Stability Analysis of Discrete-Time Recurrent Neural Networks

Nikita E. Barabanov and Danil V. Prokhorov, *Member, IEEE*

***Abstract*—We address the problem of global Lyapunov stability of discrete-time recurrent neural networks (RNNs) in the unforced (unperturbed) setting. It is assumed that network weights are fixed to some values, for example, those attained after training. Based on classical results of the theory of absolute stability, we propose a new approach for stability analysis of RNN with sector-type monotone nonlinearities and nonzero biases. We devise a simple state-space transformation to convert the original RNN equations to a form suitable for our stability analysis (without compositions of nonlinearities). We then write appropriate linear matrix inequalities (LMIs) to be solved to determine whether the system under study is globally exponentially stable. Unlike previous treatments, our approach readily permits us to account for nonzero biases usually present in RNN for improved approximation capabilities. We show how recent results of others on stability analysis of RNN can be interpreted as special cases within our approach. We illustrate how to use our approach with examples. Though illustrated on stability analysis of recurrent multilayer perceptrons (RMLPs), the approach proposed can also be applied to other forms of time-lagged RNN.**

***Index Terms*—Bias weight, discrete-time recurrent neural network, exponential stability, linear matrix inequality (LMI), Lyapunov stability, recurrent multilayer perceptrons (RMLPs), recurrent neural network (RNN), sector monotone nonlinearity, state-space transformation.**

I. Introduction and Problem Statement

RECURRENT neural networks (RNNs) have shown their promise and power in the variety of important applications [1], [6]. RNN are dynamic systems, and, as such, they frequently need to be analyzed for stability. Such analysis is particularly important when designing closed-loop control systems. We often have to deal with stability analysis of various closed-loop systems composed entirely of neural networks (feedforward or recurrent) which necessarily involve global feedback.

In this paper we consider the problem of global Lyapunov stability of discrete-time RNN. We assume that training of RNN has been somehow accomplished, and its weights are fixed during stability analysis. For training we favor powerful methods based on the extended Kalman filter (EKF) algorithm and backpropagation through time (BPTT) (see, e.g., [6]), but other methods of training can certainly be used.

Manuscript received March 31, 2000; revised October 12, 2000 and October 26, 2001. This work is supported in part by the URP 99-105.

N. Barabanov is with the Department of Software Engineering, St. Petersburg Electrotechnical University, Russia.

D. Prokhorov is with the Group of Artificial Neural Systems and Fuzzy Logic, Ford Research Laboratory, Dearborn, MI 48121 USA (e-mail: dprokhor@ford.com).

Publisher Item Identifier S 1045-9227(02)01798-8.

We discuss global Lyapunov stability analysis of layered recurrent networks, also known as recurrent multilayer perceptrons (RMLPs). Let n be the number of layers and $x_j^k (1 \leq j \leq n)$ be the state vector of the layer j at time step k. Forward propagation of signals through an RMLP with global feedback can be described by

$$\begin{aligned} x_1^{k+1} &= \tanh(W_1 x_1^k + V_n x_n^k + b_1) \\ x_2^{k+1} &= \tanh(W_2 x_2^k + V_1 x_1^{k+1} + b_2) \\ x_3^{k+1} &= \tanh(W_3 x_3^k + V_2 x_2^{k+1} + b_3) \\ &\cdots \\ x_n^{k+1} &= \tanh(W_n x_n^k + V_{n-1} x_{n-1}^{k+1} + b_n) \end{aligned} \tag{1}$$

where W_j, V_j are fixed matrices of weights, b_j are fixed vectors of biases. The hyperbolic tangent $\tanh$ in (1) may be replaced by any function ϕ such that functions $\phi(s)$ and $s - \phi(s)$ are monotone increasing. We analyze Lyapunov stability of an equilibrium point of *unforced* or *unperturbed* dynamical systems; therefore all external inputs to the layers are set to zero and do not appear in (1).

Problem Under Consideration: What are sufficient conditions on weight matrices and biases which guarantee that *all* trajectories of (1) converge to a (unique) equilibrium point?

This paper consists of the following sections. Section II discusses important classical results from the theory of absolute stability that laid the foundation to many approaches to stability analysis, including ours. Section III introduces a system transformation we devised to begin our stability analysis. Section IV clarifies the relationship between our approach to stability analysis of RNN and three recent results obtained by others including NL_q theory. In Section V, we show how nonzero biases usually present in RNN can be readily accounted for within the framework of our approach. The steps of our approach are summarized in Section VI. We illustrate our approach on examples in Section VII. Section VIII concludes the paper with remarks on the conservatism of all stability criteria and future directions of research.

II. Some Results From the Theory of Absolute Stability

In this section we describe some approaches and results known from the theory of absolute stability. They form the basis of our approach for stability analysis of RNN.

One of the most efficient methods of stability theory of dynamical systems is the method of absolute stability theory. A system to be analyzed for stability with this method is written in the form

$$x^{k+1} = Ax^k + B\xi^k, \quad \sigma^k = \Theta x^k + b = \begin{pmatrix} \sigma_1 \\ \dots \\ \sigma_n \end{pmatrix}$$
$$\xi^k = \begin{pmatrix} \xi_1 \\ \dots \\ \xi_n \end{pmatrix}, \quad \xi_i = \varphi_i(\sigma_i), \quad i = 1, \dots, n \tag{2}$$

where A, B, and Θ are some constant matrices, and b is a constant vector. It should be emphasized that the nonlinearities φ_i do not include explicit compositions of functions of different variables σ_i.

The main of the theory of absolute stability is the frequency theorem [7], [8]. Let F be a quadratic form with arguments x, ξ for the system (2). Assume that the pair (A, B) is stabilizable.

Theorem 1: Assume that the following conditions hold.

1) For any solution of the system (2) there exists a sequence of positive numbers $N_j \to \infty$ such that the inequality $\sum_{k=0}^{N_j} F(x^k, \xi^k) \geq 0$ is fulfilled for all j.
2) There exists a positive number ϵ such that the inequality $\mathrm{Re}\{F(z, \eta)\} \leq -\epsilon(| z |^2 + | \eta |^2)$ holds for all pairs (z, η) such that $Az + B\eta = e^{i\omega} z$ for some $\omega \in [0, \pi]$.
3) There exists a matrix D such that $F(y, Dy) \geq 0$ for all vectors y, and the matrix $A + BD$ is stable.

Then the equilibrium point of the system (2) is globally exponentially stable.

If the condition 1 holds, then the system (2) is said to satisfy an integral quadratic constraint with the form F. If $F(x, \xi) \geq 0$ for all x and $\xi = \mathrm{col}(\varphi_j(\sigma_j))$, then the system (2) is said to satisfy a local quadratic constraint with the form F. Satisfaction of the local quadratic constraint evidently implies satisfaction of the integral constraint (but not vice versa).

The condition 3 is usually called the condition of minimal stability. For almost all cases it is checked easily with $d = \mathrm{diag}\{\mu_j\}\Theta$ and some appropriate choice of the numbers μ_j.

The main condition 2 is usually called the frequency domain condition. The problem of checking this condition is reduced to checking for positive definiteness of some parameter-dependent matrix. This complicated problem is equivalent to the problem of existence of a Hermitian solution to

$$(Ax + B\xi)^* H(Ax + B\xi) - x^* Hx + F(x, \xi) < 0 \tag{3}$$

for all $(x, \xi) \neq 0$. Inequality (3) is a linear matrix inequality (LMI) with respect to the components of the matrix H. Furthermore, the quadratic form F is usually of the following type:

$$F = \sum_{j=1}^{r} \tau_j F_j \tag{4}$$

where τ_j are arbitrary positive numbers, and F_j are the quadratic forms with fixed coefficients, each of the forms describing some property of a nonlinearity. In this case the inequality (3) becomes an LMI also with respect to the parameters τ_j. These parameters and the matrix H may be found by an efficient interior point algorithm of the convex optimization [9].

One of the useful types of quadratic forms F_j in (4) is the sector quadratic form known from the early 1960s. It utilizes the fact that plots of nonlinear functions φ_j lie in some sectors $[\nu_j, \mu_j]$

$$\varphi_j(s)/s \in [\nu_j, \quad \mu_j] \quad \text{for all } s.$$

In particular, for the function $\varphi_j = \tanh$ we have $\nu_j = 0, \mu_j = 1$.

The corresponding quadratic form is

$$F_j = (\xi_j - \nu_j\sigma_j)(\mu_j\sigma_j - \xi_j).$$

Note that a function for which $F_j \geq 0$ may be nonmonotone and time varying. Thus, if the conditions of the Theorem 1 holds with the form F_j then the system (2) would be stable even if the function φ_j is nonmonotone and time varying. It is required only that its plot lie in the sector $[\nu_j, \mu_j]$. The frequency domain condition 1 with a quadratic form (4) that guarantees stability of all systems (2) with nonlinear functions φ_j satisfying the local quadratic constraint $F_j \geq 0$ for all $j = 1, \dots, r$, is called the circle criterion. It is not very specific in terms of the amount of information used about nonlinear functions; therefore the circle criterion is necessarily conservative when used for stability analysis of systems with nonlinear functions of a particular kind. To improve the circle criterion one can use additional information about the nonlinear function $\tanh$, e.g., its monotonicity or information about its derivative, $d(\tanh(s))/ds \in [0, 1]$.

In this paper, we will utilize two kinds of quadratic forms

$$F = \xi^*\Gamma(M\Theta x - \xi)$$

and

$$F = (\xi - N\Theta x)^*\Gamma(M\Theta x - \xi).$$

Here diagonal matrices M and N represent sector bounds for nonlinearities; $M = \mathrm{diag}\{\mu_j\}$ and $N = \mathrm{diag}\{\nu_j\}$. Γ is a matrix satisfying certain properties to be discussed in the next sections. This matrix is another argument of the LMI (3), which is to be solved with respect to *both* H and Γ. To solve LMIs, we will resort to the popular LMI Toolbox in MATLAB [10]. It is the state-of-the art tool for solving various LMIs arising in system and control theory.

III. System Transformation

To apply the method of absolute stability theory to stability analysis of RNN, it is necessary to transform the (1) to the form (2). Hence, the first step of our approach is a transformation step discussed in this section. In (2), Θ becomes a matrix of blocks of W_j and V_j, and b becomes a vector of biases. (A proper treatment of the vector b for stability analysis is explained in Section V.) Without loss of generality we henceforth assume that all nonlinearities are hyperbolic tangents, or $\tanh$, although the minimum requirement is to have all nonlinearities of a sector type (see Section II). A recurrent network (1) containing just one layer is already cast in the form (2). An RMLP with n layers without global feedback $(V_n = 0)$ can be analyzed for stability using (2) layer by layer. However, the (1) with $V_n \neq 0$ must be modified to fit the form (2). To transform these equations to (2) we propose to use a special state-space extension method. We would like to start with an illustration and then continue with a general case of the state-space extension method.

We consider a two-layer RMLP with global feedback described by

$$x_1^{k+1} = \tanh\left(W_1 x_1^k + V_2 x_2^k + b_1\right)$$
$$x_2^{k+1} = \tanh\left(W_2 x_2^k + V_1 x_1^{k+1} + b_2\right). \tag{5}$$

The original system above can be transformed into

$$\begin{aligned} x_{11}^{k+1} &= \tanh\left(W_1 x_{12}^k + V_2 x_{21}^k + b_1\right)\\ x_{12}^{k+1} &= x_{11}^k\\ x_{21}^{k+1} &= \tanh\left(W_2 x_{22}^k + V_1 x_{11}^k + b_2\right)\\ x_{22}^{k+1} &= x_{21}^k. \end{aligned} \tag{6}$$

The system (6) can now be written in the form suitable for our stability analysis

$$\begin{aligned} x^{k+1} &= Ax^k + B\xi^k, \quad \xi^k = \mathrm{col}\left(\xi_1^k, \xi_2^k\right)\\ \xi_j^k &= \tanh\left(\sigma_j^k\right), \quad j = 1, 2, \quad \sigma^k = \Theta x^k + b\\ \Theta &= \mathrm{col}(\Theta_1, \Theta_2), \quad b = \mathrm{col}(b_1, b_2)\\ \Theta_1 &= [0, W_1, V_2, 0], \quad \Theta_2 = [V_1, 0, 0, W_2]. \end{aligned} \tag{7}$$

where $x = \mathrm{col}(x_{11}, x_{12}, x_{21}, x_{22})$, and A and B are shown below

$$A = \begin{pmatrix} 0 & 0 & 0 & 0\\ I & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & I & 0 \end{pmatrix}, \quad B = \begin{pmatrix} I & 0\\ 0 & 0\\ 0 & I\\ 0 & 0 \end{pmatrix}.$$

We would like to demonstrate that the system (6) may be converted to two independent systems which are counterparts of the original system (5). We can write

$$\begin{aligned} x_{12}^{k+2} &= \phi_1\left(x_{12}^k, x_{21}^k\right)\\ x_{21}^{k+2} &= \phi_2\left(x_{21}^k, \phi_1\left(x_{12}^k, x_{21}^k\right)\right). \end{aligned} \tag{8}$$

We note that this system is identical to the original system (5) since one time step of the system (5) corresponds to two time steps of the system (6). (Indeed, let us choose $x_{12}^0 = x_1^0, x_{21}^0 = x_2^0$, then $x_1^1 = x_{12}^2, x_2^1 = x_{21}^2$, etc.) For the second counterpart, we have

$$\begin{aligned} x_{22}^{k+2} &= \phi_2\left(x_{11}^k, x_{22}^k\right)\\ x_{11}^{k+2} &= \phi_1\left(x_{11}^k, \phi_2\left(x_{11}^k, x_{22}^k\right)\right). \end{aligned} \tag{9}$$

The system (9) corresponds to the following system:

$$\begin{aligned} x_2^{k+1} &= \tanh\left(W_2 x_2^k + V_1 x_1^k + b_2\right)\\ x_1^{k+1} &= \tanh\left(W_1 x_1^k + V_2 x_2^{k+1} + b_1\right). \end{aligned} \tag{10}$$

It is easy to see that the system (10) is like the original system (5) for which the forward propagation begins with the second layer and ends with the first layer [it is the other way around for the system (5)].

For each state vector x_l of a layer l of an n-layer RMLP we introduce n copies

$$\begin{aligned} x_{l1}^{k+1} &= \tanh\left(W_l x_{ln}^k + V_{l-1} x_{l-1,1}^k + b_l\right)\\ &\qquad\qquad l = 1, \ldots, n (\mathrm{mod}\, n)\\ x_{l2}^{k+1} &= x_{l1}^k\\ x_{l3}^{k+1} &= x_{l2}^k\\ &\cdots\cdots\cdots\cdots\\ x_{\mathrm{ln}}^{k+1} &= x_{l,n-1}^k. \end{aligned} \tag{11}$$

It can be seen that (11) has the form (2). Equation (11) can be interpreted as describing n independent processes in the RMLP (1), and it is the general form illustrated in the example above for $n = 2$. Each process represents a system that is a counterpart of the original system (1). It differs from systems represented by other processes. For instance, a system with the state vector $\mathrm{col}(x_{1n}, x_{2,n-1}, x_{3,n-2}, \ldots, x_{n1})$ can be written as

$$\begin{aligned} x_{1n}^{k+n} &= \phi_1\left(x_{1n}^k, x_{2,n-1}^k, x_{3,n-2}^k, \ldots, x_{n1}^k\right)\\ x_{2,n-1}^{k+n} &= \phi_2\left(x_{1n}^k, x_{2,n-1}^k, x_{3,n-2}^k, \ldots, x_{n1}^k\right)\\ &\cdots\cdots\cdots\cdots\\ x_{n1}^{k+n} &= \phi_n\left(x_{1n}^k, x_{2,n-1}^k, x_{3,n-2}^k, \ldots, x_{n1}^k\right) \end{aligned} \tag{12}$$

where vector functions ϕ_i may be compositions of the functions $\tanh$ with appropriate arguments. However, there are no compositions of nonlinear functions in (11), which is important for our stability analysis.

Setting $k + n$ instead of $k + 1$ as the time index of the left-hand side of (12), we arrive at a system which is an *exact* equivalent of the system (1) since $x_{1n} = x_1, x_{2,n-1} = x_2, x_{3,n-2} = x_3, \ldots, x_{n1} = x_n$ in this case. It is also natural to count the forward propagation of signals sequentially from the first layer of (1). For other $n - 1$ processes with state vectors $\mathrm{col}(x_{11}, x_{2n}, x_{3,n-1}, \ldots, x_{n2}), \mathrm{col}(x_{12}, x_{21}, x_{3,n}, \ldots, x_{n3}),$ $\ldots, \mathrm{col}(x_{1,n-1}, x_{2,n-2}, x_{3,n-3}, \ldots, x_{nn})$, their systems differ from (1) only by the order in which signals propagate through the layers. The first layer to propagate through is denoted by the second index $*$ of the last component x_{n*} of the state vector. The propagation order is usual (sequential) for all the systems.

Stability is established at once for all n processes (12), counterparts of the original system (1). If this is possible (by solving an LMI system to be discussed further), then the global stability of the original system (1) and the system (11) follows immediately. [In the example discussed earlier in this section, proving stability for both (8) and (9) means doing so for (5) and (6).] Furthermore, recognizing the independence between the processes has important consequences for simplifying the computational aspects of our stability analysis (see Section VII-B for an illustration).

IV. On Existing Criteria for Lyapunov Stability of RNNs

This section discusses three recently published criteria for Lyapunov stability analysis of recurrent neural networks. We show that each of these is a special case of the criterion based on Theorem 1.

In [1], NL_q-systems are introduced to describe the dynamics of RNNs. A NL_q system (without external inputs) has the form

$$p_{k+1} = P_1 Q_1 P_2 Q_2 \ldots, P_q Q_q p_k \tag{13}$$

where $p_k \in \mathbf{R}^n$ is a state vector, Q_j are known matrices, andP_j are diagonal matrices. The problem is to check Lyapunov stability of system (13) for all matrices P_j satisfying the condition $0 \le P_j \le I$.

The systems (1) and (13) have different forms, but the system (1) can be transformed into the form of (13)

$$\begin{aligned} x^{k+1} = &\begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & f_q \end{pmatrix} \\ &\cdot \begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & 0 \\ & & & & V_{q-1} & W_q \end{pmatrix} \\ &\cdot \begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & f_{q-1} & \\ & & & & & I \end{pmatrix} \\ &\cdot \begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & 0 & \\ & & & V_{q-2} & W_{q-1} & \\ & & & & & I \end{pmatrix} \cdots \\ &\cdot \begin{pmatrix} f_1 & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{pmatrix} \\ &\cdot \begin{pmatrix} W_1 & 0 & \cdots & 0 & 0 & V_q \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{pmatrix} x^k \end{aligned}$$

where $f_j(s) = \tanh(s)/s, x^k = \mathrm{col}(x_1^k, \ldots, x_q^k)$, and all off-diagonal blocks not shown are set to zero. This system has the form (13) with

$$P_1 = \begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & f_q \end{pmatrix}, \ldots$$

$$P_q = \begin{pmatrix} f_1 & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{pmatrix}$$

$$Q_1 = \begin{pmatrix} I & & & & & \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & 0 \\ & & & & V_{q-1} & W_q \end{pmatrix}, \ldots$$

$$Q_q = \begin{pmatrix} W_1 & 0 & \cdots & 0 & 0 & V_q \\ & I & & & & \\ & & \cdots & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{pmatrix}.$$

Let us consider a special case of the system (1) with $b_j = 0, j = 1, \ldots, q (n \equiv q)$. The following stability criterion was published in [1, p. 127]:

If there exist diagonal positive–definite matrices D_j such that

$$\| D_j Q_j D_{j+1}^{-1} \| < 1 \quad \text{for all } j = 1, \ldots, q (\operatorname{mod} q) \tag{14}$$

then the system (13) [consequently, the network (1)] is stable.

Assuming diagonal positive definite matrices $D_j = \mathrm{diag}(D_{j,1}, \ldots, D_{j,q})$, the conditions (14) take the form

$$\begin{gathered} D_{1,1} < D_{2,1} < D_{3,1} < \cdots < D_{q,1} \\ D_{q,2} < D_{1,2} < D_{2,2} < \cdots < D_{q-1,2} \\ \cdots\cdots\cdots\cdots \\ D_{2,q} < D_{3,q} < D_{4,q} < \cdots < D_{1,q} \\ \left\| D_{q,1} V_q D_{1,q}^{-1} \quad D_{q,1} W_1 D_{1,1}^{-1} \right\| < 1 \\ \left\| D_{q-1,2} V_1 D_{q,1}^{-1} \quad D_{q-1,2} W_2 D_{q,2}^{-1} \right\| < 1 \\ \cdots\cdots\cdots\cdots \\ \left\| D_{1,q} V_{q-1} D_{2,q-1}^{-1} \quad D_{1,q} W_q D_{2,q}^{-1} \right| < 1. \end{gathered}$$

We can show how to obtain a stronger stability criterion using the approach from the theory of absolute stability. As described in Section III, we extend the state space as follows:

$$
\begin{aligned}
x_{11}^{k+1} &= \xi_1^k = \tanh\left(W_1 x_{1q}^k + V_q x_{q1}^k\right)\\
x_{12}^{k+1} &= x_{11}^k\\
&\cdots\cdots\cdots\cdots\cdots\\
x_{1q}^{k+1} &= x_{1,q-1}^k\\
x_{21}^{k+1} &= \xi_2^k = \tanh\left(W_2 x_{2q}^k + V_1 x_{11}^k\right)\\
x_{22}^{k+1} &= x_{21}^k\\
&\cdots\cdots\cdots\cdots\cdots\\
x_{2q}^{k+1} &= x_{2,q-1}^k\\
&\cdots\cdots\cdots\cdots\cdots\\
x_{q1}^{k+1} &= \xi_q^k = \tanh\left(W_q x_{qq}^k + V_{q-1} x_{q-1,1}^k\right)\\
x_{q2}^{k+1} &= x_{q1}^k\\
&\cdots\cdots\cdots\cdots\cdots\\
x_{qq}^{k+1} &= x_{q,q-1}^k.
\end{aligned} \tag{15}
$$

For all positive definite diagonal matrices Γ_j the following quadratic inequalities must be satisfied on trajectories of (15):

$$
\begin{aligned}
&\xi_1^*\Gamma_1(W_1 x_{1q} + V_q x_{q1} - \xi_1) \geq 0\\
&\xi_2^*\Gamma_2(W_2 x_{2q} + V_1 x_{11} - \xi_2) \geq 0, \ldots\\
&\xi_q^*\Gamma_q(W_q x_{qq} + V_{q-1} x_{q-1,1} - \xi_q) \geq 0.
\end{aligned} \tag{16}
$$

It can be shown that (3) has a solution with diagonal matrix H if there exists a solution to the coupled inequalities shown at the bottom of the page, with respect to diagonal positive definite matrices $H_{i,j}$ and $\Gamma_i$where the symbol # in expressions $H_{i,\#}$ stands for i; with additional constraints

$$
\begin{aligned}
H_{q,q} &\leq H_{q-1,q-1} \leq H_{q-2,q-2} \leq \cdots \leq H_{2,2}\\
H_{2q,2q} &\leq H_{2q-1,2q-1} \leq H_{2q-2,2q-2} \leq \cdots \leq H_{q+2,q+2}\\
&\cdots\cdots\cdots\cdots\\
H_{qq,qq} &\leq H_{qq-1,qq-1} \leq H_{qq-2,qq-2} \leq \cdots\\
&\leq H_{q(q-1)+2,q(q-1)+2}.
\end{aligned}
$$

Each of the inequalities above can be transformed to

$$
\begin{pmatrix} -I & 0 & X^* \\ 0 & -I & Y^* \\ X & Y & -I \end{pmatrix} < 0
$$

which is equivalent to

$$
\|X \quad Y\| < 1.
$$

where X and Y are some matrices. We obtain

$$
\begin{aligned}
&\left\| (\Gamma_1 - H_{11})^{-\frac{1}{2}}\Gamma_1 V_q \left(H_{q(q-1)+1,\#} - H_{q(q-1)+2,\#}\right)^{-\frac{1}{2}} \Big/ 2 \right.\\
&\qquad \left. (\Gamma_1 - H_{11})^{-\frac{1}{2}}\Gamma_1 W_1 H_{q,q}^{-\frac{1}{2}} \Big/ 2 \right\| < 1\\
&\left\| (\Gamma_2 - H_{q+1,q+1})^{-\frac{1}{2}}\Gamma_2 V_1 (H_{11} - H_{22})^{-\frac{1}{2}} \Big/ 2 \right.\\
&\qquad \left. (\Gamma_2 - H_{q+1,q+1})^{-\frac{1}{2}}\Gamma_2 W_2 H_{2q,2q}^{-\frac{1}{2}} \Big/ 2 \right\| < 1\\
&\qquad\cdots\cdots\cdots\cdots\\
&\left\| \left(\Gamma_q - H_{q(q-1)+1,\#}\right)^{-\frac{1}{2}} \Gamma_q V_{q-1} \right.\\
&\qquad \times \left(H_{q(q-2)+1,\#} - H_{q(q-2)+2,\#}\right)^{-\frac{1}{2}} \Big/ 2\\
&\qquad \left. \left(\Gamma_q - H_{q(q-1)+1,\#}\right)^{-\frac{1}{2}} \Gamma_q W_q H_{qq,qq}^{-\frac{1}{2}} \Big/ 2 \right\| < 1.
\end{aligned} \tag{17}
$$

Comparing appropriate terms to the left and to the right of matrices V_i and W_j we notice that

$$
\begin{aligned}
D_{q,1} &= (\Gamma_1 - H_{11})^{-\frac{1}{2}}\Gamma_1/2 = (H_{11} - H_{22})^{\frac{1}{2}}\\
D_{1,q} &= \left(\Gamma_q - H_{q(q-1)+1,\#}\right)^{-\frac{1}{2}} \Gamma_q/2\\
&= \left(H_{q(q-1)+1,\#} - H_{q(q-1)+2,\#}\right)^{\frac{1}{2}}\\
D_{11} &= H_{q,q}^{\frac{1}{2}}\\
D_{q-1,2} &= (\Gamma_2 - H_{q+1,q+1})^{-\frac{1}{2}}\Gamma_2/2\\
&= (H_{q+1,q+1} - H_{q+2,q+2})^{\frac{1}{2}}\\
D_{q,2} &= H_{2q,2q}^{\frac{1}{2}}\\
&\cdots\cdots\cdots\cdots\\
D_{2,q-1} &= \left(\Gamma_{q-1} - H_{q(q-2)+1,\#}\right)^{-\frac{1}{2}} \Gamma_{q-1}/2\\
&= \left(H_{q(q-2)+1,\#} - H_{q(q-2)+2,\#}\right)^{\frac{1}{2}}\\
D_{2,q} &= H_{qq,qq}^{\frac{1}{2}}.
\end{aligned}
$$

$$
\begin{pmatrix} H_{q(q-1)+2,\#} - H_{q(q-1)+1,\#} & 0 & V_q^*\Gamma_1/2 \\ 0 & -H_{q,q} & W_1^*\Gamma_1/2 \\ \Gamma_1 V_q/2 & \Gamma_1 W_1/2 & H_{11} - \Gamma_1 \end{pmatrix} < 0
$$

$$
\begin{pmatrix} H_{22} - H_{11} & 0 & V_1^*\Gamma_2/2 \\ 0 & -H_{2q,2q} & W_2^*\Gamma_2/2 \\ \Gamma_2 V_1/2 & \Gamma_2 W_2/2 & H_{q+1,q+1} - \Gamma_2 \end{pmatrix} < 0
$$

$$
\begin{pmatrix} H_{q+2,q+2} - H_{q+1,q+1} & 0 & V_2^*\Gamma_3/2 \\ 0 & -H_{3q,3q} & W_3^*\Gamma_3/2 \\ \Gamma_3 V_2/2 & \Gamma_3 W_3/2 & H_{2q+1,2q+1} - \Gamma_3 \end{pmatrix} < 0
$$

$$\cdots\cdots\cdots\cdots$$

$$
\begin{pmatrix} H_{q(q-2)+2,\#} - H_{q(q-2)+1,\#} & 0 & V_{q-1}^*\Gamma_q/2 \\ 0 & -H_{qq,qq} & W_q^*\Gamma_q/2 \\ \Gamma_q V_{q-1}/2 & \Gamma_q W_q/2 & H_{q(q-1)+1,\#} - \Gamma_q \end{pmatrix} < 0
$$

It can be seen that for any set of diagonal positive definite $D_{i,j}$ satisfying (14) there always exists an appropriate set of diagonal positive definite $H_{i,i}$ and Γ_i, but not vice versa since the total number of $H_{i,i}$ and Γ_i is larger than the total number of $D_{i,j}$. Thus, (14) is a special case of (3). In general, matrix H in (3) may have more complicated structure (e.g., blocks $H_{i,j}$ may be nondiagonal), and the matrix Γ in (3) may be nondiagonal. Hence the criterion based on Theorem 1 is stronger than that in [1].

Another stability criterion published in [1] (diagonal dominance criterion, see p. 128) does allow the use of nondiagonal matrices D, but it is handicapped due to a mistake in the proof. To prove positive definiteness of a matrix $N^{-1}SN$, where S is a symmetric matrix, and N is a diagonal positive definite matrix, the authors have only proven that all eigenvalues of $N^{-1}SN$ lie in the right half plane. This is not sufficient because matrix $N^{-1}SN$ may be *nonsymmetric*. Indeed, consider the matrix

$$S = \begin{pmatrix} 1 & 4 \\ 4 & 20 \end{pmatrix}, \quad N = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

then

$$G = N^{-1}SN = \begin{pmatrix} 1 & 1 \\ 16 & 20 \end{pmatrix}$$

which has only positive real eigenvalues. But it is not positive definite since, e.g., $x^*Gx < 0$ for vector $x = \mathrm{col}(-5, 1)$. In fact, the proof in [1] only holds for the case of $N = \delta I$, where δ is a positive number.

A stability criterion is proven in [2] for the system

$$x^{k+1} = \tanh(Wx^k) \tag{18}$$

which describes the dynamics of a fully recurrent layer with m nodes. This stability criterion is summarized below:

A system (18) is stable if there exists a Hermitian, positive definite matrix H and a symmetric matrix $\Gamma = \{\gamma_{jk}\}$ such that the following conditions hold.

1) An inequality $\xi^*H\xi - x^*Hx + \xi^*\Gamma(Wx - \xi) < 0$ holds for all $(x, \xi) \neq 0; \xi = \tanh(Wx)$.
2) The following inequalities hold for all $j = 1, \ldots, m$:

$$\gamma_{jk} < 0 \quad \text{for } j \neq k, \quad \sum_{k=1}^{m} \gamma_{jk} > 0 \tag{19}$$

Now we show that this criterion is also a particular case of the criterion of Theorem 1.

The monotone property of the function $\tanh$ implies the inequality (for all $i, j = 1, \ldots, m$)

$$F_{ij} = (\xi_i - \xi_j)(W_i x - W_j x - (\xi_i - \xi_j)) \geq 0 \tag{20}$$

where W_j is row number j of the matrix $W; \xi_j = \tanh(W_j x)$. Adding these inequalities with nonnegative weights $(-\gamma_{ij} = -\gamma_{ji})$ [due to (19)] to nonnegative $\sum_{j=1}^{m} \xi_j (W_j x - \xi_j)(\sum_{k=1}^{m} \gamma_{jk})$ we obtain

$$F = \xi^*\Gamma(Wx - \xi) \geq 0. \tag{21}$$

Hence the criterion [2] is also a special case of the criterion of the Theorem 1, with a particular form for the matrix Γ.

Another approach to stability analysis of RNN is offered in [3]. A neural network with recurrence should be written in a form of linear differential inclusions [11]. The overall complexity of the method is on the order of 2^m, where m is the number of nodes in the network. For illustration, we discuss it for the neural network (18). The following stability criterion has been proven.

If there exists a Hermitian positive definite matrix H such that the following inequalities hold:

$$W^*M_i^*HM_iW - H < 0 \quad \text{for all } i = 1, 2, \ldots, r \tag{22}$$

then the system (18) is Lyapunov stable. (A method to solve (22) for a unique H is proposed in [4].)

Here M_i are various matrices reflecting the sector bounds on nonlinearities. In the worst case, the number of coupled inequalities r is equal to 2^m. However, it can be made much smaller than 2^m, if the system is essentially a feedforward neural network with one or several time-delayed feedback loops (as in the NARX recurrent neural networks [12]). Nevertheless, the necessity of constructing a large number of matrices M_i and then choosing the "right" set of r matrices (called vertex matrices) makes the approach too cumbersome and not practical for networks with several layers and even a moderate number of nodes.

Inequality (3) with a sector condition implied in (22) can be written as

$$\xi^*H\xi - x^*Hx + \xi^*\Gamma(Wx - \xi) < 0 \quad \text{for all } (x, \xi) \neq 0$$

where $\Gamma = \mathrm{diag}\{\gamma_{jj}\} > 0$, or

$$\begin{pmatrix} H - \Gamma & W^*\Gamma/2 \\ \Gamma W/2 & -H \end{pmatrix} < 0. \tag{23}$$

This is a simple LMI to be solved for $H > 0$ and $\Gamma > 0$. Its size is $2m \times 2m$, which is clearly advantageous compared to the need to solve up to 2^m inequalities of the size $m \times m$ each in (22).

As in the case of the criterion of [1], if the criterion (22) is met, then the system (18) is stable with any transfer function (not just $\tanh$) such that its plot lies in the sector $[0, 1]$. In contrast, the criterion in [2] is a little more specific on the properties of nonlinearities it accounts for (monotonicity); therefore it is less restrictive than the two other criteria discussed.

In this section we have shown that three recently proposed criteria for stability of RNN are special cases of the criterion based on Theorem 1. Our approach uses Theorem 1 and (3) directly and in a form more general than any of the stability criteria published thus far (to the best of our knowledge). In the next section we extend our approach to a much broader group of RNN, namely those with nonzero biases. We also show how to incorporate the criterion of [2] into our approach.

V. Accounting for Nonzero Biases in Stability Analysis of RNN

In this section we show how to address stability analysis of neural networks containing biases [$b_j \neq 0, j = 1, \ldots, n$, in (1)]. To the best of our knowledge, proper accounting for nonzero biases and its consequences for stability analysis have not been given adequate consideration in previously published studies including those discussed in Section IV. It has been observed that ignoring biases [$b_j = 0, j = 1, \ldots, n$, in (1)] not only severely limits the mapping capabilities of neural networks (to the point of rendering them useless) but also almost always results in extremely conservative stability criteria. As it turns out, incorporating biases into our approach to stability analysis is easy and natural and allows us to analyze stability of neural networks with biases the same way it is done for the networks without biases.

Consider an RNN in the form (2) with nonzero biases

$$x^{k+1} = Ax^k + B\xi^k, \quad \sigma^k = \Theta x^k + b, \quad \xi^k = \tanh(\sigma^k). \tag{24}$$

Let z be its equilibrium point, where

$$z = Az + B\tanh(\Theta z + b).$$

Consider the affine transformation $y = x - z$. Denote $c = \Theta z + b$. Then

$$\begin{aligned} y^{k+1} &= Ay^k + B\eta^k, \quad \eta^k = \tanh(\sigma^k + c) - \tanh(c) \\ \sigma^k &= \Theta y^k. \end{aligned} \tag{25}$$

System (25) has the same form as system (24), but an equilibrium point of this system is at the origin. The components of the nonlinear vector function η

$$\eta_j(\sigma_j) = \tanh(\sigma_j + c_j) - \tanh(c_j)$$

are different if coordinates c_j are different. Furthermore, all new nonlinear functions η_j are still monotone and their plots lie in the sector $[0, 1]$.

Hence we can apply the stability criterion of Theorem 1 to (25) with quadratic form (21). It is remarkable that, for nonzero biases b_j, the widths of the sectors for each function η_j are often substantially reduced. Indeed, the upper bound of a sector, where the plot of a function $\varphi_j(s) = \tanh(s + c_j) - \tanh(c_j)$ lies is equal to unity only if $b_j = 0$. In *all* other cases, it is less than unity.

Let us introduce the notation

$$\mu_j = \max\{\varphi_j(s)/s : s \neq 0\}.$$

The derivative of the function $\varphi_j(s)/s$ has at least one root. Hence, given c_j, it is easy to find μ_j.

Denote $M = \operatorname{diag}\{\mu_j\}$, then for the system (25) we have the following local quadratic constraint:

$$\eta^*\Gamma(M\Theta y - \eta) \geq 0 \tag{26}$$

where Γ is any positive definite diagonal matrix. Condition 2 of the Theorem 1 and (3) changes correspondingly.

It is easy to see that $\mu_j \to 0$ if $|c_j| \to \infty$. Therefore for large numbers $|c_j|$ the quadratic form (26) becomes close to $-\eta^*\Gamma\eta$, and the inequality (3) has a solution. Therefore, it is always possible to stabilize the system (25) by choosing biases such that the values $|c_j|$ are sufficiently large. Of course, in such a case the equilibrium z will be far from the origin of the coordinate system associated with x.

Now we consider the lower bounds of the sectors. For all cases discussed thus far this bound was at zero. However, we show that the slope of this line may be increased, which should improve the stability criteria. Indeed, according to formulas (24) and particular form of matrices A, B, it follows that the absolute values of each coordinate of the vector x^k are less or equal to one. Therefore one can obtain

$$|\sigma_j + c_j| \leq \sum_{i=1}^{n} |\Theta_{ji}| + |b_j| = r_j \tag{27}$$

for all $j = 1, \ldots, n$. The following lemma establishes new lower bounds for the sectors, where plots of an effective part of nonlinear functions φ_i reside.

Lemma 1: If $|s + c_j| \leq r_j$, then

$$\varphi_i(s)/s \geq \nu_j = (\tanh(r_j) - \tanh(|c_j|))/(r_j - |c_j|) \tag{28}$$

(if $|c_j| = r_j$ then $\nu_j = d(\tanh(s))/ds$ at $s = r_j$).

Proof of Lemma 1: If $c_j = 0$, then the result follows from the monotonic property of the odd function $\varphi_j(s)/s$ on the interval $(0, \infty)$.

Let $c_j > 0$ (the case $c_j < 0$ is similar) and $c_j < r_j$. Then $\varphi_j(s)/s \geq \nu_j$ for $s \in (0, r_j - c_j]$ because this function is decreasing monotonically on this interval. The function $\varphi_j(s)/s$ has a single maximum on the interval $(-\infty, 0)$. Denote the argument of this maximum by s_j. On the interval $[s_j, 0)$ the function $\varphi_j(s)/s$ is increasing, therefore $\varphi_j(s)/s \geq \nu_j$ if $s \in [s_j, 0)$. On the interval $[-\infty, s_j]$ the function $\varphi_j(s)/s$ is decreasing. Therefore, $\varphi_j(s)/s \geq \varphi_j(-r_j - c_j)/(-r_j - c_j)$ if $s \in [-r_j - c_j, s_j]$. We have $\tanh(r_j)/r_j \leq \tanh(c_j)/c_j$ because $r_j > c_j$. Hence,

$$\begin{aligned} &(r_j + c_j)[\tanh(r_j) - \tanh(c_j)] \\ &\quad \leq (r_j - c_j)[\tanh(r_j) + \tanh(c_j)] \end{aligned}$$

and

$$\frac{\tanh(r_j) - \tanh(c_j)}{r_j - c_j} \leq \frac{\tanh(r_j) + \tanh(c_j)}{r_j + c_j}$$

or

$$\nu_j \leq \frac{\tanh(r_j) + \tanh(c_j)}{r_j + c_j}.$$

Thus,

$$\varphi_j(s)/s \geq \nu_j \quad \text{for all } s \in [-r_j - c_j, r_j - c_j].$$

The case $r_j = c_j$ is obtained as a limit if we consider $c_j < r_j, c_j \to r_j$. Lemma 1 is thus proven.

Denote $N = \operatorname{diag}\{\nu_j\}$, then for all trajectories of the system (25), all $k \geq 1$ and positive definite diagonal matrix Γ the following inequality holds:

$$(\eta^k - N\Theta y^k)^*\Gamma(M\Theta y^k - \eta^k) \geq 0. \tag{29}$$

This local quadratic constraint leads to a better stability criterion than that for (26) due to positive definiteness of the matrix N.

Fig. 1 illustrates what happens with a plot of the nonlinearity due to nonzero bias, and it also shows the sector bounds.

Let us now discuss what effect $c_j \neq c_i$ for $j \neq i$ has on the criterion proposed in [2] (see Section IV). The local quadratic constraint (21) with a nondiagonal (filled) matrix Γ, satisfying condition (19), in this case is not valid, because (20) is not satisfied if $c_j \neq c_i$. These inequalities, however, may be slightly changed to become valid. Specifically, a symmetric matrix $G = \{g_{ij}\}$ replaces the matrix Γ of (19), and

$$\eta^* G(\Theta y - \eta) \geq 0. \tag{30}$$

The matrix G is such that, for all $j = 1, \ldots, m$

$$g_{ij} = -\alpha_{ij} \quad \text{for } i \neq j, \quad g_{ii} - \sum_{j=1, j\neq i}^{m} \beta_{ij}\alpha_{ij} > 0 \tag{31}$$

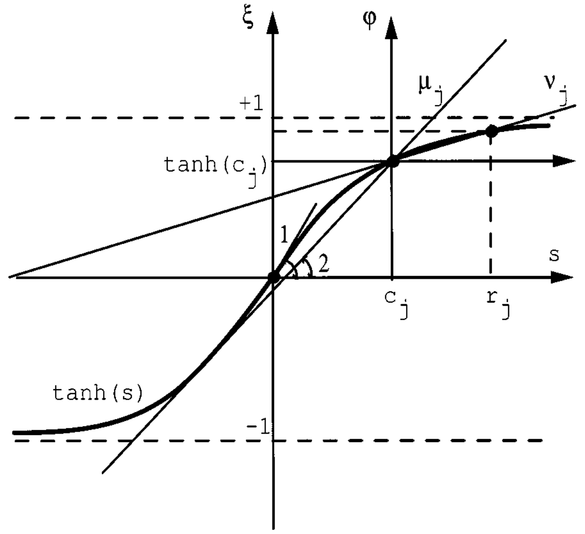

Fig. 1. The nonlinear function $\tanh(s)$ is shown in the original system of coordinates (s, ξ) with the upper bound of the sector denoted as the angle 1. Due to a nonzero bias b_j, the new coordinate system (s, φ) (its origin is at $(c_j, \tanh(c_j))$) features a smaller upper bound μ_j denoted as the angle 2. In addition, a graphical justification for ν_j of (28) is provided.

where $\alpha_{ij} = \alpha_{ji} \geq 0$, and $\beta_{ij} \geq 1$ are some numbers determined by c_i and c_j (see Appendix A). The nonequal biases result in increasing diagonal elements of a matrix G as compared to the case of systems without biases, i.e., the matrix G becomes even more diagonally dominant. If $c_j \equiv \text{const}$ (and, in particular, if $c_j \equiv 0$) we have $\beta_{ij} \equiv 1$ and the quadratic form (30) with conditions (31) coincides with the quadratic form (20) with conditions (19) (for $\Theta \equiv W$).

Thus, our approach for stability of RNN without biases is also applicable to RNN with biases, but with modifications pointed out in this section. □

VI. Steps of Our Approach

In this Section, we summarize the steps of our approach for stability analysis of RNN.

1) The dynamics of RNN should be described by the equations of the form (24). One can analyze the structure of RNN and apply the state-space extension method as described in Section III. The equations obtained after this step has a form suitable for stability analysis based on the absolute stability theory.
2) It is necessary to find out a equilibrium point z of the system obtained in Step 1. (Of course, we can only *hope* that the equilibrium point found is stable, before the stability analysis has been completed *and* conclusively confirmed its stability.) If the stationary point of RNN is determined during training, it becomes an equilibrium point of the system (24) taking into account the extended state space. Otherwise, one can use a simple procedure of calculating a few trajectories of the system until the state vector converges to the equilibrium point.
3) The state vector should be shifted in such a way that the equilibrium point moves to the origin. The nonlinear transfer functions of nodes should be altered accordingly (see Section V). This shift results in a vector c [see (25)].
4) For each transformed transfer function [which has a form $\varphi(s) = \tanh(s+c) - \tanh(c)$] it is necessary to calculate the upper bound of a sector in which the plot of this function lies. It may be done using the MATLAB's function fzero for calculating the (single) zero of a derivative of the slope of this function. Namely, for the function φ, it is sufficient to find a root of the scalar function $g(s) = s \cdot (1-(\tanh(s+c))^2) - \tanh(s+c) + \tanh(c)$. Then one can proceed with calculating the lower bounds. For each transfer function $\tanh$ it is necessary to find the maximum value of an argument utilizing the fact that the output of $\tanh$ lies in the interval $[-1, 1]$. We can use Lemma 1 (Section V) to find out the lower bounds for all functions φ.
5) One can construct the sector quadratic form [like that in (29)]. Then, using information about the independent processes in the extended system (11) (see Section III), one can construct (3) in a matrix form. It should be emphasized that the set of n independent processes, counterparts the original system (1), subsumes all possible sequences in (11). Hence establishing global stability for all the processes (by solving LMI) means that both (1) and (11) are proven to be stable. We need to show stability for n processes, with their appropriate state vectors constituting the (total) state vector of system (11). Interactions between different processes are absent due to independence of the processes (as discussed in Sections III and VII-B). This allows us to zero out blocks in the matrix H which correspond to the parts of the state vector of (11) from different processes. Setting such blocks to zero considerably reduces the number of unknown parameters without adding any conservatism to our stability criteria. In general, for a system with n layers we replace one matrix of $n(n+1) \times n(n+1)$ blocks by n matrices of $(n+1)\times(n+1)$ blocks. (An example of such replacement is given in Section VII-B.) We obtain the linear matrix inequalities with respect to the components of matrices H and Γ. It is also useful to add the inequality $\Gamma > I$.
6) We can use the interior point method to solve linear matrix inequalities. To analyze stability of a particular RNN, one can use the MATLAB LMI Control Toolbox [10]. If there is a solution to the matrix inequalities, then the system (24) (and therefore the corresponding RNN) is stable. Moreover, we can obtain a Lyapunov function for this system. It is sufficient to extract those blocks from the matrix H which correspond to the parts of the state vector of the same process. A quadratic form built using these blocks becomes a Lyapunov function for this RNN. For different processes we obtain different quadratic forms. For global asymptotic stability, the equilibrium point of RNN must be asymptotically stable in the small (and it is easy to check in simulations); therefore all different quadratic forms (Lyapunov functions) should be positive definite. It should be noted that for an RMLP with n layers and global feedback all n Lyapunov functions can be obtained simultaneously in a single run of the LMI solver.

Stability of an RMLP without global feedback may be analyzed layer by layer, and all Lyapunov functions (one for each layer) can be obtained in independent runs of the LMI solver. For example, for an RMLP with two fully recurrent hidden layers without global feedback one should first determine whether the first (preceding) hidden layer is stable. This is done by running the LMI solver to solve LMI for this layer alone. If a solution is feasible, then we can analyze stability of the second (succeeding) layer. Due to (generally) nonzero connection matrix between the first and the second layers and nonzero outputs of the first layer at its equilibrium, one should adjust biases of the second layer accordingly before invoking the LMI solver for the second layer. If a solution for the first layer is not feasible, then nothing can be said about stability of the second layer either. Generalization to cases with more than two layers is straightforward.

The steps 1)–5) may be done for any RNN, and it is possible to automate them. But final LMI [see step 6)] may still have no solutions. It does not mean that RNN is unstable, but it clearly demands application of a more sophisticated stability criterion. In such a case, we first calculate the numbers β_{ij} in the inequalities (A1) (see Appendix A). Then we can construct an additional quadratic form (30) with constraints (31). We use the sum of forms (29) and (30) in (3) to get new LMI. New constraints become less restrictive than previous ones, and LMI may have a feasible solution.

The fundamental limitation of all methods based on theory of absolute stability is in its lack of specificity on information about nonlinear functions which is used to derive stability criteria. The criterion which uses only the sector quadratic constraint (21), the criterion in [1], and the circle criterion remain valid even if each nonlinear function is time-varying, with its plot lying in the sector $[0,1]$. The criterion with a filled matrix Γ, and the criterion in [2] utilize only the monotonic property of the transfer function $\tanh$. If any of these stability criteria is met, then the system (24) with arbitrary monotone functions φ such that $d\varphi(s)/ds \in [0,1]$ for all s is stable. In particular, all (or some of) the nonlinear functions of nodes may be replaced by linear functions residing in the same sectors. Therefore if any of the criteria mentioned above holds, then a RNN with linear functions for any number of nodes or a linearized RNN would also have to be stable. But sometimes *none* of the stability criteria may be satisfied. It happens when a linearized system corresponding to (24) is not stable for some linear functions satisfying the quadratic inequalities. It mainly takes place for a set of linear functions corresponding to the upper bounds of the sectors. In any case, the first (linear) approximation of the right-hand side of the system (24) at the origin (Jacobian) should be stable, i.e., should have all its eigenvalues within the unit circle. (This may also be used to check if the system (24) and the corresponding LMI are correctly constructed).

VII. Examples

In this section, we would like to illustrate applications of our approach. First, we analyze the stability of individual layers of RNN. Next, we show how to deal with stability analysis of two-layer RNN with global feedback.

A. One-Layer System

In this section, we analyze stability of a simple RMLP with one hidden layer without global feedback. The network has two recurrent nodes of bipolar sigmoids (nonlinearity $\tanh(\lambda x)$, where $\lambda = 1/2$). The weight matrix W and the bias vector b are

$$W = \begin{pmatrix} 0.0333 & -0.0355 \\ 1.1882 & -2.2687 \end{pmatrix}, \quad b = \begin{pmatrix} -1.0092 \\ 3.5970 \end{pmatrix}.$$

The candidate equilibrium point is located at $z = \mathrm{col}(-0.4809, 0.6501)$. We write the LMI similar with the expression (23)

$$\begin{pmatrix} H - \Gamma - G & \frac{((M+N)W)^*\Gamma + \lambda W^* G}{2} \\ \frac{\Gamma(M+N)W + \lambda G W}{2} & -H - W^* N^* \Gamma M W \end{pmatrix} < 0 \qquad (32)$$

where $\Gamma > 0, G$ is a symmetric matrix satisfying (31), M is a diagonal matrix reflecting upper bounds of the sectors, and N is a diagonal matrix of lower bounds of the sectors (see Section V). Such assignment for the matrices M and N is called assigning the full sector. The matrices M and N are $\mathrm{diag}\{0.4688, 0.4383\}$ and $\mathrm{diag}\{0.3816, 0.0633\}$, respectively (the largest possible upper bound corresponds to 0.5, and it is equal to the slope of the bipolar sigmoid at zero).

The LMI is easily solved with positive definite matrix H even with $G = 0$. Thus, we prove that the equilibrium point is globally exponentially stable. It is interesting that solutions of the LMI are also feasible for $N = 0$, suggesting that stability holds with a wide margin. However, the two-node network is not stable if its bias weights are set to zero (one of the two eigenvalues of the matrix $W/2$ is outside the unit circle), and the LMIs solution is clearly infeasible in this case. This confirms the importance of taking into account nonzero biases for stability analysis.

B. Two-Layer System

In this section, we analyze stability of an equilibrium of the two-layer system introduced in Section III. According to (25), we write the following LMI:

$$(Ay + B\xi)^* H (Ay + B\xi) - y^* H y + (M\Theta y - \xi)^* \Gamma (\xi - N\Theta y) < 0, \quad \Gamma > 0 \qquad (33)$$

where $y = \mathrm{col}(y_{11}, y_{12}, y_{21}, y_{22})$, and $M = \mathrm{diag}(M_j), N = \mathrm{diag}(N_j), j = 1, 2$, are positive definite diagonal matrices specifying the sector constraints ($N_j \leq M_j$). Given matrices A, B, Θ, M and N, we would like to solve it for positive definite matrices H and diagonal $\Gamma = \mathrm{diag}\{\Gamma_j\}$. The system (6) (and, consequently, the original system (5)) is considered Lyapunov stable globally, if the LMI above have a solution.

In order to determine if the LMI have a solution using the MATLAB LMI Toolbox, one can write the first LMI as one large matrix L for the quadratic form $\mathrm{col}(y,\xi)^* L\, \mathrm{col}(y,\xi) < 0$. Here we do not provide the expression for L. Instead, we illustrate a more appealing alternative. We begin by spelling out the term with Γ in the first inequality of (33)

$$\begin{aligned}(M\Theta y-\xi)^*\Gamma(\xi-N\Theta y)\\ &=(M_1(W_1y_{12}+V_{21}y_{21})-\xi_1)^*\Gamma_1(\xi_1-N_1\\ &\quad\times(W_1y_{12}+V_{21}y_{21}))+(M_2(W_2y_{22}+V_{12}y_{11})-\xi_2)^*\Gamma_2\\ &\quad\times(\xi_2-N_2(W_2y_{22}+V_{12}y_{11})).\end{aligned}\tag{34}$$

We can see that vector pairs (y_{12},y_{21}) and (y_{11},y_{22}) do not have cross terms, which reflects independence of the two systems (8) and (9), as discussed in Section III. This suggests that the structure of matrix H can be simplified. The symmetric matrix H has a block form $H=\{H_{lj}\}_{l,j=1}^4$, and we can make half of its blocks equal to zero matrices to remove all the cross terms that do not appear in (34). Instead of a single 6×6-block matrix L, we end up with two 3×3-block matrices L_1 and L_2 for the quadratic forms $\mathrm{col}(y_{12},y_{21},\xi_1)^*L_1\mathrm{col}(y_{12},y_{21},\xi_1)<0$ and $\mathrm{col}(y_{11},y_{22},\xi_2)^*L_2\mathrm{col}(y_{11},y_{22},\xi_2)<0$, respectively, as shown in (35) and (36) at the bottom of the page. Here both matrices are symmetric, so we chose to denote all appropriate blocks located in the upper triangles by the symbol $(*)$.

Decomposing the original LMI in two sets of LMI coupled through blocks of the matrix H is a promising implementation alternative. [In fact, existence of a solution for the coupled LMI (35) and (36) is the necessary condition for solvability of (33).] Based on our experience with the MATLAB LMI Toolbox, such decomposition appears to be the only way one can implement and solve the appropriate LMI for RNN with many nodes using this Toolbox. How many nodes depends on the RNN architecture in each particular case, but a three-layer RNN with five nodes in each layer is a good ballpark to begin using the decomposition.

Before invoking the LMI solver for (35) and (36), we need to specify the matrices M and N. Recall that M is a diagonal matrix reflecting upper bounds of the sectors, where all nonlinearities reside, whereas N is a diagonal matrix of lower bounds of the sectors (see Section V).

When the LMI solver finishes up returning a feasible solution, we obtain two Lyapunov functions $\mathcal{P}_1$ and $\mathcal{P}_2$, each of which corresponds to one of the two systems (8) and (9). The functions are $\mathcal{P}_1=\mathrm{col}(y_{12},y_{21})^*P_1\mathrm{col}(y_{12},y_{21})$ and $\mathcal{P}_2=\mathrm{col}(y_{11},y_{22})^*P_2\mathrm{col}(y_{11},y_{22})$, where

$$P_1=\begin{pmatrix}H_{22}&H_{23}\\H_{32}&H_{33}\end{pmatrix},\quad P_2=\begin{pmatrix}H_{11}&H_{14}\\H_{41}&H_{44}\end{pmatrix}.$$

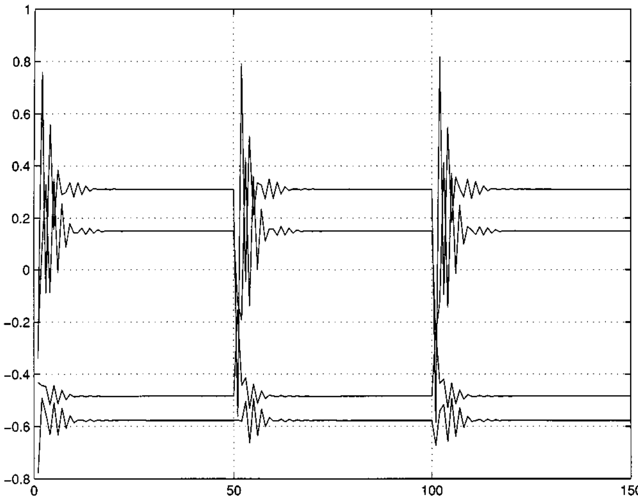

Fig. 2. Three typical trajectories for the two-layer RNN of Section VII-B. All state variables are randomly initialized at $t=0, t=50$ and $t=100$. They exhibit oscillations followed by convergence to the equilibrium point.

We choose the two-layer system (5) with two $\tanh(x/2)$ nodes in each layer. The weight matrices are shown below

$$W_1=\begin{pmatrix}-1.3482&-1.8825\\-0.7464&-0.5695\end{pmatrix},\quad W_2=\begin{pmatrix}0.4904&-0.7599\\-1.4697&-1.4608\end{pmatrix}$$
$$V_{12}=\begin{pmatrix}2.0000&0.2000\\-0.5000&1.4000\end{pmatrix},\quad V_{21}=\begin{pmatrix}1.5000&0.5000\\1.2000&-0.1000\end{pmatrix}.$$

The bias vectors for the first and the second layers are $\mathrm{col}(0.3000,-0.5000)$ and $\mathrm{col}(-1.0000,1.0000)$, respectively. This system is suspected to be stable, with the candidate equilibrium point at $z_1=\mathrm{col}(0.1493,-0.4851)$ for the first layer and $z_2=\mathrm{col}(-0.5775,0.3103)$ for the second layer. Fig. 2 illustrates a few typical transient processes all converging to the same equilibrium point. We calculate the upper and lower bounds for the sectors of all four nonlinearities, i.e., the matrices M_1, M_2, N_1 and N_2. We determine that $M_1=\mathrm{diag}\{0.4972,0.4682\}, M_2=\mathrm{diag}\{0.4532,0.4876\}, N_1=\mathrm{diag}\{0.1612,0.2091\}$, and $N_2=\mathrm{diag}\{0.1275,0.1318\}$. As in the previous example, the matrices N_1 and N_2 are obtained using Lemma 1 with Θ_1 and Θ_2 from (7).

We invoke the LMI solver of MATLAB LMI Toolbox to solve (35) and (36) for $H=\{H_{lj}\}_{l,j=1}^4$ and $\Gamma_1>I$ and $\Gamma_2>I$. The solver returns a feasible solution which proves global exponential stability of this RNN.

We wish to examine whether a feasible solution of (35) and (36) also exists for expanded sectors. First, we consider the case when $N_1=\mathrm{diag}\{0.0,0.0\}$ and $N_2=\mathrm{diag}\{0.0,0.0\}$ (the lower

$$\begin{pmatrix}(-H_{22}-W_1^*M_1^*\Gamma_1N_1W_1)&(*)&(*)\\(-V_{21}^*N_1^*\Gamma_1M_1W_1-H_{32})&H_{44}-H_{33}-V_{21}^*N_1^*\Gamma_1M_1V_{21}&(*)\\(\Gamma_1(M_1+N_1)W_1/2)&H_{14}+\Gamma_1(M_1+N_1)V_{21}/2&H_{11}-\Gamma_1\end{pmatrix}<0\tag{35}$$

$$\begin{pmatrix}(H_{22}-H_{11}-V_{12}^*M_2^*\Gamma_2N_2V_{12})&(*)&(*)\\(-W_2^*M_2^*\Gamma_2N_2V_{12}-H_{41})&-H_{44}-W_2^*M_2^*\Gamma_2N_2W_2&(*)\\(H_{32}+\Gamma_2(M_2+N_2)V_{12}/2)&\Gamma_2(M_2+N_2)W_2/2&H_{33}-\Gamma_2\end{pmatrix}<0.\tag{36}$$

bounds of the sectors are at their minimum). It turns out that no feasible solution exists in this case, although a feasible solution can be found when only N_1 is a zero matrix. Next, we consider what happens when both N_1 and N_2 remain as before (determined from Lemma 1), but the upper bounds of the sectors are raised to their maximum levels, $M_1 = \mathrm{diag}\{0.5, 0.5\}$ and $M_2 = \mathrm{diag}\{0.5, 0.5\}$. We find that, as in the first case, no feasible solution exists when both M_1 and M_2 are at their maximum. We are able to obtain a solution for $M_1 = \mathrm{diag}\{0.5, 0.5\}$ keeping $M_2 = \mathrm{diag}\{0.4532, 0.4876\}$ as before. This example illustrates importance of having proper bounds of the sectors N and M for stability analysis.

Concluding this section, we wish to point out what changes one would have to make to (35) and (36) in order to apply the combined criterion utilizing both (29) and (30) to stability analysis. The last row of (35) should be replaced with $\Gamma_1(M_1 + N_1)W_1/2 + G_1W_1/4, H_{14} + \Gamma_1(M_1 + N_1)V_{21}/2 + G_1V_{21}/4$ and $H_{11} - \Gamma_1 - G_1$ (from left to right), and the last row of (36) is to be replaced with $H_{32} + \Gamma_2(M_2 + N_2)V_{12}/2 + G_2V_{12}/4, \Gamma_2(M_2 + N_2)W_2/2 + G_2W_2/4$ and $H_{33} - \Gamma_2 - G_2$ (from left to right). Additional constraints (31) on elements of symmetric nondiagonal matrices G_1 and G_2 must be respected as well.

VIII. Conclusion and Future Directions

In this paper we proposed a new approach for stability theory of RNN based on classical results of the theory of absolute stability. This approach features a state space extension method and the use of LMI. Our approach accounts for nonzero biases usually present in RNN for improved approximation properties. Accounting for nonzero biases is done naturally, and it aids to satisfy the stability criteria. This is another important feature of our approach missing from pertinent results published elsewhere. We showed that those results are special cases within our approach. Though all our examples deal with stability analysis of RMLP, the approach developed can be readily applied to other forms of time-lagged recurrent neural networks. [Such a network differs from RMLP by allowing a more general pattern of connections between its nodes (see [6]).] Indeed, a given time-lagged recurrent neural network may be analyzed for stability within our approach as long as its equations can be transformed into the form (2).

We emphasize the fundamental difficulty with all modern stability criteria. All of these are based on a reduction of a given nonlinear system to a linear system and judging stability of that linear system instead. Frequently, a full reduction of such kind is not possible due to severe performance degradation of the closed-loop system. After all, we resort to nonlinear systems because only they deliver the required performance sought in the first place! To achieve a proper performance, a neural network usually needs to use as many of its nonlinearities as possible. Furthermore, it uses as much information about nonlinearities as available, e.g., a bipolar sigmoid as a very specific kind of nonlinearity. In contrast, current stability methods can not distinguish between any two types of nonlinearities as long as these nonlinearities belong to the same sector, monotone and time-invariant. One avenue for future research is to develop methods enhancing discriminatory capabilities of the stability criteria for RNN, i.e., to permit them to distinguish between nonlinearities of different kinds, thereby mitigating their conservatism.

We showed that stability analysis of trained recurrent neural networks is done via solving a set of LMI. A solution for the full sector is frequently infeasible, but feasibility of solution for a very narrow sector is almost always guaranteed for all locally stable RNN (e.g., for those discussed in [6]). This follows directly from the necessary condition for asymptotic stability (local stability of an RNN linearized around its equilibrium point). This also suggests that one possible way of achieving feasibility of the LMI is to keep the sector as narrow as possible, which translates into the requirement to increase biases and decrease other weights. It is clear that straightforward implementations of this requirement can be harmful to performance of RNN, and one should develop other implementations to achieve a reasonable compromise between stability and performance. While we are aware of original experiments to tradeoff performance of RNN with stability carried out in [1], details on how to incorporate stability encouragement into the EKF-based training process taking into account the approach proposed in this paper are to be clarified in future work.

Appendix A

For each pair (i, j) let β_{ij} be the minimum number such that for all numbers s_i, s_j the following inequality holds:

$$(\varphi_i(s_i) - \varphi_j(s_j))(s_i - s_j - (\varphi_i(s_i) - \varphi_j(s_j))) + (\beta_{ij} - 1) \times [\varphi_i(s_i)(s_i - \varphi_i(s_i)) + \varphi_j(s_j)(s_j - \varphi_j(s_j))] \geq 0. \quad \text{(A1)}$$

Note that $\varphi_i(s_i)(s_i - \varphi_i(s_i)) \geq 0$ for all i and s_i. The numbers β_{ij} are determined by the biases c_i, c_j and may be calculated using a simple optimization procedure. [For each pair (i, j) one has to find such β_{ij} that to satisfy (A1) in the two-dimensional region $|s_i + c_i| \leq r_i, |s_j + c_j| \leq r_j$ (see (27)]. In practice, a search on a coarse enough grid often suffices.) For $i = j$ the first term in (A1) disappears and we have $\beta_{ii} = 1$ for all i. If $c_i \neq c_j$, then $\beta_{ij} = \beta_{ji} > 1$.

The inequality (A1) may be rewritten as follows:

$$\beta_{ij}[\varphi_i(s_i)(s_i - \varphi_i(s_i)) + \varphi_j(s_j)(s_j - \varphi_j(s_j))] - \varphi_i(s_i)(s_j - \varphi_j(s_j)) - \varphi_j(s_j)(s_i - \varphi_i(s_i)) \geq 0. \quad \text{(A2)}$$

Adding these inequalities with nonnegative weights $\alpha_{ij} = \alpha_{ji}$ to nonnegative $\sum_{j=1}^{m} \eta_j(\Theta_j y - \eta_j)(\alpha_{jj} - \sum_{k=1, k\neq j}^{m} \beta_{jk}\alpha_{jk})$ we arrive at (30).

We illustrate how to obtain (30) based on (31) for a recurrent layer with two nodes. We assume $c_1 \neq c_2$, so that $\beta_{12} = \beta_{21} > 1$. As in (A2), we write

$$\beta_{12}[\eta_1(\Theta_1 y - \eta_1) + \eta_2(\Theta_2 y - \eta_2)] - \eta_1(\Theta_2 y - \eta_2) - \eta_2(\Theta_1 y - \eta_1) \geq 0.$$

We multiply this equation by nonnegative α_{12}, then add to it nonnegative $(\alpha_{11} - \beta_{12}\alpha_{12})\eta_1(\Theta_1 y - \eta_1)$ and

$(\alpha_{22} - \beta_{21}\alpha_{21})\eta_2(\Theta_2 y - \eta_2)$. Keeping in mind that $\alpha_{12} = \alpha_{21}$, the result is easily reduced to the quadratic form (30).

ACKNOWLEDGMENT

The bulk of research for this work was carried out while N. Barabanov was a visiting fellow of the SABIT program at Ford Research Laboratory. The authors would like to express our deep gratitude to organizers of this program. The authors are also very grateful to L. Feldkamp and anonymous reviewers for valuable comments.

REFERENCES

[1] J. A. K Suykens, J. P. L. Vandwalle, and B. L. R. De Moor, *Artificial Neural Networks for the Modeling and Control of Non-Linear Systems*. Boston, MA: Kluwer, 1996.

[2] Y.-C. Chu and K. Glover, "Bounds of the induced norm and model reduction errors for systems with repeated scalar nonlinearities," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 471–483, 1999.

[3] K. Tanaka, "An approach to stability criteria of neural-network control systems," *IEEE Trans. Neural Networks*, vol. 7, pp. 629–642, May 1996.

[4] L. Vandenberghe and S. Boyd, "A polynomial-time algorithm for determining quadratic Lyapunov functions for nonlinear systems," in *Proc. Europ. Conf. Circuit Theory Design*, 1993, pp. 1065–1068.

[5] N. E. Barabanov, "Frequency domain criteria for stability and instability in the large for the stationary sets of nonlinear systems of differential equations with a monotone nonlinearity," *Siberian Math. J.*, vol. XXVIII, no. 2, pp. 21–34, 1987.

[6] L. A. Feldkamp and G. V. Puskorius, "A signal processing framework based on dynamic neural networks with application to problems in adaptation, filtering and classification," *Proc. IEEE*, vol. 86, pp. 2259–2277, Nov. 1998.

[7] V. A. Yakubovich, "The method of matrix inequalities in the stability theory of nonlinear control systems," *Automat. Remote Contr.*, vol. 25–26, Apr. 1967.

[8] Y. Tsypkin, "Frequency criteria for the absolute stability of nonlinear sampled data systems," *Automat. Remote Contr.*, vol. 25, pp. 261–267, 1964.

[9] Y. Nesterov and Y. Nemirovsky, *Interior-Point Polynomial Methods in Convex Programming*. Philadelphia, PA: SIAM, 1994, vol. 13. of *Studies in Applied Mathematics*.

[10] *LMI Control Toolbox User's Guide*, Mathworks, Inc., Natick, MA, 1995.

[11] S. Boyd, L. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994, vol. 15. of *Studies in Applied Mathematics*.

[12] S. Chen, S. A. Billings, and P. M. Grant, "Nonlinear system identification using neural networks," *Int. J. Contr.*, vol. 51, no. 6, pp. 1191–1214, 1990.

Nikita Barabanov received M.S. degree and Ph.D. degree in 1976 and 1979, respectively, from Leningrad University, Soviet Union, and the Doctor of Science degree from Kiev Institute of Cybernetics, Soviet Union, in 1990.

Since 1979, he has been with St. Petersburg Electrotechnical University, where he is currently Full Professor. He is the author of more than 50 papers in peer-reviewed journals. His research interests include stability of differential and discrete equations, robust stability, H-infinity control theory, as well as signal processing and pattern recognition.

Dr. Barabanov received the Leningrad Mathematical Society Award in 1980.

Danil Prokhorov (S'95–M'97) received the M.S. degree with Honors from the State Academy of Aerospace Engineering (former LIAP), St. Petersburg, Russia, in 1992, and the Ph.D. degree in electrical engineering from Texas Tech University, Lubbock, in 1997.

Since joining the staff of Ford Research Laboratory, Dearborn, MI, in 1997, he has been studying reinforcement learning, recurrent network training algorithms, support vector machines and regularization, methods of encouraging robustness of neurocontrollers, stability of recurrent networks, unsupervised dynamic classification, and cluster-weighted modeling. His research interest is in various machine learning approaches and their applications to problems in system modeling, control, and signal processing.

Dr. Prokhorov is a recipient of the 1999 Young Investigator Award of International Neural Network Society. He acted as the IJCNN'01 Technical Program Chair.