

ON STABILITY AND MODEL ORDER REDUCTION OF PERTURBED NONLINEAR NEURAL NETWORKS

Marissa Condon¹ and Georgi G. Grahovski^{1,2}

¹ School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, IRELAND,

² Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences,
72 Tsarigradsko chaussée, 1784 Sofia, BULGARIA

Email: marissa.condon@dcu.ie (M. Condon) grah@eeng.dcu.ie (G. Grahovski)

KEYWORDS

model reduction, balanced truncation, neural networks, perturbed nonlinear systems

ABSTRACT

In this paper, the qualitative theory of large-scale dynamical systems is surveyed. In particular, the focus is the Hopfield Neural networks both with and without perturbations. Properties relating to asymptotic and exponential stability and instability are detailed. A model reduction technique based on balanced truncation is applied to the neural networks. Its effect on the stability properties of the networks is then examined. A numerical test illustrates some important points.

1. INTRODUCTION

Nonlinear neural networks such as the Hopfield neural network [Hopfield (1982)], [Hopfield et al. (1986)], [Denker (1986)] are powerful computational systems for a wide variety of applications [Borisjuk et al. (2005)], [Edelstein-Keshet (1986)]. This lies in the fact that they are extremely robust to malfunctions. Their dynamic behaviour exhibits stable states and this is advantageous. For example, a time evolution of an array of neuron-like elements towards equilibrium points can be viewed as the evolution of an imperfect pattern towards a correct stored pattern. As stated in [Hinton et al. (Eds)], this is similar to the storage of information in an associative memory [Kohonen (1984)].

In most of the applications involving neural networks, the model equations form a large-scale system (see e.g. [Pomerlau et al. (1988)], [Iwata et al. (1990)], [Griffin et al. (1991)] and [Wawrzynek et al. (1993)]) and consequently, this leads to costly and inefficient computations. Therefore, model reduction is of paramount importance. However, its effect on the stability properties of the system is crucial in determining its usefulness. The reduced model must mirror the properties of the original system if it is to be of practical utility. Furthermore, the behaviour of the reduced system subject to perturbations must also match that of the original system.

Nonlinear model reduction has increasingly become a focus of research as in general, linear models are inadequate to describe real-world processes. While nu-

merous approaches for linear model reduction have been proposed [Antoulas (2003)], there is a dearth of effective nonlinear model reduction techniques. Balanced truncation, as pioneered by Moore [Moore (1981)], is a very effective linear model reduction technique and consequently, it has been extended by several authors for nonlinear systems. For example, Scherpen introduced the notion of controllability and observability functions to generalise the controllability and observability grammians which characterise linear systems [Scherpen (1993)]. However, their calculation is computationally expensive and their use is hence restricted [Scherpen (1993); Gray et al. (1998)]. To counteract this, empirical grammians have been proposed by several authors [Lall et al. (2002)], [Hahn et al. (2002a)] and [Condon et al. (2004)]. It is the technique in [Condon et al. (2004)] that is adopted in this work. In particular, the paper will examine the effect of the application of empirical balanced truncation on the stability properties of both unperturbed and perturbed neural networks.

The present paper looks at the robustness of neural networks to perturbations and examines if the related properties are preserved with empirical balanced truncation.

The paper is organized as follows: In section 2, the Hopfield neural network model is described briefly. In Section 3, the stability properties of neural networks are sketched. Then, in Section 4, the robustness properties of perturbed neural networks are reviewed, while the model reduction technique is reviewed in Section 5. The preservation of the robustness qualities in a reduced model is addressed in Section 6. These results are illustrated with a numerical example in Section 7.

2. NONLINEAR NEURAL NETWORK

Consider the following system of non-linear ODE's (known as Hopfield models [Hopfield (1982)], [Hopfield et al. (1986)] in the form:

$$\dot{u}_i = -b_i u_i + \sum_{j=1}^n A_{ij} G_j(u_j) + U_i(t), \quad (1)$$

where $i = 1, \dots, n$, A_{ij} and b_i are constants, $U_i(t)$ are functions of the time variable t . Model order reduction is applied to the model equations (1), and the paper studies the qualitative behaviour of the solutions of the reduced

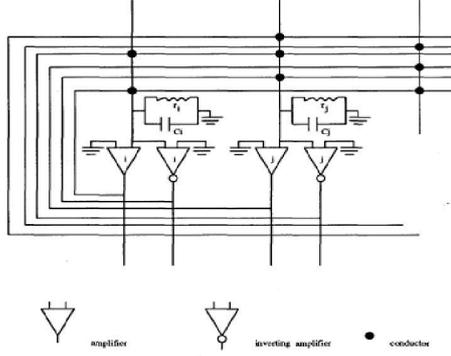


Figure 1: Hopfield neural network model

perturbed model (see eqn.(17) below) near equilibrium points (the positions where $\dot{u}_i = 0$, $i = 1, \dots, n$). By setting the external inputs $U_i(t)$, $i = 1, \dots, n$, equal to zero, $u^* = [u_1, \dots, u_n]^T \in \mathbb{R}^n$ is defined to be an equilibrium for eqn. (1). The locations of such equilibria in \mathbb{R}^n are determined by the interconnection pattern of the neural network (i.e., by the parameters A_{ij} ; $i, j = 1, \dots, n$) as well as by the parameters b_i , and the nature of the nonlinearities $G_i(\cdot)$, $i = 1, \dots, n$.

It is frequently assumed also, that

$$G_i(u_i) = a_i \arctan(\lambda_i u_i), \quad (2)$$

for some $a_i > 0$ and $\lambda_i > 0$.

In neural network applications, usually the input currents $I_i(t)$ are held constant over a time interval of interest, so it is customary to define an equilibrium as a point $u^* \in \mathbb{R}^n$ having the following property:

$$-b_i u_i^* + \sum_{j=1}^n A_{ij} G_j(u_j^*) + c_i = 0, \quad (3)$$

where c_i , $i = 1, \dots, n$ are time-independent constants, not necessarily equal to zero. Note that in this case, the locations of such equilibria in \mathbb{R}^n are determined by the interconnection pattern of the neural network (i.e., by the parameters A_{ij} ; $i, j = 1, \dots, N$), by the parameters b_i , the nature of the nonlinearities $G_i(\cdot)$, and the constant inputs c_i , $i = 1, \dots, N$.

In this study of the stability properties of such models, it will be assumed also that a given equilibrium point u^* is an isolated equilibrium point for eqn. 1, i.e., there exists an $r > 0$ such that in the neighbourhood $B(u^*, r) = \{u(t) \in \mathbb{R}^n : |u(t) - u^*| < r\}$ no equilibrium for eqn. 1, other than $u = u^*$, exists. It is shown in [Mitchel et al. (1989)], [Li et al. (1986)] that this assumption is a reasonable one for the case of the systems considered herein. When analyzing the stability properties of a given equilibrium point, it is possible to assume, without loss of generality, that this equilibrium is located at the origin of \mathbb{R}^n . By appropriate coordinate transformation, one can always map the equilibrium point u^* into the origin.

It is often convenient to view a system in the form of (1) as an interconnection of N free subsystems (or isolated subsystems) described by equations of the form:

$$\dot{p}_i = -b_i p_i + A_{ii} G_i(p_i^*) + U_i(t), \quad (4)$$

where $i = 1, \dots, n$. From this viewpoint, the terms

$$g_i(x_1, \dots, x_n) = \sum_{\substack{j=1 \\ j \neq i}}^n A_{ij} G_j(x_j^*), \quad (5)$$

make up the interconnecting structure of the system in (3).

3. STABILITY OF NONLINEAR NEURAL NETWORKS

3.1 Stability of Free Subsystems

Prior to describing the stability properties of the entire neural network, one will consider a brief sketch of the stability results concerning the individual free subsystems, described by eqn. (4).

Below, there is a sketch of the basic results on the various stability properties of the equilibrium $p_i = 0$ of the the model equations (4) for the case of vanishing external inputs $U_i(t) = 0$. The precise $\epsilon - \delta$ definitions of such concepts can be found, e.g., in [Miller et al. (1982)].

For any p_{i0} , sufficiently close to $p_i = 0$, if the solutions $\phi_i(t, t_0, p_{i0})$ of (4) with zero external inputs, remain close enough to the equilibrium $p_i = 0$ of (4) will be *stable* for $U_i(t) = 0$, $i = 1, \dots, n$.

If $p_i = 0$ is not stable, then it is said to be *unstable*.

If $p_i = 0$ is stable and if, in addition, $|\phi_i(t, t_0, p_{i0})|$ tends to zero as $t \rightarrow \infty$ whenever $p_i \in \mathcal{D}_i$, where \mathcal{D}_i is a subset of \mathbb{R} containing the origin $p_i = 0$, then $p_i = 0$ is said to be *asymptotically stable* and \mathcal{D}_i is called the *domain of attraction* for $p_i = 0$. If $\mathcal{D}_i \equiv \mathbb{R}$, then $p_i = 0$ is said to be *globally asymptotically stable*.

If $p_i = 0$ is asymptotically stable and if, in addition, $|\phi_i(t, t_0, p_{i0})|$ tends to zero exponentially, then $p_i = 0$ is said to be *exponentially stable*.

The *direct method of Lyapunov* enables the stability properties of the origin $p_i = 0$ for (4) to be determined without the actual determination of the solutions $\phi_i(t, t_0, p_{i0})$. In this method, continuously differentiable scalar-valued functions, $v_i(p_i)$, called Lyapunov functions, are employed. If a positive definite function $v_i(p_i)$ can be found such that the rate of change of $v_i(p_i)$ with respect to time along the solutions of (4), denoted $Dv_i(p_i)$, is negative semi-definite, then the equilibrium p_i for (4) will be stable. If a positive definite $v_i(p_i)$ can be found such that $Dv_i(p_i)$ is negative definite, then the equilibrium $p_i = 0$ will be asymptotically stable. If, in addition, both $v_i(p_i)$ and $Dv_i(p_i)$ are quadratic forms, then $p_i = 0$ will be exponentially stable. A more precise and complete treatment of Lyapunov's direct method is given in [Miller et al. (1982)].

3.2 Stability of Neural Networks

The equilibrium $u = 0$ of the neural network (1) is exponentially stable if: 1) for the system (1) all the external inputs are zero: $U_i(t) \equiv 0$, $i = 1, \dots, n$; 2) the interconnections satisfy the estimate: $u_i A_{ij} G_j(u_j) \leq |u_i| |a_{ij}| |u_j|$, for all $|u_i| < r_i$, $|u_j| < r_j$ and $i, j = 1, \dots, n$, where a_{ij} are real constants;

3) there exists a vector $\alpha \in \mathbb{R}_+^N$ (i.e. $\alpha^T = (\alpha_1, \dots, \alpha_N)$ and $\alpha_i > 0$, $i = 1, \dots, n$) such that the test matrix $\mathbf{S} = [s_{ij}]$

$$s_{ij} = \begin{cases} \alpha_i(-b_i + a_{ii}) & i = j \\ \frac{1}{2}(\alpha_i a_{ij} + \alpha_j a_{ji}) & i \neq j \end{cases} \quad (6)$$

is negative definite.

The Lyapunov function of the neural network (1) $v(\mathbf{u}) = \frac{1}{2} \sum_{i=1}^n \alpha_i u_i^2$ is considered as a weighted sum of the Lyapunov functions of the free subsystems with $U_i(t) \equiv 0$. The weighting factor $\alpha > 0$ is chosen so as to emphasize the qualitative properties of the individual subsystems.

Finally, it should be noted that the parameters A_{ij} need not form a symmetric matrix for the definitions of stability given previously to hold.

To conclude, the books by [Miller et al. (1982)], [Mitchel et al. (1977)] give a more detailed survey on stability analysis.

4. NONLINEAR NETWORKS WITH PERTURBATIONS

Consider the perturbed version of the Hopfield model (1)

$$\dot{u}_i = -\tilde{b}_i u_i + \sum_{j=1}^n \tilde{A}_{ij} \tilde{G}_j(u_j) + U_i(t), \quad (7)$$

where $\tilde{b}_i = b_i + \Delta b_i$, $\tilde{A}_{ij} = A_{ij} + \Delta \tilde{A}_{ij}$ (note, that $\Delta \tilde{A}_{ij}$ does not need to be a symmetric) and $\tilde{G}_j(u_j) = G_j(u_j) + \Delta \tilde{G}_j(u_j)$ ($\Delta \tilde{G}_j(u_j) \in C^2[\mathbb{R}, (-1, 1)]$), $1 \leq i \leq n$.

The system (1) is said to be *robust*, if for every asymptotically stable equilibrium u_e of (1), and for every $\varepsilon > 0$, there is a $\delta > 0$, such that for any perturbed system (7), as long as

$$\max \{ |\Delta b_i|, |\Delta A_{ij}|, |\Delta G_j(u_e)|, |\Delta G'_j(u_e)|, |U_i| \} < \delta$$

where

$$\Delta G'_j(u_j) = \frac{d(\Delta G_j)(u_j)}{du_j},$$

there is an *asymptotically stable* equilibrium \tilde{u}_e , of system (7), such that $|u_e - \tilde{u}_e| < \varepsilon$.

To summarise, robustness means that the system (1) is not overly sensitive to small perturbations. This is very important from a practical viewpoint as robustness ensures that small errors encountered in practical implementations of associative memories will not affect in an adverse manner the accuracy of the desired stored memories. Robustness ensures that the locations of the desired

asymptotically stable equilibria, which are used as memories in the neural networks, are not affected adversely by small perturbations.

The equilibrium u_e of the system (7) is *exponentially stable* iff [Wang et al. (1994)]: 1) u_e is an equilibrium point of both (7) and (1); 2) the matrix $-\mathbf{b} + \mathbf{A}\mathbf{G}'$ is Hurwitz stable; 3)

$$\max \{ |\Delta b_i|_\infty, |\Delta A_{ij}|_\infty, |\Delta A_{ij}|_1, |\Delta G'_j(u_e)|_\infty \} < K_0,$$

where $\frac{1}{K_0} = 2|\mathbf{P}|_\infty(1 + |\mathbf{A}|_\infty + |\mathbf{G}'|_\infty)$, and $\mathbf{P} = \mathbf{P}^T$ is a positive definite symmetric matrix, determined by $\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P} = -\mathbf{E}$, where $\mathbf{E} \in \mathbb{R}^{n \times n}$ is the identity matrix, $\mathbf{A} = [A_{ij}]_{i,j=1}^n$, $\mathbf{b} = \text{diag}(b_1, \dots, b_n)$, and where

$$\mathbf{G}'(u) = \left(\frac{dG_1(u_1)}{du_1}, \dots, \frac{dG_n(u_n)}{du_n} \right)^T.$$

Note that K_0 is a positive number, that is determined by system (1) and is independent of the system perturbations. It is an admissible bound for robust stability.

The equilibrium \tilde{u}_e of the system (7) is *exponentially stable* iff [Wang et al. (1994)]: 1) the matrix $-\mathbf{b} + \mathbf{A}\mathbf{G}'$ is Hurwitz stable; 2)

$$\max \{ |\Delta b_i|_\infty, |\Delta A_{ij}|_\infty, |\Delta A_{ij}|_1, |\Delta G'_j(u_e)|_\infty \} < K,$$

where $\frac{1}{K} = 4|\mathbf{P}|_\infty(1 + |\mathbf{G}'|_1 + |\mathbf{A}|_\infty)$, where again $\mathbf{P} = \mathbf{P}^T$ is a positive definite symmetric matrix, determined by $\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P} = -\mathbf{E}$, ($\mathbf{E} \in \mathbb{R}^{n \times n}$ is the identity matrix), $|\mathbf{G}'|_1 = \sup_{|x-x_e|_\infty < \varepsilon} |\mathbf{G}'|_\infty$, $|\tilde{u}_e - u_e| < \min(\varepsilon, 1/(4|\mathbf{G}'|_2|\mathbf{A}|_\infty|\mathbf{P}|_\infty))$, $\mathbf{A} = [A_{ij}]_{i,j=1}^n$ is a $n \times n$ - matrix, and $\mathbf{b} = \text{diag}(b_1, \dots, b_n)$.

Note that in the definitions given above, it is supposed that there exists an equilibrium of the perturbed system (7) that remains close to the corresponding equilibrium of the unperturbed system (1) for $t > 0$. For small perturbations, these assumptions hold true [Mitchel et al. (1989)], [Mitchel et al. (1977)]. [Wang et al. (1994)] give a detailed study of the stability conditions relating to perturbed neural networks.

In specific applications involving adaptive schemes for learning algorithms in neural networks, the interconnection patterns (and external inputs) are changed to yield an evolution of different sets of desired asymptotically stable equilibrium points with appropriate domains of attraction. One can derive a series of conditions, that can be used as constraints to guarantee that the desired equilibria always have the desired stability properties [Mitchel et al., 1989].

5. MODEL REDUCTION OF NONLINEAR SYSTEMS: EMPIRICAL GRAMIANS AND BALANCED TRUNCATION

Let us consider nonlinear systems of the most general form:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(t, \mathbf{x}(t)) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{h}(t, \mathbf{x}(t)) \end{aligned} \quad (8)$$

where $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^q$ are non-linear functions and the function $\mathbf{u}(t) \in \mathbb{R}^n$ is regarded as an input signal to the system (8) and the function $\mathbf{y}(t) \in \mathbb{R}^q$ is an output signal. In this case, $\mathbf{f}(t, \mathbf{x}(t))$ is called a dynamical term (or drift term) and $\mathbf{B}(t)\mathbf{u}(t)$ is called a source term (or diffusion term).

The unperturbed Hopfield network model (1) corresponds to (8) if the components of $\mathbf{f}(t, \mathbf{x}(t))$ are given by

$$f_i(t, \mathbf{x}(t)) = -b_i x_i + \sum_{j=1}^n A_{ij} G_j(x_j)$$

and the source term in (8) is given by the external input $U(t)$. In a similar manner, one can identify the different terms in the perturbed Hopfield model (7).

Suppose that the equilibrium point is reached when $\mathbf{u}(t) = 0$. Consider the vicinity of an isolated asymptotically stable equilibrium point (steady-state solution) which is supposed to be a constant solution and is chosen for simplicity at $\mathbf{x} = 0$, i.e. $\mathbf{f}(t, 0) \equiv 0$. It is also assumed that the system does not leave the region of attraction of this equilibrium point when the input is applied for the initial data used. If the system exhibits multiple steady-state solutions, then the analysis may be applied separately in the vicinity of each solution provided that extra care is taken to ensure that the system does not leave the region of attraction of the corresponding (asymptotically stable) equilibrium point.

Let also $\mathbf{x}^{ilm}(t)$ be the solution of (8) with $\mathbf{u} \equiv 0$:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad \mathbf{x}^{ilm}(0) = c_m T_l e_i. \quad (9)$$

It is assumed that the initial condition in (9) does not take the system outside the region of attraction of the equilibrium point $\mathbf{x} = 0$. Then the 'state-space average' of the 'nonlinear' fundamental solution may be defined as:

$$\langle \Theta(t) \rangle = \frac{1}{rs} \sum_{m=1}^s \sum_{l=1}^r \sum_{i=1}^n \frac{1}{c_m} x^{ilm}(t) e_i^T T_l^T \quad (10)$$

where $\mathbf{M} \equiv \{c_1, c_2, \dots, c_s\}$ is the set of s positive constants, $\mathbf{T}^n \equiv \{T_1, T_2, \dots, T_r\}$ – is the set of r orthogonal $n \times n$ matrices and $\mathbf{E}^n \equiv \{e_1, e_2, \dots, e_n\}$ is the set of standard unit vectors in \mathbb{R}^n . So for the system in (8), the *nonlinear* controllability gramian is defined as:

$$P = \int_0^\infty \langle \Theta(-\tau) \rangle^{-1} B(-\tau) B^T(-\tau) \langle \Theta(-\tau) \rangle^{-1T} d\tau \quad (11)$$

where $\langle \Theta(t) \rangle$ is as described in (10) and the *nonlinear* observability gramian is defined as [Condon et al. (2004)]:

$$Q = \int_0^\infty z^T(\tau) z(\tau) d\tau, \quad z(t) = \frac{1}{rs} \sum_{i,l,m} \frac{1}{c_m} y^{ilm}(t) e_i^T T_l^T \quad (12)$$

$y^{ilm}(t)$ is the output which corresponds to an initial state $x^{ilm}(0) = c_m T_l e_i$ and a zero source term.

Let T be the matrix that transforms both P and Q into diagonal form S as follows:

$$TPT^* = S, \quad (T^{-1})^* QT^{-1} = S, \quad (TPQT^{-1} = S^2)$$

The states of the system are then ordered according to decreasing values of the diagonal entries in S . Once balanced, a Galerkin projection $\Pi = [I, 0]$, where Π is $k \times n$ projection matrix and I is $k \times k$ unit matrix, is then employed to project the transformed system onto the states corresponding to the k largest singular values (i.e. the k largest values of the diagonal matrix S where k is the desired dimension of the reduced-order model).

The final reduced nonlinear model that corresponds to (8) takes the form:

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \Pi T \mathbf{f}(t, T^{-1} \Pi^* \mathbf{z}(t)) + \Pi T \mathbf{B}(t) \mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{h}(t, T^{-1} \Pi^* \mathbf{z}(t)) \end{aligned} \quad (13)$$

6. PERTURBED NONLINEAR NETWORKS AND MODEL REDUCTIONS

The model reduction technique of Section 5 will now be applied to the perturbed neural networks.

6.1 The case of free subsystems

In the case of vanishing $U_i(t)$, the reduced nonlinear model that corresponds to (4) is given by

$$\dot{z}_i = -\bar{b}_i z_i + \bar{A}_{ii} \bar{G}_i(\bar{p}_i), \quad (14)$$

or in vector form

$$\dot{\mathbf{z}} = -\bar{\mathbf{B}} \mathbf{z} + \bar{\mathbf{A}} \bar{\mathbf{G}}(\bar{\mathbf{p}}). \quad (15)$$

where $\bar{\mathbf{B}} = \bar{B} \mathbb{E}$, $\bar{B} = (\bar{b}_1, \dots, \bar{b}_n)^T$ (here, again \mathbb{E} is the $n \times n$ unit matrix) and the reduced model parameters are expressed through the initial ones as follows:

$$\begin{aligned} \bar{\mathbf{A}} &= \Pi T \tilde{\mathbf{A}} T^{-1} \Pi^*, & \mathbf{z} &= \Pi T \mathbf{p} \\ \bar{\mathbf{B}} &= \Pi T \tilde{\mathbf{B}} T^{-1} \Pi^*, & \bar{\mathbf{G}} &= \Pi T \tilde{\mathbf{G}}. \end{aligned} \quad (16)$$

Here, Π is $k \times n$ Galerkin projection matrix and T is the transformation matrix that casts both the gramians (11) and (12) into diagonal form (cf. Section 5).

If it is possible to force the solutions $\phi_i(t, t_0, p_{i0}, \varepsilon)$ of (4) with $U_i(t = 0)$, $i = 1, \dots, n$ to remain as close as desired to the stable equilibrium $p_i = 0$ for all $t > t_0 > 0$, by choosing p_{i0} to be sufficiently close to $p_i = 0$, then it will be possible to force the solutions $\tilde{\phi}_i(t, t_0, p_{i0}) = \Pi T \tilde{\phi}_i(t, t_0, p_{i0})$ of (15) to remain as close as desired the same stable equilibrium point for all $t > t_0 > 0$. So the equilibrium $p_i = 0$ of (4) with zero external inputs, will be *also* a stable equilibrium point for (14). The proof can be done in a similar manner to those in [Condon et al. (2008)].

Analogously, if $p_i = 0$ is an unstable equilibrium point for eqn. (4) with zero external inputs, then it will be unstable equilibrium point for eqn. (14) as well.

Finally, if $p_i = 0$ is exponentially stable for eqn. (4) with zero external inputs, then it will be exponentially stable equilibrium point for eqn. (14) as well.

The proofs are in a similar manner as in [Condon et al. (2008)].

6.2 The case of neural networks

Now let us consider the case of *interconnected systems* (*neural networks*).

Applying again model order reduction (as it is outlined in Section 5), we get a reduced nonlinear model, that corresponds to (7) in the form

$$\dot{z}_i = -\bar{b}_i z_i + \sum_{j=1}^n \bar{A}_{ij} \bar{G}_j(\bar{p}_j) + \bar{U}_i(t). \quad (17)$$

or in vector form

$$\dot{\mathbf{z}} = -\bar{\mathbf{B}}\mathbf{z} + \bar{\mathbf{A}}\bar{\mathbf{G}}(\bar{\mathbf{p}}) + \bar{\mathbf{U}}(t), \quad (18)$$

and the reduced model parameters are expressed through the initial ones (using the Galerkin projection matrix Π and the transformation matrix T) as follows:

$$\begin{aligned} \bar{\mathbf{A}} &= \Pi T \tilde{\mathbf{A}} T^{-1} \Pi^* & \bar{\mathbf{U}} &= \Pi T \tilde{\mathbf{U}}, \\ \bar{\mathbf{B}} &= \Pi T \tilde{\mathbf{B}} T^{-1} \Pi^*, & \bar{\mathbf{G}} &= \Pi T \tilde{\mathbf{G}}. \end{aligned} \quad (19)$$

If the equilibrium point $\mathbf{p} = 0$ for eqn. (7) is asymptotically stable in the case of zero external input $U_i(t) = 0$, then it is asymptotically stable equilibrium point for eqn. (17) as well.

Indeed, if there exist constants $\sigma_{i1} > 0$ and $\sigma_{i2} > 0$ for some $\tilde{r}_i > 0$, such that $\sigma_{i1} p_i^2 < p_i G_i(p_i) < \sigma_{i2} p_i^2$, $|p_i| < r_i$, then one can find constants $\tilde{\sigma}_{i1} > 0$ and $\tilde{\sigma}_{i2} > 0$, such that:

$$\tilde{\sigma}_{i1} z_i^2 < z_i \tilde{G}_i(z_i) < \tilde{\sigma}_{i2} p_i^2, \quad |z_i| < \tilde{r}_i. \quad (20)$$

If the interconnections for eqn. (4) satisfy the estimate: $u_i \tilde{A}_{ij} \tilde{G}_j(x_j) \leq |u_i| |\tilde{a}_{ij}| |u_j|$, for all $|u_i| < r_i$, $|u_j| < r_j$ and $i, j = 1, \dots, n$, where \tilde{a}_{ij} are real constants, then the interconnections for eqn. (17) will satisfy the estimate: $z_i \tilde{A}_{ij} \tilde{G}_j(z_j) \leq |z_i| |\tilde{a}_{ij}| |z_j|$, for all $|z_i| < \tilde{r}_i$, $|z_j| < \tilde{r}_j$ and $i, j = 1, \dots, n$, where \tilde{a}_{ij} are suitably chosen real constants. Secondly, if the test matrix $\mathbf{S} = [s_{ij}]$ for eqn. (4) is negative definite, then the test matrix $\tilde{\mathbf{S}} = [\tilde{s}_{ij}]$ for eqn. (17) will be negatively definite as well.

Detailed proofs for the stability results of neural networks can be found in [Mitchel et al. (1989)] and for the effect of model reduction on the stability of the Hopfield type network models – in [Condon et al. (2008)].

Summarizing, if $\mathbf{u} = 0$ is an exponentially stable equilibrium of the perturbed neural network (7), then it is an exponentially stable equilibrium of the reduced neural network (18), for the case when all the external inputs are zero: $U_i(t) \equiv 0$, $i = 1, \dots, n$.

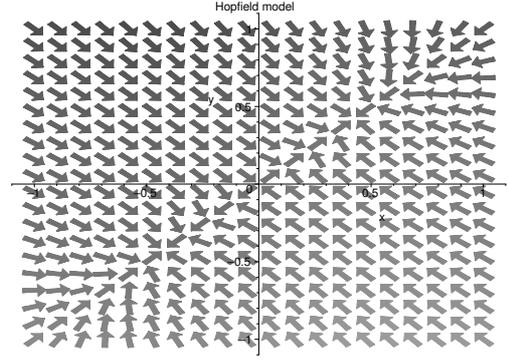


Figure 2: Phase portrait for the unperturbed Hopfield neural network model

7. EXAMPLE

For illustration of some of the stability results given above, we will take a specific test case, taken from [Hopfield (1984); Mitchel et al. (1989)]. For this case

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad g_i(x_i) = \frac{2}{\pi} \arctan\left(\frac{\lambda\pi}{2} x_i\right),$$

where $\lambda = 1.4$. Choosing the specific values of b_1 and b_2 to be $b_1 = b_2 = 1.1$, the model equation (1) takes the form:

$$\begin{aligned} \dot{u}_1 &= -1.1u_1 + g_2(u_2) + U_1(t) \\ \dot{u}_2 &= -1.1u_2 + g_1(u_1) + U_2(t). \end{aligned} \quad (21)$$

In the case of zero input $U_1(t) = U_2(t) = 0$ we have three equilibrium points for eqn. (21): $(-0.454, -0.454)$, $(0, 0)$ and $(0.454, 0.454)$ (as it is plotted in Fig.2).

By using the direct Lyapunov methods, one can prove that the equilibrium point $(0, 0)$ is not globally asymptotically stable. The precise proof can be found in [Mitchel et al. (1977)] and [Mitchel et al. (1989)].

The translated nonlinearities $G_i(w_i)$, $i = 1, 2$ for the equilibrium points $(\pm 0.454, \pm 0.454)$ are given by:

$$\begin{aligned} G_i(w_i) &= \frac{2}{\pi} \arctan\left(\frac{\lambda\pi}{2}(w_i \pm 0.454)\right) \\ &+ \frac{2}{\pi} \arctan\left(\frac{\lambda\pi}{2}(\mp 0.454)\right) \end{aligned} \quad (22)$$

In this case, we have $0 < G_i(w_i)/w_i < \sigma_{i2}$, $\sigma_{i2} < b = 1.1$, when $|w_i| < 0.454$ and the test matrix S takes the form

$$S = \begin{pmatrix} \frac{1.1}{\sigma_{i2}} & -1 \\ -1 & \frac{1.1}{\sigma_{i2}} \end{pmatrix}.$$

Since all successive principal minors of S are positive, the equilibrium points $(\pm 0.454, \pm 0.454)$ are asymptotically stable.

Consider the perturbed system, that corresponds to (21) in the form:

$$\begin{aligned} \dot{u}_1 &= -1.1u_1 + (1 + \epsilon)g_2(u_2) + U_1(t) \\ \dot{u}_2 &= -1.1u_2 + (1 + \epsilon)g_1(u_1) + U_2(t), \end{aligned} \quad (23)$$

where ϵ is a perturbation parameter.

Eqn. (23) can be linearised by using a Taylor series expansion of $\arctan x$ about the origin: $\arctan x = x - \frac{x^3}{3} + o(x^4)$. For sufficiently small ϵ , one of the eigenvalues of the linearised system is positive while the other is negative and hence the origin is a saddle point for both the perturbed and unperturbed systems. For the other equilibrium points when small perturbations are applied, the equilibrium points remain asymptotically stable.

In general, however, as shown in [Wang et al. (1994)], the perturbed neural network system may not have the same asymptotically stable points as the unperturbed system.

For large scale systems, when empirical balanced truncation is applied, the same conclusions can be made.

8. CONCLUSIONS

The paper has examined the stability properties of large-scale dynamical systems. In particular, it has focussed on nonlinear neural networks with perturbations. For this type of system, the properties of both the individual neurons and the interconnecting structure are employed to determine results for the system. The paper proceeds to ascertain the effect of model reduction technique on the resultant robustness of the equilibrium points of the system.

Further publications will explore the effect of model order reduction on the stability properties of perturbed nonlinear neural networks with feedback.

Acknowledgments

The authors wish to acknowledge the financial support of Science Foundation Ireland for this research.

REFERENCES

- A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems* (Advances in Design and Control Series), SIAM, Philadelphia (2003).
- Borisyuk A., Friedman A., Ermentrout B. and Terman D., *Tutorials in Mathematical Biosciences I: Mathematical Neuroscience*, Lecture Notes in Mathematics **1860**, Springer-Verlag, Berlin, 2005.
- Condon M. and Ivanov, R. Nonlinear systems-algebraic gramians and model reduction. *J. Nonl. Sci.* 2004, Vol.14, 2004, pp 405-414.
- M. Condon, G. G. Grahovski, Stability and Model Order Reductions of Nonlinear Systems with Applications to Neural Networks I: Hopfield Type Models, Preprint 2008.
- John S. Denker, Ed., *Neural Networks for Computing*, American Institute of Physics Conf. **151**, Snowbird, UT, 1986.
- Leah Edelstein-Keshet, *Mathematical Models in Biology*, Classics in Applied Mathematics **46**, SIAM, Philadelphia, 2005.
- Gray WS, Scherpen JMA. Hankel Operators and Gramians for Nonlinear Systems. *Proceedings of the 37th IEEE Conference on Decision and Control (CDC'98)*, Tampa, FL, USA 1998, pp. 1416-1421.
- M. Griffin, G. Tahara, K. Knorpp, R. Pinkham, and B. Riley, An 11- million transistor neural network execution engine, in *ISSCC91, Dig. Tech. Papers, Feb. 1991*, pp. 180-181.
- Lj. T. Grujic, A. A. Martynyuk, and M. Ribbens-Pavella, *Stability of Large-Scale Systems Under Structural and Singular Perturbations*, Nauka Dumka, Kiev, USSR, 1984.
- Hahn J, Edgar TF. An Improved Method for Nonlinear Model Reduction Using Balancing of Empirical Gramians. *Computers and Chemical Engineering* 2002; Vol. 26(10), pp. 1379-1397.
- G. E. Hinton and J. A. Anderson, Eds., *Parallel Models of Associative Memory*, Erlbaum, 1981.
- J. J. Hopfield *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proc. NatL Acad. Sci. USA **79** (1982), 2554-2558.
- J. J. Hopfield, *Neurons with graded response have collective computational properties like those of two-state neurons*, Proc. Nat. Acad. Sci. USA **81** (1984), 3088-3092.
- J. J. Hopfield and D. W. Tank, Computing with neural circuits: A model, *Science*, vol. 233, pp. 625-633, 1986.
- A. Iwata, Y. Yoshida, S. Matsuda, Y. Sato, and N. Suzumura, An artificial neural network accelerator using general purpose 24 bits floating point digital signal processors, in *Int. Joint Conf. Neural Networks, 1990, vol. II, pp. 43-46*.
- T. Kohonen, *Self-Organization and Associative Memory*, New York: Springer-Verlag, 1984.
- Lall S, Marsden JE, Glavaski S. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control* 2002; Vol. 12, pp. 519-535.
- J.-H. Li, A. N. Michel, and W. Porod, Qualitative analysis and synthesis of a class of neural networks, *ZEEE Trans. Circuits Syst.*, 533-541, 1986.
- R. K. Miller and A. N. Michel, *Ordinary Differential Equations*, New York: Academic, 1982.
- A. N. Michel and R. K. Miller, *Qualitative Analysis of Large Scale Dynamical Systems*, New York: Academic, 1977.
- A. N. Michel, J. A. Farrell, W. Porod, Qualitative Analysis of Neural Networks, *IEEE Trans. Circ. Sys.*, Vol. 36, No. 2, (1989), 229-243.
- Moore B. Principal Component analysis in linear systems: Controllability, Observability and model reduction. *IEEE Trans. on Automatic Control* 1981; AC-26(1).
- Phillips JR. Projection-based approaches for model reduction of weakly nonlinear, time-varying systems, *IEEE Transactions on computer-aided design of integrated circuits and systems* 2003, Vol. 22, No. 2.

D. Pomerleau, G. Gusciora, D. Touretzky, and H. Kung, Neural network simulation at warp speed: How we got 17 million connections per second, in *IEEE Int. Conf. Neural Networks, 1988, vol. II*, pp. 143–150.

Rewienski M and White J. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices, *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems* 2003, Vol. 22, No. 2, pp. 155-170.

Scherpen JMA. Balancing of nonlinear systems. *Systems and Control Letters* 1993; Vol. 21, pp.143-153.

K. Wang and A. N. Michel, Robustness and Perturbation Analysis of a Class of Nonlinear Systems with Applications to Neural Networks, *IEEE Trans. Circ. Sys, Vol. 41, No. 1, (1994)*, 24–32.

J. Wawrzyniek, K. Asanovic, and N. Morgan, The design of a neuromicroprocessor, *IEEE Trans. Neural Networks, vol. 4, pp. 394–399, May 1993*.

AUTHOR BIOGRAPHIES

MARISSA CONDON is a lecturer at DCU. Her research interests include the development of simulation techniques for nonlinear and high-frequency circuits and mathematical modelling of biological systems. Her webpage is <http://www.eeng.dcu.ie/~condonm.html>.

GEORGI GRAHOVSKI is a post-doctoral researcher at DCU. He has research interests in integrable systems, solitons, Lie groups and Lie algebras, differential geometry and biophysical models in addition to nonlinear circuit simulation techniques and nonlinear neural networks.