Limit cycles of a perceptron

1998 J. Phys. A: Math. Gen. 31 2967

(http://iopscience.iop.org/0305-4470/31/13/006)

J. Phys. A: Math. Gen. **31** (1998) 2967–2974. Printed in the UK PII: S0305-4470(98)90055-5

Limit cycles of a perceptron

M Schröder and W Kinzel

Institut für Theoretische Physik, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

Received 16 December 1997

Abstract. An artificial neural network can be used to generate a series of numbers. A Boolean perceptron generates bit sequences with a periodic structure. The corresponding spectrum of cycle lengths is investigated analytically and numerically; it has similarities with properties of rational numbers.

In the last 15 years models and methods of statistical physics have successfully been used to understand emergent computation of neural networks. Several properties of infinitely large attractors and multilayer networks could be calculated analytically. Such systems of simple units interacting by synaptic weights can be used as associative memory and classifiers; they are trained by a set of examples, detect unknown rules and structures in high-dimensional data, and store patterns in a distributed and content addressable way (Hertz *et al* 1991, Watkin *et al* 1993, Opper and Kinzel 1996).

Another important application of neural networks is time-series analysis (Weigand 1993). But only recently has statistical physics been used to model training and prediction of bit sequences by a perceptron (Eisenstein *et al* 1995, Schröder *et al* 1996). A neural network is trained by a sequence of numbers; after the training phase the network makes predictions on the rest of the sequence. In analogy to generalization the training and test data are generated by a neural network, as well. It turns out that the generation of sequences of numbers by a perceptron or multilayer network is already an interesting problem which should be understood before prediction is investigated (Eisenstein *et al* 1995, Kanter *et al* 1995).

This problem is a special case of the *neuronic equations* of Caianiello (1961). These equations, which were suggested to model a neuron including time dependency, can only be solved in special cases. Most work was done for one input and a memory back into time with couplings that decrease exponentially $w_i = a^i$ ($a < 1$). Several analytic results for the transients and the limit cycles of the resulting dynamics have been achieved for this case (for example, Caianiello and Luca 1965, Cosnard *et al* 1988a). Not much work has been done for other weights (for example, Cosnard *et al* 1988b, 1992). Our motivation for studying this recursion equation is to examine the generalization ability and in a first step the ability of a perceptron to generate time series. Hence we have no restrictions on the weight vector *a priori*.

Numerical analysis of a perceptron with random weights generating sequences of numbers shows that the sequences are related to the Fourier modes of the weight vector. Therefore it is useful to study weight vectors with a single mode only. In this case an

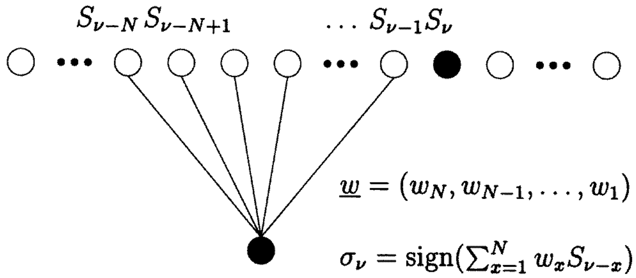

Figure 1. A perceptron learning a time series. The desired output of the perceptron (marked) is the next bit of the series and therefore part of other input patterns as well.

analytic solution of a stationary sequence could be derived for large frequencies (Kanter *et al* 1995).

This solution holds for continuous odd transfer functions, for example $\tanh(\beta x)$. As a function of the slope β a phase transition to a nonzero sequence occurs. The phase of the weight vector results in a frequency shift of the attractor. In this paper we wish to extend this solution to infinite slope β and general frequencies, that is we derive an analytic solution for the bit generator. We find a much richer structure of the bit sequences generated by a Boolean perceptron compared with sequences of continuous ones.

A bit generator (figure 1) is defined by the equation

$$S_\nu = \mathrm{sign}\sum_{j=1}^{N} w_j S_{\nu-j} \qquad (\nu \in \{0, 1, 2, \ldots\}) \tag{1}$$

where $\underline{w} \in \mathbb{R}^N$ is the weight vector of the perceptron of size N. Given an initial state $(S_{-N}, \ldots, S_{-1}) \in \{-1, 1\}^N$, equation (1) defines a binary sequence $(S_0, S_1, \ldots)$ which has to run into a periodic cycle of length $L \leqslant 2^N$. We try to find an analytic solution of the periodic attractor. As mentioned, it is useful to restrict the weights to one single mode

$$w_j = \cos\left(2\pi q\frac{j}{N} + \pi\phi\right) \tag{2}$$

with a frequency $q \in \mathbb{N}$ and a phase $\phi \in [0, 1[$. Equation (1) may be expressed in terms of the local fields $h_\nu = \frac{1}{N}\sum_{j=1}^{N} w_j S_{\nu-j}$:

$$h_\nu = \frac{1}{N}\sum_{j=1}^{N}\cos\left(2\pi q\frac{j}{N} + \pi\phi\right)\mathrm{sign}(h_{\nu-j}). \tag{3}$$

We have to solve this self-consistent equation for the function h_ν. Since simulations show that limit cycles are dominated by one frequency, we assume $\mathrm{sign}(h_\nu)$ is a periodically alternating step function with frequency $k+\tau$ (with $k \in \mathbb{N}$, $\tau \in [0, 1[$), where the frequency is defined for the variable ν/N:

$$\mathrm{sign}(h_\nu) = \mathrm{sign}\left(\sin\left(2\pi\frac{(k+\tau)}{N}\nu\right)\right). \tag{4}$$

In the case that N is a multiple of $2(k+\tau)$, i.e. for integer wavelengths, we found an analytic solution of equation (1). It follows, using equation (4), that the right-hand side of equation (3) is a periodic function with a period of length $N/(k+\tau)$ and that

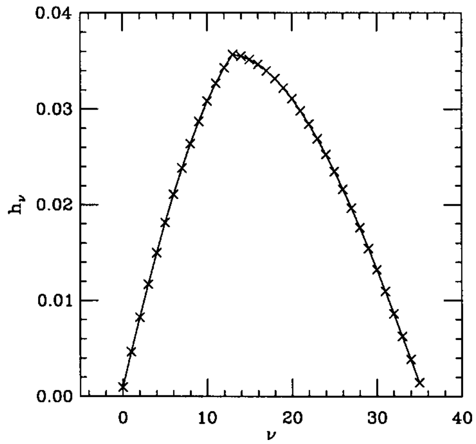

Figure 2. Internal field of the bit-generator, $N = 1021$, $q = 11$, $\phi = 0.25$, $k = 14$, $\tau = 0.180\,556$.

$h_\nu = -h_{\nu+N/(2(k+\tau))}$ for $\nu \in \{0, \ldots, \frac{N}{2(k+\tau)} - 1\}$. Our main result is

$$h_\nu = \frac{1}{N\sin(\frac{\pi q}{N})} \frac{\sin(\frac{1}{2}(T+1)(\frac{\pi q}{k+\tau} + \pi))}{\cos(\frac{\pi q}{2(k+\tau)})} \sin\left(2\pi q \frac{\nu}{N} + \phi\pi + \frac{T}{2}\left(\frac{\pi q}{k+\tau} + \pi\right) + \frac{\pi q}{N}\right) - \begin{cases} 0 & \text{for } T \text{ odd} \\ +\frac{2}{N}\cos(\phi\pi) + \frac{\sin(\phi\pi - \frac{\pi q}{N})}{N\sin(\frac{\pi q}{N})} & \text{for } T \text{ even} \end{cases} \tag{5}$$

where we have abbreviated $T = [(k+\tau)(2 - 2\nu/N)]$ using the *Gaussian bracket* $[x]$ that denotes the closest integer less than x. For $\phi = \tau = 0$ and $k = q$ this results in

$$h_\nu = \frac{2k}{N\sin(\frac{\pi k}{N})} \sin\left(\frac{2\pi k\nu}{N} + \frac{\pi k}{N}\right). \tag{6}$$

A sample function is plotted in figure 2. Note that it consists of two parts with the same frequency q. In the limit $N \to \infty$ these parts are connected continuously. Equation (3) gives the condition

$$h_\nu \geqslant 0 \qquad \nu \in \left\{0, \ldots, \frac{N}{2(k+\tau)} - 1\right\}. \tag{7}$$

Figure 3 shows the possible frequencies that satisfy condition (7) within our ansatz. Only values of $k + \tau = N/(2i)$ with integers i are possible with our ansatz.

For $N \to \infty$ a necessary condition for equation (7) is $h_0 = 0$. For $k \geqslant q$ this is sufficient so the nontrivial ($h_\nu \not\equiv 0$) solutions are given by the values of τ, k that fulfil the equation $\sin(\phi\pi + (2k+1)(\pi q/(2(k+\tau)) + \pi/2)) = 0$ which is equivalent to $q(2k+1) = (k+\tau)(2z - 2\phi - 1)$ with an integer z. The frequencies $k + \tau$ that are allowed from this condition are also shown in figure 3.

We see that the analytic solution of the sequence generator with a continuous transfer function (Kanter *et al* 1995) cannot just be extrapolated to the case of the bit generator. The continuous generator, close to the transition point and for $k \gg 1$, has $k = q$ and $\tau = \phi$, whereas we find a spectrum of solutions with $k \geqslant q$ and $\tau(q, \phi, k)$, as shown in figure 3.

Up to now we have considered only integer wavelengths. Now we wish to discuss the general case of arbitrary values of q and ϕ. We wish to address two questions.

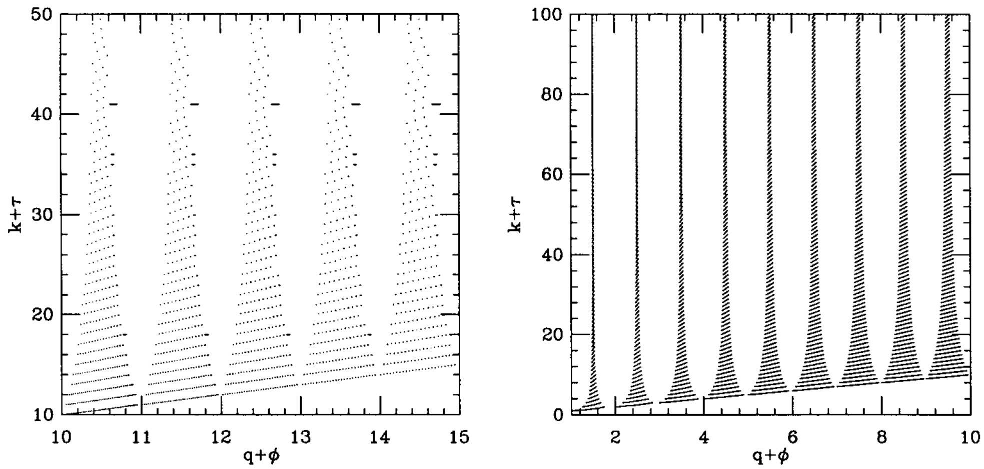

Figure 3. Solutions of equation (3). Frequency $k+\tau$ of the solution versus frequency plus phase shift of the couplings $q+\phi$. Left-hand side: $N = 10\,000$, $k+\tau = N/(2i)$ $(i = 1, \ldots, N/2)$. Right-hand side: the limit $N \to \infty$.

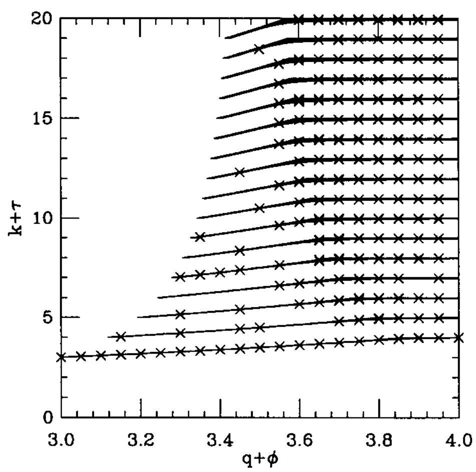

Figure 4. Possible solutions of the bit generator. Frequency $k+\tau$ of the solution versus frequency plus phase shift of the couplings $q+\phi$. Comparison of the simulation ($\times$: $N = 443$) with the extrapolated analytic solution (lines).

(i) Do additional solutions consisting of one frequency exist?

(ii) What are the properties of the bit sequences?

We assume that there are solutions of the form (4) with general frequencies $k+\tau$ for a given system size N. As a function of $q+\phi$ we numerically scan the output frequency $k+\tau$ and determine the frequency of the limit cycle when the system was started with a sequence of frequency $k+\tau$. Some of these initial states stay at stable states with almost the same frequency; other ones run to the lowest branch ($k = q$). Random initial states lead to the lowest branch with a very high probability. Figure 4 shows that the results of this simulation are in agreement with the extension of our equations (5) and (7) to general $k+\tau$ which leads to allowed regions for for $q+\phi$ as a function of $k+\tau$. For the lowest

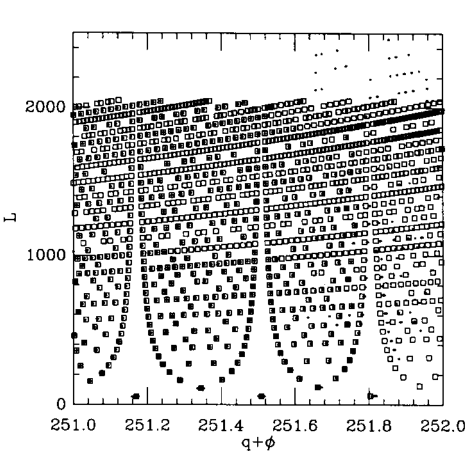

Figure 5. Cycle length L of the BG as a function of the frequency $q + \phi = r/s$ for $N = 1024$ (dots) and numerator of $(Ns)/r$ (squares). For $k = 251$, $\phi \in [0, 1[$; only a part of the cycle lengths shown since the highest observed cycle length was 8000.

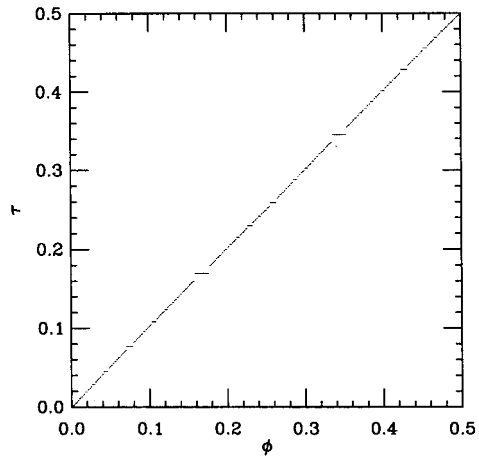

Figure 6. τ as a function of ϕ for $N = 1024$, $k = q = 251$.

branch $k = q$ the phase ϕ of the weights results in a frequency shift τ of the bit sequence, with $\tau \approx \phi$ for $q \gg 1$ similar to the continuous case.

The next problem is to understand the length L of the stationary cycle generated by the finite bit generator with one frequency in the couplings and a random initial vector. We consider the case $q = k$, only. Figure 5 shows the results of a numerical calculation of equation (1) for $N = 1024$. Obviously L has a rich structure as the function of the phase ϕ of the weights. For $0 \leqslant \phi \leqslant \frac{1}{2}$ each cycle has only one maximum in the Fourier spectrum with frequency $q + \tau(\phi)$. The numerical results show that in this case L is bounded by the value $2N$. Each value of L belongs to a whole interval in the ϕ-axis, but only to a single value of τ. Hence, the function $\tau(\phi)$ has a step-like structure; τ is locked at rational numbers as shown in figure 6. The size of the steps decreases with increasing N and r.

Can this structure of $L(\tau)$ be understood from the extension of the analytic solution (5)? For arbitrary values of τ the sequence S_l is quasiperiodic, in general with an infinite period L. However, if $k + \tau$ is rational, $k + \tau = r/s$ with integers r and s which are relatively prime, then the period L is given by the numerator of Ns/r. This means, that L is the smallest multiple of the wavelength $N/(k + \tau)$ that is an integer. In fact, in figure 5 we have plotted all of these values of L for $q + \phi = r/s$ and $L < 2N$. For $L < N$ all of these L values correspond to a cycle of the bit generator. For $N < L < 2N$ the bit generator produces only even values of L whereas the analytic argument gives all integers L.

For $\frac{1}{2} < \phi < 1$ the bit-generator essentially produces the same structure of L values as for $0 < \phi < \frac{1}{2}$. However, there are always a few solutions which are mixtures of several modes $k_1 + \tau_1$ and $k_2 + \tau_2$, and which yield periods with $L > 2N$. For $0 < \tau < \frac{1}{2}$ we never observed such mixtures of modes.

Note that the structure of $L(q + \tau)$ is essentially determined by the properties of rational numbers, which might be discussed in high-school mathematics. If the numerator p is plotted for each rational number $x = p/r$ in the unit interval, we obtain figure 7 ($r < 800$). Hence, above each rational number p/r a ceiling opens, below which no other values of p appear. Each ceiling has the form $1/|p - rx|$. Low values of r have a wide ceiling. These

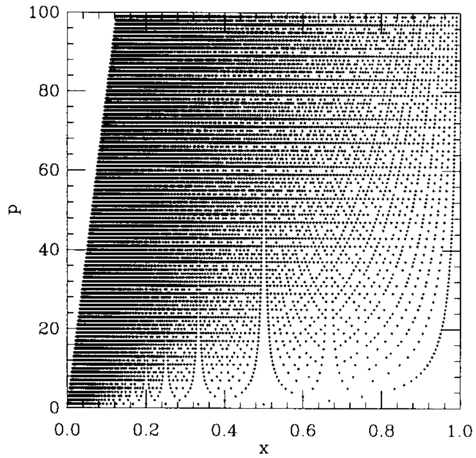

Figure 7. Numerator p of the reduced fraction $x = p/r$ ($r < 800$).

results, which determine the structure of the cycle lengths of the bit generator, may be well known in number theory and nonlinear dynamics (circle map, winding number), but they have skipped our attention so far.

The upper bound $2N$ of the cycle length L can be understood as follows. The analytic solution (5), extended to general values of $k+\tau$, yields quasiperiodic bit sequences with infinite cycle lengths $L = \infty$ for irrational $k+\tau$. However, each cycle length has to be limited by the number of input strings for the deterministic bit generator, equation (1), which gives $L < 2^N$. The last argument opens a different possibility to calculate L from equation (5). Let us start with the sequence $(S_0, S_1, \ldots, S_{N-1})$ given by the equation (5), of the analytic solution. If a bit generator tries to follow this solution it can do so only if each input string $(S_l, S_{l+1}, \ldots, S_{l+N-1})$ has not occurred before. Hence, the first appearance of a previous sequence

$$(S_l, S_{l+1}, \ldots, S_{l+N}) = (S_{l+L}, S_{l+L+1}, \ldots, S_{l+L+N-1}) \tag{8}$$

defines a length L of a cycle in agreement with figure 5.

More insight can be achieved by examining the continued fraction expansion of $2(k+\tau)/N$:

$$2\frac{k+\tau}{N} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} \qquad a_i \in \mathbb{N}. \tag{9}$$

We define the expansion to order i to be s_i and l_i to be the denominators the s_i:

$$
\begin{aligned}
s_0 &= 0 & l_0 &= 1\\
s_1 &= \frac{1}{a_1} & l_1 &= a_1\\
s_2 &= \cfrac{1}{a_1 + \cfrac{1}{a_2}} & l_2 &= a_1 a_2 + 1\\
s_3 &= \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}} & l_3 &= (a_1 a_2 + 1) a_3 + a_1\\
&\vdots & &\vdots\\
& & l_i &= a_i l_{i-1} + l_{i-2}.
\end{aligned}
\tag{10}
$$

If $2(k+\tau)/N$ is of the form $s_1 = 1/a_1$ the length L of the cycles as a function of N can be given easily:

$$
L = \begin{cases} 1 & \text{for } N < a_1 \\ 2a_1 & \text{for } N \geqslant a_1. \end{cases}
\tag{11}
$$

L is obviously limited by 2N.

For general $2(k+\tau)/N$ the continued fraction expansion reveals a hierarchy of 'defects', each having a period of l_i, in the periodic structure of the resulting sequence. This leads to the the fact that for any length scale introduced by N, two identical subsequences of length N with a distance less than $2N$ can be found.

As a consequence the perceptron locks, for a given frequency $k+\tau$, in cycles that correspond to frequencies given by the continued fraction expansion of $2(k+\tau)/N$, truncated at a certain depth. This explains the steps in figure 6.

Finally we point out similarities to the case of bit generators with exponentially decaying weights and additional bias (Cosnard *et al* 1988a). In this case one finds cycles which are limited by $N+1$. All of the cycles can be classified by rational numbers r/L where L is the length of the cycle, $L \leqslant N+1$ and r is the number of positive bits in the cycle.

In summary, we have obtained an analytic solution for the cycles of a bit-generator with periodic weight vectors. We found a whole spectrum of periodic attractors; the frequencies $k+\tau$ depend in a complex way on the frequency q and phase ϕ of the weight vector of the perceptron.

Numerical simulations showed that the bit sequences relax into cycles with lengths L, which are smaller than $2N$. The structure of L as a function of $k+\tau$ has been analysed in terms of number theory. An analytic solution was given for certain frequencies; the extension to the measured frequencies results in a similar structure of cycles.

Acknowledgments

This work has been supported by the Deutsche Forschungsgemeinschaft. We thank Ido Kanter for valuable discussions. We also thank a referee for pointing out the interesting papers of Cosnard *et al* to us which are related to our work.

References

Caianiello E R 1961 *J. Theor. Biol.* **2** 204
Caianiello E R and De Luca A 1965 *Kybernetik* **3** 33
Cosnard M, Tchuente M and Tindo G 1992 *Complex Syst.* **6** 13
Cosnard M, Goles Chacc E and Moumida D 1988 *Discrete Appl. Math.* **21** 21
Cosnard M, Moumida D, Goles E and St Pierre T 1988 *Complex Syst.* **2** 161
Eisenstein E, Kanter I, Kessler D A and Kinzel W 1995 *Phys. Rev. Lett.* **74** 6
Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City, CA: Addison-Wesley)
Kanter I, Kessler D A, Priel A and Eisenstein E 1995 *Phys. Rev. Lett.* **75** 2614
Opper M and Kinzel W 1996 *Statistical Mechanics of Generalization (Physics of Neural Networks III)* ed E Domany, J L van Hemmen and K Schulten (New York: Springer)
Schröder M, Kinzel W and Kanter I 1996 *J. Phys. A: Math. Gen.* **29** 7965
Watkin T L H, Rau A and Biehl M 1993 *Rev. Mod. Phys.* **65** 499
Weigand A S and Gershenfeld N A (ed) 1993 *Time Series Prediction, Forecasting the Future and Understanding the Past* (Santa Fe, NM: Santa Fe Institute)