

¿POR QUE UNA RED DE NEURONAS DE ESTADO CONTINUO PUEDE EVOLUCIONAR HACIA UN MINIMO GLOBAL DE ENERGIA?

J. P. Cheret y E. Ferrán

Departamento de Física, Facultad de Ingeniería, Universidad de Buenos Aires,
Paseo Colón 850, 1063 Buenos Aires

En la presente comunicación analizamos cómo se modifican las características de las trayectorias de la evolución dinámica de una red continua de Hopfield en el espacio v de estados neuronales al variar el parámetro λ que controla la ganancia del procesamiento neuronal. Mostramos que, debido a la no linealidad de la función $g(u_i)$, dichas trayectorias no tienen la dirección del gradiente de la energía E ni del gradiente de la función de Lyapunov \mathcal{L} definida por Hopfield. Determinamos los ángulos promedio $\langle \alpha \rangle$ y $\langle \beta \rangle$ entre los vectores \mathbf{V} y $\nabla \mathcal{L}$ y los vectores \mathbf{V} y ∇E , respectivamente, para distintos valores de λ , realizando un muestreo estadístico del espacio v (mediante un proceso Monte Carlo) en redes que tienen diferentes cantidades de neuronas y/o puntos fijos. También mostramos cómo varían temporalmente α y β siguiendo diversas trayectorias en dicho espacio, en el caso particular de una red construida para resolver el problema del viajante de comercio según el método de Hopfield y Tank. En todos los casos analizados se observa un valor menor de $\langle \alpha \rangle$ si λ es alto (neuronas de estado cuasi-discreto).

En 1984, J. J. Hopfield² propone un modelo de red neuronal donde cada neurona posee una función de activación $g(u_i)$ continua, monótona cre

$$V_i = \frac{1}{2} + \frac{1}{\pi} \arctan \left(\lambda u_i / 2 \right) \quad (1)$$

donde V_i es la frecuencia promedio de emisión de potenciales de acción, u_i el potencial de membrana del soma de la neurona y λ un parámetro de control que indica la ganancia del procesamiento neuronal. Las ecuaciones que definen la dinámica de estos sistemas son:

$$C_i \frac{du_i}{dt} = \sum_j T_{ij} V_j - u_i / R_i + I_i \quad (2)$$

$$V_i = g_i(u_i) \quad (3)$$

donde C_i y R_i son respectivamente, la capacitancia y la resistencia de la membrana de la i -ésima neurona, T_{ij} es la eficacia sináptica de la conexión que va de la j -ésima neurona a la i -ésima neurona e I_i es la entrada exterior constante. Hopfield demostró que el sistema posee la siguiente función de Lyapunov, la cual asegura la convergencia de la evolución dinámica del espacio de la red a puntos fijos en el espacio v :

$$\mathcal{L} = E + \sum_i \frac{1}{R_i} \int_0^{V_i} g_i^{-1}(V) dV \quad (4)$$

donde:

$$E = \frac{1}{2} \sum_i \sum_j T_{ij} V_i V_j - \sum_i I_i V_i \quad (5)$$

es la energía del vidrio de spin equivalente de la red neuronal.

El modelo de Hopfield de neuronas de estado continuo ha sido aplicado con bastante éxito a la resolución de problemas de complejidad combinatoria, en los cuales se desea obtener el mínimo global de una función de muchas variables. Para ello se ponen en correspondencia la función a minimizar con la energía E . En cambio, las redes compuestas por neuronas de estado discreto no producen soluciones adecuadas para estos problemas de optimización.

Derivando en (4) y (5), es sencillo demostrar que:

$$\frac{d\mathcal{L}}{dV_i} = - \left(\sum_j T_{ij} V_j - u_i / R_i + I_i \right) = - C_i \frac{du_i}{dt}, \quad (6)$$

por lo que:

$$\frac{dV_i}{dt} = \frac{dV_i}{du_i} \frac{du_i}{dt} = - \frac{1}{C_i} \frac{dV_i}{du_i} \frac{d\mathcal{L}}{dV_i} \quad (7)$$

La ecuación (7) muestra que las trayectorias en el espacio V no siguen caminos colineales con el gradiente de \mathcal{L} ya que dV_i/du_i es, en general, distinto para las diferentes componentes de \mathbf{V} .

El parámetro λ juega un papel preponderante en la dinámica de la red, ya que no sólo define el apartamiento de V_i con respecto a su valor máximo

permitido (una vez lograda la convergencia)² sino que, además, interviene en la relación entre las componentes del gradiente de \mathcal{L} y las de dV/dt , como puede verse en (7).

Para estudiar el grado de apartamiento de las trayectorias en v con respecto a las direcciones de los gradientes $\nabla\mathcal{L}$ y ∇E , hemos definido los siguientes ángulos:

$$\alpha = \arccos \frac{\nabla\mathcal{L} \cdot \vec{V}}{|\nabla\mathcal{L}| |\vec{V}|} \quad (8)$$

$$\beta = \arccos \frac{\nabla E \cdot \vec{V}}{|\nabla E| |\vec{V}|} \quad (9)$$

En primer lugar, realizamos una evaluación de los valores de α y β promediados sobre distintos puntos del espacio v , elegidos al azar, en función de la ganancia λ . La figura 1 muestra dicha variación para una red diseñada para resolver el problema del viajante de comercio (PVC) de acuerdo con el método propuesto por Hopfield y Tank³. En particular, la red que hemos analizado consta de 100 neuronas (permite resolver un PVC de 10 ciudades). En los estudios numéricos efectuados hemos utilizado los siguientes valores de los parámetros de la energía: $A = B = 500$, $C = 200$, $D = 150$ y $n = 13$ (para una definición de estos parámetros ver referencia 3). Las mismas curvas se obtienen para $T_{ii} = 0$ y para $T_{ii} = -C$. De lo observado en estas curvas podemos concluir que, para ganancias pequeñas, es factible encontrar trayectorias de v que se alejan bastante del camino de gradiente ∇E . Hemos encontrado el mismo comportamiento para α en redes cuyos atractores fueron fijados siguiendo la regla de Cooper para memorias direccionales por su contenido^{1,2}.

Por otro lado, analizamos como varían los ángulos α y β en función del tiempo, siguiendo una trayectoria real en el espacio v . En las figuras 2 y 3 se muestran cómo evolucionan α y β para la red del PVC. La figura 2 corresponde al caso de $\lambda = 32$ y la figura 3 al caso $\lambda = 0.1$. En ambos casos, se puede observar que durante las primeras iteraciones (en las que el sistema aún no ha definido a qué atractor finalmente caerá), ambos ángulos pueden llegar a valores de hasta 90 grados, estabilizándose finalmente en un valor menor (en la mayoría de los casos se observa que la red converge a una matriz

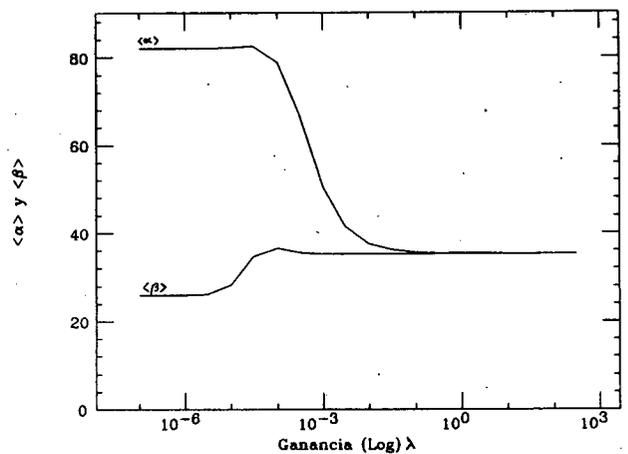


Figura 1: Variación de los ángulos promedio $\langle \alpha \rangle$ y $\langle \beta \rangle$ en función del parámetro λ , para una red que resuelve un PVC de 10 ciudades.

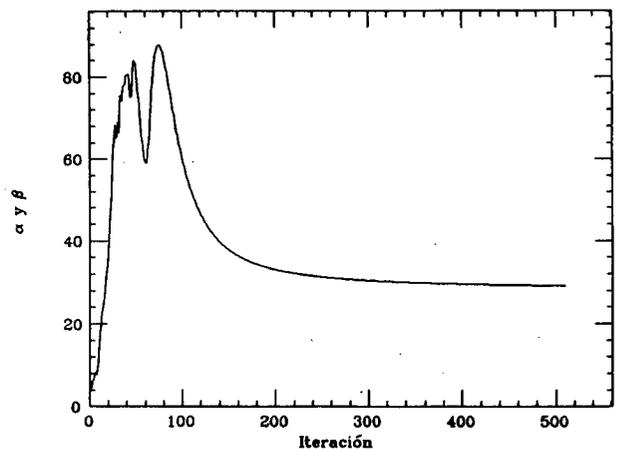


Figura 2: Variación de los ángulos promediados $\langle \alpha \rangle$ y $\langle \beta \rangle$ en función del tiempo, para una red que resuelve el PVC de 10 ciudades, cuando $\lambda = 32$.

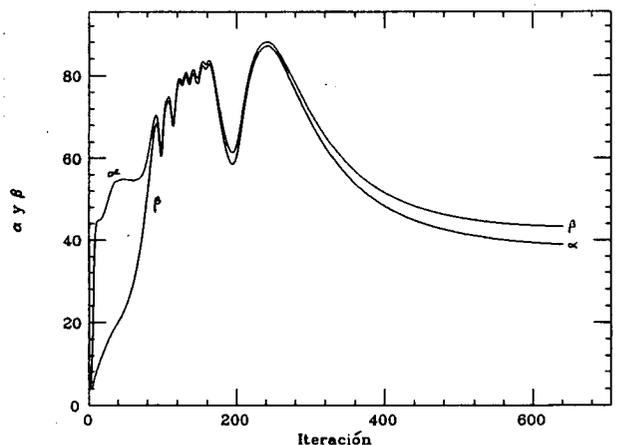


Figura 3: Idem figura 2 para el caso $\lambda = 0,1$

de permutación la cual, además, corresponde a una de las soluciones casi óptimas del PVC). En la figura 2, como la ganancia λ es alta, α y β son casi coincidentes. Hemos obtenido resultados análogos variando todavía más el rango de λ , observándose que la red converge al mismo atractor aún cuando la ganancia se modifica en cuatro órdenes de magnitud -de 0.3 a 300- (para la misma posición de las ciudades del PVC y con condiciones iniciales idénticas). La convergencia tampoco se vió afectada por el hecho que los T_{ij} no sean nulos.

Estos estudios sugieren que la mayor efectivi-

dad de las redes neuronales de estado continuo para hallar mínimos globales, se debe a que su evolución dinámica se aparta bastante de la seguida típicamente en un método de gradiente.

REFERENCIAS

1. Hopfield, Proc. Natl. Acad. Sci. (USA) **79**, 2554-2558, 1982.
2. Hopfield, Proc. Natl. Acad. Sci. (USA) **81**, 3088-3092, 1984.
3. Hopfield y Tank, Biol. Cyb. **52**, 141-152, 1985.