Neural networks with prescribed large time behaviour

1998 J. Phys. A: Math. Gen. 31 9555

(http://iopscience.iop.org/0305-4470/31/47/015)

J. Phys. A: Math. Gen. **31** (1998) 9555–9570. Printed in the UK PII: S0305-4470(98)90286-4

Neural networks with prescribed large time behaviour

S A Vakulenko† and P V Gordon‡

Institute for Problems of Mechanical Engineering, Russian Academy of Science, 199178, St Petersburg, VO Bolshoy pr. 61, Russia

Received 5 January 1998

Abstract. The generalized Hebb rule (with a non-symmetrical synaptic matrix) allows us to create simple neural networks with complicated large time behaviour. These networks can simulate, in a sense, any dynamics and, in particular, can generate any hyperbolic attractors and invariant sets.

The explicit mathematical algorithm allows us, by adjusting the network parameters (the neuron number, coupling matrix and thresholds) to obtain a network with given large time dynamics.

1. Introduction

In this paper, neural networks are considered

$$x_i(t+1) = \sigma\left(\sum_{j=1}^{M} K_{ij}x_j(t) + \theta_i\right) \qquad i = 1, \ldots, M. \tag{1.1}$$

Over the last decade, a number of works have focused on system (1.1) playing a crucial role as a simple model of attractor neural networks [1–5]. Besides the case of continuous states ($x_i \in \mathbb{R}$), we shall study models with discrete states (spins) $s_i \in \{-1, 1\}$

$$s_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^{m} K_{ij}s_j(t) + \theta_i\right) \qquad i = 1, \ldots, M. \tag{1.2}$$

In (1.1), let us fix a sigmoidal function $\sigma \in C^\infty$ with fast decreasing derivative σ' and satisfying $\sigma(+\infty) > \sigma(-\infty)$. Denote by $\boldsymbol{P}$ the system parameters (the number of neurons M, the coupling matrix $\boldsymbol{K}$ and the thresholds θ_i): $\boldsymbol{P} = \{M, \boldsymbol{K}, \theta\}$.

The attractors of (1.1) with symmetric K are also well studied [1–3]. Other cases have been investigated, mainly by computer simulations [4–7], see also [8] and references therein. Here we consider the fully asymmetric SK model where the synaptic matrix $\boldsymbol{K}$ is defined by the following 'generalized' Hebb rule

$$K_{ij} = \sum_{l=1}^{p} A_{il}B_{lj} \qquad 0 < p < M \tag{1.3}$$

and suggest an analytic approach to this attractor problem (computer simulations are used, mainly to check the analytic results). This small generalization of the classical Hebb rule $\boldsymbol{K} = \boldsymbol{A}\boldsymbol{A}^T$ makes the appearance of many new physical effects in systems (1.1) and (1.2)

† E-mail address: vakul@microm.ipme.ru
‡ E-mail address: gordon@mech.ipme.ru

0305-4470/98/479555+16$19.50

possible. Moreover, generalization (1.3) allows us to construct exactly solvable analytical models of chaotic neural networks.

The aim of this paper is to show, by (1.3), that systems (1.1) and (1.2) possess wonderful properties. They can simulate any prescribed C^1-smooth dynamical systems (maps) and generate all structurally stable chaotic (for example, hyperbolic) attractors. More precisely, for a given structurally stable attractor, there exists a suitable choice of the parameters $\boldsymbol{P}$ such that network (1.1) generates the same (up to a homeomorphism) attractor.

One can also construct the network with a given topology of the structurally stable local attractors and their attraction basins.

Another analytical approach to chaotic behaviour in the neural networks has been suggested in a recent paper [9]. It gives a special kind of chaos connected with the so-called snap-back repellers (which occur in maps with special stretching properties). Detailed mathematical definition of such repellers is sufficiently complicated and we refer to [9] and references therein. Here we just mention the main properties of these repellers. Systems which possess them generate an infinite number of periodic cycles. Moreover, these repellers are robust: they do not vanish under small smooth perturbations of the dynamical system. (Let us note that hyperbolic invariant sets and attractors also have analogous properties.)

The model [9] can be considered as a complicated variant of (1.1). On the other hand, it is easy to show (see below) that any snap-back repellers can be obtained by (1.1).

Let us turn to case (1.2). Given map $q \to G(q)$, we can create network (1.2) simulating this map (see section 2). For discrete networks (1.2) chaotic attractors are impossible since all the trajectories are periodic (maximal period 2^M). Instead of chaos, we observe complicated periodic trajectories. The results depend on properties of a given map G and the number of spins M. Under condition (1.3) (one-dimensional case, $p = 1$), one can prove that the maximal possible period is M. Let us note here, that in some other models of asymmetric neural networks cycle lengths have exponentially large order [8]. We have carried out computer simulations for the classic chaotic maps. They have shown that approximations (1.3) with different $M \gg 1$ give a number of periods within the interval $[1, M]$. For fixed M, we usually observe a family of stable attracting periodic cycles. Periods depend on M in a very complicated and intriguing manner (see section 4 and figure 2).

This picture can be explained by recent work [10] and classical results [11]. Network (1.2), simulating map G, generates some perturbed trajectories (pseudo-orbits) of G. At each iteration, the corresponding error ϵ has the order M^{-1}.

Can such pseudo-orbits correspond to some actual trajectory of the dynamical system G? Much attention was given to this problem (see, for example, [11, 12]). It was shown that, in the neighbourhood of hyperbolic attractor or invariant sets, there holds a so-called shadow property, i.e. a trajectory can be found close to the pseudo-orbit for all i.

Our calculations and analytic results yield that (1.2) can generate some periodic pseudo-orbits (corresponding to actual orbits coexisting in the hyperbolic attractor). However, to create a number of orbits, we should use *simultaneously* a number of networks (1.2) with different M (one can imagine, for example, that we turn on and turn off some neurons). It is remarkable that, for a fixed M, the periodic cycles are very stable (it can give a new approach to the problem of periodic trajectory stabilization [13]). The average cycle length is M^ρ where the exponent ρ depends on the chaotic attractor properties (see [10]).

To conclude this introduction, let us note that models (1.1) and (1.2) can be useful for pattern recognition and associative memory devices (as the optimal control systems) for the 'neural identification' of the dynamical systems and other applications. For example, superpositions of network (1.2) and similar networks generate any complicated prescribed spatial-temporal spin patterns (see sections 2.1 and 3.3).

2. Main mathematical results

2.1. Continuous states

We begin with the case of continuous states (1.1). The main assertion can be formulated as follows. Define quantities q_i (the 'hidden' collective coordinates) by $q_i = \sum_{j=1}^{M} B_{ij}x_j$, $i = 1, \ldots, p$.

Then the 'hidden' q-dynamics will be given by

$$q_l(t+1) = \sum_{k=1}^{M} B_{lk}\sigma\left(\sum_{r=1}^{p} A_{kr}q_r(t) + \theta_k\right) = F(q, \boldsymbol{P}) \tag{2.1}$$

and the neuron dynamics is governed by q:

$$x_i(t+1) = \sigma\left(\sum_{l=1}^{p} A_{il}q_l(t) + \theta_i\right) \qquad i = 1, \ldots, M. \tag{2.2}$$

Thus, the spins form some 'coherent pattern' evolving together with $q(t)$ and they are strongly correlated (see below).

The family of maps (2.1), depending on the parameters $\boldsymbol{P}$, has the following key property (which can be called 'absorbtion with approximation').

Theorem 1. Let Q be the p-dimensional unit cube $Q = \{q : |q_i| \leqslant 1\}$ and $q \to G(q)$ any prescribed C^1-mapping that maps Q inside Q. Then, for any $\delta > 0$, there is a such choice of the parameters $\boldsymbol{P}$ that

$$|G(q) - F(q, \boldsymbol{P})| \qquad |G'_q(q) - F'_q(q, \boldsymbol{P})| < \delta \qquad (q \in Q) \tag{2.3}$$

('approximation') and, moreover, for any point $q \in \mathbb{R}^p$, iterations $q, F(q), F^2(q) = F(F(q)), \ldots, F^n(q) \ldots$ enter for the cube Q and remain in Q i.e. this cube is an absorbing set ('absorbtion'). This means that $F^n(q) \in Q$ for any $n > n_0(\delta, q)$.

In other words, this assertion can be explained as follows.

For given map G (such that $G(Q) \subset Q$), we can construct the neural network with 'hidden' q-dynamics simulating this map G for large iterations (times). Indeed, the F iterations quickly enter for the cube Q. In this cube, any prescribed G can be approximated by $F(q, \boldsymbol{P})$ with any given precision δ.

The proof of this assertion is a development of the well known results on so-called multilayered neural networks, see [14–16]. An outline of the proof can be found in section 3, mathematical details are given in appendix A.

Note that the absorbtion property plays an important role. In fact, approximation (2.3) may hold only in compact domains. Thus, we should be sure that any iteration sequence attains, at some step, such a compact domain.

This result has the following basic consequences.

(1) For any given $T > 0$ and $\epsilon > 0$, choosing sufficiently small $\delta(\epsilon, T)$, one can ϵ-approximate any families of G-iterations $q, G(q), G^2(q), \ldots, G^T(q)$. This means that, for any positive integer $j < N(\epsilon)$ and any $q \in Q$, one has $|F^j(q, \boldsymbol{P}) - G^j(q)| < \epsilon$. Thus, we can globally control families of trajectories (of any finite length).

(2) The second basic corollary is that if the prescribed map G has some robust (rough) local chaotic attractor (or invariant set) (for example, hyperbolic attractor Γ with the shadow property [8]), we can obtain the neural network with the hidden q dynamics generating a topologically equivalent local attractor (invariant set) $\tilde{\Gamma}$. (This equivalency means that there exists a homeomorphism h that maps the G-trajectories inside Γ onto the F-ones inside $\tilde{\Gamma}$).

Thus, we can say that all robust chaotic regimes, occurring in dynamics, can be realized by network (1.1).

Examples of robust maps are given, for instance, by the A-axiom Smale diffeomorphisms (which can have a number of local attractors and their basins), the Anosov maps and others. The famous simple example is given by the Arnold map: if $(x, y) \bmod 1$ denotes a point on a two-dimensional torus then $G(x, y) = (2x + y, x + y)$.

All homoclinic situations, under some suppositions, can also be realized by the networks (since, due to the Birkhoff–Smale theorem, they are connected with hyperbolic sets). Inside such hyperbolic sets the infinite number of unstable cycles coexist and periods of these cycles can be extremely large. The behaviour of the map iteration inside these sets, in a sense, is completely random (see [11, 12]) and can be described by the so-called Bernoulli shifts [11].

All the snap-back repellers (see [9]) can also occur in dynamics (1.1). Indeed, it is easy to see, by the definition (see [9]), that these repellers are robust. Such repellers also born infinite collections of periodic cycles with large periods [9].

A very simple criterion of chaos onset in one-dimensional maps was suggested by [17, 18]: a cycle with period three entails the chaos existence. If a given map G has the period 3, then, for small δ, the corresponding neural map F also has period 3 (for small δ), thus also generating chaos.

Remark. Interesting chaotic maps are given by piecewise linear maps [19]. The simplest classical example is $G(q) = 2q \bmod 1$, where $q \in [0, 1]$. By piecewise linear maps we can create dynamical systems simulating the Bernoulli shifts (i.e. completely 'random'). For such maps, inequality (2.3) can be satisfied anywhere excluding small neighbourhoods of the break points.

(3) The third point is a possibility to create, by superpositions of (1.1) (or (1.2)) and some additional network, *arbitrary prescribed spatial-temporal spin x patterns*.

To see this, first let us note that although the SK model (1.1) can generate any attractors it, nonetheless, cannot generate arbitrary x patterns.

To see it, let us study the spin–spin correlators. First let us note that each chaotic dynamics $F(q, \boldsymbol{P})$ generates, in a canonical manner, a invariant measure $\mu(q)$ on a chaotic attractor (the so-called Bowen–Ruelle–Sinai (BRS) invariant measure). Theorem 1 yields, roughly speaking, that systems (2.1) should generate all such measures ('equilibrium' q-densities). Indeed, since all the hyperbolic sets can occur in dynamics (2.1), all the BRS measures can also appear, therefore, by adjusting the network parameters, we can change $\mu(q)$. Note that the limit measures (densities) μ can also depend on initial distributions $\rho_{in}(q) = \rho(q, 0)$. In fact, a number of hyperbolic local attractors Γ_i can coexist in dynamics (2.1) and each attractor has the corresponding attraction basin $\mathcal{B}(\Gamma_i)$. If, at initial moment, the support of ρ_{in} is contained in $\mathcal{B}(\Gamma_k)$ then μ is a BRS measure μ_k induced by Γ_k. If this support has intersections with some basins, we obtain μ as a linear combination of different μ_k. Thus, in this case we have a set of different invariant measures.

Given $\mu(q)$ one can calculate the corresponding invariant x-distribution by (2.2). This yields that the mean correlation between ith and jth spins is defined by

$$\langle x_i x_j \rangle = \int \dots \int_Q \sigma\left(\sum_{l=1}^{p} A_{il} q_l + \theta_i\right) \sigma\left(\sum_{l=1}^{p} A_{jl} q_l + \theta_j\right) \mu(q)\, \mathrm{d}q_1 \dots \mathrm{d}q_p \tag{2.4}$$

where the brackets denote time averaging. Thus, we can conclude that, although the spin dynamics is chaotic, there exist strong spin correlations (since M spins are ruled by p hidden modes q_i and usually $p \ll M$).

To solve the problem of simulation of any spin patterns (with arbitrary correlations) let us consider, in addition, the following network:

$$z_i(t) = \operatorname{sgn}\left(\sum_{j=1}^{M} J_{ij}x_j(t)\right) \qquad (i = 1, 2 \ldots N) \tag{2.5}$$

where the quantities $z_j \in \{1, -1\}$ are some 'outputs' and the x_i are some 'inputs'.

The key observation is that combinations of (2.5) and network (1.1) or (1.2) (generating inputs x for (2.5)) allow us to create *practically any prescribed output spatial-temporal patterns*. We consider this problem in detail below, for the case of discrete states (1.2) (see section 3.3).

2.2. Discrete states

For (1.2) we define q_i as $q_i = \sum_{j=1}^{M} B_{ij}s_j$. These quantities lie in a discrete set $\mathcal{B}$. After the first iteration, q enter for $\mathcal{B}$ and forever remains in it. We have

$$q_l(t+1) = \sum_{k=1}^{M} B_{lk}\sigma\left(\sum_{r=1}^{p} A_{kr}q_r(t) + \theta_k\right) = F(q, \boldsymbol{P}). \tag{2.6}$$

Thus we are dealing with the two dynamics: piecewise constant map $F : \mathcal{B} \to \mathcal{B}$ from (2.6) and discrete map (1.2). They are connected by relation (2.2) (where $\sigma = \operatorname{sgn}$).

For $p = 1$ equation (2.6) can be simplified. We can assume, without any loss of generality, that $A_{il} = 1$ and $\theta_i > \theta_{i+1}$, $\theta_i \in (0, -1)$ and denote $B_{li} = b_i$. Then equation (2.6) takes the form

$$q(t+1) = \sum_{k=1}^{M} b_k\sigma(q(t) + \theta_k) = f(q(t), b, \theta, m). \tag{2.7}$$

Let be $q \in [0, 1]$ and f maps this interval into itself. There exists a natural partition of $[0, 1]$ in a number of subintervals $\beta_i = [-\theta_{i+1}, -\theta_i]$. The behaviour of iterations (2.7) is completely defined by some matrix α_{ij} which is 1, if the image $f_i = f(\beta_i)$ lies in β_j, and 0 otherwise. We can have only M different values f_i in $\mathcal{B}$, thus, dynamics (2.7) generates only periodic cycles and the maximal period is M. In particular, the matrix α_{ij} allows us to determinate all the possible cycle lengths appearing in the system (2.7). In order to extract these lengths we calculate matrix iterations $\boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^M$. Non-zero diagonal elements of the matrix $\boldsymbol{A}^k$ correspond to the cycle with period k.

Suppose that $f(q)$ from (2.7) approximates some smooth or piecewise smooth $G(q)$ (below we shall show that such approximations, of any precision, actually exist; see theorem 2).

Thus, this partition $[0, 1]$ by subintervals $\{\beta_j\}_{j=1}^{M}$ gives some 'round-error' scheme generating pseudo-orbits of G.

Let us turn to the spin dynamics. Beginning with $t = 1$, all s-patterns are M vectors $s(t) = (1, 1, \ldots, 1, -1, \ldots, -1)$, where the sign break is located at ith coordinate if $q(t)$ lies between $-\theta_i$ and $-\theta_{i+1}$.

Let us take into consideration some initial spin distributions $\rho_{in}(\boldsymbol{s})$. This spin distribution leads to some initial q density. If the spin number M is large, one can expect that this density is strongly localized. In fact, let the spins be distributed randomly and independently. Then, for large M, the density is a Gaussian peak. Calculations (see below) show that coefficients b_i have order M^{-1}. Thus, initial q density $\rho_{in}(q)$ is centred at 0 and the peak width is $O(M^{-1/2})$.

If we take other spin distributions, we can obtain peaks localized at points $q^0 \neq 0$.

However, this 'strong' localization is actually strong only for non-chaotic maps G. If iterations $G^T(q^0)$ give an isolated stable orbit, for example, with period T_0, then, for almost all initial spin configurations, spin dynamics (1.3) also will give a periodic orbit with the same period T_0. But if, on the contrary, the point q^0 lies in some chaotic attractor, then any domain of size $O(M^{-1/2})$ contains, for large M, a number of small subintervals $\beta_j = O(M^{-1})$. Thus, there should exist a number of spin trajectories with large periods. It was observed in computer simulations (see section 4).

Let us formulate now a discrete variant of theorem 1.

Theorem 2. Let Q be the p-dimensional unit cube $Q = \{q : |q_i| \leqslant 1\}$, $q \to G(q)$ be any prescribed C^0-mapping that maps Q in Q. Then, for any $\delta > 0$, there is a choice of the parameters $\boldsymbol{P}(\delta)$ such that

$$|G(q) - F(q, \boldsymbol{P})| < \delta \tag{2.8}$$

('approximation') and, moreover, starting with any point $q \in \mathbb{R}^p$, iterations $q, F(q), F^2(q) \ldots$ enter for the cube Q and remain in Q i.e. this cube is absorbing ('absorbtion').

Naturally, the spin number $M(\delta)$ tends to ∞ as the approximation precision $\delta \to 0$. It is easy to see that property (1) again holds. However, (2) is now invalid (as was discussed above).

In section 3.3 we shall show that point (3) holds for discrete networks i.e. we can create a superposition of two networks generating arbitrary prescribed time periodic (with any given period T), by a suitable parameter choice.

3. Construction of network with the prescribed large time and spacetime behaviour

3.1. Continuous states

3.1.1. Outline of proof. Let us turn to the key ideas in proving theorem 1. The method is *constructive and robust* and allows us to find the network simulating given dynamics $\boldsymbol{G}$.

First of all, let us note that, if we remove the 'absorbtion condition' (that iterations F^n enter for the cube Q, together with the second inequality (2.3)), then our assertion is well known from the theory of multilayered neural networks [14–16].

To obtain our (stronger) assertion, we should combine these methods with new approaches.

Let G be some map defined in Q mapping any points of this cube in its interior. Let us extend (it is possible) this map by some C^1-map $\tilde{G}$ on all $\mathbb{R}^p$ in a special way. All the iterations $\tilde{G}^n$ shall enter for Q at some step (see appendix A) since this G can be chosen as a contraction (for q outside of the cube Q).

After this we find a special approximation of this new $\tilde{G}$ in all space $\mathbb{R}^p$. To find it, we first use standard ideas [16] that reduce (3.1) to the one-dimensional case ($q \in \mathbb{R}$). Then G becomes a function of q. Moreover, one can, without any loss of generality, assume that this function G is smooth and fast decreasing (from the Schwartz class). This reduced approximation problem can be solved explicitly, by the classical ideas from the wavelet theory [17]. We know that the wavelet methods are robust and effective in signal denoising and compressing. Thus, this method is robust (see appendix A).

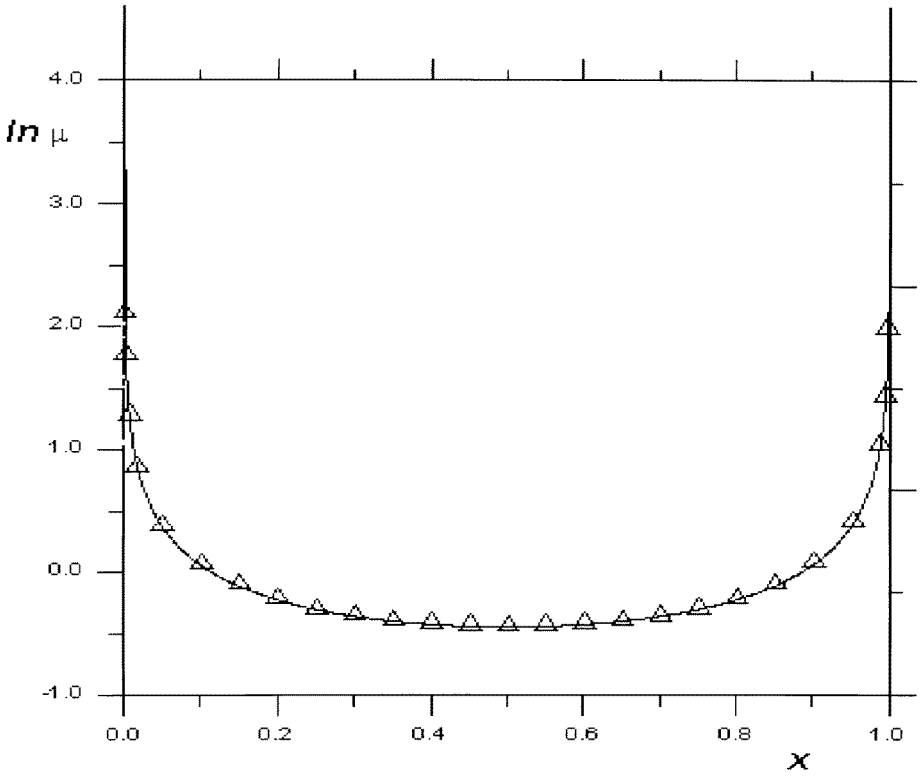

Figure 1. Invariant measure μ for $x \to 4x(1-x)$ (full curve) and its neural approximation by three neurons with precision $\delta = 10^{-5}$ ($\Delta\Delta\Delta$).

3.2. Discrete states

Here proof also proceeds into some steps. Following section 3.1, we note that one can only consider the one-dimensional case: $G(q) \in \mathbb{R}$, $q \in \mathbb{R}$. The approximation satisfying (2.8) can have, for example, the form $f(q) = \sum_{k=1}^{M} b_k \mathrm{sgn}(q + \theta_k)$ and $q \in [0, 1]$. Clearly, the function f has constant values inside each interval $[-\theta_{k+1}, -\theta_k]$, where $\theta_k > \theta_{k+1}$.

Let us require that functions $f(q)$ and $G(q)$ should coincide at M points $q_k = -\frac{1}{2}(\theta_{k+1} + \theta_k)$ then b_k can be chosen by $b_1 = \frac{1}{2}(G(q_M) + G(q_1))$, $b_k = \frac{1}{2}(G(q_k) - G(q_{k-1}))$.

Then for smooth map G the approximation error has order $\max_k |\theta_{k+1} - \theta_k|$. (Of course, there exist many other approximations, for example, one can take $q_k = -\theta_k$ etc.)

In the next section, we describe results of calculations with these approximations. To conclude this section, let us note that a natural approximation of the piecewise linear map $G(q) = 2q - 1 \bmod 1$ leads to the following simplest mapping of the discrete set $\mathcal{B} = \{1, 2 \ldots M\}$. This map transforms i to $j(i) = 2i - 1 \bmod M$. Here the periods can be calculated analytically that allows us to explain figure 1. In fact, it is easy to show that such a map has a cycle of long T (where $T > 0$ is integer) only and only if $2^T = 1 \bmod p$ where $p \neq 2$ is a prime divisor of M. For example, if $M = 28$ then the period is 3 since $28 = 2 \times 2 \times 7$ and $2^3 = 1 \bmod 7$, if $M = 1026$ then there occurs a cycle with period 18 ($1026 = 2 \times 3^3 \times 19$ and $2^{18} = 1 \bmod 19$ according to the 'little' Fermat theorem).

In general, the periods T are values of the Euler function $\phi(p)$ of odd prime divisors of the spin number M.

This simple example helps us to see the strong irregularity of the periods (see figure 1). In fact, for $M = 1024$ we obtain $T = 1$; nonetheless for close value $M = 1018$ we have $T = 508$, since 509 is prime!

In the following we shall use this example.

3.3. Generation of arbitrary patterns by two discrete networks

Let $z_i(t)$ be an arbitrary output pattern consisting of 1 or -1 (where $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$). Let us consider the following systems of linear equations for unknown connections J_{ij}

$$z_i(t) = \sum_{j=1}^{M} J_{ij} s_j(t)) \qquad i = 1, \ldots, N \qquad t = 1, 2 \ldots T \tag{3.1}$$

where $s_j(t)$ and $z_i(t)$ are given. Let us set $N > T$. Then systems (3.1) can be resolved if T patterns s_j are linearly independent (i.e. there exist no coefficients C_t such that $\sum_{t=1}^{T} C_t s_i(t) = 0$ for all i).

It is not difficult to check this condition for concrete A, B, θ and M. For example, we can take these parameters so that network (1.2) will simulate the map $q \to 2q - 1 \bmod 1$ (see the example from section 3.2). Then system (1.2) can generate a cycle of long $T_0 > T$ where all s patterns for different $t = 1, 2, \ldots, T_0$ are linearly independent. This follows from section 2.2. Indeed, the vectors $s(t)$ are $s(t) = (1, 1, \ldots, 1, -1, \ldots, -1)$ where we obtain the different numbers of 1 for different $t = 1, 2, \ldots, T_0$. This simple observation completes the proof.

4. Computer simulations

We discuss here the following three problems: (1) approximation of given G by F (see (2.3)), (2) closeness of iterations, and (3) invariant measures for G and F.

We focus our attention on the three classical one-dimensional maps: (a) the map $q \to rq \exp(-\lambda q)$, $q > 0$, (b) the quadratic map $q \to rq(1 - q)$, $0 < r \leqslant 4$ and (c) $q \to rq \bmod 1$.

4.1. Continuous case

For (a) ($r = 13$, $\lambda = 0.1$) we, using only 7–8 neurons, can obtain an approximation F giving the precision $\delta = 5.26 \times 10^{-2}$ within q-interval $[0, 40]$. We observe a stable attracting cycle of length 4. The relative differences between the iterations G^T and F^T, even for $T > 1000$, do not exceed 5×10^{-2}.

We see a similar situation for the quadratic map, while we are dealing with stable cycles. All stable periodic cycles of (b) and its neural approximations F are very close.

It is well known that one can observe the chaotic behaviour for (b) if r, for instance, lies within $[3.8, 4.0]$. In this case trajectories of G and the approximation F are close but only for a small number of iterations: eventually, the trajectories diverge exponentially. For example, let the approximation precision be at 10^{-7} and $r = 4$. Then we have observed, during 50–150 (depending on initial data q) iterations, the relative difference between $F^T(q)$ and $G^T(q)$ at 2–3%.

However, the statistical properties of F and $G(q) = rq(1-q)$ are always very similar. In fact, the invariant measure μ for $r = 4$ can be calculated analytically (seldom is the case!) [21]. It is $\mu = 1/\pi[x(1-x)]^{-1/2}$.

The measure of the neural approximation can be calculated by computer (the algorithm see [21]). The accordance is excellent: curves coincide anywhere excluding very small neighbourhoods of boundary points, see figure 1.

4.2. Discrete case

Here we have considered mainly maps (b) and (c). The most interesting example is (b), since this map shows the period doubling bifurcations when the parameter r takes values at $r = 3.5, 3.56, 3.57\ldots$ [21]. If r at 4, we observe the chaos [21].

While we are dealing with stable periodic cycles, discrete approximations work quite well, however now we should take about 500–1000 neurons. To check it, we have prepared the two independent programs based on two different algorithms (one of them we have briefly discussed in section 2.2). These programs calculate the periods appearing in dynamics (2.7). Suppose we take $M = 600$ neurons (spins) and approximate the map $q \to rq(1-q)$ (that gives the precision δ in (2.5) of the order 0.001). We have seen that actually (2.6) exhibits the same bifurcation sequence (periodic cycles with period 2, 4, 8, 16). Here results are not sensitive to details of approximation scheme (see section 3.2).

For the chaotic domain ($r = 4$) we observe another picture. The dependence of the period on the neuron number M becomes very non-trivial. For case (c) it can be described analytically (for special approximations, see section 3.3).

For example, if we use approximation (2.7) of map (b) with $M = 200$, we observe only the single cycle with period $P = 9$. If $M = 500$, we have the three different cycles with periods $P_1 = 4$, $P_2 = 4$ and $P_3 = 11$. For $M = 800$ periods $P_1 = 2$, $P_2 = 3$, $P_3 = 11$ have been received, for $M = 900$ we have obtained $P_1 = 3$, $P_2 = 5$, $P_3 = 19$ etc. The period lengths (as well as the number of periods) is extremely sensitive to M and to the approximation choice. For example, when $M = 4125$ we have found 5 periods: $P_1 = 3$, $P_2 = 5$, $P_3 = 8$, $P_4 = 21$, $P_5 = 66$, for $M = 4126$ $P_1 = 3$, $P_2 = 20$, $P_3 = 45$ and for $M = 4127$ $P_1 = 3$, $P_2 = 110$. In the last case the overwhelming majority of initial configurations evolves to a cycle with $P = 110$, and it is typical. However, even a small change of the matrix $\boldsymbol{K}$ can drastically change the output periodic spin trajectories.

This complicated period spectrum can be illustrated by figure 2.

We have also calculated the spin–spin correlators for discrete approximations. Here the accordance between computer results and our predictions is better for such M where we have the long cycles. If this cycle is relatively short (length < 40), the actual correlator values may differ strongly from (2.4) (for example, one can obtain 0.13 or -0.40 whereas the theory foresees -0.18). These results can be easily explained. In contrast to the continuous case (see section 4.1), statistical properties of G and its discrete neural approximations F are very different. In fact, for discrete case attractor of F is a union of limit cycles and equilibria. Number of these cycles, its lengths and attraction basins are very sensitive to approximation choice and number of neurons, see figure 3. The analogous behaviour we already observed in the example of section 3.2 was studied completely analytically.

5. Conclusion

In this paper we have established the possibility of generation of *any* structurally stable kinds of chaos (including all hyperbolic attractors and invariant sets) by the simplest neural network: the fully asymmetric SK model. The superposition of two such networks can generate any spatial-temporal patterns.

Naturally such assertions can be proved only analytically. The key role here is played

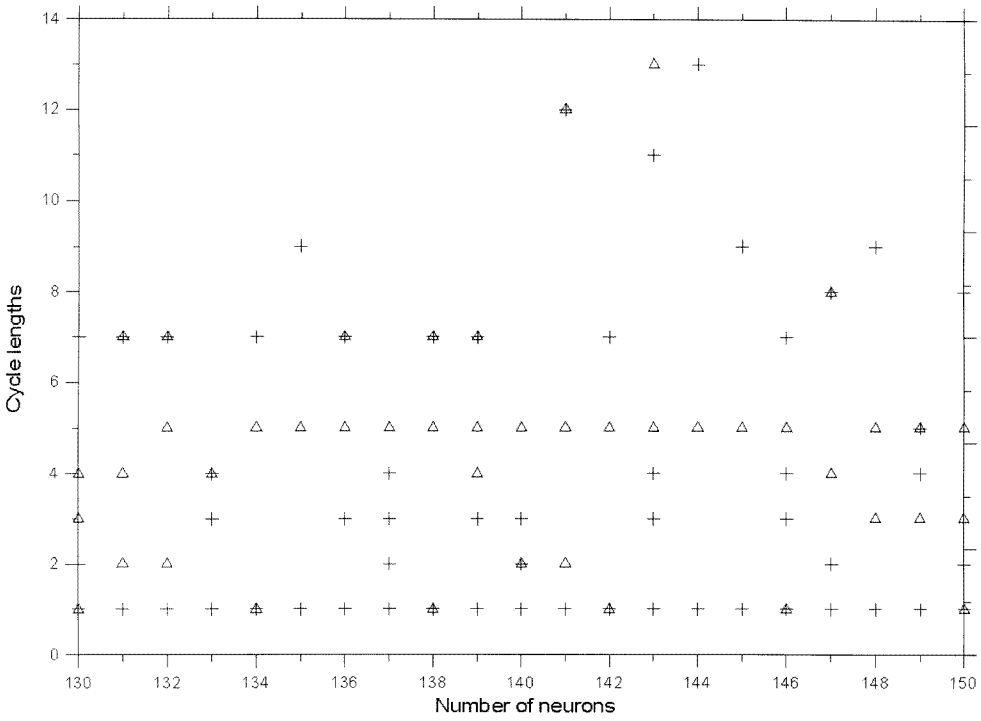

Figure 2. Cycle length appearing in two different neural approximations of $x \to 4x(1-x)$ with same number of neurons. Δ corresponds to one approximation, $+$ to another.

by the generalized Hebb rule (1.3) that gives some explicit choice of the synaptic matrix. We have not considered learning methods here.

We suppose that the most effective (from a calculative point of view) learning method should be based on a gradient descent or a combination of the Hebb rule and gradient methods.

To conclude, we would like to note some key points. The great advantage of such a model as the Hopfield model with symmetrical synaptic matrix is that it can be studied analytically because its properties remind us of those of a common statistical mechanical system possessing an energy (mathematically, Lyapunov functional decreasing along trajectories). In nature, however, we observe non-symmetrical neuron connections that lead to periodic and chaotic behaviour. In such a real situation, it is impossible to find an analogue of the energy (Lyapunov functional).

The generalized Hebb rule (1.3) allows us to overcome this difficulty. From the physical view point, such a rule leads to the appearance of the natural 'order parameters' q_i (see section 2.1) governing the dynamics of the network.

The main result of such an approach is as follows. For any prescribed global geometry of an attractor, if this attractor is structurally stable (robust), we can, by (1.3), create analytically solvable simple models of neural networks with topologically equivalent attractor geometry.

Hebb rule (1.3) also helps to find a number of the coupled oscillator and reaction–diffusion systems with analogous complicated large time behaviour [22–24]. The simple

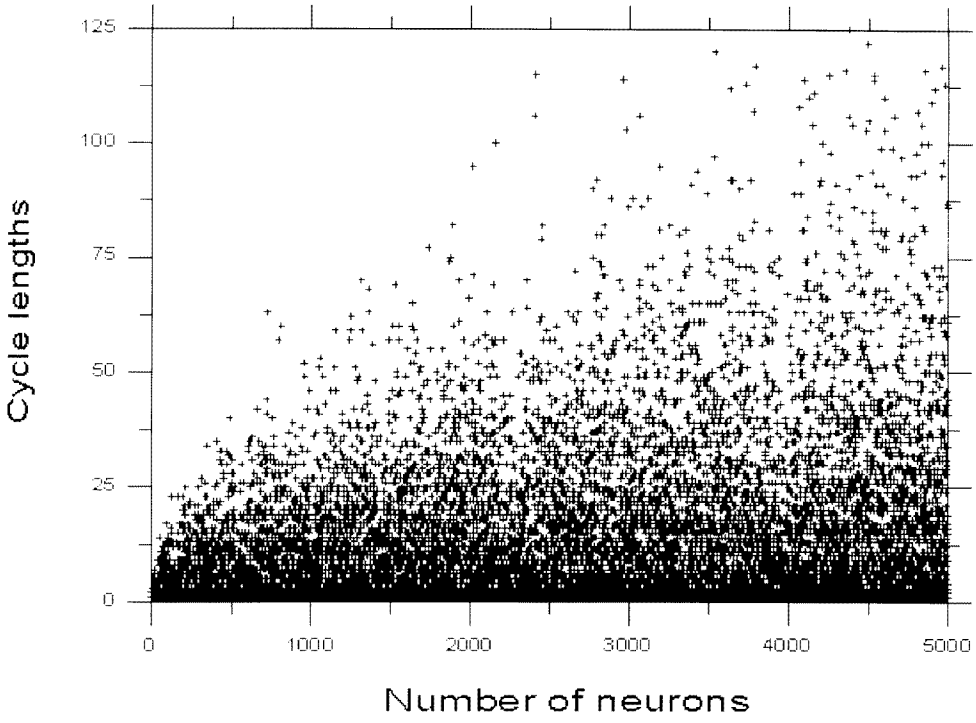

Figure 3. Length of cycles appearing in neural approximations of $x \to 4x(1-x)$.

example can be given by the classical Hopfield system [25]

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \sum_{j=1}^{M} K_{ij}\sigma(q_j) - bq_i + \theta_i \qquad b \geqslant 0. \tag{5.1}$$

The important point is a wonderful connection between (1.3) and the Lax pair representation. The substitution (1.3) for $\boldsymbol{K}$ leads to a remarkable system [23, 24]. This system for $b = 0$ can be rewritten in the Lax form (see appendix B).

Acknowledgment

This work was supported by Russian Foundation of Basic Research, grant no 97-01-01123.

Appendix A. Proof of theorem 1

This proof is standard, mainly follows [16, 23] and proceeds in a number of steps.

Step 1. Prolongation of map G. Suppose G is defined in Q and $G(Q) \subset Q$. Then we can extend this map on all $\mathbb{R}^p$ so that this extended map $\tilde{G}$ will satisfy:

(a) $\tilde{G} \in C^1(\mathbb{R}^p)$ and, inside the cube Q, $\tilde{G} = G$;

(b) $|\tilde{G}| < a|q|$ for any $q \in \mathbb{R}^p$ which lies outside of some small neighbourhood V of Q. The constant $0 < a < 1$ (this means that outside of Q the $\tilde{G}$ is a contracting map). Moreover, let be $\tilde{G}_i(q) \equiv 0$ if $|q| > 2$.

Such $\tilde{G}$ can be obtained in a simple way, and we omit the details. Notice that V and a can be chosen so that all the iterations $\tilde{G}^n$ will converge to Q.

Step 2. Approximation of the map $\tilde{G}$ by averaging. Let us approximate this new map $\tilde{G}$ uniformly in $\mathbb{R}^p$ in the norm C^1 with accuracy $\delta/10$ by a new $C^\infty(\mathbb{R}^p)$-map H from the Schwartz class $S(\mathbb{R}^p)$ of fast decreasing functions. This approximation can be obtained by averaging [23]

$$H_r(q) = \int_{\mathbb{R}^n} \tilde{G}(q - q')\omega_r(q')\,\mathrm{d}^p q' \tag{A.1}$$

where ω_r is a cut-off positive C^∞-function with the support in the ball of radius r. We have, $|H_r - \tilde{G}|_{C^1}$ tends to 0 as $r \to 0$. Using it, we take a small r and set $H = H_r$.

Step 3. Reformulation and simplification of the approximation problem. Instead of (2.3), we shall consider the following new problem: *to find parameters $\boldsymbol{P}$ such that*

$$|H(q) - F(q, \boldsymbol{P})|, \qquad |\partial H(q)/\partial q_i - \partial F(q, \boldsymbol{P})/\partial q_i| < \tfrac{1}{4}\delta(1 + |q|) \qquad (q \in \mathbb{R}^p) \tag{A.2}$$

for this new field H from the Schwartz class $S(\mathbb{R}^p)$. We notice that (A.2) implies (2.3). Moreover, for small δ, since H is contracting outside of Q and $a < 1$, the neural approximation F also is contracting outside of Q. Thus, the solution of (A.2) resolves both our approximation and absorbtion problems from theorem 1 for the original map G.

Remark. We can, without any loss of generality, suppose that $H(q)$ in (A.2) is *a scalar function* of q. In fact, we can combine, in a special way, the separate approximations of each component H_j (see final step). Thus, let us assume $H \in \mathbb{R}$. Below we omit the component index j (up to final step).

Step 4. Representation of scalar function H by sum of 'plane waves' and one-dimensional reformulation of the problem. It follows [16]. The function H can be approximated, for large L, by the sum

$$\bar{H}(q) = L^{-1} \sum_{l=1}^{L} \eta_l(x_l) \qquad x_l = \langle q, e^l \rangle \tag{A.3}$$

where the vectors e^l are uniformly distributed among the unit sphere and the brackets denote the standard scalar product. Representation (A.3) can be considered as a sum of 'plane waves' $\eta_l(x)$. The following step gives approximations of $\eta_l(x)$.

Step 5. Approximation of functions $\eta_l(x)$. We temporarily omit index l. Denote $a = (a_1, a_2, \ldots, a_M)$ and $X = (X_1, \ldots, X_M)$. Let us define

$$\xi(x, a, X, M) = \sum_{i=1}^{M} c_i \sigma(a_i(x - X_i)). \tag{A.4}$$

Moreover, let us set

$$p(x, a, c, X) = \eta(x) - \xi(x, a, X, M). \tag{A.5}$$

In this section our aim is to resolve, for any $\tilde{\delta} > 0$, the following approximation problem for $\eta(z)$: to find an integer M and numbers a_i, c_i and X_i such that

$$|p(x, a, c, X)| < \tilde{\delta}(1 + |x|) \qquad (x \in \mathbb{R}) \tag{A.6}$$

$$\left|\frac{\partial p(x, a, c, X)}{\partial x}\right| < \tilde{\delta} \qquad (x \in \mathbb{R}). \tag{A.7}$$

Due to steps 1–4, the solution of (A.6) and (A.7) allows us to find the solution of our original approximation problem.

Let us focus attention on (A.7). Our aim is to show that, for any $\tilde{\delta} > 0$, there exists a vector $a(\tilde{\delta})$ with components $a_i > 0$ and vectors $c(\tilde{\delta})$ and $X(\tilde{\delta})$ such that for the derivatives of η and ξ the following estimate holds:

$$|\nu(x) - \xi_x'(x, a, c, X)| < \tilde{\delta} \qquad (x \in \mathbb{R}) \qquad \nu(x) = \eta_x'(x). \tag{A.8}$$

It is clear that (A.8) implies (A.6) and (A.7). In fact, integrating $\eta' - \xi'$ one has

$$\int_0^x \eta_x' \, \mathrm{d}x = \sum_{i=1} a_i^{-1}(c_i(\sigma(a_i(x - X_i)) - \sigma(-a_i X_i)) + s(x) \qquad |s(x)| < \tilde{\delta}|x| \tag{A.9}$$

that gives

$$|\eta(x) - \sum_{i=0} \tilde{c}_i(\sigma(a_i(x - X_i))| < \tilde{\delta}(1 + |x|) \tag{A.10}$$

where $a_0 = 0$, $\tilde{c}_0 = \eta(0)\sigma^{-1}(0)$ and one supposes (without any loss of generality) that $\sigma(0) \neq 0$.

Let γ_k, z_k (where $k = 1, 2, 3$) be numbers such that $z_1 < z_2 < z_3$ and

$$\gamma_1 + \gamma_2 + \gamma_3 = 0 \qquad \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 = 0. \tag{A.11}$$

Then the function

$$\rho(z) = \sum_{i=1}^{3} \gamma_i \sigma'(z - z_i) \tag{A.12}$$

is a 'wavelet-type' function such that

$$\int_{-\infty}^{\infty} \rho(z) \, \mathrm{d}z = 0 \qquad \int_{-\infty}^{\infty} z\rho(z) \, \mathrm{d}z = 0. \tag{A.13}$$

Note that

$$\int_{-\infty}^{\infty} \nu(x) \, \mathrm{d}x = 0. \tag{A.14}$$

Let us suppose, in addition, that

$$\int_{-\infty}^{\infty} x\nu(x) \, \mathrm{d}x = 0. \tag{A.15}$$

If the approximation problem (A.8) is solved for such $\nu = \eta'$, then it is resolvable in general case (when only (A.14) holds). In fact, one can always write

$$\tilde{\nu} = \nu - \tilde{c}\sigma'(x). \tag{A.16}$$

Since $\sigma(+\infty) \neq \sigma(-\infty)$, we can choose $\tilde{c}$ so that $\tilde{\nu}$ satisfies (A.15).

To solve the approximation problem (A.8), let us write the well known Calderón identity (playing a key role in the wavelet theory)

$$\nu(z) = C_\rho \int_0^\infty \int_{-\infty}^{\infty} \lambda^{-3} C(\lambda, X)\rho(\lambda^{-1}(z - X)) \, \mathrm{d}X \, \mathrm{d}\lambda \tag{A.17}$$

where

$$C(\lambda, X) = \int_{-\infty}^{\infty} \nu(z)\rho(\lambda^{-1}(z - X))\,\mathrm{d}z \tag{A.18}$$

and C_ρ is a constant depending on the function ρ. These relations hold, since ν is a fast decreasing function ($\nu \in S(\mathbb{R})$).

The next step is now a transformation of (A.17) into in a discrete sum. To do this, let us observe the following. If we can appoximate integral (A19) by

$$\nu(r, R_1, R, z) = C_\rho \int_r^{R_1} \int_{-R}^{R} \lambda^{-3} C(\lambda, X)\rho(\lambda^{-1}(z - X))\,\mathrm{d}X\,\mathrm{d}\lambda \tag{A.19}$$

(where R_1, R and $r > 0$) with any precision δ_1, then the approximation problem (A.8) is solved.

In fact, integral (A.19) can be replaced, with any precision, by a finite sum (the Riemaniann sum). Namely, for any $\epsilon > 0$ one has

$$\sup_{z \in I} |\nu(z) - \sum_i \sum_j {\lambda_i}^{-3} m_{ij} C(\lambda_i, X_j)\rho(\lambda_i^{-1}(z - X_j))| < \epsilon \tag{A.20}$$

for some λ_i, X_j and m_{ij}. By (A.12) the sum in estimate (A.20) can be rewritten as $\sum_{i=1}^{M} c_i \sigma'(a_i(x - X_i))$.

Finally, to solve the approximation problem, one needs to check that the contributions J_k of type

$$\begin{aligned} J_1 &= \int_0^r \left(\int_{-\infty}^{\infty} \lambda^{-3} C(\lambda, X)\rho(\lambda^{-1}(z - X))\,\mathrm{d}X \right) \mathrm{d}\lambda \\ J_2 &= \int_r^{\infty} \left(\int_R^{\infty} \lambda^{-3} C(\lambda, X)\rho(\lambda^{-1}(z - X))\,\mathrm{d}X \right) \mathrm{d}\lambda \end{aligned} \tag{A.21}$$

satisfy

$$|J_k| < \kappa(r, R) \tag{A.22}$$

where $\kappa \to 0$ as $R \to \infty$ and $r \to 0$.

In order to investigate J_k, let us estimate $C(\lambda, X)$. First, let us notice that, due to condition (A.13) and the Taylor expansion, one has, for $0 < \lambda < 1$, that

$$|C(\lambda, X)| < \lambda^3 \int_{\mathbb{R}} y^2 \nu''(X + s\lambda y)\rho(y)\,\mathrm{d}y \qquad s \in [0, 1].$$

Using $\nu \in S(\mathbb{R})$, one obtains

$$|C(\lambda, X)| < C_N \lambda^3 (1 + |X|)^{-N}$$

and, therefore, $|J_1| < c(C_N + 1)r$.

Thus estimate (A24) holds for $k = 1$.

Consider J_2. One observes that $|C(\lambda, X)| < c_N |1 + \lambda^{-1} X|^{-N}$ for any $N > 0$ and $0 < \lambda < 1$. The estimate of J_2 follows now from inequalities

$$\int \lambda^{-3}\,\mathrm{d}\lambda \int_R^{\infty} (1 + \lambda^{-1}X)^{-N}\,\mathrm{d}X < \int \lambda^{-2}\,\mathrm{d}\lambda (1 + \lambda^{-1}R)^{-N} < \int_0^1 \mathrm{d}t\,(1 + tR)^{-N}.$$

It implies that

$$|J_2| < c\left(\int_R \lambda^{-3}\,\mathrm{d}\lambda + R^{-1} \right) \to 0 \qquad (R \to \infty).$$

Finally, it completes the fifth step and thus the approximation problem for η_j is solved.

Final step. This step is a combination of the previous steps and the 'plane wave' representation.

Let us consider approximations ξ from (A.4) and substitute it into the right-hand side of (A.2). It gives the following representation for components H_i of H (where we again introduce into consideration indecies: j for the components of the field H and l for η_l):

$$H_i(q, C, a, X, m_1, \ldots, m_L) = \sum_{l=1}^{L} \sum_{k_1=1}^{m_l} C_{ik_1l}\sigma(a^l_{ik_1}(\langle q, e^l\rangle - X_{ik_1})) + h_i(q) \tag{A.23}$$

where the corrections h_i and its derivatives do satisfy estimates $|h_i| < 0.25(1 + |q|)$.

Let us make some trivial observations. We note that, without loss of generality: (1) one can set $m_l = m_1$; (2) one can suppose that $a^l_{ik_1}$ and X_{ik_1} do not depend on the component index i: $a^l_{ik_1} = a^l_{k_1}$ and $X_{ik_1} = X_{k_1}$.

In fact, these coeffients arise as a result of the discretization of the Calderón integral (A.17). This discretization could be done in the same manner for any i.

Now let us set $m = m_1L$. We can replace the pairs of indecies (k_1, l) by an index k. Then, let us define vector X and matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ by relations

$$B_{ik} = c_{ik_1l} \qquad A_{kr} = a^l_{k_1}e^l_r \qquad X_k = X_{k_1}$$

where e^l_r is rth component of the unit vector e^l. This completes the proof.

Appendix B. Generalized Hebb rule and Lax pair

Let us consider system (5.1) with the matrix $\boldsymbol{K}$ from (1.3). Then (5.1) takes the form

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \sum_{k=1}^{p} A_{ik}\Phi_k(q) - bq_i + \theta_i \qquad b \geqslant 0 \qquad (i = 1, 2, \ldots, M) \tag{B.1}$$

where Φ_k are defined by $\Phi = \sum_{j=1}^{M} B_{jk}\sigma(q_j)$. This system has appeared in [19–21]. For $b > 0$ such a system is dissipative. In the $b = 0$ system (B.1) becomes conservative. If, moreover, $\theta_i = 0$ we can rewrite it in the Lax form.

Let us consider the Lie algebra generated by operators $B_j = \partial/\partial x_j$, $j = 1, 2, \ldots, p$ and $L_i = \exp(\sum_s A_{is}x_s)$. These operators satisfy the following commutation relations:

$$[L_i, B_j] = A_{ij}L_i.$$

By setting $L(q) = \sum_{i=1}^{m} \exp(q_i)L_i$ and $B(q) = \sum_{j=1}^{n} \Phi_j(q)B_j$ one finds that (B.1) can be rewritten as

$$\frac{\mathrm{d}L}{\mathrm{d}t} = [B, L]$$

i.e. by the Lax pair.

Thus, one can expect that (B.1) can be resolved. It can be done by some substitution [21], even for $\theta_i \neq 0$ and $b \neq 0$. It yields that system (5.1) can generate (for $b > 0$) any hyperbolic attractors, under a suitable choice of $\boldsymbol{K}$, M and σ.

References

[1] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[2] Gardner E 1989 *J. Phys. A: Math. Gen.* **22** 1969
[3] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **2** 337
[4] Cessac B 1995 *J. Physique* I **5** 409

[5] Sompolinsky H, Crisanti A and Sommers H-J 1988 *Phys. Rev. Lett.* **61** 259
[6] Derrida B 1989 *Helv. Phys. Acta* **62** 512
[7] Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675
[8] Bastilla U and Parisi G 1997 *J. Phys. A: Math. Gen.* **30** 5613
[9] Chen L and Aihara K 1997 *Physica* D **104** 286
[10] Bastilla U and Parisi G 1997 *J. Phys. A: Math. Gen.* **30** 3757
[11] Bowen R 1971 *Trans. Am. Math. Soc.* **154** 577
[12] Katok A 1980 *Publ. Math. IHES* **51** 137
[13] Ott E, Grebogi C and Yorke J A 1990 *Phys. Rev. Lett.* **64** 1196
[14] Cybenko G 1989 *Math. Control Signals Syst.* **2** 303
[15] Funahashi K and Nakamura Y 1993 *Neural Netw.* **6** 801
[16] Barron A R 1993 *IEEE Trans. Inf. Theory* **39** 930
[17] Sharkovski A 1964 *Ukr. Math. J.* **16** 61
[18] Li T and Yorke J A 1975 *Am. Math. Monthly* **82** 985
[19] Dahlkvist P 1997 *J. Phys. A: Math. Gen.* **30** L351
[20] Meyer Y 1990 *Ondelettes and Operators* (Paris: Hermann)
[21] Lichtenberg A J and Lieberman M A 1984 *Regular and Stochastic Motion* (New York: Springer)
[22] Vakulenko S A 1994 *J. Phys. A: Math. Gen.* **27** 2335
[23] Vakulenko S A 1997 *Ann. Inst. Henri Poincarè Phys. Théor.* **66** 373
[24] Vakulenko S A 1997 *CRAS Paris* **324** 509
[25] Hopfield J J 1982 *Proc. Natl Acad., USA* **79** 2554