Neural Networks 74 (2016) 58–72

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Stability of discrete time recurrent neural networks and nonlinear optimization problems

Jayant Singh *, Nikita Barabanov

Department of Mathematics, North Dakota State University, Fargo, ND-58102, USA

ARTICLE INFO

Article history:
Received 20 April 2015
Received in revised form 21 August 2015
Accepted 30 October 2015
Available online 7 November 2015

Keywords:
Global Lyapunov stability
Optimization
Recurrent neural networks

ABSTRACT

We consider the method of Reduction of Dissipativity Domain to prove global Lyapunov stability of Discrete Time Recurrent Neural Networks. The standard and advanced criteria for Absolute Stability of these essentially nonlinear systems produce rather weak results. The method mentioned above is proved to be more powerful. It involves a multi-step procedure with maximization of special nonconvex functions over polytopes on every step. We derive conditions which guarantee an existence of at most one point of local maximum for such functions over every hyperplane. This nontrivial result is valid for wide range of neuron transfer functions.

1. Introduction and problem setting

In this paper, we study existence of points of local maxima for function $f(x) = \sum_{i=1}^{n} c_i\phi(x_i)$, where $\phi(\cdot)$ is a nonlinear function, over a hyperplane. This problem arises in stability analysis of nonlinear dynamical systems (Barabanov & Prokhorov, 2003), for example Discrete Time Recurrent Neural Networks (DTRNN's). DTRNN's have various applications such as image processing (Chen, Hung, Chen, Liao, & Chen, 2006), time series analysis (Gicquel, Anderson, & Kevekidis, 1998), etc. They are essentially nonlinear systems of specific structure. Implementation of such systems without rigorous stability analysis is hazardous. Therefore, it is necessary to analyze stability of these systems before using them for practical applications. A typical DTRNN can be described by the following systems of equations;

$$\begin{aligned} x_1^{k+1} &= \phi(W_1x_1^k + V_nx_n^k + b_1), \\ x_2^{k+1} &= \phi(W_2x_2^k + V_1x_1^{k+1} + b_2), \\ &\dots \\ x_n^{k+1} &= \phi(W_nx_n^k + V_{n-1}x_{n-1}^{k+1} + b_n), \end{aligned} \quad (1)$$

where n is the number of layers, $\phi(\cdot)$ is the activation function, x_j^k is the state vector of the layer j at time step k, W_j and V_j are fixed weight matrices, and b_j is a fixed vector representing bias.

System (1) describes the dynamics of a standard multilayered DTRNN without delays. Such RNN's may be used to solve a number of control problems (Prokhorov, Puskorius, & Feldkamp, 2001). An implementation of Back Propagation Through Time (BPTT) algorithm on the learning step results in determining coefficients of matrices W_i, V_i, and vectors b_i, for which system (1) matches input–output pairs and behaves like globally asymptotically stable system. For implementation of this DTRNN, however, it is necessary to prove global asymptotic stability (GAS) of the unique equilibrium state.

The strongest stability criteria for system (1) (except for Barabanov & Prokhorov, 2003), to the best of our knowledge, are provided by the theory of absolute stability (Barabanov & Prokhorov, 2002). Corresponding approach assumes construction of integral quadratic constraints (IQC), which use properties of activation function ϕ, and application of LMI/KYP lemma to find a quadratic Lyapunov function. We describe this approach and some other approaches in more detail later in this paper.

Unfortunately, criteria for absolute stability are not sufficiently strong to cover practically important RNN's. This is true even for one layered low dimensional systems. All available (millions) of numerical solutions of RNNs obtained after learning process tend to the same equilibrium state, but none of absolute stability criteria is applicable. It occurs due to the following reason. Classical criteria for GAS take into account only certain properties of nonlinear functions (monotonicity, sector condition). But among systems (1) with the same parameters but different function ϕ satisfying same properties, there might be an unstable system. In this case, criteria for absolute stability are not applicable.

An alternative approach was needed, and it has appeared in Barabanov and Prokhorov (2003). The method of reduction of

* Corresponding author. Tel.: +1 701 200 2555.
E-mail addresses: Jayant.Singh@ndsu.edu, jsingh@member.ams.org (J. Singh), nikita.barabanov@ndsu.edu (N. Barabanov).

http://dx.doi.org/10.1016/j.neunet.2015.10.013
0893-6080/

dissipativity domain (MRDD) is applicable to all systems (1), which have convex Lyapunov functions. One of deficiencies of methods based on a search of quadratic Lyapunov functions is the fact that the level sets of these functions are positive homogeneous (i.e. ellipses with the same parameters). But nonlinear function close to zero and far from zero belong to essentially different sectors, which requires ellipses with different parameters. All standard methods fail in such cases, but if system (1) has a convex (nonquadratic) Lyapunov function, then the MRDD works. This is very common due to large number of nonlinear functions in system (1).

Certainly, there exists a cost for such a generalization. It concerns a necessity to solve number of optimization problems of the same kind. This paper is devoted to this subject.

There are a number of interesting papers about GAS of DTRNN with several delays and/or several activation functions and/or nonzero linear parts (see for example Gao & Chen, 2007; Liu, Wang, Serrano, & Liu, 2007; Song, Gao, & Zheng, 2009). All stability criteria in these papers are based of a construction of certain local or integral quadratic constraints and application of quadratic Lyapunov functions. The MRDD may be also used for such systems. This may be a subject of future investigations.

Next, we consider two favorite approached for global asymptotic stability of systems (1).

In Sukyens, Vandwalle, and De Moor (1996), a stability criterion has been developed using NL_q approach. A typical NL_q system (without external inputs), is of the form,

$$p_{k+1} = P_1Q_1P_2Q_2 \dots P_qQ_qp_k \tag{2}$$

where $p_k \in \mathbb{R}^n, P_i = (\text{diag}(\overline{p_j}))_{j=1}^n$, and Q_i is a constant matrix. Here $\overline{p_j}$ depends on p_k continuously. The problem under consideration is to check the stability of system (2), with matrices P_i satisfying the relation $\|P_i\| \le 1$. The stability criterion using NL_q approach says that if there exists diagonal positive definite matrices D_j such that $\|D_jQ_jD_{j+1}^{-1}\| < 1$ for all $j = 1, \dots, q \pmod q$, then the system (2) is globally asymptotically stable. Using a suitable method (Barabanov & Prokhorov, 2002), the RNN defined in (1) can be transformed to form (2). Therefore the above criterion can be used to check the stability of systems of form (1). The NL_q approach gives sufficient conditions for stability of nonlinear systems. However, there exist nonlinear stable systems, for which the NL_q stability criterion is not satisfied. These nonlinear systems, for example, RNN, have shown promise in various applications (Feldkamp & Puskorius, 1998).

Another stability criterion was developed using theory of absolute stability (Aizerman & Gantmakher, 1963; Barabanov, 1987; Barabanov & Prokhorov, 2002; Narendra & Taylor, 1973; Yakubovich, 1967). A system to be analyzed for stability using this approach should be written in the automatic control form:

$$\begin{aligned} x^{k+1} &= Ax^k + B\psi^k, \\ \sigma^k &= Cx^k, \\ \psi_i &= \phi_i(\sigma_i), \quad i = 1 \dots m, \end{aligned} \tag{3}$$

where, A, B, C are matrices of suitable size, $\psi^k = (\psi_1, \dots, \psi_m)$ is the input vector at step k, $\sigma^k = (\sigma_1, \dots, \sigma_m)$ is the output vector at step k, and $\{\phi_i(\cdot)\}_{i=1}^m$ are nonlinear functions.

Before analyzing the stability of system (1) using theory of Absolute stability, it needs to be transformed to (3). State Space Extension method has been introduced in Barabanov and Prokhorov (2002) to transform RNN to (3).

One of the significant contribution of theory of absolute stability is the frequency domain criterion (Szegö & Kalman, 1963; Yakubovich, 1964, 1971, 1973). Frequency domain criterion gives necessary and sufficient conditions for the existence of a quadratic Lyapunov function for class of systems (3) with functions $\phi(\cdot)$ satisfying given local quadratic constraint. One of the most common constraints used for stability analysis of nonlinear systems is sector constraint, and the corresponding stability criterion is known as circle criterion. It has been shown in Barabanov and Prokhorov (2002) that stability criterion given by NL_q approach is weaker than the circle criterion.

The circle criterion gives sufficient condition for stability of nonlinear systems, with nonlinear function $\phi(\cdot)$ satisfying sector constraint. It only utilizes the fact that the nonlinear function $\phi(\cdot)$ satisfies a given sector condition. It might happen that given a sector, defined by function $\phi(\cdot)$, there exists a nonlinear function satisfying sector condition, such that the corresponding system is unstable. Additional information about the nonlinear function can be used to check stability of nonlinear systems of particular kind, for example RNN. A modified stability criterion using additional information about the nonlinear function, (e.g. monotonicity) has been developed in Barabanov and Prokhorov (2002). But this criterion has been shown to be essentially sufficient for systems with large number of nonlinear functions. In addition, this criterion is not applicable to some practically stable systems, for instance RNN's.

The stability criterion given by theory of absolute stability (Lankaster, Ran, & Rodman, 1986; Molinari, 1975) checks necessary and sufficient conditions for existence of Lyapunov functions of a particular kind (e.g. quadratic forms). But there exist stable systems, for which quadratic Lyapunov functions do not exist. An alternative stability criterion has been proposed in Barabanov and Prokhorov (2003).

Consider the system

$$x_{k+1} = \psi(x_k). \tag{4}$$

Let D_0 denote the whole space of vector x_k. Suppose there exist sets $\{D_k\}$ such that $D_{k+1} \subset D_k, \psi(D_k) \subseteq D_{k+1}$. If $\{D_k\} \to 0$ (in Hausdorff metric), as $k \to \infty$, then obviously system (4) is globally asymptotically stable. This approach is known as reduction of dissipativity domain.

In order to implement this approach, the sets D_k need to be defined. A possible choice of D_k is given by

$$D_{k+1} = \{x \in D_k : f_{k+1,j}(x) \le \alpha_{k+1,j},\ j = 1 \dots m_{k+1}\}$$

where m_k is the number of constraints at step k, $f_{k,j}$ is a function, and $\alpha_{k+1,j} = \max_{x \in D_k} f_{k,j}(\psi(x))$.

The set D_k is characterized by the set of pairs $(f_{k,j}, \alpha_{k,j})$ where $j \in \{1 \dots m\}$. A possible choice of $f_{k,j}(\cdot)$ is linear functions. Then D_k takes the shape of a polytope. It has been shown in Barabanov and Prokhorov (2003) that if system (4) has a convex Lyapunov function, then there exist linear functions $f_{k,j}$ such that $\{D_k\} \to 0$.

The set D_k is constructed by computing the value $\alpha_{k,j}$ for every j. Since the function $\psi(\cdot)$ is nonconcave over the set D_k, it can have multiple points of local maxima. At every step k, the points of local maxima for the function $f(\psi(\cdot))$ need to be computed.

Consider a single layer RNN with zero bias. Using substitution $y = Wx$, it can be expressed as

$$y_{k+1} = W\phi(y_k). \tag{5}$$

For the case of RNN in (5), the function $f(\phi(\cdot))$ is given by the inner product $f(x) := \langle l_j, W\phi(x)\rangle$. We need to find points of local maxima for $f(\cdot)$ over polytopes defined by matrix of constraints, $L = col(l_1, l_2, \dots, l_m)$. It has been seen that, in all the cases, the function $f(\cdot)$ has points of local maxima on the boundary of the polytope. We will first locate the points of local maxima for $f(\cdot)$ on an arbitrary hyperplane. The subject of this paper is the solution to the following problem.

Problem Setting. Consider the hyperplane, $P = \{x : l^Tx = b\}$ where l is a unit normal vector and $b \in \mathbb{R}$. How many points of local maxima does the function $f(x) = \sum_{i=1}^n c_i\phi(x_i)$, $c_i \neq 0$ for all i, have on P? Here, $\phi(\cdot)$ is a standard neuron transfer function.