International Journal of Pure and Applied Mathematics

Volume 39 No. 4 2007, 545-562

DYNAMIC PROPERTIES OF RECURRENT NEURAL NETWORKS AND ITS APPLICATIONS

Leong-Kwan Li[1], S. Shao[2 §]

[1]Department of Applied Mathematics
The Hong Kong Polytechnic University
Hung Ham, Kowloon, HONG KONG, P.R. CHINA
e-mail: malblkli@polyu.edu.hk

[2]Department of Mathematics
Cleveland State University
2121 Euclid Avenue, RT 1527, Cleveland, OH 44115, USA
e-mail: s.shao@csuohio.edu

Abstract: We study the dynamics of the leaky integrator recurrent neural network. Our results show that there exists at least one equilibrium point of the system, and the set of solutions of the leaky integrator recurrent neural dynamics is positive invariant and attractive. The globally exponential stability property of the system has been discussed. Our examples show that the leaky integrator recurrent neural network together with the state space search algorithm can be an effectively tool for many applications including data compression and learning the short-term foreign exchange rates.

AMS Subject Classification: 37C70, 37C75, 68T05, 68T20
Key Words: leaky integrator, recurrent neural networks, positive invariant, attractive set, global exponentially stability

1. Introduction

If brain tissue is properly stained and examined under a light microscope, it is possible to visualized the individual neuron which compose the brain that is a vast and intricate network of neurons [6]. Neural networks are trainable analytic tools that attempt to mimic information processing patterns in the brain, they can be used effectively to automate both routine and ad hoc tasks.

Received: August 1, 2007

§Correspondence author

Hence, neural networks models have been created by both scientists and engineers to provide a general class of flexible nonlinear models in which they have been successfully applied in many different fields. Since the human brain is a recurrent neural network (RNN)–a network of neurons with feedback connections, RNNs are biologically more plausible and computationally more powerful than other adaptive models such as hidden Markov models (no continuous internal states), feedforward networks and support vector machines (no internal states at all). Moreover, RNNs have richer dynamical structures, it can learn extremely complex temporal patterns to yield good results.

The simplest continuous-time model of the neuron in frequent use is the leaky integrator model, which has become popular in the simpler applications of neural networks which choose analog VLSI (very-large-scale integrated) for their implementation. The leaky integrator model uses the 'firing rate' (to mimic the biological measure of the number of spikes traversing the axon in some recent interval: but the artificial neuron need not involve any explicit spike generation) as a continuously varying output measure of the cell's activity, in which the internal state of the neuron is described by a single variable, see [1]. The leaky integrator dynamics are common in computational neuroscience and have been studied by many researchers in the field in the past few decades (see [9], [10], and [16]). In this paper, we study the dynamics of the leaky integrator RNN from the dynamical system's point of view. Our result shows that there exists at least one equilibrium point of leaky-integrator RNN system, and the set of solutions of leaky-integrator RNN system is a positive invariant and attractive set. The globally exponential stability condition of the system has been obtained. We show, by examples, that with the powerful learning algorithm– the state space search learning algorithm (SSSA), the leaky integrator RNN system can be an effective tool to perform data compression and learn short term foreign exchange rates. The experimental examples demonstrate some extremely promising results for the leaky integrator recurrent neural dynamic approach to be applied to these areas.

The organization of the paper is as follows. In Section 2, we first highlight the leaky integrator RNN model, then we present some rigorous proofs of the characters of the solutions and the stability property of leaky integrator RNN system. The approximation of the leaky integrator RNN model and the SSSA are presented in Section 3. Applications and examples are discussed in the concluding Section 4. Some final remarks are given in Section 5.

2. Dynamics of the Leaky Integrator RNN Model

Consider the continuous-time leaky integrator model of the RNNs of the system of nonlinear equations described by

$$\frac{dx_i}{dt} = -a_i x_i + b_i \sigma(\sum_{j=1}^{n} w_{ij} x_i + \theta_i)\,, \qquad i = 1, 2, 3, \ldots, n, \tag{2.1}$$

where x_i represents the internal state of the i-th neuron. $W = [w_{ij}]_{n\times n}$ is the synaptic connection weight matrix with w_{ij} being the synaptic weight of a connection from neuron n_j to neuron n_i. A =diag$[a_1, a_2, \ldots, a_n]$ and B =diag$[b_1, b_2, \ldots, b_n]$ are diagonal matrices with *positive diagonal entries.* σ is a neuronal activation function that is bounded, differentiable and monotonic increasing on [-1, 1]. We assume that $\sigma(z) = \tanh(z)$, which is the symmetric sigmoid logistic function. $\theta = [\theta_1, \theta_2, ..., , \theta_n]^T$ is the input bias or threshold vector of the system. For convenience, we rewrite system (2.1) in the following matrix form

$$\frac{dx}{dt} = -Ax + B\sigma(Wx + \theta), \tag{2.2}$$

where $\sigma = \text{diag}[\sigma_1, \sigma_2, ..., \sigma_n]$ with σ_i is defined by

$$\sigma_i(Wx + \theta) = \sigma(\sum_{j=1}^{n} w_{ij} x_i + \theta_i), \quad i = 1, 2, ..., n. \tag{2.3}$$

System (2.1) is a continuous-time model of the leaky integrator RNNs, which has been studied in many applied areas of sciences (see [9], [16], [11], [13]). However, the solution behavior of the system (2.1) was less understood from the dynamical system point of view. By studying the solution behavior and the stability property of system (2.1), we shall be better in understand the characters of the network behavior in general, which will provide more hints in network design and improving learning algorithms. In the spirit of [16], we show in the following theorems some solution behavior and the stability property of system (2.1). The first few results show that there exists at least one equilibrium point of the (2.1), the set of solutions of (2.1) is a positive invariant and attractive set.

Before we proceed to discuss the main results, we show that the RNN model (2.2) has bounded solution trajectory for any initial point in R^n.

By Peano's Local Existence Theorem (see [3]) for solutions of ordinary differential equations, given any $x_0 \in R^n$, there exists a positive number $t^*(x_0)$ such that the system (2.2) has a solution $x(t, x_0) \in R^n$ for $t \in [0, t^*(x_0))$, which is the maximal right existence interval of the solution $x(t;x_0)$ satisfying $x(0,x_0)$. Now we present our first theorem.

Theorem 1. *The solution $x(t;x_0)$ of (2.1) with initial condition $x(0,x_0) = x_0$ is bounded for all $t \in [0,t^*(x_0))$. In addition, this local existence of the solution $x(t;x_0)$ exists globally. That is, $t^*(x_0) = +\infty$ for any $x_0 \in R^n$.*

Proof. There are many ways can be employed to prove the desired results. Our proof is in the spirit of [16].

Since $f(x) = -Ax + B\sigma(Wx+\theta)$ is continuously differentiable, initial value problem of the nonlinear autonomous system of the RNN model (2.2) has the solution $x(t) = x(t;x_0)$ satisfying

$$x(t) = e^{-At}x_0 + \int_0^t e^{-A(t-s)}B\sigma(Wx(s)+\theta)ds. \tag{2.4}$$

Since $\|\sigma(y)\| = \|\tanh(y)\| \le n^{\frac{1}{2}}$, $\|B\| < M$ for some positive $M > 0$, where $\|\cdot\|$ is the Euclidean norm defined by $\|y\| = [\sum_{i=1}^n {y_i}^2]^{\frac{1}{2}}$. Let $a_{min} = \min_i\{a_i\}$, where a_i's are the entries of the diagonal matrix A in (2.2). Then we have $a_{min} > 0$ since $a_i > 0$ for all $i = 1,2,...,n$. Hence we obtain

$$\begin{aligned}\|x(t)\| &\le e^{-a_{min}t}\|x_0\| + Mn^{\frac{1}{2}}\int_0^t e^{-a_{min}(t-s)}ds \\ &\le e^{-a_{min}t}\|x_0\| + \frac{Mn^{\frac{1}{2}}}{a_{min}}(1-e^{-a_{min}t}) \\ &\le \max\{\|x_0\|, \frac{Mn^{\frac{1}{2}}}{a_{min}}\},\end{aligned} \tag{2.5}$$

for all $t \in [0,t^*(x_0))$. Therefore, the solution $x(t;x_0)$ is bounded for $t \in [0;t^*(x_0))$. By the continuation theorem of the solutions of ordinary differential equations (see [3]), the local existence of the solution $x(t;x_0)$ of the RNN model (2.2) and its bounded property guarantee the global existence of the solution. That is, $t^*(x_0) = +\infty$ for any $x_0 \in R^n$. We finish the proof. □

Now let $\Omega \in R^n$ be the solution set of (2.1) defined by

$$\Omega = [-\frac{a_1}{b_1}, \frac{a_1}{b_1}] \times [-\frac{a_2}{b_2}, \frac{a_2}{b_2}] \times, \dots, \times [-\frac{a_n}{b_n}, \frac{a_n}{b_n}], \tag{2.6}$$

where all a_i's and b_i's are defined in (2.1) and they are positive numbers.

We shall show in the following Theorem 2 that Ω is a *positive invariant* and *attractive set* of the RNN model (2.1). That is, for any solution trajectory of the RNN model (2.1) starting from Ω, the solution trajectory *cannot* escape from Ω. Moreover, for any solution trajectory of the RNN model (2.1) starting from the outside of Ω, it will converge to Ω.

Theorem 2. *Ω is a positive invariant and attractive set of the RNN model (2.1).*

Proof. We first rewrite (2.1) in the following form

$$\frac{1}{a_i}\frac{dx_i}{dt} = -x_i + \frac{b_i}{a_i}\sigma(\sum_{j=1}^{n} w_{ij}x_i + \theta_i)\,, \qquad i = 1, 2, 3, \dots, n. \tag{2.7}$$

We notice that for each $i = 1, 2, ..., n$, if $x_i \geq \frac{b_i}{a_i}$ then $\frac{dx_i}{dt} < -a_i \cdot (\frac{b_i}{a_i}) + b_i \cdot 1 = 0$ holds. Similarly, if $x_i \leq \frac{-b_i}{a_i}$ then $\frac{dx_i}{dt} > -a_i \cdot (\frac{-b_i}{a_i}) - b_i \cdot 1 = 0$ holds. Therefore, Ω is a positive invariant se of the RNN (2.1).

Given any $\epsilon > 0$, define

$$\Omega_\epsilon = [-\frac{b_1}{a_1} - \epsilon, \frac{b_1}{a_1} + \epsilon] \times [-\frac{b_2}{a_2} - \epsilon, \frac{b_2}{a_2} + \epsilon] \times \cdots \times [-\frac{b_n}{a_n} - \epsilon, \frac{b_n}{a_n} + \epsilon]. \tag{2.8}$$

It is also obvious that Ω_ϵ is an attractive and positive invariant set of the RNN (2.1). Moreover, for any solution $x(t; x_0)$ with initial point $x_0 \notin \Omega_\epsilon$, it will enter into Ω_ϵ in a finite time. Because of the positive invariant property of Ω_ϵ, $x(t; x_0)$ will stay in Ω_ϵ after a *finite* period of time. Let $\epsilon \to 0^+$, we have $\lim_{\epsilon \to 0^+} \Omega_\epsilon = \Omega$. Therefore, we can conclude that Ω is a positive invariant and attractive set of the RNN model (2.1). Theorem 2 follows. □

To provide the stability property of system (2.1) and its equivalent form (2.2), we introduce the following lemma.

Lemma 1. *Let $\beta = \max\{\frac{b_i}{a_i} | i = 1, 2, \dots, n\}$, then there exist at least one equilibrium point $y^* \in [-\beta, \beta]^n$ of (2.2) such that*

$$y^* = A^{-1}B\sigma(Wy^* + \theta) \tag{2.9}$$

for each θ.

Proof. Define

$$g_i(y) = \frac{b_i}{a_i}\sigma(\sum_{j=1}^{n} w_{ij}y_j + \theta_i) = y_i\,, \qquad i = 1, 2, ..., n. \tag{2.10}$$

For $y \in [-\beta, \beta]^n$, since $|\sigma(z)| = |\tanh(z)| \leq 1$, we have

$$y_i = g_i(y) \equiv |\frac{b_i}{a_i}||\sigma(\sum_{j=1}^{n} w_{ij}y_j + \theta_i)| \leq \frac{b_i}{a_i}\,, \qquad i = 1, 2, ..., n. \tag{2.11}$$

That implies $g(y) \in [-\beta, \beta]^n$. Since function σ is continuous, we can conclude that for any given θ and the connection weight matrix W, $g(y) : [-\beta, \beta]^n \to [-\beta, \beta]^n$ is a continuous function. By the Brouwer's Fixed Point Theorem (see [7]), $g(y)$ has a fixed point $y^* \in [-\beta, \beta]^n$ with

$$y^* = A^{-1}B\sigma(Wy^* + \theta).$$

Now we define the set $S^* = \{y^* \in R^n | y^* = A^{-1}B\sigma(Wy^* + \theta)\}$ of equilibrium points of (2.2). For simplicity, we let matrix $A = \text{diag}[a_1, a_2, ..., a_n] = I_{n\times n}$ in system 2.1), where $I_{n\times n}$ is the $n \times n$ identity matrix, then the system (2.1) becomes

$$\frac{dx_i}{dt} = -x_i + b_i\sigma(\sum_{j=1}^{n} w_{ij}x_i + \theta_i), \qquad i = 1, 2, 3, \ldots, n, \tag{2.12}$$

which can also be written as the following matrix form

$$\frac{dx}{dt} = -x + B\sigma(Wx + \theta). \tag{2.13}$$

We are now ready to present the following theorem about the stability property of system (2.12) and its equivalent form (2.13).

Theorem 3. *If W is invertible, (WB) is negative semi-definite, and if S^* is a singleton, then the leaky integrator RNN model (2.2) (or 2.1) is globally exponentially stable. That is, there exist two positive constants $p \geq 1$ and $q > 0$ such that for any $x_0 \in R^n$ and $t \in [0, \infty)$*

$$\|x(t; x_0) - x^*\| \leq p\|x_0 - x^*\| \exp(-qt), \tag{2.14}$$

where $x(t; x_0)$ is the solution of (2.13) with initial condition $x(0; x_0) = x_0$, and x^* is an equilibrium point of (2.13) in S^* with $A = I$. That is,

$$x^* = B\sigma(Wx^* + \theta). \tag{2.15}$$

Proof. We notice that if W is invertible and S^* is singleton, then RNN model (2.1) has a unique equilibrium point. Making the changing of variable $y = Wx$, the system (2.13) can be written as the following system

$$\frac{dy}{dt} = -y + WB\sigma(y + \theta), \tag{2.16}$$

with $y^* = Wx^*$ as its equilibrium point. Set another new variable $z = y - y^* = W(x - x^*)$, then (2.16) can be reduced to the system of the form

$$\frac{dz}{dt} = -z + \hat{W}f(z)\,, \tag{2.17}$$

with $z = 0$ as its equilibrium point, where $\hat{W} = WB$, $f(z) = (f_1(z_1), f_2(z_2), \ldots, f_n(z_n))^T = (\sigma(z_1 + y_1{}^* + \theta_1) - \sigma(y_1{}^* + \theta_1), \sigma(z_2 + y_2{}^* + \theta_2) - \sigma(y_2{}^* + \theta_2), \ldots, \sigma(z_n + y_n{}^* + \theta_n) - \sigma(y_n{}^* + \theta_n)) \in R^n$ for $z = (z_1, z_2, ..., z_n) \in R^n$. Since $\sigma(\rho) = \tanh(\rho)$ for all $\rho \in R$, we have

$$0 \leq [f_i(\rho)]^2 \leq \rho f_i(\rho) \leq (\rho)^2, \quad \forall \rho \in R. \tag{2.18}$$

For system (2.12) (or 2.13), the solution $x(t; x_0)$ is uniquely determined by its initial condition $x(0; x_0) = x_0 \in R^n$ for all $t \geq 0$ because the right-hand side of (2.13) is a globally *Lipschitz* continuous function. Let $z(t) = z(t; z_0) = W(x(t; x_0) - x^*)$ be the unique solution of (2.17) with the initial condition $z(0; z_0) = z_0 = W(x_0 - x^*) \in R^n$.

We now define the Lyapunov function of the generalized Luré-Postnikov type, see [2]

$$V(z) = z^T z + k \sum_{i=1}^{n} \int_0^{z_i} f_i(\rho) d\rho, \quad \forall z \in R^n \tag{2.19}$$

with $k = \lambda_{max}(\hat{W}^T\hat{W}) > 0$, where $\lambda_{max}(\hat{W}^T\hat{W})$ denotes the maximal eigenvalues of the symmetric matrix $\hat{W}^T\hat{W}$. Then we have

$$\|z\|^2 \leq V(z) \leq \|z\|^2 + k \sum_{i=1}^{n} \frac{{z_i}^2}{2} \leq s\|z\|^2, \quad \forall z \in R^n, \tag{2.20}$$

where $s = 1 + \frac{k}{2}$.

To compute the derivative of $V(z)$ along the solution $z(t)$ of (2.17), we have

$$\begin{aligned}
\frac{dV(z)}{dt} &= (2z + kf(z))^T(\frac{dz}{dt}) = (2z + kf(z))^T(-z + \hat{W}f(z)) \\
&= -2z^T z + 2z^T\hat{W}f(z) - k(f(z))^T z + k(f(z))^T\hat{W}(f(z)) \\
&\quad = -z^T z + [-z^T z + 2z^T\hat{W}f(z)] - k(f(z))^T(f(z)) \\
&\quad - k[(f(z))^T z - (f(z))^T(f(z)) - (f(z))^T\hat{W}(f(z))] \\
&\quad \leq -z^T z + (f(z))^T[\hat{W}^T\hat{W}](f(z)) - k(f(z))^T(f(z)) \\
&\quad - k[(f(z))^T z - (f(z))^T(f(z)) - (f(z))^T\hat{W}(f(z))]
\end{aligned}$$

$$\leq -z^T z + (f(z))^T[\hat{W}^T\hat{W}](f(z)) - k(f(z))^T(f(z))$$
$$\leq -\|z\|^2 + [\lambda_{max}(\hat{W}^T\hat{W}) - k]\|f(z)\|^2 = -\|z\|^2 , \quad (2.21)$$

by (2.18), $\hat{W} = WB$ is negative semi-definite and $k = \lambda_{max}(\hat{W}^T\hat{W})$ (2.20) and (2.21) imply

$$\frac{dV(z(t))}{dt} \leq -\frac{1}{s^2}V(z(t)), \quad \forall t \geq 0. \quad (2.22)$$

Solving (2.22), we have

$$V(z(t)) \leq V(z_0)\exp(-\frac{t}{s}), \quad \forall t \geq 0. \quad (2.23)$$

Therefore,

$$\|z(t)\| \leq \sqrt{s}\|z_0\|\exp(-\frac{t}{2s}), \quad \forall t \geq 0. \quad (2.24)$$

Since $z(t) = W(x(t; x_0) - x^*)$ and $z_0 = W(x_0 - x^*)$, we have

$$\|x(t; x_0) - x^*\| \leq \text{cond}(W)\sqrt{s}\|x_0 - x^*\|\exp(-\frac{t}{2s}), \quad \forall t \geq 0 , \quad (2.25)$$

where cond(W)=$[\frac{\lambda_{max}(W^TW)}{\lambda_{min}(W^TW)}]^{\frac{1}{2}} \geq 1$ is the conditional number of the matrix W. Since W is invertible, which implies the minimum eigenvalues $\lambda_{min}(W^TW)$ of the symmetric matrix W^TW greater than zero, this guarantees the well-definiteness of cond(W). Therefore, we have

$$\|x(t; x_0) - x^*\| \leq p\|x_0 - x^*\| \exp(-qt)$$

with $p = (\text{cond}W)\sqrt{s} \geq 1$, $q = \frac{1}{2s}$ and $s = 1 + \frac{\lambda_{max}[(WB)^T(WB)]}{2}$. Theorem 3 follows. □

Remarks. We notice that inequality (2.25) shows the global exponential convergence rate of the continuous-time leaky integrator RNN model (2.12). If we let matrix $A = \text{diag}[a_1, a_2, ..., a_n] = aI_{n\times n}$, then

$$\|x(t; x_0) - x^*\| \leq p\|x_0 - x^*\| \exp(-qt)$$

with $p = (\text{cond}W)\sqrt{s} \geq 1$, $q = \frac{1}{2sa}$, and $s = 1 + \frac{\lambda_{max}[(WB)^T(WB)]}{2a^2}$. If a is small, the leaky integrator RNN model (2.1) will be more sensitive to input noise and round-off errors in numerical simulations and implementation. We will show in the next section the approximation of the leaky integrator RNN model and the SSSA.

3. The Approximation of the Leaky Integrator RNN Model and the SSSA

We have presented three important characters of the solutions of system (2.1) in Section 2. It is known that the continuous-time model (2.1) (or its equivalent form (2.2)) and its numerical discretization of system (3.1) need not share the same dynamical behavior. However, the discrete-time system (3.1) of system (2.1) will inherit the same dynamics of system (2.1) when the step size is "small" (see [21]). For practical purpose, we approximate system (2.1) by Euler's method to obtain the dynamics of a discrete-time leaky integrator RNN model

$$x(t+1) = (I - HA)x(t) + HB\sigma(Wx(t) + \theta), \tag{3.1}$$

where $H = \text{diag}[h_1, h_2, ..., h_n]$ is a diagonal matrix with positive entries h_i for each i, which h_i is the step size of Euler's discretization. The two systems (3.1) and (2.2) share the same dynamical behavior when $h \to 0$. We will show in the this section a powerful learning algorithm – the state space search algorithm (SSSA), see [14], for the leaky integrator RNN model.

The SSSA is a learning algorithm for the discrete-time leaky integrator RNN (3.1). In neural networks, learning is a process of changing the network parameters so that the system output will approach to the target trajectory. By searching in the neighborhood of the target trajectory in the state space for each iteration, the SSSA performs nonlinear optimization learning process and provides the best feasible solution for the nonlinear optimization problem, see [14].

Consider a given trajectory $y(t) \in R^n$, we use (3.1) to approximate the target trajectory $y(t)$ with the error function E defined by

$$E(W, h, \theta) = \| x(t, W, h, \theta) - y(t) \|^2 = \sum_{i=1}^{n} \sum_{t=1}^{m} [x_i(t, W, h, \theta) - y_i(t)]^2 \tag{3.2}$$

for some positive integer m. Our goal is to minimize the error between the target trajectory $y(t)$ and the system output $x(t)$. To simplify the analysis, we let $A = B = I_{n\times n}, h = h_i, i = 1, 2, ..., n$. We wish to find the optimal connection weight matrix $W^* = [w_{ij}]_{n\times n}$. That is, for a fixed h and θ, we solve the nonlinear least square problem

$$E(W^*) = \min_{W^+} \sum_{i=1}^{n} \sum_{t=1}^{m} [x_i(t, W^+) - y_i(t)]^2\,, \tag{3.3}$$

where W^+ is a feasible solution defined by

$$W^+ = \sigma^{-1}[\frac{1}{h}(C_{t+1} - D_t) - D_t]{D_t}^T(D_t{D_t}^T)^{-1}, \tag{3.4}$$

where matrices $C_{t+1} = [x(t+1)..., x(2)]$ and $D_t = [x(t)..., x(1)]$.

If the network is exactly capable, that is, $E = 0$, the optimal solution W^* of (3.3) can be reached using a simple optimization strategy by vary h and repeated solve W^+ by (3.4), then we have $W^* = W^+$. If $E \neq 0$, (3.3) has no exact solution, which implies y is not reachable for this h.

The SSSA is developed to carry out the global optimization techniques to provide the best feasible solution of (3.3). The idea is that instead of moving in the parameter space of W, we search the class of the x-convex set $C_y(R^n)$, see [14], in the state space for each iteration. Here is how the SSSA defines: for each k, we define sequence of feasible points $\{Y_k\}$ by

$$Y_{k+1} = \alpha_{k+1}y + (1 - \alpha_{k+1})x(W_{k+1}), \tag{3.5}$$

with

$$\begin{aligned} & x(W_{k+1}) = Y_k, \\ & 0 < \alpha_k < 1, \\ & E(W_{k+1}) \leq E(W_k), \\ & \alpha_{k+1} = \alpha_k \quad \text{if} \quad E(W_{k+1}) < E(W_k). \end{aligned} \tag{3.6}$$

The SSSA performs nonlinear optimization learning process and provides the best feasible solution for the nonlinear optimization problem. The convergence analysis shows that the network convergence to the desired solution is guaranteed and the stability of the algorithm is depending on $\{\alpha_i\}$'s, see [14], and we check if the weight matrices W_k for each k for the stability property, that is, if ${W_k}^{-1}$ exists. In the next section, we show how the SSSA and the leaky integrator RNN work together to be applied in different fields.

4. Applications and Examples

In this section, we illustrate, by examples, how the leaky integrator RNN approach can be effectively applied to the areas of data compression and learning short-term exchange rates.

To demonstrate how the leaky integrator RNN system with the SSSA be effectively applied to different applications, we assume that matrices A and B

of (3.1) are the identity matrices for simplicity. Now we have system

$$x(t+1) = x(t) + h[-x(t) + \sigma(Wx(t) + \theta)]. \tag{4.1}$$

(A) *Signal Compression.* Although the fast Fourier transform and the wavelet transform are well-known and powerful tools available in the recent years, neural networks are also considered to be very suitable for nonlinear data processing problems because of their inherently nonlinear nature (see [8], [23], [10], [12]).

Given a signal $z(t)$ of length m (or with m data), the leaky integrator RNN system (4.1) approach for signal compression can be carried out by the following stages:

(1) Segmentation. The given signal $z(t)$ of length m is first partitioned into n equal segments with the same length p.

(2) Cycling. For each segment $z_k(t), k = 1, 2, ..., n$, we define a new sequence $\{z1_k(t)\}$ has the property of mean$(z1_k(t)) = 0$ and replace $z1_k(t)$ by $z_k(t)$ for all k in the remaining discussion.

(3) Smoothing. Because inherent in the collection of data taken over time is some form of random variation, there exist methods of smoothing for reducing of canceling the effect due to random variation. There are many different smoothing techniques. The moving average smoothing technique or double integration smoothing technique can be used to do the task. These methods are simple and easy to be implemented.

(4) Normalization. To normalize $z_k(t)$ within the range -1 to 1 of σ instead of 0 to 1, the parameters θ can be assign to be zero vector to save memory space. Again, this setting has very little effect on the efficiency of data compression, see [9].

(5) Approximation. Neurons of $x_1(t), x_2(t), ..., x_n(t)$ of the leaky integrator RNN system (4.1) with network size n can be used to generated the compressed data set $X(t)$ via (4.1) with given values of the RNN parameters h and W. By varying these RNN quantities systematically, we obtain the regenerated the given spectrum to have the smallest discrepancies with respect to the original spectrum.

(6) Optimization. As the network size n is fixed, we need to optimize the parameters W and h so that we can minimize the discrepancy between $z(t)$ and the compressed data set $X(t)$ within the tolerance. That is to minimize the mean square error

$$E = \sum_{i=1}^{n} \sum_{t=1}^{p} [\{x_i(t) - z((i-1)p + t)\}^2 / m]^{\frac{1}{2}}, \tag{4.2}$$

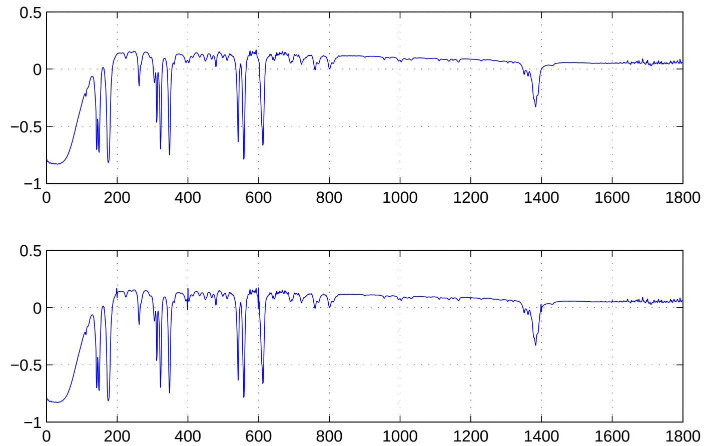

Figure 1: Brbenzen

with the SSSA. As the optimized RNN quantities together with $X(1)$ achieved as the compressed data set, it will be used to regenerate the original signal. Some error analysis of this compression method can be found in [10].

We choose four different infra-red spectra (brbenzen, cyclo6br, cl2pheno and no4pheno) to illustrate the effectiveness of this approach. The recurrent neural nets used are of size 9 with first 1,800 point each sample. The results are plotted in Figure 1, 2, 3 and 4. For each figure, the top diagram is the original signal, the bottom diagram is the recovered signal using the RNN of the dynamical system approach. The least square errors for brbenzen, cyclo5cl, cl2pheno and no4phone are 0.024306, 0.072161, 0.306229 and 0.293058, respectively. They converge within less than 5 seconds using *MATLAB 7* running in a reasonable current PC system with less than 358 iterations. We can see the results are very promising in these experiments of signal compression.

(B) *Learning the Short-term Exchange Rates.* As opposed to the traditional trading methods by human decisions, neural networks offer a simple but automatic system in the trading off foreign exchange markets. We assume that there exist short-term trends in our foreign exchange series, and we could use neural network techniques to model the short-term trend movement of the foreign exchange rates and to make predictions. Our assumption is based on the result of Yao and Tan [22], which they show that statistically the foreign exchange series do not support the random walk hypothesis [13]. In this example, we collected 701 days of data consisting of the daily quotes of four major currencies with

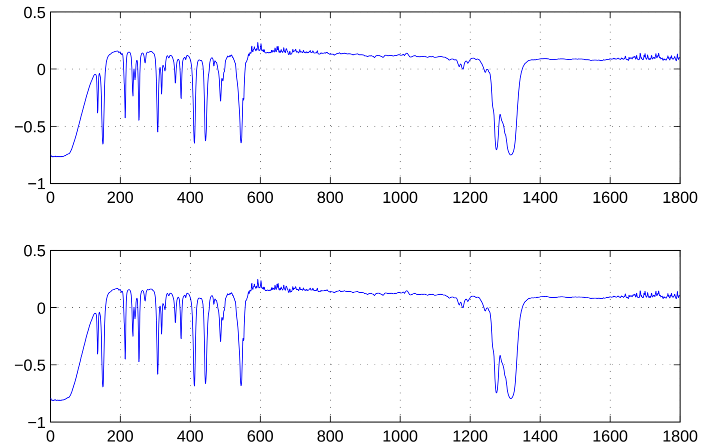

Figure 2: Cyclo6br

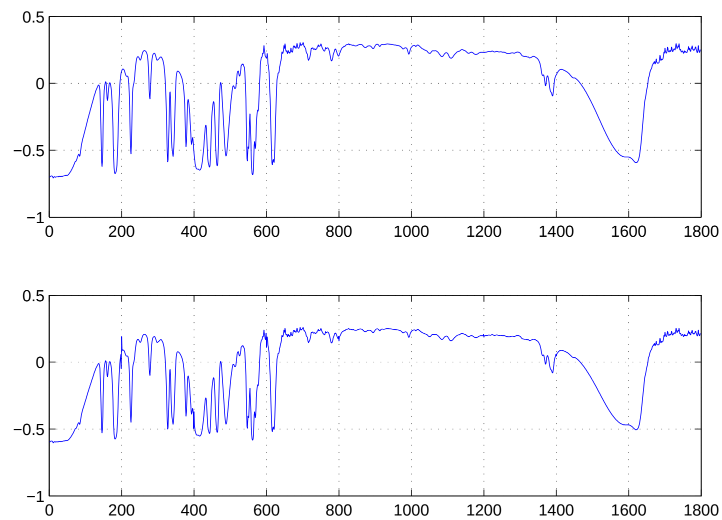

Figure 3: Cl2pheno

the US dollar. They are: (1) Euro (EUR)/US; (2) Yen (JPY)/US; (3) Sterling (GBF)/US; (4) Swiss Franc (CHF)/US, starting from January 1999 to May 2002.

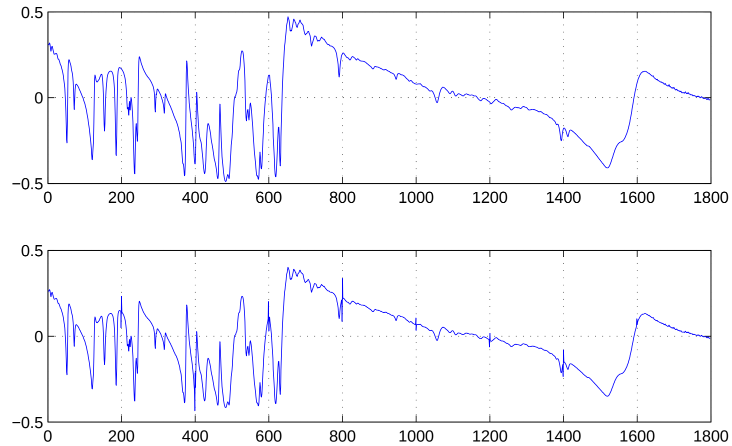

Figure 4: No4pheno

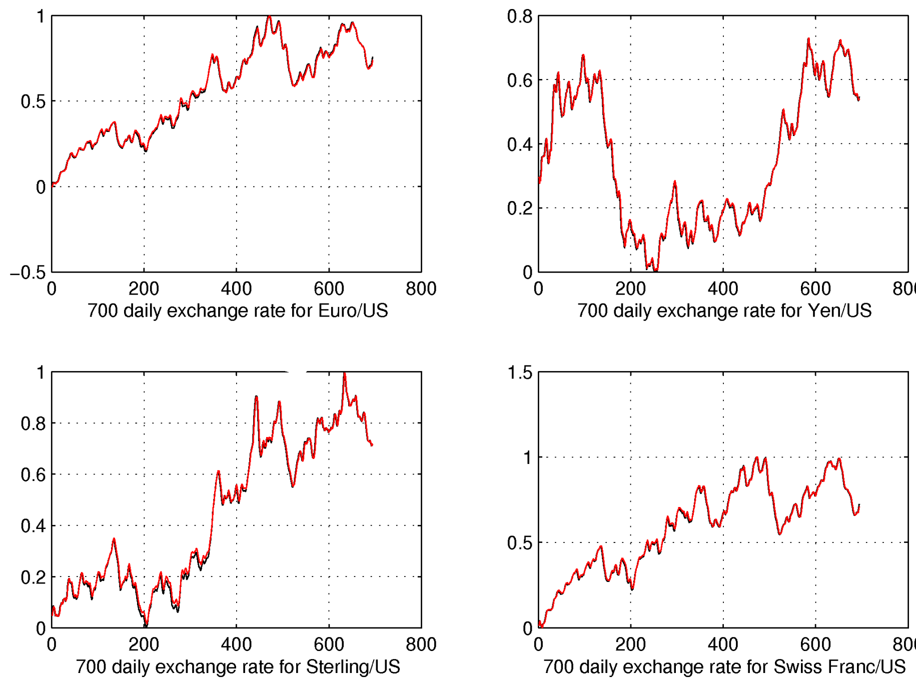

Figure 5: Figures 5 through 8: 700 daily exchange rates

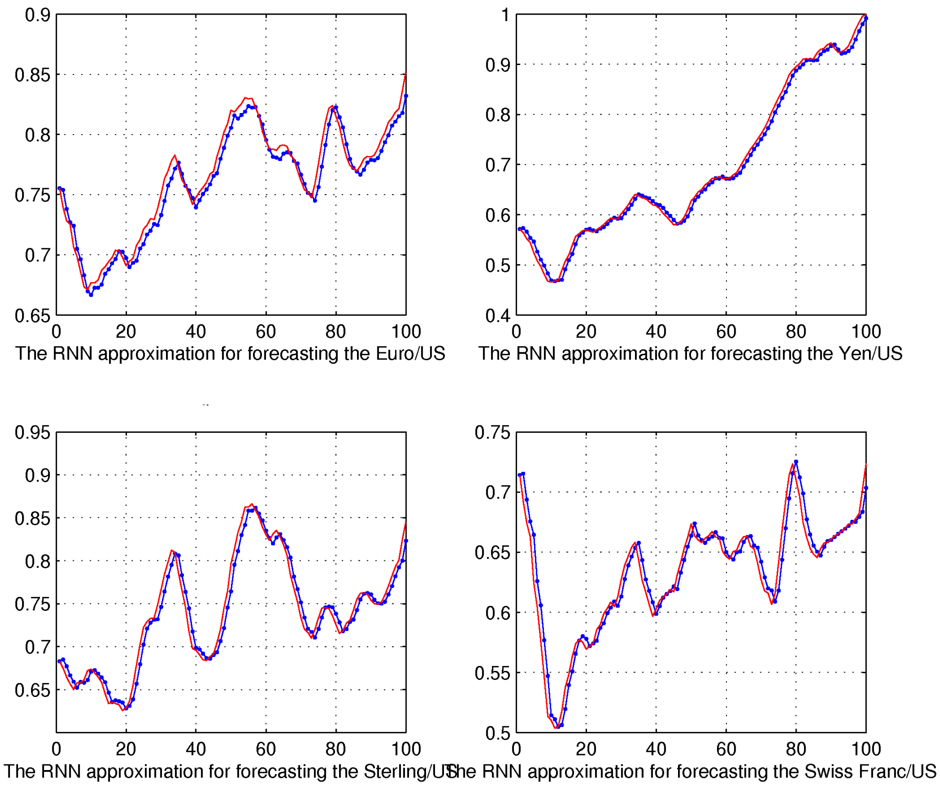

Figure 6: Figures 9 through 12: Learning short-term exchange rates

We use the first 701 days observations to train and validate the neural network. Before the training process, we employ a normalization for each component series in $x(t)$. This normalization is given by

$$y(t_i) = [\frac{x(t_i) - \min\{x(t_i)\}}{\text{Range of} \quad x(t_i)} - 0.5] \times 1.90\,, \qquad t = 1, 2, 3, ..., 701, \quad i = 1, 2, 3, 4, \tag{4.3}$$

as in as in [13].

After the training using the state space search algorithm and validation stages, we then used the resulting neural network parameters to make the out-of-sample forecasts for the last 100 observations. We found that the out-of-sample forecast root mean square errors are 0.0612 (EUR), 0.099 (GBF), 0.00289 (CHF) and 5.7536 (JPY), which measures how good our model is in terms of its prediction abilities. The learning dynamic used for the discrete-time model is system (4.1). We show the 700 daily exchange rates of the four currencies in the following Figures 5, 6, 7 and 8 respectively, where the solid lines are the moving average series of order 50 while the black lines are the best leaky integrator RNN approximations. In Figures 9 through 12, we show the actual moving average series and the learning forecast moving average values generated by the leaky integrator RNN for the last 100 days' out-of sample forecasts. The blue lines are the best leaky integrator RNN approximations while the red lines are the original data. From these figures, we see that the learning forecast values generated by the leaky integrator RNN model follow closely with the actual observations for each currency.

5. Concluding Remarks

We study the dynamic properties of theleaky integrator RNNs. The characters of the solutions of dynamical system (2.1) has been analyzed and its applications to data compression and learning short-term foreign exchange rates are discussed. Our results show that the set of solutions of the leaky integrator RNN system is a positive invariant and attractive set. The stability properties for system (2.1) has been obtained. Examples are given to demonstrate some extremely promising results for its applications to signal compression of infra-red spectrum and to learn the short-term exchange rates in the currency markets.

Acknowledgements

This research was supported by the Hong Kong Polytechnic University's Research Project Award A-PE95.

References

[1] E. Fiesler, R. Beale, *Handbook of Neural Computation*, New York (1969).

[2] M. Forti, A. Tesi, New conditions for global stability of neural networks with applications to linear and quadratic programming problems, *IEEE Trans. Circuits Syst. I*, **42** (1995), 354-366.

[3] J.A. Hale, *Handbook of Neural Computation*, Oxford University Press, New York (1997).

[4] D. Hankerson, G.A. Harris, P.D. Johnson, *Introduction to Information Theory and Data Compression*, CRC Press (1998).

[5] R.A. Horn, C.A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge (1985).

[6] J. Hertz, A. Krogh, R. Palmer, *Introduction to the Theory of Neural Computation*, Perseus Books Group (1991).

[7] L. Jin, N. Nikifork, M.M. Gupta, Absolute stability conditions for discrete-time neural networks, *IEEE Tranc. Neural Networks*, **5**, No. 6 (1994), 954-964.

[8] A. Lapedes, R. Farber, Nonlinear signal processing using neural networks, *Prediction and System Modeling*, Technical Report LA-UR87-2662, Los Alamos National Laboratory (1987).

[9] L.K. Li, Approximation theory and recurrent networks, In: *Proceedings of IJ CNN-Baltimore-92*, **2** (1992), 266-271.

[10] L.K. Li, Data compression by recurrent neural dynamics, In: *Proceedings of 1996 IEEE Region Ten Conference*, **1**, No. 2 (1996), 96-101.

[11] L.K. Li, Fixed point analysis for discrete-time recurrent neural networks, In: *Advances in Data Mining and Modeling* (Ed-s: W.K. Ching, K.P. Ng), World Science (2003), 107-115.

[12] L.K. Li, F.T. Chau, K.M. Leung, Compression of Ultraviolet-visible Spectrum with recurrent neural network, *Chemometrics and Intelligent Laboratory Systems*, **52**, No. 2 (2000), 135-143.

[13] L.K. Li, W.K. Pang, W.T. Yu, Forecasting short-term exchange rates: A recurrent neural network approach, *Neural Networks in Business Forecasting* (Ed. G. Peter Zhang), Idea Group Publishing (2004), 195-212.

[14] L.K. Li, S. Shao, A state space search algorithm and its application in learning short-term exchange rates, *Preprint*.

[15] L.K. Li, K.M. Yeung, Discrete-time recurrent neural networks and trajectory approximation, *J. of Appl. Mathematics*, **13** (2003), 23-33.

[16] X.B. Liang, J. Wang, A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints, *IEEE Tranc. Neural Networks*, **11**, No. 6 (2000), 1251-1262.

[17] M.J. Perez-Ilzarbe, Convergence analysis of a discrete-time recurrent neural network to perform quadratic real optimization with bound constraints, *IEEE Tranc. Neural Networks*, **9**, No. 6 (1998), 1344-1351.

[18] T. Tillson, *Smoothing Techniques for More Accurate Signals*, TASC, January (1998), 57.

[19] H. Tuy, *Convex Analysis and Global Optimization*, Kluwer Academic Publisher, Dordrecht (1998).

[20] P. Tenti, Forecasting foreign exchange rates using recurrent neural networks, *Appl. Artificial Intelligence*, **10** (1996), 150-170.

[21] X. Wang, E.K. Blum, Discrete-time versus continuous-time models of neural networks, *J. of Comp. Syst. Sci.*, **45** (1992), 1-19.

[22] J. Yao, C.L. Tan, A case study on using neural networks to perform technical forecasting of forex, *Neural Computation*, **34** (2000), 79-98.

[23] A. Zaknich, Y. Attikiouzel, Application of artificial neural networks to nonlinear signal processing, *Computational Intelligence: A Dynamic System Prespective* (Ed-s: M. Palaniswami et al), IEEE Press (1995).