ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

# XOR and Backpropagation Learning: In and Out of the Chaos? *

K.Bertels, L. Neuberg
University of Namur
Rempart de la Vierge 8
5000 Namur, B

S. Vassiliadis
T.U. Delft
Mekelweg 4
2628 CD Delft, Nl

G. Pechanek
IBM Microelectronics Div.
Research Triangle Park
North Carolina 27709, USA

**Abstract.** In this paper, we investigate the dynamic behavior of a backpropagation neural network while learning the XOR-boolean function. It has been shown that the backpropagation algorithm exhibits chaotic behavior and this implies an highly irregular and virtually unpredictable evolution. We study the chaotic behavior as learning progresses. Our investigation indicates that chaos appears to diminish as the neural network learns to produce the correct output. It is also observed that for certain values of the learning rate parameter the network does not converge and it appears as it may not arrive at producing the correct output.

## 1. Introduction

It has been observed that the delta rule bears a structural resemblance between the Verhulst equation, known for its chaotic behavior. The delta rule for the output node is

$$\Delta_p \tilde{W}_{ji} = \beta(t_{pj} - O_{pj})O_{pi}(1 - O_{pi})O_{pi}^M$$

and for the hidden nodes

$$\Delta_p W_{ji} = \beta(\sum_{k=1}^{n} \delta_{pk} W_{kj})O_{pi}(1 - O_{pi})O_{pi}^M$$

In both equations, the expression $O_{pi}(1 - O_{pi})$ is found which is also present in the Verhulst equation which is as follows : $x_{t+1} = \alpha x_t(1 - x_t)$.

Given the similarities of the equations it would seem natural to observe similar phenonema such as bifurcations and chaos as was found for the Verhulst equation. Indeed the chaotic behavior has been observed in [VM90]. A bifurcation diagram for the backpropagation algorithm has been created using the

*Research partially supported by IBM-Belgium.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

sum of the absolute values of the weights for different learning rates. The parameter regime indicates that bifurcations occur for learning rates inferior to 2.3 and a window appears between $\beta$=2.9 and $\beta$=3.2. For learning rates superior to 3.3, the weights grow exponentially. Futhermore, in [VM90] the presence of chaos has been illustrated by means of a phase space diagram, a power spectrum and the calculation of Lyapunov exponents. It has also been emphasized that for the full range of values the network succesfully learned. This implies that even for values in the chaotic regime, a neural network can succesfully learn.

A number of other papers discuss chaotic and dynamic aspects of neural networks other than backpropagation networks. In [AI90] a simple one neuron neural network is analysed that has a number of properties of biological neurons, such as the squid giant axons. The authors find a similar behavior of alternating periodic and chaotic sequences of neuron responses. In [DM88], it is shown that feed forward neural networks in general have a chaotic behavior because the distance between two arbitrarily close configurations always increases, which may be interpreted as sensitivity to initial conditions. Similar results are discussed in [SC88] where a continuous time dynamic model of a nonlinear network with random assymetric couplings is studied. For these networks, phenomena such as oscillations, bifurcations and chaos have been observed.

In this paper, we are mainly concerned by the role of chaos in the learning process of backpropagation neural networks. The main contributions of the paper can be summarized by the following : 1) as the learning the XOR-boolean function progresses, the order of chaos in the backpropagation network appears to diminish, 2) as the number of iterations used for learning increases, we find convergence for a wider range of learning rates ($\beta$), 3) the backpropagation neural network converges faster for some values of the learning rate $\beta$ and there are instances of $\beta$ for which the network appears not to converge within 500.000 iterations[1] and we do not know whether it will ever converge.

The paper is organized in function of the different calculations and visualizations made of the learning process. We first graphically expose the presence of chaos and how it evolves as learning progresses by mean of the bifurcation diagram and the corresponding phase spaces. We then compute on the basis of these data the Fourier transform and the Kolmogorov Entropy.

## 2. Bifurcation Diagrams

In Figures 1 and 2, we show the bifurcation diagrams for the XOR-function where the outputvalues of the network are plotted against different $\beta$-values for a temperature equal to 0.6. The bifurcation diagrams, constructed at different moments during the learning process, allow us to observe how the order of chaos

[1] We have chosen to restrict the maximum number of iterations during learning to 500.000. This restriction was imposed by our computational capabilities.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

| iterations/$\beta$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40K | | | | | | | | * | | * | | | | | |
| 200K | | | | | | | | * | * | * | | | | | |
| 400K | | | | | | | | * | * | * | * | * | | | |
| 500k | | * | * | * | * | * | | * | * | * | * | * | * | * | * |

Table 1: Evolution of convergence (*=convergence)

evolves over time. For different values of the learning rate $\beta$, we allowed the network to learn during a limited number of iterations, ranging from 40.000 to 500.000. For each run, we kept a record of the last 200 output values generated by the network. These values are then plotted against the corresponding $\beta$-value, resulting in a bifurcation diagram.

From these diagrams, we can make the following observations. First, from Figures 1 and 2, it is clear that we can distinguish between three possible states in which the neural network can be for each value of $\beta$ : 1) convergence of the network resulting in correct output values equal to 1 and 0 (e.g. in Figure 1 : for $\beta$=8.2 to 8.4), 2) finite periodicity corresponding with a finite number of values which do not correspond with the desired output(e.g. in Figure 1 : for $\beta < 8.2$), 3) chaos having an unperiodic series of output values obviously not corresponding with the desired output (e.g. in Figure 1 : for $\beta > 10.5$).

A second observation is that, as the number of iterations increases, the neural network converges for more $\beta$-values. After 40.000 iterations, we only have convergence for $\beta$-values around 8 and 10 whereas, after 400.000 iterations, convergence is achieved for values between 8 and 12. We also observe that after 500.000 iterations, which is the maximum number of iterations allowed, there is still no convergence for $\beta$'s around 1 and 7. We do not know whether, for these values, the network will ever converge. This evolution is summarized in Table 1.

Thirdly, as the number of iterations for the learning process increases, the order of chaos diminishes. When we observe Figures 1 and 2, we see that, even in the chaotic zone, more values are correctly computed. We will see that this observation is supported by the computation of the Kolmogorov entropy.

## 3. Phase Space

The phase space is a space in which each possible state of the system is represented unequivocally by a point in that space where each coordinate corresponds with a state variable of the system. If a dynamic system has a strange attractor, a phase space diagram will reveal its presence if the dimension of the attractor is less than the projection dimension. Some kind of structure will appear whenever some kind of deterministic system is involved, which of course is the case for

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

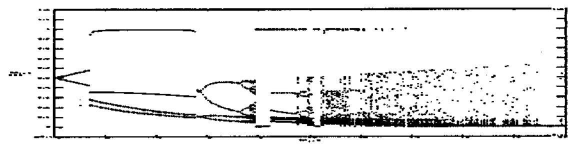

Figure 1: XOR Bifurcation diagram for temp=0.6 : 40.000 iterations.

Figure 2: XOR Bifurcation diagram for temp=0.6 : 400.000 iterations.

the backpropagation algorithm. However, this phase plan projection is nothing but an indication of chaos and its topographical characteristics are difficult to interpret.
In Figures 3 and 4, we show phase spaces for different values of $\beta$. For $\beta$=2, which corresponds with the non chaotic zone, a relatively regular geometric object emerges. For $\beta$=9, corresponding with the the chaotic parameter regime, we find a much more irregular object. This is considered to be an indication of chaos.

# 4. Fourier Power Spectrum

Chaotic systems are characterized by a broadband Fourier power spectrum in which no particular frequency can be found. We therefore expect to see for values inside the chaotic zone a broadband spectrum. We computed this power

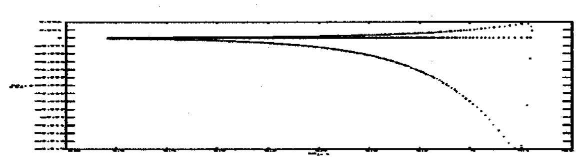

Figure 3: Phase space for temp=0.6 and $\beta$=2

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

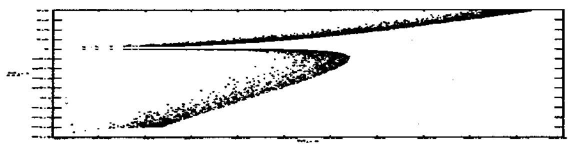

Figure 4: Phase space for temp=0.6 and $\beta$=12

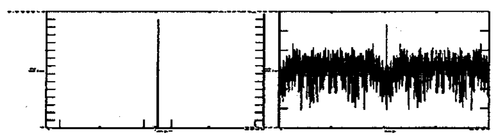

Figure 5: Power spectrum for temp=0.6 and a) $\beta$=2 and b) $\beta$=12

spectrum for $\beta$-values 2 and 12. As can be seen in Figure 5, in the periodic case the five peaks refer to the period 4 (Figure 5 a)) and the continuous broadband spectrum to chaos (Figure 5 b)).

# 5. Kolmogorov Entropy

We finally computed the Kolmogorov Entropy for the XOR-function using the approximation proposed by Grassberger and Procaccia ([GP83]). The entropy should be finite but non-zero in order to have a chaotic system. Not only did we compute the entropy for three $\beta$-values (2, 9 and 12) for a temperature of 0.6, but we also calculated the evolution of the entropy as the network learns. The results are shown in Table 2 and clearly indicate that the process becomes less chaotic. For $\beta$=12, there is initially a slight decrease in the entropy, but then this stabilises, implying that not all of the chaotic behavior disappears. This is a confirmation of what we visually could observe in the different bifurcation diagrams (Figures 1 and 2).

# 6. Conclusion

In this paper, we have investigated the XOR and the backpropagation algorithm as a non linear dynamic system having a number of interesting behavioral characteristics. While in [VM90], the presence of chaos has been established, we have additionnally shown the following : 1) for some values of the learning rate $\beta$, the neural network requires much less time (in terms of number of iterations)

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 69-74

| iterations | $\beta = 2$ | $\beta = 9$ | $\beta = 12$ |
|---|---|---|---|
| 0-10K | 0 | 0.08 | 0.09 |
| 10-20K | 0 | 0.05 | 0.08 |
| 20-30K | 0 | 0.03 | 0.08 |
| 30-40K | 0 | 0.01 | 0.08 |
| 40-50K | 0 | 0 | 0.08 |

Table 2: Evolution of Kolmogorov Entropy

to converge than for other values; 2) as the number of iterations increases, there appears to be convergence for a larger range of $\beta$-values; 3) the order of chaos appears to diminish as the number of iterations increases and the neural network learns.

Although we have to be very careful in advancing major statements, there appears to be some indication that during the learning process, the neural network gets into a chaotic state but is still capable of learning and consequently of getting out of this chaotic state. The next issue, which has not been adressed in this paper, is : what are the implications of chaos for the use of neural nets ? Are we dealing with an innocent side-effect of learning or does it constitute a fundamental property of it ? In biology, there is evidence that chaotic attractors are needed for learning [SF87] and therefore this may be an interesting line of research to see whether this also holds for artificial neural networks.

# References

[AI90] Aihara K., Takabe T., Toyoda M.;*Chaotic Neural Networks*, Physics Letters A, vol. 144, nr.6,7, 1990, pp. 333-340.

[DM88] Derrida B., Meir R.;*Chaotic behavior of a layered neural network*, Physical Review A, vol. 38, nr. 6, sept. 1988, pp.3116-3119.

[SF87] Skarda C., Freeman W.;*How brains make chaos in order to make sense of the world*, Behavioral and Brain Sciences, vol. 10, 1987, pp.161-195.

[GP83] Grassberger P., Procaccia I.; *Measuring the Strangeness of Strange Attractors*, Physica 9 D, 1983, pp. 189-208.

[SC88] Sompolinsky H., Crisanti A.;*Chaos in random neural networks*, Physical Review Letters, vol. 61, nr. 3, 1988, pp.259-262.

[VM90] Van der Maas H., Verschure P., Molenaar P.; *A note on chaotic behavior in simple neural networks*, Neural Networks, 1990, vol. 3, pp.119-122.