Communicated by Francoise Fogelman-Soulie

Dynamics of Discrete Time, Continuous State Hopfield Networks

Pascal Koiran
Laboratoire de l'Informatique du Parallélisme,
Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 07, France

The dynamics of discrete time, continuous state Hopfield networks is driven by an energy function. In this paper, we use this tool to prove under mild hypotheses that any trajectory converges to a fixed point for the sequential iteration, and to a cycle of length 2 or a fixed point for the parallel iteration. Perhaps surprisingly, it seems that no rigorous proof of these results was published before.

1 Introduction

A fundamental property of discrete time, discrete state Hopfield networks is that their dynamics is driven by an energy function (Hopfield 1982). This allows the length of a limit cycle to be bounded: the parallel iteration has cycles of length 1 or 2 only, and the sequential iteration has only fixed points. These results describe rather completely the asymptotic behavior of the network, since any trajectory enters a limit cycle after a transient period.

Discrete time, continuous state Hopfield networks are also driven by an energy function (Marcus and Westervelt 1989; Fogelman-Soulié *et al.* 1989). However, a trajectory will generally *not* enter a cycle, so that the discrete-case case argument does not apply here, and the question of the convergence to a cycle arises.

In this paper, we prove under mild hypotheses that any trajectory converges to a fixed point for the sequential iteration, and to a cycle of length 2 or a fixed point for the parallel iteration. This is a very fundamental result, which seemingly has not been established previously. Papers showing the existence of a Lyapunov function (Marcus and Westervelt 1989; Fogelman-Soulié *et al.* 1989) prove that this function converges, but do not prove the same thing for the state of the network; other works deal with *local* convergence [e.g., in the neighborhood of an attractive fixed point (Michel *et al.* 1990)], and their methods cannot yield a global convergence result.

Our proof is based on an intermediate result stating that Hopfield networks generically have a finite number of fixed points. Hence the requirements in Michel *et al.* (1990) and other papers that the fixed points be isolated are not very restrictive. This result is also of interest in the case of continuous time Hopfield networks (Hirsch 1984), because these networks have the same fixed points as the discrete time networks we consider. It implies the same convergence property as in discrete time networks (Hirsch 1989). The generic finiteness of the set of fixed points is stated as an open problem in Hirsch (1989).

This paper is nearly self-contained since we give proof sketchs for most of the results we need; we refer the reader to Fogelman-Soulié *et al.* (1989) or Goles and Martinez (1990) for more details.

Our results are theoretical in nature, but are related to two questions of practical interest:

- What is the maximum number of fixed points (or length-2 cycles) for a network of a given size?
- What is the speed of convergence to a fixed point (or a length-2 cycle)?

If the network is used for instance as an associative memory, the number of fixed points can be interpreted as the number of items that can be stored. The convergence speed gives an estimate of the time it takes to retrieve an item from a corrupted version. These questions have been studied in depth in the discrete case, but hardly anything is known for continuous networks. This is not too surprising since the simpler problems considered in this paper were not addressed successfully before. One can hope that refinements of the methods presented here will yield explicit bounds of practical interest.

2 Preliminaries

We consider a network of n interconnected neurons, whose states $x_1, \dots, x_n$ belong to $[-1, 1]$. The transition function of neuron i is $x_i \mapsto f(A_i)$. A_i is the activation of neuron i, defined by:

$$A_i = \sum_{j=1}^{n} w_{ij} x_j - b_i$$

b_i is the threshold of neuron i, and w_{ij} is the weight of the connection between neurons i and j. $b = (b_i)_{1 \le i \le n}$ is the vector of thresholds. The matrix of weights $W = (w_{ij})$ is assumed to be symmetric, with a nonnegative diagonal. f is continuous, strictly increasing on an interval $[\alpha, \beta]$ ($\alpha < \beta$ and, possibly, $\alpha = -\infty$ or $\beta = +\infty$), and constant outside

$$\begin{aligned} \forall x \le \alpha, f(x) &= -1 \\ \forall x \ge \beta, f(x) &= 1 \end{aligned}$$

It seems that Fogelman-Soulié *et al.* (1989) assumes that α and β are finite, but this is in fact not necessary. If $\alpha = -\infty$ or $\beta = +\infty$, we ask that $\lim_{x \to \pm\infty} f(x) = \pm 1$.

The hypotheses listed up to this point will be assumed throughout the paper, and will not be repeated.

We will sometimes assume that f is piecewise C^1. This means that there is an increasing sequence $c_1, \dots, c_p$ such that f is C^1 on the intervals $I_0 =]-\infty, c_1], I_1 = [c_1, c_2], \dots, I_p = [c_p, +\infty[$.

In the parallel iteration mode, all neurons change state simultaneously: for $t \in \mathbf{N}$ and $1 \le i \le n$,

$$x_i(t+1) = f(A_i(t)) \tag{2.1}$$

P denotes the function associated to this iteration mode: $x(t+1) = P(x(t))$.

In the sequential iteration mode, the neurons are updated in increasing order:

$$x_i(t+i/n) = f(A_i(t+(i-1)/n)) \tag{2.2}$$

S denotes the function associated to this iteration mode: $x(t+1) = S(x(t))$. Hence all neurons are updated during one unit of time, like in the parallel case. Assuming a specific update order for sequential iterations is in fact not necessary. It is sufficient to update each neuron an infinite number of times. Note that these iteration modes have the same fixed points.

Let F be the function associated to a given iteration mode ($F = P$ or $F = S$). A cycle of length T is a sequence $(y^0, \dots, y^{T-1})$ of distinct states such that $F(y^i) = y^{(i+1) \bmod T}$. It is naturally identified to the shifted cycle $(y^1, \dots, y^{T-1}, y^0)$. We say that a sequence $(x(t))$ of iterates converges to this cycle if for any i such that $0 \le i \le T-1$, $\lim_{t \to +\infty} x(Tt+i) = y^i$.

The ball of center a and radius ϵ is $B(a, \epsilon) = \{x \in \mathbf{R}^n, ||x-a|| < \epsilon\}$, where $||.||$ denotes the Euclidian norm on $\mathbf{R}^n$.

3 Sequential Iterations

The existence of a Lyapunov function for the sequential iteration and its consequence on the length of cycles are stated in Theorem 1 and Corollary 1. The first result can be found in Fogelman-Soulié (1989). We outline its proof, because it will be useful for establishing Theorem 2, which is with Theorem 4 the main result of this section.

Theorem 1. *Let E be defined by*

$$E(x) = -x^T Wx/2 + b^T x + \sum_{i=1}^{n} \int_0^{x_i} f^{-1}(\xi) d\xi$$

E is a Lyapunov function of the sequential iteration (2.2), i.e., if $x(t+1/n) \neq x(t)$, $E(x(t+1/n)) < E(x(t))$.

Proof. Let us first assume that some integrals $\int_0^{x_j(t)} f^{-1}(\xi) d\xi$ diverge. This implies that $\beta = +\infty$ or $\alpha = -\infty$, and $x_j(t) = \pm 1$. This is possible only for $t = 0$: if $\beta = +\infty$, $1 \notin f(\mathbf{R})$ hence $\forall i, \forall t > 0, x_i(t) \neq 1$ (the same is true if $\alpha = -\infty$). It follows that $E(x(1)) < E(x(0)) = +\infty$.

Let us now assume that the integrals converge. If x_k is updated at time t, the energy variation $\Delta E = E(x(t+1/n)) - E(x(t))$ is [see Fogelman-Soulié *et al.* (1989) for details]

$$\begin{aligned} \Delta E &= -w_{kk}(x_k(t+1/n) - x_k(t))^2/2 \\ &\quad + \int_{A_k(t-1)}^{A_k(t)} [f(A_k(t-1)) - f(\xi)]\, d\xi \end{aligned} \tag{3.1}$$

The first term is clearly negative.

- If $A_k(t-1)$ and $A_k(t)$ are both smaller than α or both greater than β, $x_k(t-1+1/n) = x_k(t+1/n)$. But $x_k(t-1+1/n) = x_k(t)$ since x_k has not been updated between time $t-1+1/n$ and time t. Hence $x_k(t+1/n) = x_k(t)$.
- If the previous condition is not true, either $A_k(t) = A_k(t-1)$ and $x_k(t+1/n) = x_k(t)$ again, or $A_k(t) \neq A_k(t-1)$. In the latter case, $x_k(t+1/n) \neq x_k(t)$ and $E(x(t+1/n)) < E(x(t))$ since f is strictly increasing on $[\alpha, \beta]$. □

Corollary 1. *Any cycle of the sequential iteration is a fixed point.*

This is a standard consequence of the existence of a Lyapunov function; the proof will be omitted. In order to establish our main result, we need the following hypothesis:

(H) The network has a finite number of fixed points.

Theorem 2. *Under the additional hypothesis* (H), *the sequential iteration converges to a fixed point from any starting point* $x^0 \in [-1, 1]^n$.

Proof. Since E is a nonincreasing function, we can define

$$E_{\min} = \lim_{t \to \infty} E(x(t))$$

Let K be the set of the accumulation points of $(x(t))_{t \in \mathbf{N}}$, and let $y^0 \in K$, $y^1 = S(y^0)$. Since $E(x)$ is a continuous function of x, $E(y^0) = E_{\min}$. y^1 also belongs to K, hence $E(y^1) = E_{\min}$. According to Theorem 1, $y^0 = y^1$: the elements of K are fixed points. Thus K is finite according to hypothesis (H).

If $K = \{x^1, \ldots, x^k\}$, the following property holds:

$$\forall \epsilon > 0, \exists N, \forall t \geq N, x(t) \in \bigcup_{1 \leq i \leq k} B(x^i, \epsilon)$$

Assume instead that an infinite sequence of terms $x(t)$ does not belong to this union. This sequence has at least one accumulation point, which cannot belong to $\{x^1, \ldots, x^k\}$. This is a contradiction.

Assume that $k > 1$, and choose ϵ so that the sets $B(x^i, \epsilon)$ are disjoint. Let $t \geq N$ such that $x(t) \in B(x^1, \epsilon)$. There is a $t' \geq t$ such that $x(t') \in B(x^1, \epsilon)$ and $x(t'+1) \notin B(x^1, \epsilon)$ (otherwise, $K = \{x_1\}$). Since $t'+1 \geq N$, $x(t'+1) \in B(x^i, \epsilon)$ for some $i \neq 1$. This is impossible for ϵ small enough, because S is continuous, and x^1 is a fixed point of S. Hence $k = 1$, and $\lim_{t \to +\infty} x(t) = x^1$. □

This theorem is in fact a general result on dynamic systems driven by a Lyapunov function: the specific form of the iterated function or of the energy function is not important.

We now give a second proof of Theorem 2, assuming that f is piecewise C^1. The first one is simpler and more general; however, the second one is not completely useless, because of the following lemma, which is of independent interest.

Lemma 1. *Let $M = \|f'\|_\infty$, with f piecewise C^1. If x_k is updated at time t,*

$$|\Delta x| \leq 2\sqrt{M|\Delta E|}$$

with $\Delta x = |x(t+1/n) - x(t)|$ and $\Delta E = |E(x(t+1/n)) - E(x(t)|$.

This result can be read as "ΔE small $\Longrightarrow$ Δx small," which is a generalization of the well-known property "$\Delta E = 0 \Longrightarrow \Delta x = 0$."

Lemma Proof. Since $w_{kk} \geq 0$, a partial integration in 3.1 yields (set $u' = 1$ and $v = f(A_k(t-1)) - f(\xi)$ in the integral):

$$|\Delta E| \geq \int_{A_k(t-1)}^{A_k(t)} (A_k(t) - \xi) f'(\xi) d\xi$$

Assume for instance that $A_k(t) > A_k(t-1)$. Let $\epsilon \in (0, A_k(t) - A_k(t-1)]$:

$$\begin{aligned} |\Delta E| &\geq \int_{A_k(t-1)}^{A_k(t)-\epsilon} (A_k(t) - \xi) f'(\xi) d\xi \\ &\geq \epsilon \int_{A_k(t-1)}^{A_k(t)-\epsilon} f'(\xi) d\xi = \epsilon \left[f(A_k(t) - \epsilon) - f(A_k(t-1)) \right] \\ |\Delta E| &\geq \epsilon \left[f(A_k(t)) - f(A_k(t-1)) - M\epsilon \right] = \epsilon \left[x_k(t+1) - x_k(t) - M\epsilon \right] \end{aligned}$$

hence $0 \leq x_k(t+1) - x_k(t) \leq M\epsilon + |\Delta E|/\epsilon$.

As a function of ϵ, the right side is greater than $2\sqrt{M|\Delta E|}$ (value obtained for $\epsilon = \sqrt{|\Delta E|/M}$). On the one hand, if

$$A_k(t) - A_k(t-1) \geq \sqrt{|\Delta E|/M}$$

then

$$0 \leq x_k(t+1) - x_k(t) \leq 2\sqrt{M|\Delta E|}$$

On the other hand, if

$$A_k(t) - A_k(t-1) \leq \sqrt{|\Delta E|/M}$$

then

$$x_k(t+1) - x_k(t) = f(A_k(t)) - f(A_k(t-1)) \leq M\sqrt{|\Delta E|/M} = \sqrt{M|\Delta E|}$$

In both cases, $|x_k(t+1) - x_k(t)| \leq 2\sqrt{M|\Delta E|}$. This result still holds if $A_k(t) < A_k(t-1)$, and of course if $A_k(t) = A_k(t-1)$. □

Proof of Theorem 2. $E_{\min} > -\infty$, because $\inf\{E(x), x \in [-1,1]^n\} > -\infty$. This is due to the fact that

- $-x^T Wx/2 + b^T x$ is bounded on $[-1,1]^n$.
- either f^{-1} is bounded in a neighborhood of 1, or $\lim_{x\to 1} f^{-1}(x) = +\infty$.
- either f^{-1} is bounded in a neighborhood of -1, or $\lim_{x\to -1} f^{-1}(x) = -\infty$.

Consequently, $\lim_{t\to+\infty} |E(x(t+1)) - E(x(t))| = 0$ hence $\lim_{t\to+\infty} ||x(t+1) - x(t)|| = 0$ according to Lemma 1. It follows that K is connected. K is finite (see first proof) and connected, therefore it has a single element. □

Hypothesis (H) seems to be highly natural, and we are indeed going to prove that for "almost every" network, the set of fixed points is finite. We will make use of an elementary version of the *parametric transversality theorem* (Hirsch 1976):

Theorem 3. *Let U and V be two open sets of $\mathbf{R}^q$ and $\mathbf{R}^n$, respectively. Let $F : U \times V \to \mathbf{R}^n$ be a C^1 function. If 0 is a regular value of F, then it is a regular value of $F(\lambda, .)$ for λ in an open dense set.*

Recall that 0 is a regular value of F if for any $y \in U\times V$ such that $F(y) = 0$, the Jacobian matrix $DF(y)$ is of rank n. If 0 is a regular value of $F(\lambda, .)$, all zeros of $F(\lambda, .)$ are isolated.

Theorem 4. *When f is piecewise C^1, the network has a finite number of fixed points for (W, b) in an open dense set.*

Proof. Showing that the fixed points are isolated is sufficient, because they all belong to the compact $[-1,1]^n$.

Assume first that f is C^1 on $\mathbf{R}$. Theorem 3 is applied to $F(\lambda, x) = P_\lambda(x) - x$, where P_λ is the parallel iteration of network λ. We take $V = \mathbf{R}^n$ and $U = \mathbf{R}^q$, with $q = n + n(n+1)/2$ ("thresholds plus weights"). We are in fact going to prove that for any (λ, x), $DF(\lambda, x)$ is of rank n, i.e., all values of F are regular (not only 0). For a fixed (λ, x), let us find a regular $n \times n$ submatrix of $DF(\lambda, x)$. F_i being the ith component of F,

$$\begin{cases} \dfrac{\partial F_i}{\partial b_j} = -f'(A_i)\delta_{ij} \\ \dfrac{\partial F_i}{\partial x_j} = w_{ij}f'(A_i) - \delta_{ij} \end{cases}$$

Without loss of generality, we may assume that $f'(A_i) = 0 \iff 1 \leq i \leq r$, for a given r (possibly, $r = 0$). The submatrix composed of the columns $\left(\partial F_i/\partial x_j\right)_{1\leq i\leq n}$ for $1 \leq j \leq r$ and $\left(\partial F_i/\partial b_j\right)_{1\leq i\leq n}$ for $r + 1 \leq j \leq n$ is lower triangular, and its diagonal elements are nonzero. It is thus regular. Note that this argument remains valid if each neuron i uses a different output function f_i. This remark will be helpful in the end of the proof.

In the general piecewise C^1 case, we divide the space of activation vectors $A = (A_i)_{1\leq i\leq n}$ in $(p+1)^n$ "boxes." Each box is a product

$$B = \prod_{i=1}^{n} I_{k_i}$$

of intervals I_{k_i} on which f is C^1. $A_\lambda(x)$ is the value of the activation vector A of network λ in state x. For a given box B and λ in an open dense set O_B, any element of $X_B = \{x, x = P_\lambda(x) \wedge A_\lambda(x) \in B\}$ is isolated in X_B. Indeed, let $(f_i)_{1\leq i\leq n}$ be a sequence of functions of class C^1 on $\mathbf{R}$ such that $f_i(x) = f(x)$ for $x \in I_{k_i}$. An element of X_B is a fixed point of the network using the output functions f_i instead of f. According to our intermediate result, these fixed points are isolated for λ in an open dense set.

It follows that for λ in the open dense set $O = \bigcap_B O_B$, any fixed point in a box B_i is isolated in X_{B_i}. In order to prove that it is in fact isolated in the set of all fixed points $X = \bigcup_B X_B$, the following remark is sufficient:

If a fixed point is not isolated, X is infinite. Hence there is a box B such that X_B is infinite. We have just proved that this is impossible for $\lambda \in O$. □

4 Parallel Iterations

The existence of a Lyapunov function for the parallel iteration and its consequence on the length of cycles are stated in Theorem 5 and Corollary 2. These results can also be found in Theorems 1 and 2 of Fogelman-Soulié *et al.* (1989). We outline the proof of Theorem 5 because it will be useful for establishing Theorem 6, which is the main result of this section.

Note that we do not use the same Lyapunov function as Theorem 1 of Fogelman-Soulié *et al.* (1989). However, the two functions are equal up to a constant, according to Corollary 1 of the same paper.

Theorem 5. *Let V be defined by*

$$V(x,y) = -x^T Wy + b^T(x+y) + \sum_{i=1}^{n} \left[\int_0^{x_i} f^{-1}(\xi)d\xi + \int_0^{y_i} f^{-1}(\xi)d\xi \right]$$

$E(x) = V(x, P(x))$ is a Lyapunov function of the parallel iteration (2.1), i.e., if $x(t+2) \neq x(t)$, then $E(x(t+1)) < E(x(t))$.

Proof. One of the integrals may be divergent only at $t = 0$, as for sequential iterations. In this case, $E(x(1)) < E(x(0)) = +\infty$.

Let us now assume that all integrals are convergent. At time t, the energy variation $\Delta E = E(x(t+1)) - E(x(t))$ is (see Fogelman-Soulié *et al.* 1989 for details):

$$\Delta E = \sum_{i=1}^{n} \int_{A_i(t-1)}^{A_i(t+1)} [f(A_k(t-1)) - f(\xi)]\, d(\xi) \tag{4.1}$$

- If i is such that $A_i(t-1)$ and $A_i(t+1)$ are both smaller than α or both greater than β, $x_i(t+2) = x_i(t)$.
- If i is such that the previous condition is not true, either $A_i(t+1) = A_i(t-1)$ and $x_i(t+2) = x_i(t)$ again, or $A_i(t+1) \neq A_i(t-1)$. In the latter case, $x_i(t+2) \neq x_i(t)$ and $E(x(t+1)) < E(x(t))$ since f is increasing on $[\alpha, \beta]$. □

Corollary 2. *Any cycle of the parallel iteration is of length 1 or 2.*

This is a standard consequence of the existence of a Lyapunov function; the proof will be omitted. We now proceed to the parallel counterpart of Theorem 2. We need the following hypothesis:

(H') The network has a finite number of cycles.

Theorem 6. *Under the additional hypothesis (H'), the parallel iteration converges to a cycle of length 1 or 2 from any starting point $x^0 \in [-1, 1]^n$.*

Sketch of Proof. The proof is exactly the same as in the sequential case, with S replaced by $P \circ P$ and K replaced by the set K_0 of the accumulation points of $(x(2t))_{t \in \mathbf{N}}$. □

The counterpart of Lemma 1 is as follows:

Lemma 2. *Let $M = \|f'\|_\infty$, with f piecewise C^1. $|\Delta x| \leq 2\sqrt{M|\Delta E|}$, with $\Delta x = |x(t+2) - x(t)|$ and $\Delta E = |E(x(t+1)) - E(x(t))|$.*

Sketch of Proof. A partial integration in 4.1 yields:

$$\Delta E = \sum_{i=1}^{n} \int_{A_i(t-1)}^{A_i(t+1)} (A_i(t) - \xi) f'(\xi) d\xi$$

In the same way as for Lemma 1, one can prove that

$$|x_i(t+2) - x_i(t)| \leq 2\sqrt{M \int_{A_i(t-1)}^{A_i(t+1)} (A_i(t) - \xi) f'(\xi) d\xi}$$

whence the result. □

It follows that K_0 is connected; as in the sequential case, this gives a second proof of the convergence theorem.

5 Five Open Problems

A number of interesting questions related to the results of this paper deserve more study. We propose below five problems that are still unresolved to the best of our knowledge. As mentioned in the introduction, problems 3 and 5 have the greatest practical significance.

1. Prove that the number of length-2 cycles for the parallel iteration is finite for almost every network. We were able to solve this problem for fixed points (in Theorem 4) because the Jacobian matrix has a very special structure, which it no longer has when length-2 cycles are considered.

2. Give a condition on f ensuring that *every* network has a finite number of fixed points. This property does not hold for any f. For example, the identity function of $[-1, 1]^n$ can be realized by a network using an output function f such that $f(x) = x$ for $-1 \leq x \leq 1$, $f(x) = 1$ for $x \geq 1$, and $f(x) = -1$ for $x \leq -1$. In Theorem 4 we showed only that the finiteness property holds for *almost every* network.

3. Give an upper bound on the number of fixed points or length-2 cycles (when it is finite). In view of possible applications to associative memories, it would also be interesting to estimate the size of the basins of attraction.

4. Do the iterations still converge for a (hypothetical) network having an infinite number of cycles? In Theorems 2 and 6, convergence was proved for networks having finitely many cycles.

5. Study the speed of convergence to a limit cycle. For discrete state networks, this amounts to bounding the transient length (Fogelman-Soulié *et al.* 1989); for continuous state networks, this has been done by Michel *et al.* (1990) at the neighborhood of an attractive fixed point, but no global result is known.

References

Fogelman-Soulié, F., Mejia, C., Goles, E., and Martinez, S. 1989. Energy functions in neural networks with continuous local functions. *Complex Syst.* **3**, 269–293.

Goles, E., and Martinez, S. 1990. *Neural and Automata Networks: Dynamical Behavior and Applications. Mathematics and Its Applications.* Kluwer Academic Publishers, Boston.

Hirsch, M. W. 1976. *Differential Topology.* Springer-Verlag, Berlin.

Hirsch, M. W. 1989. Convergent activation dynamics in continuous time networks. *Neural Networks* **2**, 331–349.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558.

Hopfield, J. J. 1984. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci.* **81**, 3088–3092.

Marcus, C. M., and Westervelt, R. M. 1989. Dynamics of iterated-map neural networks. *Phys. Rev. A* **40**(1), 501–504.

Michel, A. N., Farrell, J. A., and Sun, H.-F. 1990. Analysis and synthesis techniques for Hopfield type synchronous discrete time neural networks with application to associative memory. *IEEE Transact. Circuits Syst.* **37**(11), 1356–1366.

Received November 30, 1992; accepted July 12, 1993.