ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

Exponential Stability of Stochastic, Retarded Neural Networks

Mark P. Joy
School of Mathematics, Kingston Univeristy,

Kingston-upon-Thames, Surrey KT1 2EE, UK.
m.joy@kingston.ac.uk

Abstract. The stability analysis of neural networks is important in the applications and has been studied by many authors. However, only recently has the stability of stochastic models of neural networks been investigated. In this paper we analyse the global asymptotic stability of a class of neural networks described by a stochastic delay differential equation. It can be argued that such a model is as comprehensive as one would like to be when studying perturbations of neural networks since delay siganalling and noise are accounted for. We present a convergence theorem and discuss some examples of its use.

1 Introduction

The mathematical foundation of the stability analysis of neural networks has advanced considerably in the last ten years or so (see [1],[2], [3], [4], for example). Recently however the theoretical foundations of this subject has broadened to include stochastic differential equation models for neural networks. The papers [5], [6] study the global asymptotic stability of stochastic, pure-delay neural networks (of the Marcus-Westervelt type, see [7]) and have laid the foundations of this subject. In this paper we study a stochastic functional differential equation model of a neural network containing terms involving instantaneous *and* delayed signals – thus our model includes stochastic, delayed Cellular Neural Networks. Our model is

$$dx(t) = (-Px(t) + A_0 f(x(t)) + A_1 f(x(t-\tau)))dt + g(x(t), x(t-\tau))dB(t), \quad (1)$$

where A_0 and A_1 are fixed $(n \times n)$ real matrices, P is a diagonal matrix of positive reals, f is the usual diagonal mapping consisting of sigmoidal functions, $B(t)$ is an n-dimensional Brownian motion and $\tau > 0$ The model (1) gives rise to an Itô process under standard conditions which we elucidate below.

Two dynamics dominate the theory of continous-time neural networks:

a) the so-called *content addressable memory* dynamic introduced in [2] and utilised in Hopfield networks;

b) *global asymptotic stability*, we refer to such a dynamic as GAS.

In (a) it is usual for a finite set of equilibria to attract all trajectories of the dynamical system. Roughly speaking such *point attractor* networks recall memory states, since solutions flow from an intial condition to one of the stored memories. In (b) a unique stable equilibrium p exists and for all initial conditions x, $\omega(x) = p$. Here the location of p usually depends on some external input to the network, thus the network may be viewed as a non-linear classifier of inputs. It is this dynamic which we study here, our goal is to seek conditions on the parameters on (1) which ensure that the stochastic neural network is almost surely GAS.

Other authors have commented on the neccessity of incorporating noise into neural network models (see [8], p. 309). In one context the importance of a model such as (1) is clear: if neural networks are to be sucessfully fabricated as non-linear circuits then delays are unavoidable due to the finite switching speed of amplifiers. Such an environment clearly also contains thermal noise. It is important therefore to enquire whether a (usually deterministic) dynamic such as GAS is robust to stochastic perturbations.

One last comment is relevant. One can argue that choosing τ to vary among the neurons in the network, and even setting $\tau = 0$ for some subset of neurons, exhibits our model as a subtype of the neural network presented in [7]. However, casting the neural network equation in the form of (1), allows for two distinct processing modes: through the medium of the feedback term supported by the matrix A_0, instantaneous signals are propagated by all cells in the network. In addition, through the medium of A_1 all neurons propagate delayed signals. This flexibility is important in the applications and is implicit in the Cellular Neural Network paradigm; see [9], for example. From a theoretical perspective, the model (1) approaches the widest possible form of functional differential equation model involving *integrated* delays. Here we 'approximate' an integral term with delays at distinct points in time.

2 Background Material On Stochastic Differential Equations

In this section we present the notation used throughout this paper and outline in brief some fundamental theory. We refer the reader to [10] for proofs of the results in this section. We denote by $C = C([-\tau, 0], \mathbf{R}^n])$, the Banach space of continuous functions $\phi\colon [-\tau, 0] \to \mathbf{R}^n$ with the norm $\|\phi\| = \sup_{-\tau \leq \theta \leq 0} \mid \phi(\theta) \mid$. If A is a vector or a matrix A^T stands for the transpose of A, trace(A) is the sum of the diagonal elements of A. If A is a real, symmetric matrix $\lambda_{\max}(A)$ denotes the largest eigenvalue of A.

Let $B(t) = (B_1(t), \ldots, B_m(t))$ be an m-dimensional Brownian motion defined on a complete measure space $(\Omega, \mathcal{F}, \mathcal{P})$; this process defines the natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ ($\mathcal{F}_t$ is the σ-subalgebra generated by $B(s)$ for $0 \leq s \leq t$). Further define

- $L^2{}_{\mathcal{F}_t}([-\tau, 0], \mathbf{R}^n])$, the family of all $\mathcal{F}_t$-measurable $C([-\tau, 0], \mathbf{R}^n])$-valued random variables ϕ such that $\|\phi\|^2_{L^2} = \sup E \mid \phi(\theta) \mid^2 < \infty$;

In the sequel SFDE will stand for stochastic functional differential equation.

ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

Consider the SFDE (1). Let $g\colon \mathbf{R}^n \times C \to M_{m\times n}(\mathbf{R})$, so that g maps onto a $(n \times m)$ real matrix. Throughout this paper we will set $m = n$, such a scheme can represent noise injected into individual neurons independentley of all other neurons. We assume that the neuron activation function, $f\colon \mathbf{R} \to \mathbf{R}$, is a non-decreasing C^1-function which satisfies $xf(x) > 0$, $f(0) = 0$ and

$$f'(x) \leq 1. \tag{2}$$

If we assume that g is locally Lipschitz continuous and satisfies the linear growth condition then standard theory shows that given initial data $\phi \in C$ a unique solution $x(t;\phi)$ exists defined on $t \geq 0$, see [10]. $x(t;\phi) \in L^2{}_{\mathcal{F}_t}([-\tau, 0], \mathbf{R}^n])$.

3 GAS of Stochastic Neural Networks

In this section we will prove the main result of this paper, a theorem guaranteeing the almost sure GAS of a stochastic neural network. We will establish our result using a Lyapunov functional $V\colon C \to \mathbf{R}$. Since solutions to (1) are processes with values in the Banach space C, we need to find the *(weak) infinitesimal generator*, $\mathcal{L}$, of the semiflow on C. Such an object allows us to find the variation of functionals such as V along solutions of (1). Using (5.1) of [11] we obtain

$$\begin{aligned} \mathcal{L}V(x_t) &= \nabla V(x_t(0)) \cdot (-Px_t(0) + A_0 f(x_t(0)) + A_1 f(x_t(-\tau))) + \\ &\qquad \frac{1}{2}\sum_{i,j} V_{x_i x_j}(x_t(0))\sigma_{ij}(x_t), \end{aligned} \tag{3}$$

provided that V has a continuous second derivative with respect to $x_t(0)$; $V_{x_i x_j}$ denotes a second partial derivative. We are now able to state and prove the main result of this paper.

Theorem *Consider the stochastic functional differential equation (1). Assume that there exist symmetric nonnegative definite matrices C_1, C_2, C_3 and C_4 with C_2 diagonal, $C_2 = \Delta(\delta_1, \ldots, \delta_n)$ such that*

$$\mathrm{trace}\ [g(x,y)^T \cdot g(x,y)] \leq x^T C_1 x + y^T C_2 y + f(x)^T C_3 f(x) + f(y)^T C_4 f(y), \tag{4}$$

where $x = x_t(0)$, $y = x_t(-\tau)$ and f is the diagonal mapping $f(x_1, \ldots, x_n) = (f(x_1), \ldots, f(x_n))^T$. Assume further that there exists a diagonal matrix $D = \Delta(d_1, \ldots, d_n)$ $(d_i > 0)$ such that the symmetric matrix

$$Q = \begin{pmatrix} -2P + C_1 + C_2 + D & A_0 & A_1 \\ A_0^T & C_3 - D & 0 \\ A_1^T & 0 & C_4 - D \end{pmatrix}$$

is negative definite.

Then the trivial solution of (1) is almost surely exponentially stable. In other words $x(t;\phi) \to 0$, $t \to \infty$, a.s.

Proof. We fix our initial condition $\xi \in C$ arbitrarily and agree that $x(t;\xi) = x(t)$. Let us define $V\colon C \to \mathbf{R}$ by

$$V(\phi) = \mid \phi(0) \mid^2$$

Clearly $V \geq 0$. Since V has a continuous second derivative with respect to $\phi(0)$ the operator $\mathcal{L}V$ satisfies

$$\begin{aligned}\mathcal{L}V(\phi) &= 2\phi(0)^T(-P\phi(0) + A_0 f(\phi(0)) + A_1 f(\phi(-\tau))) \\ &\qquad + \frac{1}{2}\text{trace }[g(x,y)^T \cdot g(x,y)]\end{aligned}$$

for all $\phi \in C$. Now let $\phi = x_t$, and set $x_t(0) = x$ and $x_t(-\tau) = y$. We have

$$\mathcal{L}V(x_t) = -2x^T Px + 2x^T A_0 f(x) + 2x^T A_1 f(y) + \frac{1}{2}\text{trace }[g(x,y)^T \cdot g(x,y)]$$

whence,

$$\begin{aligned}\mathcal{L}V &= -2x^T Px + 2x^T A_0 f(x) + 2x^T A_1 f(y) + x^T C_1 x + y^T C_2 y \\ &\qquad + f(x)^T C_3 f(x) + f(y)^T C_4 f(y)\end{aligned}$$

by hypothesis. We rearrange, to obtain

$$\begin{aligned}\mathcal{L}V(x_t) &= x^T(-2P + C_1 + C_2 + D)x + 2x^T A_0 f(x) + 2x^T A_1 f(y) \\ &\qquad + f(x)^T(-D + C_3)f(x) + f(y)^T(-D + C_4)f(y) \\ &\qquad + y^T C_2 y + f(x)^T D f(x) - x^T(C_2 + D)x + f(y)^T D f(y).\end{aligned}$$

It is now easy to conclude that

$$\begin{aligned}\mathcal{L}V(x_t) &\leq (x, f(x), f(y))^T Q \begin{pmatrix} x \\ f(x) \\ f(y) \end{pmatrix} + y^T C_2 y + f(x)^T D f(x) \\ &\qquad + f(y)^T D f(y) - x^T(C_2 + D)x.\end{aligned}$$

Let $-\lambda = \lambda_{\max}(Q)$, so that $\lambda > 0$, then we have

$$\begin{aligned}\mathcal{L}V(x_t) &\leq -\lambda(\mid x \mid^2 + \mid f(x) \mid + \mid f(y) \mid^2) - x^T(C_2 + D)x + y^T C_2 y \\ &\qquad + f(x)^T D f(x) + f(y)^T D f(y) \\ &\leq -\sum(\lambda + \delta_i + d_i)x_i^2 + \sum(\delta_i y_i^2 - \lambda f(y_i)^2 + d_i f(y_i)^2) \\ &\qquad - \sum(\lambda - d_i)f(x_i)^2.\end{aligned}$$

From the definition of the matrix Q we have $\lambda \geq d_i$, thus

$$\mathcal{L}V \leq -\sum(\lambda + \delta_i + d_i)x_i^2 + \sum(\delta_i - \lambda + d_i)y_i^2.$$

We are now in a position to exploit theorem (2.1) of [5]; we are able to conclude exponential stability of the trivial solution.

Remark In fact the theorem yields a bound on the sample Lyapunov exponent of the solution since thorem 2.1 of [5] gives

$$\limsup_{t\to\infty} \frac{1}{t} \mid x(t) \mid \leq -\frac{\gamma}{2}, \qquad a.s,$$

where γ is the unique root of the equation

$$\lambda_1 = \gamma + \lambda_1\lambda_2 e^{\gamma\tau}, \tag{5}$$

with

$$\lambda_1 = \min(\lambda + \delta_i + d_i), \quad \lambda_2 = \max\frac{(\delta_i + d_i - \lambda)}{(\lambda + \delta_i + d_i)}.$$

4 Example

In this section we examine an example of the use of the above theorem.

Consider the SFDE

$$\begin{aligned}\begin{pmatrix} dx_1(t) \\ dx_2(t) \end{pmatrix} &= \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \begin{pmatrix} -0.5 & 0 \\ -0.5 & 1 \end{pmatrix}\begin{pmatrix} f(x_1(t)) \\ f(x_2(t)) \end{pmatrix} dt + \\ &\begin{pmatrix} \frac{1}{2} & 0 \\ 1 & -\frac{1}{3} \end{pmatrix}\begin{pmatrix} f(x_1(t-\tau)) \\ f(x_2(t-\tau)) \end{pmatrix} dt + \begin{pmatrix} 0.2x_2(t-\tau) \\ 0.5x_1(t-\tau) \end{pmatrix}\begin{pmatrix} dB_1(t) \\ dB_2(t) \end{pmatrix},\end{aligned}$$

where $f(x) = \arctan(x)$. For this example we have $g^T g = 0.04y_2^2 + 0.25y_1^2$, hence condition (4) is satisfied with $C_1 = C_3 = C_4 = 0$ and $C_2 = \Delta(0.25, 0.04)$. In the matrix Q we set $-2P + C_2 + D = -D$, where $D = \Delta(d_1, d_2)$. This yields $d_1 = 1.875$ and $d_2 = 1.98$. It turns out that the matrix Q is given by

$$Q = \begin{pmatrix} -1.875 & 0 & -0.5 & 0 & 0.5 & 0 \\ 0 & -1.98 & -0.5 & 1 & 1 & -0.333 \\ -0.5 & -0.5 & -1.875 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1.98 & 0 & 0 \\ 0.5 & 1 & 0 & 0 & -1.875 & 0 \\ 0 & -0.333 & 0 & 0 & 0 & -1.98 \end{pmatrix},$$

which is indeed negative definite. Numerical calculations reveal that

$$\lambda = -\lambda_{\max}(Q) = -0.3198.$$

Thus the stochastic neural network is almost surely stable. Suppose, for example, that $\tau = 0.5$, then we are also able to compute the root of equation (5) as 0.3089, thus the sample Lyapunov exponent is at most -0.1545.

5 Conclusion

We have studied a stochastic delay differential equation model of a neural network and have established conditions which ensure that the neural network is almost surely GAS. We are able to provide a bound for the exponential stability of the network. Thus it becomes clear that GAS is a dynamic for hybrid neural networks of the type (1) that is robust to stochastic perturbations. After [5], this is of course to be expected.

The method of proof of the main result of this paper relies heavily on the methods of [5]. From this point of view we regard the instantaneous signalling

ESANN'2005 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 27-29 April 2005, d-side publi., ISBN 2-930307-05-6.

term present in (1) as a kind of nuisance term. It gives rise to the last term on the right hand side of (5). This is a term that must be nonnegative for the argument to work. This, in turn, relies on $\beta \leq 1$, where $\beta = f'$. Furthermore, arguing along these lines, one is able to tolerate differing gains, $\beta \geq 1$, for all operational amplifiers involved in the delayed signalling but must bound those involved in instantaneous signalling to be at most one. Thus we trade the size of the instantaneous signals.

Finally one is led to seek different forms for V. In this direction, the results of ([11]) may be utilised to yield alternative stability results. This author is currently using such techniques to investigate further stability results for stochastic neural networks.

References

[1] M. Cohen M. and S. Grossberg, *Absolute stability of global pattern formation and parallel memory storage by competitive neural networks*, IEEE Trans. Sys. Man and Cyber., vol. 13, pp. 815–826, 1983.

[2] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci., 79, pp. 2554–2558, 1982.

[3] Mark Joy, *On the global convergence of a class of functional differential equations with applications in neural network theory*, J.Math. Anal Appl., vol. 232, pp. 61–81, 1999.

[4] X. Liao, G. Chen and E. N. Sanchez, *Delay-dependent exponential stability analysis of delayed neural networks: an LMI approach*, Neural Networks, 15, 2002, pp. 855–866.

[5] Steve Blythe, Xuerong Mao, Xiaoxin Liao, *Stability of stochastic delay neural networks*, J. of the Franklin Inst., vol. 338, pp. 481–495, 2001.

[6] X. X. Liao, X. Mao, *Exponential stability and instability of stochastic neural networks*, Stochst. Anal. Appl., vol. 14,pp. 165–185, 1996.

[7] C. Marcus and R. Westervelt, *Stability of analog neural networks with delay*, Physical Review A, vol. 39, pp. 347–359, 1989.

[8] S. Haykin, *Neural Networks*, Prentice-Hall, NJ, 1994.

[9] J. A. Nossek and T. Roska, *Special Issue on Cellular Neural Networks*, IEEE Trans. Circuits and Syst. II, vol. 40, 1993.

[10] S.-E. A. Mohammed, *Stochastic Functional Differential Equations: Theory and Applications*, Longman, New York, 1986.

[11] H. J. Kushner, *On the Stability of Processes Defined by Stochastic Difference-Differential Equations*, J. Differential Equations, 4, pp. 424–443, 1968.