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

Stochastic analysis of the Abe formulation of Hopfield networks

Marie Kratz[1], Miguel Atencia[2] and Gonzalo Joya[3] *

1- SAMOS-Matisse, Université Paris I - MAP5, Université Paris V
e-mail: Marie.Kratz@math-info.univ-paris5.fr
2- Universidad de Málaga - Dept. Matemática Aplicada
Campus de Teatinos, s/n. 29071-Málaga - Spain
3- Universidad de Málaga - Dept. Tecnología Electrónica
Campus de Teatinos, s/n. 29071-Málaga - Spain

Abstract. This work studies the influence of random noise in the application of Hopfield networks to combinatorial optimization. It has been suggested that the Abe formulation, rather than the original Hopfield formulation, is better suited to optimization, but the eventual presence of noise in the connection weights of this model has not been considered up to now. This consideration leads to a model that is formulated as a stochastic differential equation. In the stochastic setting, the analysis reveals that the model is stable, and the states converge towards an attractive set, assuming the noise intensity is bounded. The relation of the attractor with that of the deterministic model requires further study.

1 Introduction

Hopfield networks comprise a variety of related models, that have become an appealing tool for combinatorial optimization [9]. They are dynamical systems, due to the existence of recurrent connections, and their stability is deduced from the definition of an appropriate Lyapunov function. Optimization with Hopfield networks is accomplished by matching a multilinear target function with the Lyapunov function. Among these networks, the performance of the continuous Hopfield formulation [3] is controversial, since it requires evolution strategies [10, 6] in order to attain a feasible solution. Instead, the Abe formulation [1] has been claimed to be better suited to optimization [6]. Recently, it has been argued the importance of studying whether the network still approaches some limit set in the presence of random environmental noise, and this analysis has been performed for the Hopfield formulation [4]. In the present work, we develop a similar stochastic analysis for the Abe formulation, establishing the conditions for convergence, and identifying the topics that require further study in order to guarantee the optimization capability of the Abe model, subject to random perturbations.

*This work has been partially supported by Project No. TIC2001-1758, funded by the Spanish Ministerio de Ciencia y Tecnología (MCYT) and FEDER funds.

2 Hopfield networks for optimization

2.1 Deterministic case

The continuous Hopfield network [3] can be defined as a dynamical system whose states $s_i(t)$ are defined by the following differential equation:

$$\frac{d\,u_i}{d\,t} = -u_i + net_i \; ; \qquad s_i(t) = \tanh\left(\frac{u_i(t)}{\beta}\right) \; ; \qquad i = 1, \ldots, n$$

In the sequel, we restrict ourselves to first order networks, although more powerful models have been proposed [5]. Besides, we omit the explicit dependence of time, for brevity. The parameter β regulates the slope of the hyperbolic tangent and the potential u_i of the i-th neuron is driven by the linear input net_i:

$$net_i = \sum_{j=1}^{n} w_{ij}\, s_j - b_i \tag{1}$$

where w_{ij} is the weight of the connection from neuron j to neuron i and b_i is the bias of neuron i. From a dynamical viewpoint, the stability of this system results from the existence of a Lyapunov function:

$$V(s) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i\,j}\, s_i\, s_j + \sum_{i=1}^{n} b_i\, s_i + \beta \sum_{i=1}^{n} \int_0^{s_i} \arg\tanh(x)\, dx.$$

Optimization with Hopfield networks is accomplished by matching the Lyapunov function with the target function. In combinatorial optimization problems, the target function is multilinear, quadratic in the simplest case, hence the presence of the integral term in the Lyapunov function V is a severe drawback. Besides, stable equilibria in Hopfield networks do not belong to the set of vertices of the hypercube, whereas a feasible solution s in an optimization problem requires $|s_i| = 1$ for all i. Note that the conditions $w_{ij} = w_{ji}$ (symmetry) and $w_{ii} = 0$ (no self-weights), imposed by Hopfield in order to prove stability, appear naturally when the weights are regarded as the coefficients of the target function. A new formulation was proposed by Abe [1], which has been claimed to be better suited to optimization [6], since the structure of its energy function coincides with the form of the target function in optimization problems. Further, interior fixed points are unstable, hence the state converges to a vertex of the hypercube which is a feasible solution of the problem. The Abe formulation is defined by the following system of ordinary differential equations:

$$\frac{d\,u_i}{d\,t} = net_i \; ; \qquad s_i(t) = \tanh\left(\frac{u_i(t)}{\beta}\right) \; ; \qquad i = 1, \ldots, n \tag{2}$$

where net_i is defined by Equation (1). Since the feasible solutions correspond to $|s_i| = 1$, which is mapped to $u_i = \pm\infty$ by the hyperbolic tangent, the analysis is simplified when expressing the system with the single set of variables s_i, by means of the chain rule of differentiation:

$$\frac{d\,s_i}{d\,t} = \frac{1}{\beta}\left(1 - s_i^2\right) net_i \, ; \qquad\qquad i = 1 \ldots n.$$

Assuming a symmetric weight matrix W and no self weights, the proof of the function V below being a Lyapunov function results from the positiveness of the derivative of the hyperbolic tangent and the identity $\frac{\partial V}{\partial s_i} = -\frac{d\,u_i}{d\,t}$:

$$V(s) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij}\, s_i\, s_j + \sum_{i=1}^{n} b_i\, s_i = -\frac{1}{2} s^\top W\, s + b^\top\, s \tag{3}$$

2.2 Stochastic stability of the Abe formulation

The question arises as to the influence on the behaviour of the network when the presence of random noise in weights and biases is considered. This noise could model either inaccuracies in the components of a hardware implementation or the discretization error due to a computer simulation. The analysis of this situation has already been dealt with, for the Hopfield formulation [4], and our work follows the same track as this reference, but some assumptions are weakened thanks to the adoption of the Abe formulation. Besides, not only the stability of the system is guaranteed, but also the same Lyapunov function as in the deterministic case exists. From the computational point of view, this is a favourable property, that results in optimization of the target function being preserved under stochastic perturbations. Hence, our aim is to prove that the function given by Equation (3) is a Lyapunov function of the system defined in Equation (4) below, where the presence of random noise is considered. This purpose requires a new, stochastic, setting [8, 7]. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, on which an m-dimensional Brownian motion or Wiener process W is defined $W = \{W(t), t \geq 0\}$ and let $\mathcal{F} = (\mathcal{F}_t, t \geq 0)$ denote the corresponding natural filtration, i.e. $\mathcal{F}_t$ is the σ-algebra generated by the processes $W(s)$ with $0 \leq s \leq t$. Then an additive stochastic perturbation, as in [4], can be considered in Equation (2), resulting in the following stochastic differential equation:

$$d\,u_i = net_i\, dt + \sigma(u)\, d\,W_t \tag{4}$$

where $\sigma(x) = (\sigma_{ij}(x))_{n\times m}$ is the noise intensity matrix. As in the deterministic case, it is convenient to rewrite Equation (4) in terms of the s_i, which results from the stochastic analogue of the chain rule, the general Itô formula:

$$\begin{aligned} d\,s_i &= \frac{1}{\beta}(1 - s_i^2)\, d\,u_i - \frac{1}{\beta^2}\, s_i\,(1 - s_i^2)\left(\sigma^\top(u)\,\sigma(u)\right)_{ii}\, d\,t, \\ &= \frac{1}{\beta}(1 - s_i^2)\left\{\left[net_i - \frac{s_i}{\beta}\sum_{k=1}^{m}\sigma_{ik}^2\,(u)\right]dt + \sum_{k=1}^{m}\sigma_{ik}\,(u)\, dW_k(t)\right\} \end{aligned} \tag{5}$$

where u must be regarded as a symbol that stands for $u(t) = \beta \operatorname{argtanh}(s(t))$ and $x^\top$ denotes the vector transpose. Indeed this formulation will be useful to prove our results in a simpler way than the one used for Hopfields SDE's. Then, the differential operator L associated to Equation (5) is defined by:

$$\begin{aligned} L =& \frac{1}{\beta} \sum_{i=1}^{n} (1 - s_i^2) \left(net_i - \frac{s_i}{\beta} \sum_{k=1}^{m} \sigma_{ik}^2 (u) \right) \frac{\partial}{\partial s_i} \\ &+ \frac{1}{2\,\beta^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - s_i^2)(1 - s_j^2) \left(\sum_{k=1}^{m} \sigma_{ik}(u)\, \sigma_{jk}(u) \right) \frac{\partial^2}{\partial s_i s_j} \end{aligned}$$

Hence, for V defined in Equation (3), we obtain

$$\begin{aligned} LV(s) =& -\frac{1}{\beta} \sum_{i=1}^{n} (1 - s_i^2)\, net_i^2 \\ & - \frac{1}{2\,\beta^2} \sum_{i \neq j} (1 - s_i^2)(1 - s_j^2)\, w_{ij} \sum_{k=1}^{m} \sigma_{ik}(u)\, \sigma_{jk}(u) \\ & + \frac{1}{\beta^2} \sum_{i=1}^{n} s_i (1 - s_i^2)\, net_i \sum_{k=1}^{m} \sigma_{ik}^2 (u) \end{aligned} \tag{6}$$

In the deterministic case, i.e. when $\sigma \equiv 0$, as recalled in Section 2.1 above, it has been proved [2] that, under some assumptions, the state s converges to the vertices: $|s_i| = 1$. Therefore, our aim is to prove, in the stochastic case defined by Equation (5), that the system still exhibits some kind of convergence to some set K, to be determined.

Guided by the deterministic case, where the condition for stability is $\frac{dV}{dt} \leq 0$, it must first be determined whether we can assure $LV \leq 0$. This is indeed possible since, for instance, $\sigma(.)$ can be assumed to satisfy the following relation:

$$|\sigma(u)|^2 \leq \frac{\beta \sum_{i=1}^{n} (1 - s_i^2)\, net_i^2}{\max_{i,j} |w_{ij}| + \max_{i} |net_i|}$$

where, given a matrix A, $|A| = \sqrt{\operatorname{trace}(A^\top A)}$ is its trace norm. The characterization of the set $\{s : LV(s) \leq 0\}$ can be summarized in the following result:

Lemma 1 *Let $(\lambda_i)_i$ be the eigenvalues of the symmetric weight matrix W and $\lambda_{min} = \min_i \lambda_i$.*

- *If $\lambda_{min} \geq 0$ and if $net_i \geq 0\ \forall i$, σ has to be chosen according to $(\sigma^\top \sigma)_{ii} \leq \beta\, net_i$ in order to satisfy $LV \leq 0$.*
- *If $\lambda_{min} \leq 0$ and if $net_i \geq -\frac{|\lambda_{min}|}{2}$, then σ has to be chosen according to $(\sigma^\top \sigma)_{ii} \leq \frac{2\beta\, net_i^2}{|\lambda_{min}| + 2\, net_i}$ in order to satisfy $LV \leq 0$.*

- *Otherwise, the condition $LV \leq 0$ results regardless the value of the random noise σ.*

The proof proceeds by straightforward computations and is omitted. Further analysis is expected to transform these conditions into a criterion usable in practical applications, which is an ongoing task.

Contrarily to the deterministic case, the condition $LV \leq 0$ alone is not enough to guarantee the convergence to the limit set $\{s : LV(s) = 0\}$, hence we need some additional technicalities to discuss the stochastic stability of the Abe formulation.

Since $(s_t)_t$ is an Itô diffusion with operator L, then (see for instance Theorem 8.7 in [8]) V satisfies the following equation:

$$V(s(t)) = V(s_0) + \int_0^t LV(s(r))dr + M_t$$

where M_t is defined by

$$M_t = -\int_0^t \sum_{i=1}^n net_i \sum_{k=1}^m \sigma_{ik}\,(u)\, dW_k(t) = -\sum_{k=1}^m \int_0^t H_k(s(r))\, dW_k(r)$$

with $H_k(s(r)) = \sum_{i=1}^n net_i\, \sigma_{ik}\,(u)$.
Therefore let us define the limit set K:

$$K = \{s = s(t,\omega) : LV = 0\} \cap \{s = s(t,\omega) : H_k(s) = 0\ , \forall\, 1 \leq k \leq m\} \qquad (7)$$

With this notation, we can now state the main result of the paper, which is the analogue of Theorem 2.1 in [4], for the Abe formulation:

Theorem 1 *Assume $LV(s) \leq 0$. Then the set K defined in Equation (7) is not empty and for any initial value $s_0 \in [-1,1]$, the solution $s(t;s_0)$ of Equation (5) satisfies*

$$\liminf_{t\to\infty} d(s(t;s_0),K) = 0 \quad a.s.$$

with $d(y,K) = \min_{z\in K} |y - z|$ and "a.s." stands for "almost sure convergence", i.e. with probability one. Moreover, if for any $s \in K$, there exists a neighbourhood of s such that every point $r \neq s$ of this neighbourhood satisfies $V(r) \neq V(s)$, then, for any initial value s_0,

$$\lim_{t\to\infty} s(t;s_0) \in K \quad a.s. \qquad (8)$$

The proof requires an involved procedure and is omitted, since it follows the same steps as Theorem 2.1 in [4], but in a simpler way since no extra restriction is required on the transfer function tanh. This is a consequence of adopting the bounded states $|s(t,\omega)| \leq 1$ instead of the variables u, since then, the Lyapunov function V is bounded: $|V(s)| \leq \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n |w_{ij}| + \sum_{i=1}^n |b_i|$.

3 Conclusions

In this contribution, the presence of random noise in the weights of the continuous Hopfield network, in the Abe formulation, is considered. The stability of the stochastic model is shown, inspired by similar analysis on the Hopfield formulation. Besides, weaker assumptions are here required due to the adoption, as system states, of the s variables, which are confined to the unitary hypercube. The used Lyapunov function is identical to that of the deterministic case, which also matches with the multilinear target function of combinatorial optimization problems. Consequently, the system states converge, with probability one, towards some limit set K, which is a subset of the fixed points of the deterministic case. From a computational point of view, the work exposes a sort of robustness of the Abe formulation, since combinatorial optimization is expected even in the presence of random perturbations, as long as this noise is bounded.

The work raises several fundamental questions that deserve further study. First, the analysis of the bound on the noise intensity should lead to usable practical criteria that allow to determine whether a particular network is stable. Secondly, it should be desirable to ascertain the exact form of the limit set K and to compare it to the set of stable fixed points of the deterministic case; in particular, determining whether interior fixed points are stable, as proved by previous work in the deterministic case, is crucial for optimization, since these points are unfeasible. Finally, the influence of discretization on a computational implementation requires careful consideration, since numerical methods for stochastic differential equations are concerned.

References

[1] S. Abe. Theories on the Hopfield Neural Networks. In *Proc. IEE International Joint Conference on Neural Networks*, volume I, pages 557–564, 1989.

[2] M. Atencia, G. Joya, and F. Sandoval. Dynamical Analysis of Continuous Higher Order Hopfield Networks for Combinatorial Optimization. *Neural Computation (submitted)*, 2005.

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

[4] S. Hu, X. Liao, and X. Mao. Stochastic Hopfield neural networks. *Journal of Physics A: Mathematical and General*, 36:2235–2249, 2003.

[5] G. Joya, M. Atencia, and F. Sandoval. Associating arbitrary-order energy functions to an artificial neural network. Implications concerning the resolution of optimization problems. *Neurocomputing*, 14:139–156, 1997.

[6] G. Joya, M. Atencia, and F. Sandoval. Hopfield Neural Networks for Optimization: Study of the Different Dynamics. *Neurocomputing*, 43(1-4):219–237, 2002.

[7] H. J. Kushner. *Stochastic Stability and Control.* Academic Press, 1967.

[8] B. Øksendal. *Stochastic Differential Equations.* Springer-Verlag, 2003.

[9] K. A. Smith. Neural networks for combinatorial optimization: A review of more than a decade of research. *INFORMS Journal on Computing*, 11(1):15–34, 1999.

[10] M. Vidyasagar. Minimum-seeking properties of analog neural networks with multilinear objective functions. *IEEE Trans. on Automatic Control*, 40(8):1359–1375, 1995.