*Neural Processing Letters* **12:** 129–144, 2000.



# Associative Memories in Infinite Dimensional Spaces

ENRIQUE CARLOS SEGURA[1] and ROBERTO P. J. PERAZZO[2]
[1]*Departamento de Computacion, Universidad de Buenos Aires, Pabellon I Ciudad Universitaria (1428) Buenos Aires, Argentina, e-mail: esegura@dc.uba.ar;* [2]*Centro de Estudios Avanzados, Universidad de Buenos Aires, Argentina*

**Abstract.** A generalization of the Little–Hopfield neural network model for associative memories is presented that considers the case of a continuum of processing units. The state space corresponds to an infinite dimensional euclidean space. A dynamics is proposed that minimizes an energy functional that is a natural extension of the discrete case. The case in which the synaptic weight operator is defined through the autocorrelation rule (Hebb rule) with orthogonal memories is analyzed. We also consider the case of memories that are not orthogonal. Finally, we discuss the generalization of the non deterministic, finite temperature dynamics.

**Mathematics Subject Classifications (2000):** 68Txx

**Key words:** associative memory, dynamical systems, Glauber dynamics, Hopfield model, infinite dimensional state space, stability.

## 1. Introduction

Since the Little model [1,2] was formulated to describe the computational ability of an ensemble of simple processing units, appeared to be necessary to reconcile the biological evidence of a truly continuum of the neural tissue with the descriptions provided by discrete models inspired in an Ising system. While the empirical evidence always shows patterns of activity or quiescence involving patches with finite sizes, the ferromagnetic approach suggests systems with discrete processing units with no finite dimensions. In spite of this simplification all the discrete models have been remarkably successful in providing descriptions of emergent processing abilities that correspond to stilized facts concerning basic elementary cognitive processes.

In [5] Hopfield introduced the well-known two-state neuron model, while in [6] he extended his model to a continuous range of activities. However, the space of states describing the patterns of activity remained discrete, in the sense that the number of units was, at most, countable.

The purpose of the present Letter is to bridge the gap between these discrete models and the case in which processing units are viewed as points of a continuous metric space. We also require this generalization not to be made blindly, in order to preserve the salient features that have made attractive all the discrete models. In spite of the fact that the corresponding state space is an infinite dimensional functional space we require that a basic simple dynamics can be defined having asymptotic, stationary solutions that can be associated to minima of an energy functional of Lyapunov type and can be taken to represent the memories that are stored in the system.

With the above conditions there are a number of questions concerning the fundamental similarities and the possible differences between the emergent processing properties of discrete and continuous models. We deal with the nature of spurious states, the extensions of the Hebb rule and its robustness against the relaxation of the basic assumption of orthogonality of the memory states, the problem of estimating the size of the basins of attraction and of the storage capacity.

There is a particular aspect that deserves special attention and this has to do with the extensions to include finite temperature fluctuations. The biological evidence indicates that the behaviour of individual neurons involve stochastic processes such as the release of neurotransmitters into the synaptic cleft. In order to take into consideration this feature in the continuous case we extend the well known prescripiton of the non deterministic Glauber dynamics [12] through a path integral approach.

An extended version of this Letter including detailed proofs will be published elsewhere. Except for particular cases, we therefore omit detailed proofs.

## 2. Preliminaries

We assume that $\mathbf{v}(x, t)$ describes the activity of a point-like neuron located in $x$ at time $t$. This pattern of activity evolves according to:

$$\frac{\partial \mathbf{v}(x,t)}{\partial t} = -\mathbf{v}(x,t) + g_\sigma\left(\int_\Omega \mathbf{T}(x,y)\mathbf{v}(y,t)\mathrm{d}y\right) \tag{1}$$

with $\mathbf{v}(x,t) : \Omega \times R_{\geqslant 0} \to R$, $\Omega \subset X$. $X$ is a metric space, $\Omega$ a compact domain, $g_\sigma$ a *sigmoid* function, i.e. $g_\sigma \in C^1(R)$, non decreasing and odd and satisfying $\lim_{x\to\pm\infty} g_\sigma(x) = \pm \mathrm{V_M}$, $\lim_{\sigma\to\infty} g_\sigma(x) = sgn(x)$ $\forall x \neq 0$, $|\, g_\sigma(x)\,| < \min\{\mathrm{V_M}, \sigma x\}$ and $g'_\sigma(0) = \sigma$.

If we call $S$ the set of all possible states $\mathbf{v}(x)$ (patterns of activity) of the system, a solution $\mathbf{v}(x, t)$ fulfilling Equation (1) is a trajectory in $S$.

From now on we assume, without loss of generality, that $\mathrm{V_M} = 1$. As for $\mathrm{T}:\Omega \times \Omega \to R$, we assume it is continuous almost everywhere (a.e.) in order to warrant that the integral is well defined. As a natural extension of the discrete case we introduce the *local field* on (or net input to) the neuron located in $x$ when the state of the system is $\mathbf{v}(y, t)$:

$$h_t^{\mathbf{v}}(x) \stackrel{\triangle}{=} \int_\Omega \mathbf{T}(x,y)\mathbf{v}(y,t)\mathrm{d}y. \tag{2}$$

In the particular case when $t = 0$, we write $h^{\mathbf{v}}(x) \stackrel{\triangle}{=} h_0^{\mathbf{v}}(x)$. Note that $h^{\mathbf{v}}$ is linear in $\mathbf{v}$.

Given an initial condition $\mathbf{v}_0^\mu(x) \stackrel{\triangle}{=} \mathbf{v}^\mu(x, 0)$ and the solution $\mathbf{v}(x, t)$ of Equation (1) that is associated to it, we say that $\mathbf{v}^\mu(x)$ is a *memory* or an *attractor* iff:

(1) $\mathbf{v}^\mu$ is an equilibrium point, i.e.

$$\mathbf{v}^\mu(x) = g_\sigma\left(\int_\Omega \mathbf{T}(x, y)\mathbf{v}^\mu(y)\mathrm{d}y\right) \quad \text{a.e.}$$

(2) For every $t_0 \geqslant 0$ and $\mathbf{v}_0$ a different initial condition that corresponds to the solution v, there exists $\delta(t_0) > 0$ such that if $\| \mathbf{v}^\mu - \mathbf{v}_0 \| < \delta$ then $\| \mathbf{v}^\mu(\cdot, t) - \mathbf{v}(\cdot, t) \| \to 0$ when $t \to \infty$.

Thus, attractors are stationary solutions of Equation (1).

Except when indicated, we assume that S=$L^2(\Omega)$ and moreover that $| \Omega | < \infty$ ($\Omega$ has finite Lebesgue measure).

We define the *energy* of the system at time $t_0$ as:

$$\mathbf{H}[\mathbf{v}(\cdot, t_0)] = -\frac{1}{2}\int_\Omega\int_\Omega \mathbf{T}(x, y)\mathbf{v}(x, t_0)\mathbf{v}(y, t_0)\mathrm{d}x\mathrm{d}y + \int_\Omega \int_0^{\mathbf{v}(x,t_0)} g_\sigma^{-1}(s)\mathrm{d}s\mathrm{d}x \tag{2}$$

where the notation $\mathbf{H}[\mathbf{v}(\cdot, t_0)]$ means that $\mathbf{v}$ is viewed as a function of $x$. Thus, each element $\mathbf{v}$ in $\mathbf{S}$ has an energy $\mathbf{H}(\mathbf{v})$ associated to it. This is a clear extension of what has been proposed in [6] for the energy of the (discrete) model with neurons with graded activation functions (Hopfield 1984). The above definition is justified by the following:

THEOREM 2.1. *If* $\mathbf{T}$ *is symmetric, then* $\mathbf{H}$ *is monotonically decreasing with* $t$ *and it reaches its minima at states* $\mathbf{v}_{t_e}(x) = \mathbf{v}(x, t_e)$ *such that*

$$\left[\frac{\partial \mathbf{v}}{\partial t}(x, t)\right]_{t_e} = 0 \tag{3}$$

a.e. in $\Omega$, i.e. given a solution $\mathbf{v}(x, t)$ the minima of $\mathbf{H}$ coincide with the equilibrium points of the system. The reciprocal is not necessarily true: from the previous theorem it does not follow that if a solution $\mathbf{v}(x, t)$ of (1) satisfies condition (3) for some $t^*$, then $\mathbf{v}(x, t^*)$ is an attractor. For example, the trivial solution $\mathbf{v} \equiv 0$ satisfies it for all $t$, but as we will soon see its stability or instability depends on the slope $\sigma$ of $g_\sigma$ at the origin. In general, the possibility to construct non-trivial memories strongly depends on such parameter.

The sigmoid function $g_\sigma$ plays an importat role in determining in which cases the system has non-trivial stationary solutions. A necessary condition is given by the following.

THEOREM 2.2. *If* $\sigma < 1/(\mathrm{M} \mid \Omega \mid)$, *being* M *such that* $\mid \mathbf{T}(x,y) \mid \leqslant \mathrm{M}$, *then the unique stationary solution of Equation (1) is* $\mathbf{v} \equiv 0$.

Besides having $\sigma \mathrm{M} \mid \Omega \mid \geqslant 1$ as required by the preceding theorem other conditions have to be fulfilled in order to warrant the effective existence of non trivial solutions.

When $\sigma \to \infty$, the attractors approach the asymptotic bounds of $g_\sigma$. This is proven by the following

THEOREM 2.3. *If for some* $\varepsilon > 0$ *holds that* $\mathbf{T}(x,y) \geqslant 0$ *when* $\mid x - y \mid < \epsilon$, *then*

$$\lim_{\sigma \to \infty} \max_{\mathbf{v} \text{ attractors},\ x \in \Omega} \min\{1 - \mathbf{v}(x), 1 + \mathbf{v}(x)\} = 0$$

## 3. Orthogonal Memories

The case of storing orthogonal memories is particularly relevant. This can be achieved through a straightforward generalization of the Hebb rule [3]. Let $\{\mathbf{v}^\mu(x)\}$ be an orthogonal set of functions in $L^2(\Omega)$, that is $(\mathbf{v}^\mu, \mathbf{v}^\nu) = 0$ if $\mu \neq \nu$, and define:

$$\mathbf{T}(x,y) = \frac{1}{\mid \Omega \mid} \sum_{\mu=1}^{p} \mathbf{v}^\mu(x) \mathbf{v}^\mu(y) \tag{4}$$

Then the following theorem holds:

THEOREM 3.1. *The system defined by Equation (1), with* $\mathbf{T}(x,y)$ *as in Equation (4), may have any finite number p of orthogonal memories taking values in* $\{\mathrm{V}_*, -\mathrm{V}_*\}$, *where* $g_\sigma(\pm \mathrm{V}_*^3) = \pm \mathrm{V}_*$.

The solutions $\mathbf{v}^\mu(x)$ look like the example shown in Figure 1. This kind of activation patterns agrees with the intuitive generalization of the attractors of an Ising-type, spin glass discrete neural network model in which patches of full activation and full quiescence alternate randomly. These can also be viewed as the vertices of an infinite (continuous) dimensional hypercube. In addition, orthogonal memories fulfill the following.

LEMMA 3.2. *If the memories are orthogonal, the distance between any two of them is always the same.*

Therefore the following corollaries hold.

COROLLARY 3.3. *The orthogonal memories are never dense under the* $L^2(\Omega)$ *norm.*

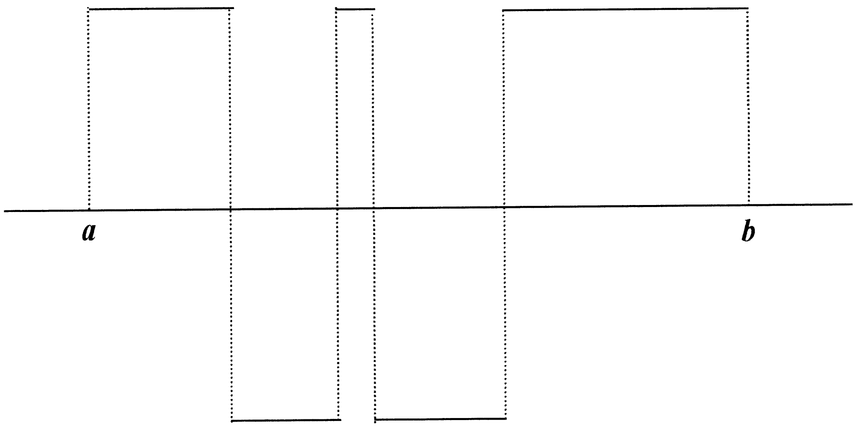


*Figure 1.* A memory in the space $S = L^2[a, b]$.

COROLLARY 3.4. *The set of orthogonal memories is at most countable.*

The question arises whether the set of orthogonal fixed points of Equation (1) can be infinite. Note firstly that it must be countable: the elements $\mathbf{v}^\mu$ as they were defined belong to $L^2(\Omega)$, a separable space; hence every orthogonal set in it must be countable. However even an infinite countable number of orthogonal fixed points is not possible while preserving the integrability of **T**. Note that if there are $k_\mu$ changes of sign in $\mathbf{v}^\mu$ (for example, if $\mathbf{v}^\mu$ as in Figure 1, $k_\mu = 4$) then each term of the form $\mathbf{v}^\mu(x)\mathbf{v}^\mu(y)$ divides the domain $\Omega \times \Omega$ in $(k_\mu + 1)^2$ square regions. Moreover, each region is separated from the next by discontinuity lines because that term takes the constant values $+V_*^2$ or $-V_*^2$. If the set of memories is infinite, the number of terms in **T** that are added is also infinite, therefore those discontinuity lines are dense at least in a neighborhood of some point, and **T** ceases to be sectionally continuous.

Most of the discussions on discrete models are made in the thermodynamic limit in which the number of neurons and of memories are taken to tend to infinity while their ratio is kept constant. This possibility is lost in the continuous limit: the number of memories may be as large as desired but has to be finite, while neurons are infinite, in fact, more than countable. This is certainly not a problem as far as the biological plausibility of the model is concerned.

### 3.1. STABILITY OF THE SOLUTIONS

In this section we investigate under which conditions the elements $\mathbf{v}^\mu$, satisfying the hypothesis of orthogonality, are stable fixed points for Equation (1) (i.e. memories of the system).

THEOREM 3.5. *The elements* $\mathbf{v}^\mu$ *are stable fixed points of Equation (1) iff*

$$g'_\sigma(V_*^3) < \frac{1}{V_*^2}$$

THEOREM 3.6. *(Stability of zero) The trivial solution* $\mathbf{v} \equiv 0$ *is stable iff* $g'_\sigma(0) = \sigma < 1/V_*^2$.

We provide a sketch of the proof of theorems 3.5 and 3.6:

The computation of the directional derivatives of $\mathbf{H}(\mathbf{v})$ at an arbitrary point yields:

$$D_w\mathbf{H}(\mathbf{v}) = -\frac{1}{|\,\Omega\,|}\sum_{\nu=1}^{p}(\mathbf{v}^\nu, \mathbf{v})(\mathbf{v}^\nu, w) + (g_\sigma^{-1}(\mathbf{v}), w) \tag{5}$$

with $w \in L^2(\Omega)$ and $\|\, w \,\| = 1$. Now, if $\mathbf{v} = \mathbf{v}^\mu$, using the orthogonality condition and noting that $\|\, \mathbf{v}^\mu \,\|^2 = V_*^2 \,|\, \Omega \,|$, it follows that $D_w\mathbf{H}(\mathbf{v}) = \mathbf{0}$.

Similarly, it is easy to check that $D_w\mathbf{H}(\mathbf{v})$ vanishes in general for any element in the set $span\{\mathbf{v}^\mu\}_{\mu=1}^p$, i.e. linear combinations of the memories, when those combinations take values on the set $\{V_*, -V_*, 0\}$. The computation of the second directional derivative yields:

$$D^2_{w^2}\mathbf{H}(\mathbf{v}) = -\frac{1}{|\,\Omega\,|}\sum_{\nu=1}^{p}(\mathbf{v}^\nu, w)^2 + \left(\frac{\partial}{\partial \mathbf{v}} g_\sigma^{-1}(\mathbf{v})w, w\right)$$

But the $\mathbf{v}^\nu$ are supposed to be orthogonal. Therefore, the use of Bessel's inequality yields:

$$\sum_{\nu=1}^{p}\frac{(\mathbf{v}^\nu, w)^2}{\|\, \mathbf{v}^\nu \,\|^2} \leqslant \|\, w \,\|^2 = 1 \Longleftrightarrow \sum_{\nu=1}^{p}(\mathbf{v}^\nu, w)^2 \leqslant V_*^2 \,|\, \Omega \,|$$

hence

$$D^2_{w^2}\mathbf{H}(\mathbf{v}) \geqslant \left(\frac{\partial}{\partial \mathbf{v}} g_\sigma^{-1}(\mathbf{v})w, w\right) - V_*^2$$

for any w in S, $\|\, w \,\| = 1$. Then, a necessary and sufficient condition for an element v in S to be a minimum of $\mathbf{H}$ is:

$$(g_\sigma^{-1\prime}(\mathbf{v})w, w) - V_*^2 > 0 \quad \forall\, w \in S, \; \|\, w \,\| = 1 \tag{6}$$

Theorem 3.6 follows directly from the last equation. Applying this equation to the case in which $\mathbf{v} = \mathbf{v}^\mu$ and keeping in mind that

$$g_\sigma^{-1\prime}(\mathbf{v}) = \frac{1}{g'_\sigma(g_\sigma^{-1}(\mathbf{v}))},$$

the above condition reduces to

$$\left(\frac{w}{g'_\sigma(g_\sigma^{-1}(\mathbf{v}^\mu))}, w\right) - \mathrm{V}_*^2 = \left(\frac{w}{g'_\sigma(\mathrm{V}_*^2\mathbf{v}^\mu)}, w\right) - \mathrm{V}_*^2 = \frac{1}{g'_\sigma(\pm\mathrm{V}_*^3)} - \mathrm{V}_*^2 > 0,$$

that, by virtue of the imparity of $g_\sigma$ can be rewritten as:

$$g'_\sigma(\mathrm{V}_*^3) < \frac{1}{\mathrm{V}_*^2} \quad \text{or} \quad g'_\sigma(\mathrm{V}_*^3)\mathrm{V}_*^2 < 1 \tag{7}$$

Let us compare the necessary and sufficient condition given by theorem 3.6 for the stability of zero with the uniqueness condition for the general case (theorem 2.2). When **T** is defined according to (4), the $\mathbf{v}^\mu$'s being stationary solutions of (1) and therefore $\mathbf{v}^\mu(x) \in \{\mathrm{V}_*, 0, -\mathrm{V}_*\}$, we have:

$$|\, \mathbf{T}(x, y) \,| \leqslant \frac{p\mathrm{V}_*^2}{|\, \Omega \,|} = \mathrm{M}.$$

In this case the condition for the origin to be the only solution is that $\sigma < 1/(\mathrm{M}\,|\,\Omega\,|) = 1/(p\mathrm{V}_*^2)$. This is more restrictive than what follows from theorem 3.6. Therefore, for the case of orthogonal memories there exists an intermediate range for the values of $\sigma$ ($\sigma \in [1/(p\mathrm{V}_*^2), 1/(\mathrm{V}_*^2)]$, which degenerates into a point if $p = 1$) for which the trivial solution $\mathbf{v} \equiv 0$ is an attractor, but not necessarily the only solution of (2.1). Note, in addition, that the conditions derived in theorems 3.5 and 3.6 are independent of $p$ (number of memories) and this is direct consequence of the orthogonality of the memories.

### 3.2. BASINS OF ATTRACTION

The preceeding results allow to measure the basins of attraction with respect to the $L^2(\Omega)$ norm. It is possible to prove that:

THEOREM 3.7. *For $p \geqslant 2$, each orthogonal memory $\mathbf{v}^\mu$, $1 \leqslant \mu \leqslant p$, has a basin of attraction of radius $k \geqslant \mathrm{V}_*\sqrt{|\,\Omega\,|/2}$. The equality holds for spherical basins.*

In other words, whenever $\|\, \mathbf{v}^\mu - \mathbf{v}_0 \,\| < k$, the distance $\|\, \mathbf{v}^\mu(\cdot, t) - \mathbf{v}(\cdot, t) \,\| \rightarrow 0$ when $t \rightarrow \infty$ (being $\mathbf{v}_0$ any initial condition for Equation (1) and $\mathbf{v}$ the corresponding solution).

The proof is based on the computation of the maximum value of $k$ for which $D_w\mathbf{H}(\mathbf{v}^\mu + kw) > 0 \;\; \forall w \in \mathrm{S}, \|\, \mathrm{w} \,\| = 1$.

Thus we have spherical basins with radius $\mathrm{V}_*\sqrt{|\,\Omega\,|/2}$ for each $\mathbf{v}^\mu$ and, consequently, also for $-\mathbf{v}^\mu$. Figure 2 shows a simplified two-dimensional sketch.

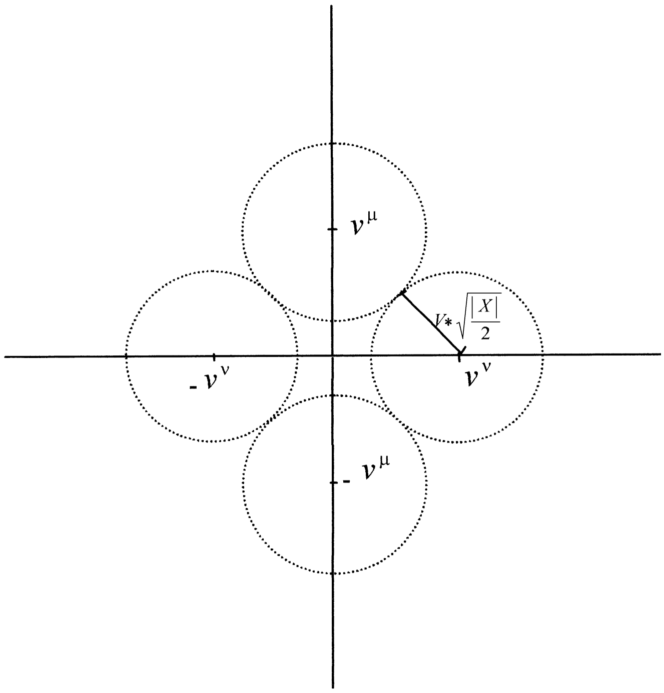


*Figure 2.* Two orthogonal memories and their inverses, each one with norm $V_*\sqrt{|\Omega|}$ and a spherical basin of attraction of radius $V_*\sqrt{\frac{|\Omega|}{2}}$.

### 3.3. SPURIOUS MEMORIES

As a consequence of the non linearity of the time evolution equation that we consider, additional, uncalled fixed points appear that are not stored in the synaptic operator **T** through the Hebb prescription. These are called *spurious states* or *spurious memories*.

There are, generally speaking, three questions concerning these memories:

(a) Are they stable?
(b) How many are they? and
(c) Where are they?

Question (b) relates to the rate at which the number of spurious states grows as more memories are stored while question (c) actually asks how near are the spurious states to the originally stored memories.

It is possible to distinguish two types of spurious states: *mixture* and *non-mixture* memories. **v** is said to be a mixture state if it can be expressed as a linear combination

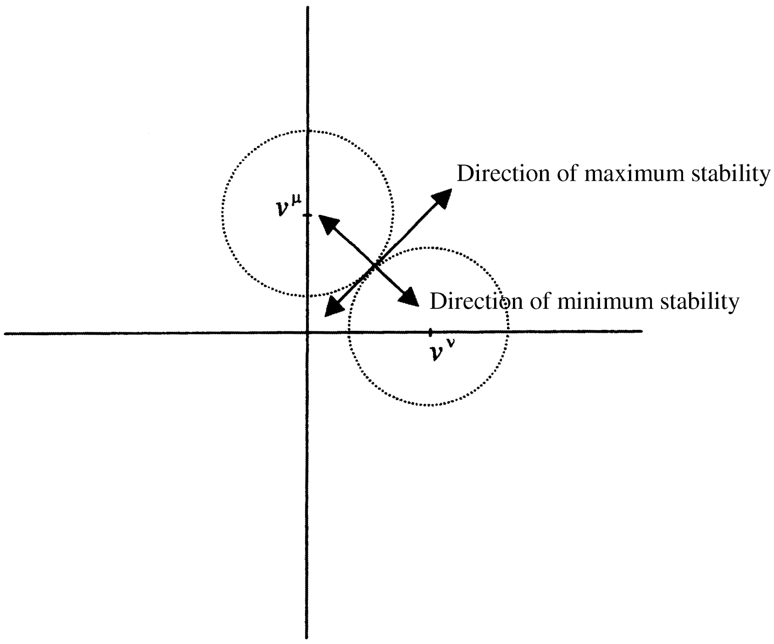


*Figure 3.* A mixture spurious state. Dotted lines indicate limits for the spherical basins of the two memories which compose the spurious state.

of the stored memories: $\mathbf{v} = \sum_{i=1}^{q} \alpha_{\mu_i} \mathbf{v}^{\mu_i}$ with $q \leqslant p$, $\mathbf{v}^{\mu_i} \in \{\mathbf{v}^{\mu}\}$ and $\alpha_{\mu_i}$ real constants. A spurious state for which the set $\{\alpha_{\mu}\}$ does not exist is called a non-mixture state.

We have already mentioned the fact that every mixture state is a fixed point if $\mathbf{v}(x) \in \{V_*, -V_*, 0\}$. This may be easily seen by using the linearity of $h^{\mathbf{v}}$:

$$h^{\mathbf{v}}(x) = \sum_{i=1}^{q} \alpha_{\mu_i} h^{\mathbf{v}^{\mu_i}}(x) = \sum_{i=1}^{q} \alpha_{\mu_i} V_*^2 \mathbf{v}^{\mu_i}(x) = V_*^2 \sum_{i=1}^{q} \alpha_{\mu_i} \mathbf{v}^{\mu_i}(x) = V_*^2 \mathbf{v}(x)$$

It also follows from the proof of theorems 3.5 and 3.6.

Thus, it is clear that only a small subset of $span\{\mathbf{v}^{\mu}\}_{\mu=1}^{p}$ contains spurious states. In particular, it follows that if $\mathbf{v}^{\mu}$ and $\mathbf{v}^{\nu}$ are memories, then $\pm\frac{1}{2}\mathbf{v}^{\mu} \pm \frac{1}{2}\mathbf{v}^{\nu}$ are fixed points of the dynamics. This implies that there exist at least $4\binom{p}{2}$ spurious (mixture) states. Fortunately, they are unstable in general. This follows from the following

THEOREM 3.8. *If* $\mathbf{v}$ *is a mixture spurious memory and there exists* $x \in \Omega$ *such that* $\mathbf{v}(x) = 0$*, then* $v$ *is a saddle point of the dynamics.*

That leads to the useful

COROLLARY 3.9. *The basins of attraction for mixture states have zero radius, in the sense of the* $L^2(\Omega)$ *norm.*

EXAMPLE. *For* $q = 2$*, all combinations of the form* $\mathbf{v} = \pm\frac{1}{2}\mathbf{v}^\mu \pm \frac{1}{2}\mathbf{v}^\nu$ *are actual spurious states and the situation can be easily illustrated (Figure 3).*

The direction of maximum unstability is given by $\mathbf{v}^\mu - \mathbf{v}^\nu$ and that of maximum stability is $\pm\mathbf{v}$ (directly towards or from the origin of coordinates).

Unlike the case of mixture memories, which can be calculated analytically, the non-mixture spurious states are very difficult to find. Indeed, at the limiting case $p \to \infty$, the following property is evident, and holds independently of the prescription used to define $\mathbf{T}$:

LEMMA 3.10. If $\{\mathbf{v}^\mu\}_{\mu=1}^{\infty}$ is complete and $p \to \infty$, there are no non-mixture spurious memories. (The set $\{\mathbf{v}^\mu\}$ is said to be *complete* in $L^2(\Omega)$ iff the minimal subspace of $L^2(\Omega)$ which contains it is the entire space $L^2(\Omega)$).

Since here we use S=$L^2(\Omega)$, a question about the sense of lemma 3.10 may arise, i.e. is there some basis of $L^2(\Omega)$ whose elements take on only two values, say $\pm V_*$? The answer is yes. The relevant example for $\Omega \subset R$ (that can be extended to $R^n$) are the *Haar wavelets*. These are an orthonormal and complete basis in $L^2(-\infty, +\infty)$. They are bi-valued, but since they are normalized, such values change from one function to another. If normality is relaxed, it is possible to force them to take values in $\{V_*, -V_*\}$. If restricted to a bounded interval $\Omega \subset R$, they form an orthogonal (but not orthonormal) and complete set in $L^2(\Omega)$. However, in case we construct $\mathbf{T}$ according to (4), this completeness can be only asymptotical: as we already saw, the number of orthogonal memories can be as large as desired, but it cannot be infinite.

## 4. Modified Hebb Rules

Until now the metrics on $X$, the space of processing units, has not been specified. This is not necessary, because the contribution of each neuron to the generation of $h_t^{\mathbf{v}}(x)$ for any $x \in \Omega$ is independent of any distance, just as in the discrete cases [5,6].

Thus, provided the memories are orthogonal, we have a memory of unbounded 'resolution' in the sense that the patches of activation and quiescence can be as small as desired. Any function $\mathbf{v}(x) : \Omega \to \{-V_*, V_*\}$ with a set of discontinuities of zero Lebesgue measure can be memorized by the system. In the case $\Omega \subset R$, this implies that the set of discontinuities must be countable. For example, the Dirichlet function, valuing 1, or $V_*$, on every rational point and 0, or $-V_*$, on every irrational point, cannot be a memory.

It is clear that a realistic model must have some lower bound for such 'resolution' at least to take into account the finite size of individual neurons. We will briefly mention two ways to tackle this problem (a more detailed description will be published soon). Both imply some modification on the Hebb rule to define the synaptic operator **T**:

(i) Introducing a probability distribution on the size of the patches of activity of the memories. A natural way of doing this is characterizing the memories by the average length $\lambda$ of their connected regions or 'patches' with equal sign and introducing a distribution in which memories with vanishing $\lambda$ have zero probability. The properties of the associative memory are well preserved, without imposing too rigorous conditions on the distribution (it suffices by requiring that the memories $v^\mu$ be independent, identically distributed and such that $E[v^\mu(x)] = 0 \quad \forall \mu$, a.e. in $\Omega$).

(ii) Redefining the synaptic operator **T**, by modulating it with a range cut-off function that decreases with the distance between neurons. It is worth mentioning that the properties of the associative memory are preserved only when such cut-off function takes both positive and negative values, as the well known *mexican hat* function that has been used to model lateral inhibition in the cortex. This fact has the pleasant feature of being in agreement with the neurophysiological evidences of the modular organization of the brain cortex [8]. This modulation in which only the interaction of neurons with its nearest neighbours is preserved causes a degradation of retrieval capacity, increasing the *crosstalk* terms between memories.

These two ways of limiting the 'resolution' are complementary in the sense that, while in the case in which memories are directly affected without changing the Hebbian operator, the stability of the system grows with $\lambda$, when modulating **T** with lateral inhibition, the effect is the inverse: retrieval gets better for increasing $\lambda$.

## 5. Finite Temperature Considerations

Until now we where mainly concerned with what is, thermodynamically speaking, a zero temperature dynamics, i.e. a deterministic law of evolution. In the present section we will formulate a finite temperature version of the continuous system that we have introduced, in the same fashion as the stochastic versions [10, 11] of the discrete Hopfield model. In that approach, which is based on the Glauber thermodynamical model [12], the probability distribution of the state of the $i$th processing unit $S_i$ at an instant $n+1$ is given by

$$\mathbf{P}(S_i = \pm 1) = \frac{1}{1 + \exp(\mp 2\beta h_i)}$$

As usual $h_i^\zeta = \sum_{i=1}^{N} \mathbf{T}_{ij}\zeta_j$ is the local field on site $i$ when all neurons have activities labelled by $\zeta$ at time $n$. The parameter $\beta$ represents the inverse temperature (we set the Boltzmann constant equal to 1).

The activity pattern of the system is therefore defined through a Markov process that can be formulated as a *Master equation* [4]:

$$\mathbf{P}(\xi, n+1) = \mathbf{P}(\xi, n) + \sum_{\zeta \neq \xi} [\mathbf{W}_\beta(\xi \mid \zeta)\mathbf{P}(\zeta, n) - \mathbf{W}_\beta(\zeta \mid \xi)\mathbf{P}(\xi, n)] \tag{8}$$

where

$$\mathbf{W}_\beta(\xi \mid \zeta) = \prod_{i=1}^{N} \frac{1}{\exp\{-\beta \xi_i h_i^\zeta\} + 1} \tag{9}$$

In our case, $\xi$ and $\zeta$ are elements in the (normed) state space S and $\mathbf{W}_\beta : \mathrm{S} \times \mathrm{S} \longmapsto [0, 1]$ is the transition probability matrix.

Being S continuous infinite dimensional it is no longer possible to construct $\mathbf{W}_\beta(\xi \mid \zeta)$ from point transition probabilities. This problem can be circumvented with a proper definition of $\mathbf{W}_\beta$. Equation (9) can be formally rewritten in the continuous time limit as the differential form of the Chapman–Kolmogorov equation [9]

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = \int_{\mathrm{S}} \{\mathbf{W}_\beta(\xi \mid \zeta)\mathbf{P}(\zeta, \mathbf{t}) - \mathbf{W}_\beta(\zeta \mid \xi)\mathbf{P}(\xi, \mathbf{t})\} \mathrm{d}\zeta \tag{10}$$

The integration has to be taken over the whole set $S$ of possible activation patterns. In the present framework in which processing units are assumed to be a continuous metric space this integration has to be expressed as a Feynmann path integral [7]. Assuming for simplicity that $\Omega = [a, b] \subset R$ and $\mathrm{S} = \mathrm{L}^2(\Omega)$, we define:

$$\int_{\mathrm{S}} \{\mathbf{W}_\beta(\xi \mid \zeta)\mathbf{P}(\zeta, t) - \mathbf{W}_\beta(\zeta \mid \xi)\mathbf{P}(\xi, t)\} \mathrm{d}\zeta$$

$$\stackrel{\triangle}{=} \int_{\zeta(a)}^{\zeta(b)} \{\mathbf{W}_\beta(\xi \mid \zeta)\mathbf{P}(\zeta, t) - \mathbf{W}_\beta(\zeta \mid \xi)\mathbf{P}(\xi, t)\} D\zeta(\tau)$$

which is a functional integral over all $\zeta \in L^2([a, b])$.

Let us express $\mathbf{W}_\beta$ as a function of the jump $r$ and of the initial state:

$$\mathbf{W}_\beta(\xi \mid \zeta) = \mathbf{W}_\beta(\zeta; r)$$

with $r = \xi - \zeta$. This leads to the following expression for the master equation:

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = \int_{\mathrm{S}} \mathbf{W}_\beta(\xi - r; r)\mathbf{P}(\xi - r, t) Dr - \mathbf{P}(\xi, t) \int_{\mathrm{S}} \mathbf{W}_\beta(\xi; -r) Dr$$

where the integrals are to be also understood as path integrals, now computed over all possible jumps $r \in \mathrm{S}$.

Now assume:

(i) Only small values of the jump contribute to $\mathbf{W}_\beta$ , i.e. $\mathbf{W}_\beta\ (\zeta; r) \approx 0$ for $\| r \| > \delta > 0$,
(ii) $\mathbf{W}_\beta$ depends smoothly on $\zeta$, that is $\mathbf{W}_\beta\ (\zeta + \Delta\zeta; r) \approx \mathbf{W}_\beta\ (\zeta; r)$ for $\| \Delta\zeta \| < \delta$ and
(iii) The solution $\mathbf{P}(\xi, t)$ of (10) also varies slowly with $\xi$.

Then we can expand the first integral of the last equation up to second order in $\xi$ [9]:

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = \int_{S} \mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)Dr - \int_{S} \{\mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)\}'(r)Dr +$$

$$+\frac{1}{2}\int_{S} \{\mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)\}''(r, r)Dr - \mathbf{P}(\xi, t)\int_{S} \mathbf{W}_\beta(\xi; -r)Dr$$

where $\{\mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)\}^{(\nu)}$ stands for the $\nu$th order derivative of $\{\mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)\}$ at $\xi$ applied to the element $(r, \ldots, r)$ in $S \times \ldots \times S = S^\nu$ (being this $\nu$th order derivative a linear application from $S^\nu$ onto $R$). The first and fourth terms of the sum cancel. Then, the equation takes on the form

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = -a_1(\xi, \mathbf{P}) + \frac{1}{2}a_2(\xi, \mathbf{P}) \tag{11}$$

with

$$a_\nu(\xi, \mathbf{P}) = \int_{S} \{\mathbf{W}_\beta(\xi; r)\mathbf{P}(\xi, t)\}^{(\nu)}(r, \ldots, r)Dr \qquad (r, \ldots, r) \in S^\nu$$

We have obtained an expression for the time evolution of the probability distribution $\mathbf{P}$ which has the form of a Fokker–Planck equation in an infinite dimensional normed space. In fact, for the case $S = R$ we get the well-known one-dimensional Fokker–Planck equation:

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = -\frac{\partial}{\partial \xi}\{a_1(\xi)\mathbf{P}\} + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}\{a_2(\xi)\mathbf{P}\}$$

with

$$a_\nu(\xi) = \int_{S} r^\nu \mathbf{W}(\xi; r)\mathrm{d}r$$

The interpretation of the Fokker–Plank equation is simple. The (functional) probability density $\mathbf{P}(\xi, t)$ is defined on the space of all possible activation patterns $\xi(x)$ and describes the evolution involved in the ‘retrieval’ process of some particular activation pattern starting from a given initial (arbitrary) configuration. The transition probability matrix $\mathbf{W}_\beta(\xi; r)$ governs this evolution through the 'moments' involved in the Fokker–Plank equation. Upon retrieval, the (stationary) probability density reached when $t$ goes to infinity is expected to be peaked in the neighborhood of the recalled memory. This final activation pattern is however not expected to be $\delta$-like due precisely to the finite temperature fluctuations. A pure deterministic case can only be recovered by letting $\beta$ go to infinity.

Note that this derivation of the finite temperature dynamics for the continuous system is a natural extension of the master equation (8) based on the transition probability matrix (9). Therefore, it could have been applied to the analog Hopfield model [6] in a direct way, since that case would have required only a simple multivariate formulation of (11) in the form

$$\frac{\partial \mathbf{P}(\xi, t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial \xi_i}\{a_1^i(\xi)\mathbf{P}\} + \frac{1}{2}\sum_{i=1,j=1}^{N} \frac{\partial^2}{\partial \xi_i \partial \xi_j}\{a_2^{ij}(\xi)\mathbf{P}\}$$

being now $a_1^i$ the usual first order moment for the $i$th coordinate, $a_2^{ij}$ the element $(i, j)$ of the covariance matrix and $N$ the number of processing units.

## 6. Conclusions

We have analyzed an extension of the Little–Hopfield model of associative memories to an infinite continuous dimensional state space. With the only assumption that the synaptic weight matrix $\mathbf{T}$ is symmetric and with non-negative diagonal elements, we derived several results that are generalizations of well known properties of discrete, Ising-type models. Two basic conditions have been analyzed: the case of orthogonal memories and the construction of the synaptic operator following the autocorrelation (Hebb) rule. The more relevant results that follow from these assumptions are:

- The number of memories is not bounded.
- The necessary and sufficient conditions for the memories and the zero to be stable, are derived in terms of the relation between the parameters of the transfer function $g_\sigma$.
- All memories have a basin of attraction with the same radius, positive in the $L^2$ norm.
- If a spurious memory vanishes at some point of the space, then its basin of attraction has zero radius in the $L^2$ norm and it is a saddle point of the dynamics.
- It is possible to impose a 'finite resolution' to the stored memories by limiting the minimum size of the activity patches. We have briefly indicated two possible alternatives for doing this.

We have also shown that a continuous model can be extended to the case in which the evolution law is non deterministic, in the same way as the Hopfield, discrete model is extended to include finite temperature effects through the Glauber dynamics. We have shown that this leads to a Fokker–Planck equation (in an infinite dimensional normed space) that governs the time evolution of the probability density distribution defined in the space of functions describing all possible activity patterns.

We believe that the system of associative memory presented and studied here is a fruitful tool for modelization in biology and neurophysiology. This model retains all the stylized facts that have made so attractive the Hopfield neural network model and its modifications, yet giving the possibility of modelization of the brain cortex as a continuous space. In fact, several neurophysiological evidences at a mesoscopic level have shown that it is not conceivable a 'discrete' regime for the activation of the cortex neurons. The realistic features of the present model are also reinforced by the robustness of the associative memory when the synaptic operator is modulated with a lateral inhibition function (*mexican hat* function). This feature is in agreement with the neurophysiological evidences in favour of the modular structure of the brain cortex.

We are confident that the results derived here can also be useful when performing something like the reverse track of what we have done, namely to reconsider the discrete case on the light of the knowledge of what happens if the state space is continuous.

## References

1. Little, W. A.: The existence of persistent states in the brain, *Math. Biosci.* **19** (1974), 101–120.
2. Little, W. A.: Analytic study of the memory storage capacity of a neural network, *Math. Biosci.* **39** (1978), 281–290.
3. Hebb, D. O.: *The Organization of Behavior: A Neuropsychological Theory,* New York, Wiley, 1949.
4. Amit, D. J.: *Modeling Brain Function,* Cambridge, Cambridge University Press, 1989.
5. Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci.* **79** (1982), 2554–2558.
6. Hopfield, J. J.: Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci.* **81** (1984), 3088–3092.
7. Feynman, R. P. and Hibbs, A. R.: *Quantum Mechanics and Path Integrals,* New York, McGraw-Hill, 1965.
8. Fodor, J. A.: *The Modularity of Mind,* Cambridge, MIT Press, 1983.
9. Van Kampen, N. G.: *Stochastic Processes in Physics and Chemistry,* Amsterdam, Elsevier, 1997.
10. Hinton, G. E. and Sejnowsky, T. J.: Optimal perceptual inference, In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (Washington, 1983),* New York, IEEE, 1983, pp. 448–453.
11. Peretto, P.: Collective properties of neural networks: A statistical physics approach, *Biological Cybernetics* **50**, 51–62.

12. Glauber, R. J.: Time dependent statistics of the Ising model, *J. of Math. Phys.* **4** (1963), 294–307.