerebral) consta de seis capas, como podemos apreciar en la Figura 8.

Las neuronas de la corteza presentan gran variedad de tamaños y formas. Raphael Lorente de Nö, estudiante de Ramón y Cajal, uso el método de *tinte de Golgi* para identificar más de 40 tipos de neuronas corticales basándose en la distribución de sus dendritas y axones. En general, las neuronas de la corteza podemos definirlas *a grosso modo* en dos grupos: neuronas de proyección e interneuronas.

Las neuronas de proyección tienen típicamente cuerpos piramidales y se encuentran localizadas mayoritariamente en las capas centrales y altas (III, V y VI). Usan el aminoácido glutamato como neurotransmisor primario, el cual es excitador. Mientras, las interneuronas locales usan el neurotransmisor inhibidor ácido γ-amino-butírico (GABA) y se localizan en todas las capas de la corteza cerebral. Los dos tipos de neuronas pueden ser observados en la Figura 9, en donde se puede observar la forma piramidal de las neuronas de proyección y los largos axones que nacen de ellas permitiéndoles transportar información entre las distintas capas de la corteza, incluso atravesándola y conectándose con otras zonas del cerebro. En cambio, las interneuronas son mas pequeñas y tienen axones más cortos ya que son neuronas locales y por eso no desarrollan conexiones con neuronas de otras zonas cerebrales.

Existe un balanceado entre estos dos tipos de neuronas a lo largo y ancho de toda la corteza cerebral que han demostrado ser de gran importancia en las funciones cerebrales [20, 3]. Son más numerosas las neuronas proyección (excitadoras) que las interneuronas locales (inhibidoras) en una relación $20\,\% - 80\,\%$ [9]. Además existe otro tipo de "balanceado" distinto del anterior pero muy relacionado con éste. Se refiere a cómo se relacionan las intensidades de los potenciales postsinápticos excitadores (EPSP) e inhibidores (IPSP), o más concretamente a cómo es la razón EPSP/IPSP a lo largo de la corteza cerebral. Se ha observado que el potencial pre sináptico es en términos absolutos de mayor intensidad en las inhibidoras con un ratio $4:1$ respecto a los potenciales pre sinápticos de las excitadoras como se ve en la Figura 10. En nuestro modelo este ratio entre conductancias sera equivalente al ratio entre las intensidades absolutas de los pesos sinápticos$\big(4|\omega_{ij}^{excitadora}| \approx |\omega_{ij}^{inhibidoras}|\big)$. Es decir, hay más neuronas excitadoras pero éstas disparan con menos intensidad justo al contrario de lo que ocurre con las neuronas inhibidoras.

2.2. Modelo

El modelo que se va a desarrollar busca incluir, dentro del modelo estándar ya descrito, el balance E/I observado y descrito en la corteza cerebral. Como hemos visto, el modelo estándar no es más que un marco teórico que auna la información obtenida de experimentos biológicos y modelos matemáticos sencillos de neuronas binarias, que es capaz de recrear la propiedad de memoria

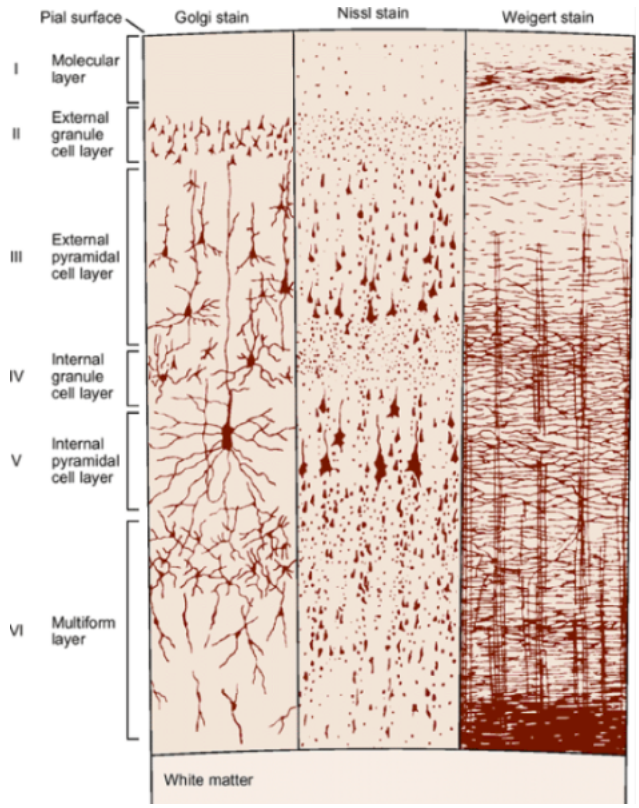

Figura 8: La imagen muestra el desarrollo vertical de la sección más general de la corteza cerebral (en cada zona de la corteza puede variar el número de capas, su grosor y su composición). En el margen izquierda de la figura se muestra la estructura vertical en capas de la corteza. En la imagen se observan tres maneras distintas de representar la sección de corteza y se diferencian por el método usado para la identificación de las neuronas. (Izquierda) *Tinte de Golgi*: revela los cuerpos neuronales y las dendritas próximas. (Centro) *Tinte de Nissl*: muestra los cuerpos neuronales y las dendritas próximas. (Derecha) *Tinte de Weigert*: Revela el patrón de distribución axional. [9]

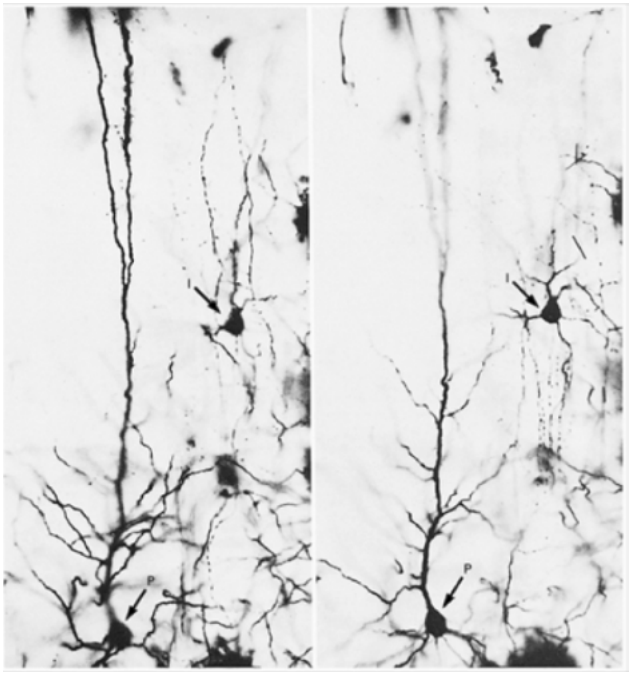

Figura 9: En ambas imágenes pueden diferenciarse los dos tipos de neuronas presentes en la corteza cerebral: las *interneuronas locales* (señalizadas con la letra **I**) y las *neuronas de proyección* (señalizadas con la letra **P**) [9]

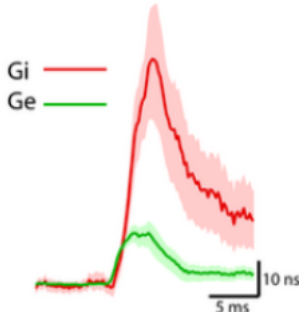

Figura 10: Diferencias entre la conductancia sináptica observada en las sinapsis excitadoras e inhibidoras (medido en una neurona estrellada situada en la capa IV de la corteza somatosensorial primaria de una rata), donde los pesos sinapticos no son mas que las conductancias maximas. Podemos decir por tanto que los pesos sinapticos no son mas que los maximos de estas curvas y en la imagen se observa como la conductancia excitadora (curva verde) es cuatro veces menor que en el caso de las inhibidoras (curva roja) [13].

asociativa característica de los cerebros más evolucionados. Aun así, no es un modelo definitivo, pero si una buena base sobre la que trabajar. En física siempre se parte de modelos simples a los que se les va añadiendo un mayor grado de rigurosidad a la vez que complejidad. Usando esta estrategia, se pretende en este trabajo construir un modelo de red neuronal, que supere los límites conceptuales del modelo estándar, que incluya la evidencia experimental del balanceado sináptico observado en el córtex, y estudiar las propiedades emergentes que de él se derivan.

Como hemos dicho el modelo estándar es incapar de emular la corteza cerebral, entre otras cosas, porque los pesos sinápticos no estan balanceados de acuerdo con las evidencias experimentales [9]. De hecho los pesos sinápticos ω_{ij} son calculados de acuerdo a la regla de Hebb que asume elegir

$$\omega_{ij}^{Hebb} \propto \sum_{\mu=1}^{M} (\xi_i^\mu - a)(\xi_j^\mu - a) \tag{4}$$

que dependerá de las características de los distintos patrones memorizados por la red. Si los patrones $\xi^\mu = \xi_i^\mu = 0, 1; i = 1, ...N$ son configuraciones aleatorias de la red, los valores que se obtengan en el sumatorio serán numeros aleatorios que nos responderán a correlación alguna. De hecho, a medida que el numero de patrones almacenados, M , aumenta y usando el teorema del límite central,[1]

la distribución de los pesos sinápticos ira tomando la forma de una distribución Gaussiana. Además, todos los términos que pueden aparecer en la distribución son simétricos en cuanto a signo como ya vimos en (3), por lo tanto la Gaussiana estará centrada en el origen. Entonces, en el modelo estándar tenemos aproximadamente el mismo número de sinapsis excitadoras ($\omega_{ij} > 0$) y sinapsis inhibidoras ($\omega_{ij} < 0$) y además las habrá de diversas intensidades, dependiendo de la varianza de la distribución Gaussiana (que a su vez dependerá de la cantidad de patrones almacenados de forma que cuantos más patrones se aprendan mayor será el rango de posibles valores que puedan tomar las ω_{ij}). Nos encontramos entonces con una situación alejada de la evidencia biológica de los balanceados descritos anteriormente. Por un lado tenemos el mismo número de sinapsis excitadoras e inhibidoras cuando sabemos que las excitadoras cuadriplican en numero a las inhibidoras.Mientras que por el otro lado, las intensidades absolutas de los pesos sinápticos son iguales en excitadoras e inhibidoras cuando sabemos que las inhibidoras disparan cuatro veces más intensamente que las excitadoras.

El modelo que se propone en este trabajo trata de incluir las evidencias experimentales observadas en la corteza cerebral elegiendo los pesos sinápticos del modelo estándar en la forma

$$\omega_{ij} = c\omega_{ij}^{Hebb} + (1 - c)\,\omega_{ij}^{B} \tag{5}$$

[1] *Si S_n es la suma de n variables aleatorias independientes y de varianza no nula pero finita, entonces la función de distribución de S_n " se aproxima bien" a una distribución normal,también llamada distribución Gaussiana.*

$$\omega_{ii} = 0 \tag{6}$$

donde con probabilidad c los pesos los elegimos de acuerdo a la regla de hebb y con probabilidad $(1-c)$ toman un valor ω_{ij}^B que es un número aletorio obtenido a partir de una distribución bimodal del tipo

$$p\left(\omega_{ij}^B\right) = \eta\mathcal{N}\left(\lambda\alpha, \sigma^2\right) + (1-\eta)\,\mathcal{N}\left(-4\lambda\alpha, \sigma^2\right) \tag{7}$$

$$\omega_{ij} = \omega_{ji} \tag{8}$$

donde η mide la intensidad relativa entre las dos Gaussianas centradas en los puntos $\lambda\alpha$ y $-4\lambda\alpha$. Estos dos puntos no son mas que los promedios de los pesos sinápticos excitadores e inhibidores, respectivamente

$$\left\langle \omega_{ij}^{excitadora} \right\rangle = \lambda\alpha \tag{9}$$

$$\left\langle \omega_{ij}^{inhibidora} \right\rangle = -4\lambda\alpha. \tag{10}$$

Esta elección nos permite conseguir el balanceo en cuanto a intensidades absolutas de los pesos sinápticos ya que las inhibidoras tendrán una "fuerza" o intensidad cuatro veces mayor que las excitadoras como indicaba la Figura 10. Por simplicidad consideramos que las dos Gaussianas tienen la misma varianza σ^2 que consideramos como un parámetro del modelo.

Para $\eta = 0.8$ se obtiene un balance entre las neuronas excitadoras e inhibidoras similar a la observada en la corteza cerebral, ya que de esta manera las neuronas excitadoras o piramidales representaran el 80 % del total de las neuronas corticales, mientras que las neuronas inhibidoras o interneuronas constituirán el 20 % restante. El factor $\alpha = M/N$ en la ecuación (5) ha sido usado por dos razones. La primera, ω_{ij}^B tiene que estar normalizado con $1/N$ ya que de lo contrario la corriente sináptica total $h_i = \sum_j \omega_{ij} s_j \varepsilon_{ij}$ divergería.. La segunda razón por la que usamos α es que ω_{ij}^B tiene que ser comparable al término de Hebb cuando el número de patrones M aumenta. Si no lo hiciéramos así, al aumentar M el término ω_{ij}^B seria despreciable respecto al termino Hebbiano y nuestro modelo quedaría muy limitado.

Como hemos dicho antes, el término hebbiano de pesos viene dado por la ecuación (1), donde $\{\xi_i^\mu = 0, 1\,;\, i = 1, \ldots, N\}$ representa los M patrones almacenados con la distribución de probabilidad

$$p\left(\xi_i^\mu\right) = a\delta\left(\xi_i^\mu - 1\right) + (1-a)\left(\xi_i^\mu\right) \tag{11}$$

donde $a = \langle \xi_i^\mu \rangle$ es el nivel medio de actividad neuronal en el patrón.

El balance de pesos sinápticos introducido en nuestro modelo se muestra de forma más detallada en la Figura 11,donde para un determinado valor de N, M y c se calculan las distribuciones de pesos sinápticos en la red neuronal cuando sólo hay término Hebiano (rojo), cuando los pesos sinápticos son son

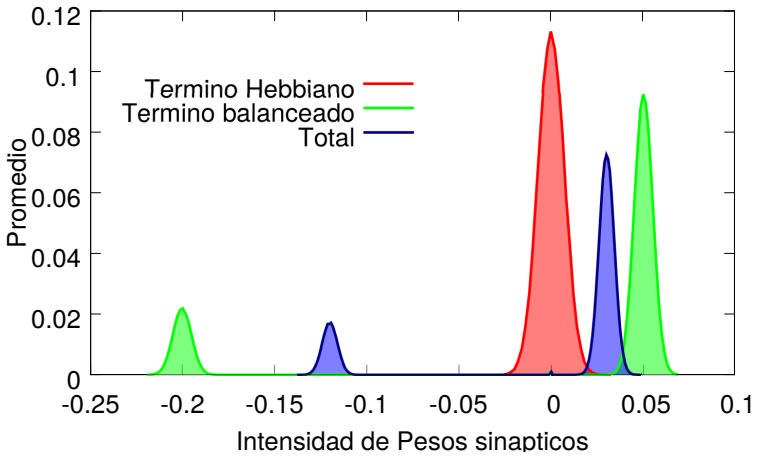

Figura 11: En la figura aparecen los histogramas de los pesos sinápticos calculados a partir de la regla de Hebb (rojo), el termino balanceado dado por la ecuación (7) (verde) y la suma de ambos que dan el peso sináptico que usaremos en el modelo presentado en este trabajo (azul). En este gráfico se han usado los siguientes parametros: $N = 1000$, $M = 50$, $c = 0.4$

sólo balanceados sin término Hebbiano (verde) y cuando consideramos ambos términos juntos como en nuestro modelo (azul).

Tenemos así una red totalmente conectada con N neuronas binarias cuyos posibles estados $s_i = 0, 1; \forall i = 1, \ldots, N$ representan a neuronas en estados silencioso y activo, respectivamente, y conectadas mediante sinapsis, excitadoras o inhibidoras cuyos pesos están distribuidos de acuerdo a una distribución bimodal balanceada como la representada en la Figura 11 (azul). Ahora definimos la evolución de estados de la red, donde cada neurona obedecerá la siguiente dinámica probabilística, paralela y síncrona [14]

$$P[s_i(t+1) = 1] = \frac{1}{2}\left\{1 + \tanh\left[2\beta\left(h_i(\mathbf{s}, t) - \theta_i\right)\right]\right\} \quad \forall i = 1, \ldots, N \tag{12}$$

donde $h_i(\mathbf{s}, t)$ es el campo local o la corriente sináptica total que llega a la neurona i , y esta definida como

$$h_i(\mathbf{s}, \mathrm{t}) = \sum_{j \neq i} \omega_{ij} s_j(t) \varepsilon_{ij} \tag{13}$$

donde ε_{ij} es la matriz de conectividad, por simplicidad hemos considerado una topología donde todas las neuronas están conectadas por lo que $\varepsilon_{ij} = 1; \forall i, j$. La variable $s_j(t)$ representa el estado actual de neurona presináptica en la localización j ,y como antes ω_{ij} es la matriz de pesos sinápticos donde cada elemento especifica la fuerza de la conexión entre las neuronas i y j. Dado que

como antes $\beta = T^{-1}$ es el inverso del parámetro de temperatura (que controlara el nivel de ruido térmico), se tiene que $\beta \to \infty$ implica una dinámica de neuronas determinista. Por ultimo definimos el término de umbral de disparo que aparece en la dinámica de la red (12), en la forma estándar en función de los pesos sinápticos, esto es $\theta_i = \frac{1}{2}\sum_j \omega_{ij}$.

En el presente modelo podremos calcular lo bien que un patrón almacenado puede ser recuperado mediante la dinámica de la red definiendo la función de *overlap (2)* ya comentada. Debido a la simetría patrón-antipatrón inherente al modelo presentado, una memoria ξ^μ (patrón) dada sera recuperada por medio de la dinámica del sistema cuando $|m^\mu(\mathbf{s})| = 1$. Podemos también medir la actividad de la red neuronal, esto es, cuantas neuronas disparan conjuntamente en un mismo instante, mediante el parámetro de orden

$$m(\mathbf{s}) \equiv \frac{1}{N}\sum_j (2s_j - 1) \tag{14}$$

este parámetro puede ser relacionado fácilmente con la actividad media de la red o *firing rate* $\nu(\mathbf{s}) = \frac{1}{N}\sum_i s_i = \frac{m(s)+1}{2}$, de tal manera que la solución con $m(\mathbf{s}) = 1$, corresponde a una población de neuronas con una alta actividad neuronal ($\nu = 1$, o estado Up) y $m(\mathbf{s}) = -1$ que corresponde a una población neuronal apagada ($\nu = 0$, o estado Down).

2.3. Resultados

La principal fenomenología emergente de nuestro modelo se muestra en las figuras 12, 13 y 14. Presentamos solo resultados correspondientes a $M = 1$, es decir el sistema almacena sólo un patrón, aunque el estudio se puede generalizar de forma sencilla a muchos más patrones, includo a $M = \alpha N$ con $N \to \infty$. La curva de magnetización es una herramienta muy potente a la hora de visualizar la tendencia que tiene el sistema para recordar la información de las memorias que tiene almacenadas els sistema para distintos valores del parámetro de agitación térmica T. Primero comprobamos que el modelo en el límite en el que los pesos sinapticos (ω_{ij}) no tienen balanceado alguno, esto es cuando la regla de aprendizaje es totalmente Hebbiana ($c = 1.0$) se recupera el modelo estándar de Hopfield. Este límite queda reflejado en la imagen inferior derecha de la figura 13, que si comparamos con la curva de magnetización del modelo estándar que hemos visto mas arriba vemos que coinciden. Ambas curvas, tienen como temperatura crítica $T = 1$ y el cambio de fase de zonas de memoria a zonas de no memoria es de segundo orden.

El modelo estándar puede ser también considerado como una buena descripción de un material magnético debido a la analogía de los estados de la neuronas $s_i = 0, 1;\ \forall i = 1, \ldots, N$ con los espines de las partículas en la materia ordinaria, ya que las interacciones entre espines pueden ser descritas mediante los pesos ω_{ij}. Como también hemos ya mencionado$T$, que es una medida del nivel de ruido en el sistema.

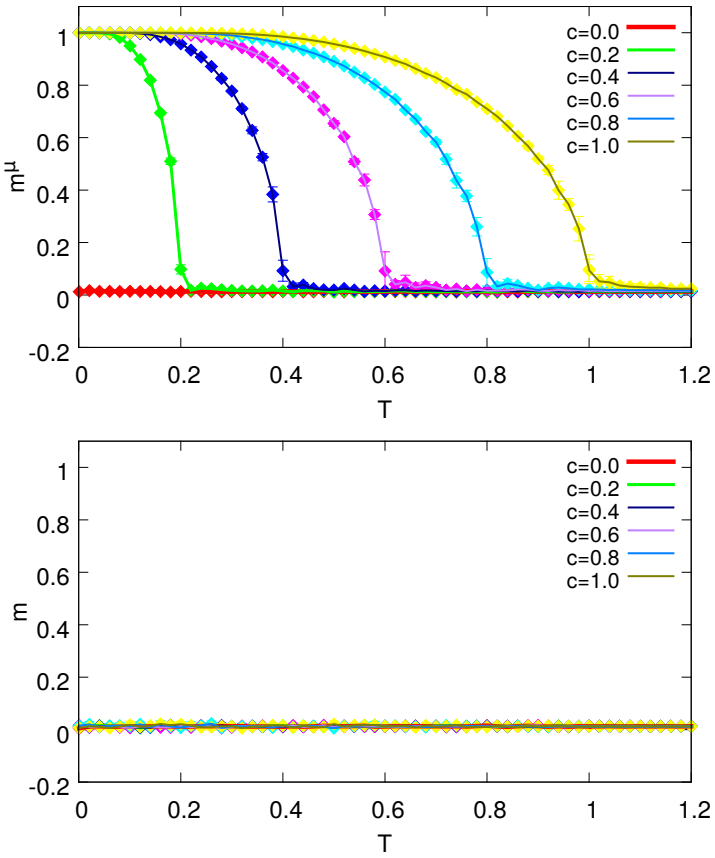

Figura 12: (Arriba) Se muestran representadas las distintas curvas de magnetización que obtenemos para distintos valores de c. Los puntos han sido calculados a partir de analisis estadístico de las distintas realizaciones del sistema consideradas para cada elección de valores de parámetros. Las barras de error son las desviaciones típicas de las distintas realizaciones respecto al valor medio. (Abajo) El mismo analisis que en la parte superior pero esta vez se analiza el parametro de orden m, en vez del m^{μ}.

En la Figura 12 en la parte superior tenemos las distintas curvas de magnetización de nuestro modelo para distintos valores de c. La figura muestra claramente no solo como la memoria desaparece al aumentar T, sino también como la memoria se ve afectada según la forma en que se han elegido los pesos sinápticos mediante el parámetro c. Cada punto de la curva de magnetización se corresponde a un promedio sobre varias realizaciones del sistema con los mismos parámetros, y las barras de error se han calculado mediante desviaciones típicas de los valores de las diferentes realizaciones respecto a esos promedios. Como claramente se observa las barras de error son casi despreciables, debido al alto número de realizaciones que se han hecho. Además este hecho indica que el sistema en todo el rango de parámetros no presente ningún tipo de metaestabilidad. Sólo cerca de la temperatura donde tiene lugar el cambio de fase las fluctuaciones respecto de la media tienden a ser importantes como era de esperar debido a que se trata de un punto crítico.

Observando el comportamiento de la curva de magnetización para distintos valores de c, es de destacar que a medida que c se hace más pequeña la temperatura crítica ,por debajo de la cual la memoria emerge, decrece también de la misma forma. de las curvas decrece de la misma forma. El porque de esta tendencia resulta bastante trivial si uno tiene en cuenta que la curva de magnetización no hace más que medir el grado de correlación entre el estado del sistema final y el patrón almacenado. De modo que para que la red tienda hacia la configuración del patrón, necesitara tener información del mismo y esto solo ocurre con el termino de aprendizaje de Hebb, ω_{ij}^{Hebb}, ya que el término balanceado ω_{ij}^{B} esta totalmente descorrelacionado con el patrón y su propósito no es otro que recrear el balanceado en cuanto a sinapsis excitadoras e inhibidoras. Por lo tanto, a medida que crece la importancia del termino balanceado respecto al de tipo Hebbiano ($c \to 0$) , estaremos descorrelacionando el peso sináptico total, ω_{ij}^{Total}, tal y como aparece en (5), es decir, estaremos introduciendo un tipo de ruido "balanceado" a nuestro modelo que poco a poco destruye memoria del sistema. Por eso lo trivial de la tendencia, porque se puede observar como la temperatura critica de cada curva T_{crit} se hace mas pequeña cuanto menor es la aportación del termino de Hebb. Nuestro sistema tendrá ahora dos tipos de ruido: el que tiene como origen la agitación térmica de las distintas neuronas y el que procede del balanceado de la red.

En la imagen de la parte inferior de la Figura 12 se muestra la evolución del paraámetro de orden m, que mide la actividad de la red, para distintos valores de temperatura y c. Para todos los valores de los parámetros (c, T) vemos que el valor de el parámetro de orden es siempre nulo $m = 0$ Como ya se ha comentado la relación entre este parámetro de orden y la actividad media de la red o mean firing rate esta dada por

$$\nu(\mathbf{s}) = \frac{1}{N}\sum_i s_i = \frac{m(s)+1}{2}$$

de manera que para todo el espacio de fases de (c, T) la red tendrá una actividad media de $\nu(\mathbf{s}) = 0.5$. Esto es, habrá el mismo numero de neuronas en

estado silencioso ($s_i = 0$) y en estado de disparo ($s_i = 1$) en cualquiera de los estados estacionarios del sistema. En las zonas de memoria esto ocurre debido a que al recuperar el patrón, la red adquiere su configuración y con ello la actividad media de los patrones. Recordando (11), elegimos los patrones como configuraciones particulares de la red con actividad media $a = \langle \xi_i^\mu \rangle = 0.5$. Por lo tanto la actividad media de los estados estacionarios ferromagnéticos o de memoria siempre deben de ser igual al valor de a previamente escogido. En la zona de no memoria, debido a la alta estocasticidad las fluctuaciones son muy grandes que hace al sistema en el estado estadionario, esté desordenado con las neuronas fluctuando en el tiempo entre sus dos posibles estados de foma que $m^\mu = 0$ y $m = 0$.

En la Figura 13 se muestran curvas de magnetización como en la Figura 12, pero esta ocasión en vez de promediar estadísticamente se han representado para cada valor de T las soluciones estacionarias que se alcanzan para cada realización de la red. Se observa claramente que todos los puntos estacionarios que se alcanzan siguen la curva del promedio estadístico. Este hecho demuestra que las soluciones estacionarias que se obtienen son las únicas soluciones estables posibles. Si esto no fuera así y tuviésemos mas soluciones estables posibles al iniciar el sistema desde distintas condiciones iniciales alguna repetición debería haber terminado en la cuenca de atracción de una de esas soluciones. Por otra parte, todas las realizaciones del sistema (distintos puntos con una misma temperatura) terminan con el mismo valor de m^μ, excepto en las zonas criticas (cerca del cambio de fase) donde aparecen mayores fluctuaciones.

Hemos comentado, que a medida que $c \to 0$, la temperatura de cambio de fase para valores decrecientes de c decrecen también y podemos entonces entender el balanceado de los pesos ω_{ij}^{Total} como un ruido que destruye memoria. Ahora bien, a pesar de poder considerarse un ruido, al igual que la agitación térmica resulta interesante diferenciar tanto donde actúan estos ruidos, como la manera en la que actúan.

En primer lugar tenemos la temperatura, formalmente aparece reflejada como su inversa ($\beta = 1/T$) en la ecuacion (12). La función $\tanh(x)$ es una función sigmoide y si nos fijamos tan solo en los valores positivos de x, tenemos dos zonas: para valores pequeños de x hasta un x_0 dado, la función crece lentamente. Mientras que a partir de x_0 la función satura y adquiere un valor constante. Lo que tenemos que entender aquí es que la zona de x bajas corresponde a las altas temperaturas, mientras que los x altos corresponde a las bajas temperaturas. En la regla de disparo, el sistema compara el campo local, h_i, de cada neurona con el valor de su umbral, θ_i, y determina si debe disparar o no. Si estamos en el régimen de bajas temperaturas, estaremos en la zona saturada de la tanh de tal manera que la respuesta sera determinista (si $h_i > \theta_i$ disparar, en caso contrario se quedara en silencio). En cambio a medida que nos acercamos a la zona de altas temperaturas, la función sigmoidea no estara saturada y aun siendo $h_i > \theta_i$, no sera suficiente para que la neurona dispare. Dado que una neurona es binaria, podemos decir que la temperatura del sistema "rompe" estocásticamente la información, ya que dejará apagadas a neuronas que deberían estar disparando y viceversa.

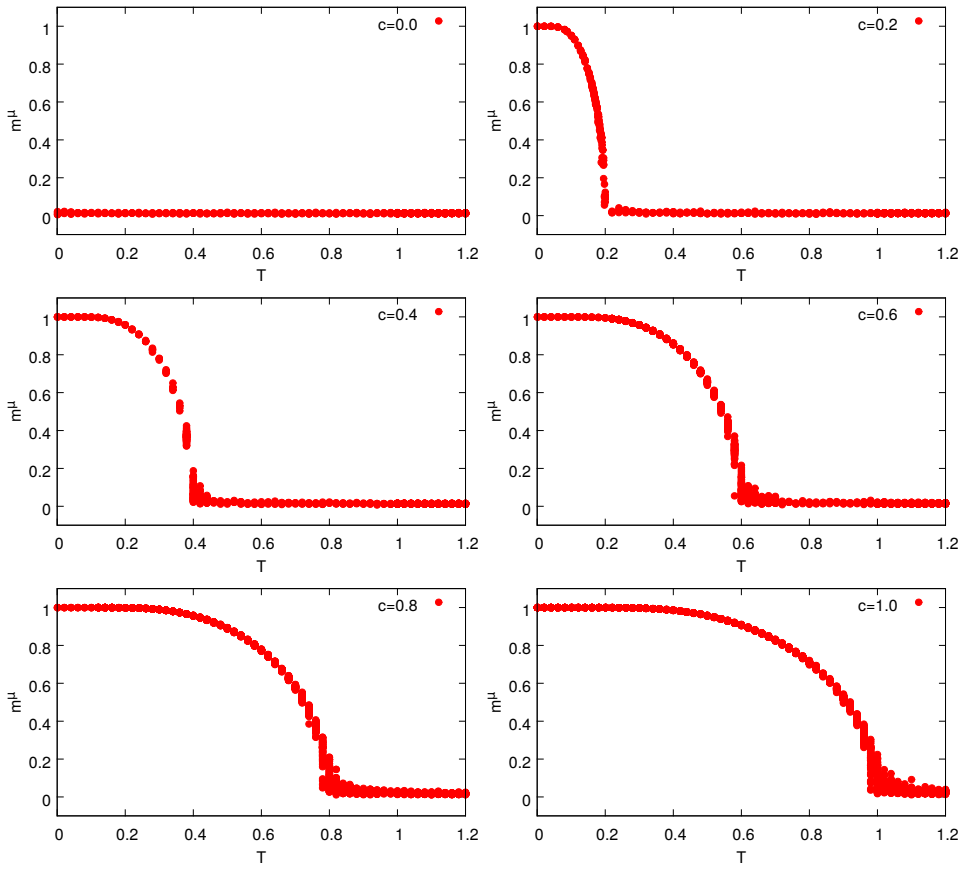

Figura 13: En las distintas imágenes aparecen curvas de magnetización para distintos valores del parámetro c. Esta vez no se ha hecho estadística y se han representado las soluciones estacionarias obtenidas en cada una de las simulaciones corridas.

Por otro lado tenemos el ruido creado por el balanceado de los pesos sinápticos. En este caso, el ruido irá implícito en los pesos sinápticos ya que estamos perturbándolos en menor o mayor medida (dependiendo del valor de c). Los pesos sinápticos son usados para computar el valor del campo local de cada neurona mediante (13) y por lo tanto si estos no son los que deberían, por ejemplo para recordar un patrón, parte de información codificada en ese patrón se estará perdiendo en cada paso de la evolución. Pero ahí esta la gran diferencia entre los ruidos: los pesos sinápticos no son binarios, de hecho son variables continuas de tal manera que el ruido balanceado puede hacer crecer o decrecer el valor del ω_{ij}, incluso pudiendo cambiarlo de signo, pero el cambio no sera tan drástico como ocurre en los elementos binarios. Otra diferencia entre los ruidos, es que el ruido térmico actúa de manera probabilística en cada actualización de la red (pudiendo a veces destruir información y en otras ocasiones dejarla intacta), el ruido balanceado en cambio esta congelado ya que en nuestro modelo la matriz de pesos sinápticos se calcula al principio y no vuelve a tocarse.

Mirando la Figura 13 se puede ver la tolerancia de la red al ruido balanceado, ya que incluso teniendo $c = 0.2$, esto es, un balanceado extremo y poca contribución del término Hebbiano el sistema es capaz de desarrollar memoria. Desde un punto de vista computacional sería interesante desarrollar un modelo que estuviera balanceado correctamente y que además siguiese desarrollando memoria asociativa, y si bien es verdad, que el intervalo de temperaturas para las cuales el sistema tiene memoria con $c \to 0$ es muy pequeña, ese intervalo existe, incluso para valores muy pequeños de c.

En la Figura 14 se representa el espacio de de fases con los parámetros libres (c, T) donde se representa en forma de mapa de colores los valores que puede alcanzar la memoria, m^μ, en el estado estacionario. Como ya se ha comentado mas arriba, al introducir un nuevo parámetro como es c, la curva de magnetización conviene representarla en tres dimensiones ya que nos interesa ver el grado de memoria que el sistema alcanza en función de T y c. En la Figura 14 se han dibujado las evoluciones de las temperaturas criticas en las que se produce un cambio de fase (T_{crit}) con resepcto a c (línea azul). A la hora de discernir en las simulaciones si el sistema tiene memoria se usa el criterio de que el sistema tiene memoria si es capaz de recuperar por encima de $m^\mu \geq 0.75$, y la curva que delimita esa zona de memoria ha sido dibujada también en la Figura 14 (curva verde). El diagrama de fases representando en la Figura 14 nos da una clara imagen global del comportamiento emergente de nuestro sistema. Se puede ver como tanto la T_{crit} como la temperatura que marca la zona a partir de la cual la memoria se empieza a deteriorar, ambas dependen de forma lineal en c (tienen distintas pendientes pero ambas son lineales). Este resultado era en cierta forma esperado dado la simplicidad del modelo propuesto, de forma que se podía facilmente intuir que cuanto mayor fuera el valor de c mayor sería la zona de memoria. Sin embargo lo que no resulta tan trivial es que la zona a partir de la cual se empieza a perder memoria y por lo tanto se empieza a intuir el cambio de fase de segundo orden es de diferente anchura en el rango de temperaturas para cada valor de c. En la Figura 13 vemos como para valores altos de c el intervalo de temperaturas desde la zona de memoria hasta

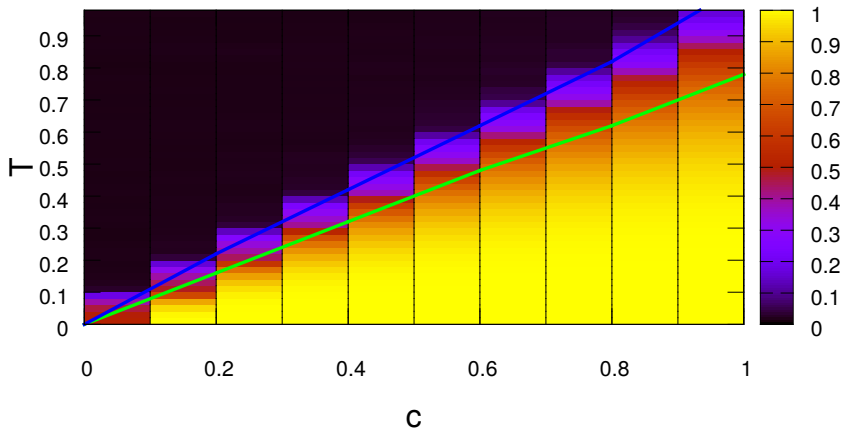

Figura 14: Diagrama de fases de los parámetros (c, T) donde los distintos colores indican los distintos valores que adquiere el solapamiento m^{μ}. La curva verde marca el limite a partir del cual se estima que no hay memoria $(m^{\mu} \geq 0.75)$. La curva azul une los distintos puntos críticos donde se produce el cambio de fase entre la zona ferromagnética y la paramagnética.

la temperatura critica donde se da el cambio de fases es mas ancho que con valores de c más pequeños. Este hecho se puede entender facilmente si se tiene en cuenta que para c más grande el término Hebbiano es más relevante por lo que se un pequeño aumento de temperatura que pueda desordenar localmente una neurona y sus vecinas apenas afecta a la memoria (ya que es una propiedad global de la red). Si embargo al disminuir c la componente hebbiana es menos importante de forma que un pequeño aumento de T induce una pérdida importante de memoria. Esta propiedad se ve claramente reflejada en la Figura 14, donde la zona de transición entre memoria y no memoria corresponde al área limitada entre las curvas verde y azul.

Parte III
Conclusiones

El trabajo que se ha presentado en esta memoria consta de dos partes. En la primera se ha hecho una breve descripción de las redes neuronales autoasociativas y en particular se ha introducido el paradigma de este tipo de redes, el llamado modelo de Hopfield, mostrando su funcionamiento y el contexto en el que se pueden usar. Con la motivación de extender este modelo a situaciones más realistas, incluyendo aspectos biológicos descritos en medios neuronales reales, hemos propuesto un modelo de red neuronal autoasociativa con pesos sinápticos que incluyen de forma natural el balanceado entre excitación e inhibición observado en las sinapsis de la corteza cerebral. Esta propuesta otorga a nuestro sistema de cierto grado de heterogeneidad sináptica funcional pues los pesos sinápticos se distribuyen de acuerdo a distribución de probabilidad bimodal, con un modo positivo (excitador) y otro negativo (inhibidor). Ademas el modelo incluye cierta información de memorias aprendidas con probabilidad c a través de una contribución de pesos sinápticos hebbiana.

Teniendo en cuenta que el objetivo era analizar las propiedades emergentes del modelo propuesto, podemos concluir que, pese a la heterogeneidad y aleatoriadad introducida por el término balanceado, siguen existiendo regiones en el espacio de parámetros donde el sistema tiene la propiedad de memoria asociativa, debido al término hebbiano. Nuestro análisis del modelo muestra que incluso para valores de c muy bajos, es decir cuando la contribución del término de información hebbiano es muy pequeñas, el sistema puede recuperar la información codificada en los patrones a temperatura suficientemente baja. Ciertas propiedades emergentes del modelo como son el decrecimiento de la T_{crit} para la aparición de memoria asociativa y la cada vez menor zona de transición en la memoria para valores decrecientes de c son bastante triviales y pueden ser explicados en términos cualitativos y sin tener que entrar en matemáticas complicadas. No por ello resulta menos interesante nuestro modelo, ya que puede ser el punto de partida de otros modelos mas específicos. Con el presente modelo hemos diseñado una red de memoria asociativa que puede modelizar de forma muy simple la corteza cerebral y debido a que en esta zona del cerebro se manifiestan muchos síntomas de enfermedades neurodegenerativas como el Alzheimer, el Autismo [15] o la enfermedad de Parkinson . Versiones más realistas de nuestro modelo, incluyendo por ejemplo neuronas de tipo integración y disparo (IF) o tipo Hodgkin-Huxley [6], o incluyendo sinapsis dinámicas, podrían ser útiles en la comprensión de estas enfermedades.

El siguiente paso para una mejor comprensión de este modelo seria el análisis de los efectos que pueden tener los parámetros relevantes en el modelo (c, M, T). Todo el análisis lo hemos reducido para la situación de un único patrón $(M = 1)$, pero sería interesante determinar la capacidad del sistema para recordar mayor número de patrones a la vez que se preserva el balanceado neuronal. Un aumento de patrones memorizados tendría una incidencia directa en la memoria (no así

en el balanceado) ya que este modelo ha sido diseñado para que el balanceado sea independiente del número de patrones haciendo así un modelo más robusto a la vez que más realista. El valor del parámetro c determina la manera en la que se mezclan los valores de los pesos sinápticos y estos si que tendrán una incidencia directa en la eficiencia de balanceado, por lo tanto habría que hacer un análisis mas exhaustivo sobre el rango de valores de c que pueden admitirse si queremos tener el mencionado balanceado.

El modelo aquí propuesto, debido a la simetría impuesta en los pesos sinápticos, describe una situación de equilibrio y por lo tanto admite un Hamiltoniano que puede ser analizado de una manera teórica usando las herramientas de la física estadística del equilibrio, lo que nos permitiría comparar con los datos de simulación.

Haciendo una pequeña variación en la elección de los pesos sinápticos puede desarrollarse una teoría de campo medio que pone sobre la mesa una fenomenología intrigante y altamente no trivial. Este modelo lo estamos desarrollando en la actualidad y en él se obtiene, entre otros fenómenos, frustración de la memoria en la zona de bajas temperaturas, fases reentrantes de memoria y estados no correlacionados con los patrones aprendidos, con una actividad neuronal alta y baja.

Este modelo a su vez puede ser dotado de mayor realismo variando por ejemplo la topología estructural de la red neuronal, usando en vez de una red completamente conectada unas redes con conexiones de tipo *pequeño mundo* o redes *invariantes de escala*, que son más realistas a la hora de modelizar el cerebro [8, 17]. También pueden incluirse variaciones en el tipo de conexiones (sinapsis) introduciendo plasticidades sinápticas de corto plazo que juegan un rol muy importante a la hora de almacenar memorias de una manera dinámica e inducir saltos entre estados de alta y baja actividad neuronal [18].

Referencias

[1] D. J. Amit. *Modeling brain function: the world of attractor neural network.* Cambridge University Press, 1989.

[2] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Statistical mechanics of neural networks near saturation. *Ann. Phys.*, 173:30–67, 1987.

[3] R. Bruno and B. Sakmann. Cortex is driven by weak but synchronously active thalamocortical synapses. *Science 16*, June 2006.

[4] J. Felipe and E. Jones. *Cajal on the cerebral cortex: an annotated translation of the complete writings.* Oxford University Press, 1988.

[5] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory.* Wiley, 1949.

[6] A. Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117:500–544, 1952.

[7] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–2558, 1982.

[8] S. Johnson, J. Marro, and J. J. Torres. Functional optimization in complex excitable networks. *EPL (Europhysics Letters)*, 83(4), 2008.

[9] E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science.* 4th edition, January 2000.

[10] J. Marro. *Física y vida: de las relaciones entre física, naturaleza y sociedad.* Editorial Crítica, 2008.

[11] J. Marro and R. Dickman. *Nonequilibrium Phase Transitions in Lattice Models.* Cambridge University Press, 1999.

[12] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 1943.

[13] M. Okun and I. Lampl. Balance of excitation and inhibition. 4(8):7467, 2009.

[14] P. Peretto. *An Introduction to the modeling of neural networks.* Cambridge University Press, 1992.

[15] R. Stoner, M. Chow, M. Boyle, S. Sunkin, P. Mouton, A. Wynshaw-Boris, S. Colamarino, E. Lein, and E. Courchsne. Patches of disorganization in the neocortex of children with autism. *The New England Journal of Medicine*, 2014.

[16] J. J. Torres and J. Marro. Brain performance versus brain transitions. *Scientific Reports*, 2015.

[17] J. J. Torres, M. A. Munoz, J. Marro, and P. L. Garrido. Influence of topology on the performance of a neural network. *Neurocomputing*, 58-60:229–234, 2004.

[18] Joaquin J. Torres and Hilbert J. Kappen. Emerging phenomena in neural networks with dynamic synapses and their computational implications. *Frontiers in Computational Neuroscience*, 7(30), 2013.

[19] M.V. Tsodyks. Associative memory in neural networks with the hebbian learning rule. *Modern Physics Letters B*, 03(07):555–560, 1989.

[20] C van Vreeswijk and H Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science (New York, N.Y.)*, 274(5293), dec 1996. PMID: 8939866.