Brief Papers

Representations of Continuous Attractors of Recurrent Neural Networks

Jiali Yu, Zhang Yi, and Lei Zhang

***Abstract*—A continuous attractor of a recurrent neural network (RNN) is a set of connected stable equilibrium points. Continuous attractors have been used to describe the encoding of continuous stimuli in neural networks. Dynamic behaviors of continuous attractors of RNNs exhibit interesting properties. This brief desires to derive explicit representations of continuous attractors of RNNs. Representations of continuous attractors of linear RNNs as well as linear-threshold (LT) RNNs are obtained under some conditions. These representations could be looked at as solutions of continuous attractors of the networks. Such results provide clear and complete descriptions to the continuous attractors.**

***Index Terms*—Continuous attractors, linear recurrent neural networks (RNNs), linear-threshold recurrent neural networks (LT RNNs), stability.**

I. Introduction

It is known that continuous attractors exist in recurrent neural networks (RNNs). A continuous attractor of an RNN is a set of connected stable equilibrium points. RNNs with continuous attractors are quite different from that of recurrent networks with discrete attractors. An RNN is said to have discrete attractors, if its equilibrium points are discretely distributed in the state space. In some conventional RNNs, such as the Hopfield RNNs [5], discrete attractors computation is a basic requirement for enabling the networks for many practical applications. However, discrete attractors may not be appropriate for patterns with continuous variability, such as the images of a 3-D object from different viewpoints. Thus, representing each object by a continuous manifold of fixed points, i.e., continuous attractors, is natural [13]. Continuous attractors are appealing to describe the encoding of continuous external stimuli, such as the orientation, the moving direction, and the spatial location of objects, or those continuous features underlying the categorization of complicated objects [20], [21].

Continuous attractors have been found in two classes of RNNs. One class is composed by RNNs with continuous distribution of infinite neurons, and the other class is composed by RNNs with finite neurons. Continuous attractors of the first class of RNNs have been widely studied by many authors; see, for example, [1], [8], [9], [14], [20], [21], and [26]. Recently, it was reported in [7] that continuous attractors can be designed by tuning the external inputs to networks that have a connectivity matrix with Toeplitz symmetry. On continuous attractors of the second class of RNNs, so far, most of the results have been reported in [6] and [10]–[12]. It should be pointed out that these two classes of networks are quite different in dynamics (for example, stability) from the mathematical point of view. This paper studies continuous attractors of the networks that belong to the second class.

In [10] and [11], line attractors of linear RNNs (linear RNNs) with finite neurons are studied. A line attractor is a special kind of continuous attractors; it is embedded in some 1-D manifold. Under some conditions, linear RNNs can possess line attractors. The theory of line attractors of linear RNNs has been used successfully to explain how the brain can keep the eyes still [11]. Generally, a continuous attractor may be embedded in some manifold with dimension larger than one. In this paper, more general cases will be studied for linear RNNs: the continuous attractors are not restricted to be in one dimension. Explicit representation for the continuous attractors of linear RNNs will be derived. By the representations, the continuous attractors can be completely described.

Continuous attractors also exist in some nonlinear RNNs with finite neurons [6], [10], [12], [23]–[25]. However, general results for continuous attractors in nonlinear RNNs with finite neurons have not been reported so far. In nonlinear RNNs, due to the nonlinearity of the transfer functions, it is not easy to explore the properties of continuous attractors. In this paper, the continuous attractors of a class of RNNs with linear-threshold (LT) transfer function will be studied. It is known that in the dynamical analysis of RNNs, the transfer functions play important role. The LT transfer function is essentially nonlinear, however, it has many interesting and good properties; see, for example, [2], [19], [23]–[25]. LT RNNs have found many applications, such as associative memory [15], winner-take-all [2], group selection [22], permitted and forbidden sets [3], [4], feature binding [19], etc. We will be interested in deriving explicit representations of continuous attractors of LT RNNs. Under some conditions, the explicit representations of continuous attractors will be obtained. These representations could be looked at as solutions of the networks.

This paper is organized as follows. Preliminaries are given in Section II. The main results on the representations of continuous attractors of linear RNNs are given in Section III. The representation of continuous attractors of LT RNNs are presented in Section IV. Simulations are given in Section V to illustrate the theory. Finally, conclusions are drawn in Section VI.

II. Preliminaries

The model of general RNNs can be described by

$$\dot{x}(t) = f(x(t)) \tag{1}$$

for $t \geq 0$, where $x \in R^n$, $f : R^n \to R^n$ is some continuous mapping that is local Lipshictz such that given any initial point $x(0) \in R^n$, the trajectory starting from $x(0)$ exists for all $t \geq 0$ and is unique.

Definition 1: A vector $x^* \in R^n$ is called an equilibrium point of (1), if it satisfies $f(x^*) = 0$.

Definition 2: An equilibrium point x^* is said to be stable, if given any constant $\epsilon > 0$, there exists a constant $\delta > 0$ such that

$$\|x(0) - x^*\| \leq \delta$$

implies that

$$\|x(t) - x^*\| \leq \epsilon$$

for all $t \geq 0$.

Manuscript received February 18, 2008; revised November 12, 2008; accepted December 02, 2008. First published January 13, 2009; current version published February 06, 2009. This work was supported by Chinese 863 High-Tech Program under Grant 2007AA01Z321.

J. Yu is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: yujiali@uestc.edu.cn).

Z. Yi is with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: zhangyi@scu.edu.cn).

L. Zhang is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: lzhang@csc.cuhk.edu.hk).

Color versions of one or more of the figures in this brief are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2008.2010771

Generally, an RNN may have more than one and even a lot of equilibrium points. The equilibrium points may be isolated or connected. The continuous attractors are related to a set of connected equilibrium points.

Definition 3: A set of equilibrium points C is called a continuous attractor if it is a connected set and each point $x^* \in C$ is stable.

Requiring stability for each equilibrium point of a continuous attractor is crucial. This is because in any practical application, only stable equilibrium points can be observed. By the above definition, without the stability, even if a set of equilibrium points is connected, it cannot be looked at as a continuous attractor. To illustrate this point, let us consider the following simple network:

$$\dot{x} + x = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} x.$$

It is easy to see that the set

$$C = \left\{ \begin{bmatrix} c \\ c \end{bmatrix} \middle| c \in R \right\}$$

is a connected set of equilibrium points. However, it is not a continuous attractor by Definition 3 since each equilibrium point of C is not stable. In fact, trajectories starting from points close to (but not in) the set C will diverge away from C.

Throughout this brief, we say a vector x is positive, denoted by $x > 0$, if each element of x is positive.

The problem of representation of continuous attractors is to derive the explicit representations of the continuous attractors. The representations could be looked at as solutions of the networks.

III. Linear Recurrent Neural Networks

In this section, we study the continuous attractors of linear RNNs. Linear RNNs have been studied by many authors. The model of linear RNNs can be described as

$$\dot{x}(t) + x(t) = Wx(t) + b \tag{2}$$

for $t \geq 0$, where $x = (x_1, \ldots, x_n)^T \in R^n$ is the state vector, W is the symmetric synaptic connection matrix, and $b = (b_1, \ldots, b_n)^T$ denotes the external input.

Since the synaptic connection matrix W is a symmetric matrix, it possesses an orthonormal eigensystem. Let $\lambda_i (i = 1, \ldots, n)$ be all the eigenvalues of W ordered by $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$. Suppose that $S_i (i = 1, \ldots, n)$ compose an orthonormal basis in R^n such that each S_i is an eigenvector of W corresponding to the eigenvalue λ_i. Let the multiplicity of λ_1 be m and denote by V_{λ_1} the eigensubspace associated with the eigenvalue λ_1. Suppose that

$$b = \sum_{i=1}^{n} \tilde{b}_i S_i.$$

Theorem 1: Suppose $\lambda_1 = 1$ and $b \perp V_{\lambda_1}$. Then, the linear network (2) has a continuous attractor and the continuous attractor can be represented by

$$C = \left\{ \sum_{i=1}^{m} c_i S_i + \sum_{j=m+1}^{n} \frac{\tilde{b}_j}{1 - \lambda_j} S_j \middle| c_i \in R (1 \leq i \leq m) \right\}.$$

Proof: Since $b \perp V_{\lambda_1}$, then $\tilde{b}_1 = \cdots = \tilde{b}_m = 0$.

First, we prove that each point of C is an equilibrium point. Given any $x^* \in C$, there exist constants $c_i (i = 1, \ldots, m)$ such that

$$x^* = \sum_{i=1}^{m} c_i S_i + \sum_{j=m+1}^{n} \frac{\tilde{b}_j}{1 - \lambda_j} S_j.$$

Then

$$\begin{aligned} Wx^* + b &= \sum_{i=1}^{m} c_i W S_i + \sum_{j=m+1}^{n} \frac{\tilde{b}_j}{1 - \lambda_j} W S_j + \sum_{i=1}^{n} \tilde{b}_i S_i \\ &= \sum_{i=1}^{m} c_i S_i + \sum_{j=m+1}^{n} \frac{\lambda_j \tilde{b}_j}{1 - \lambda_j} S_j + \sum_{i=1}^{n} \tilde{b}_i S_i \\ &= \sum_{i=1}^{m} c_i S_i + \sum_{j=m+1}^{n} \frac{\tilde{b}_j}{1 - \lambda_j} S_j = x^*. \end{aligned}$$

This shows clearly that x^* is an equilibrium point of (2). Thus, the set C is a connected set of equilibrium points.

Next, we prove that C is stable. Given any $x(0) \in R^n$, let $x(t)$ be the trajectory starting from $x(0)$. Since $S_i (i = 1, \ldots, n)$ compose an orthonormal basis of R^n, then $x(t)$ can be represented as

$$x(t) = \sum_{i=1}^{n} z_i(t) S_i$$

for $t \geq 0$, where $z_i(t) (i = 1, \ldots, n)$ are some differentiable functions.

It follows from (2) that

$$\dot{z}_i(t) = 0, \qquad (i = 1, \ldots, m)$$

and

$$\dot{z}_i(t) = (\lambda_i - 1) \cdot z_i(t) + \tilde{b}_i, \qquad (i = m+1, \ldots, n)$$

for $t \geq 0$. Then, we have

$$z_i(t) = \begin{cases} z_i(0), & 1 \leq i \leq m \\ \left(z_i(0) - \frac{\tilde{b}_i}{1 - \lambda_i} \right) e^{(\lambda_i - 1)t} + \frac{\tilde{b}_i}{1 - \lambda_i}, & m+1 \leq i \leq n \end{cases}$$

for $t \geq 0$. Thus

$$\begin{aligned} x(t) = \sum_{i=1}^{m} z_i(0) S_i &+ \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right) S_j e^{(\lambda_j - 1)t} \\ &+ \sum_{j=m+1}^{n} \frac{\tilde{b}_j}{1 - \lambda_j} S_j \end{aligned}$$

for $t \geq 0$.

Given any $\epsilon > 0$, choose a constant $\delta = \epsilon$, if

$$\begin{aligned} &\|x(0) - x^*\| \\ &= \left\| \sum_{i=1}^{m} (z_i(0) - c_i) \cdot S_i + \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right) \cdot S_j \right\| \\ &= \sqrt{\sum_{i=1}^{m} (z_i(0) - c_i)^2 + \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right)^2} \leq \delta \end{aligned}$$

then

$$\begin{aligned} &\|x(t) - x^*\| \\ &= \left\| \sum_{i=1}^{m} (z_i(0) - c_i) \cdot S_i + \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right) \cdot S_j e^{(\lambda_j - 1)t} \right\| \\ &= \sqrt{\sum_{i=1}^{m} (z_i(0) - c_i)^2 + \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right)^2 e^{2(\lambda_j - 1)t}} \\ &\leq \sqrt{\sum_{i=1}^{m} (z_i(0) - c_i)^2 + \sum_{j=m+1}^{n} \left(z_j(0) - \frac{\tilde{b}_j}{1 - \lambda_j} \right)^2} \leq \delta = \epsilon \end{aligned}$$

for all $t \geq 0$. By Definite 2, it shows that x^* is stable. By Definition 3, the set C is a continuous attractor of the network (2). The proof is complete. ■

Theorem 1 gives explicit representation of continuous attractor of the network (2) using the eigenvectors of the connection matrix W. By the representation, the continuous attractor could be looked at as being solved from the networks. It gives a complete description of the continuous attractor. Moreover, Theorem 1 provides a method to construct continuous attractors for network (2).

IV. Linear Threshold Recurrent Neural Networks

The linear threshold RNN model can be described by

$$\dot{x(t)} + x(t) = W\sigma\left(x(t)\right) + b \tag{3}$$

for $t \geq 0$, where $x = (x_1, \ldots, x_n)^T \in R^n$ is the state vector, W is the connection weight, which is symmetric, and $b = (b_1, \ldots, b_n)^T$ denotes the external input. $\sigma(x)$ is a rectification nonlinear function defined by

$$\sigma(s) = \max\{0, s\}, \qquad s \in R.$$

Definition 4: Let M be an $n \times n$ matrix, and let $P \subseteq \{1, 2, \ldots, n\}$ be an index set. The matrix M_P is said to be a submatrix of M if the matrix M_P can be constructed from M simply by removing from M all rows and columns not indexed by P.

Let $P \subseteq \{1, 2, \ldots, n\}$ be an index set with p elements, and $Z = \{1, 2, \ldots, n\} - P$. Then, the network (3) can be rewritten as

$$\begin{cases} \dot{x}_P(t) + x_P(t) = W_P \cdot \sigma\left(x_P(t)\right) + W_{PZ} \cdot \sigma\left(x_Z(t)\right) + b_P \\ \dot{x}_Z(t) + x_Z(t) = W_{ZP} \cdot \sigma\left(x_P(t)\right) + W_Z \cdot \sigma\left(x_Z(t)\right) + b_Z \end{cases} \tag{4}$$

for $t \geq 0$, where W_P and W_Z are submatrices of W, W_{PZ} is a matrix constructed from W by removing from W all rows not indexed by P and all columns not indexed by Z, and W_{ZP} is a matrix constructed from W by removing from W all rows not indexed by Z and all columns not indexed by P.

By assumption, the synaptic matrix W_P is a symmetric matrix. Let $\lambda_i^P (i = 1, \ldots, p)$ be all the eigenvalues of W_P ordered by $\lambda_1^P \geq \lambda_2^P \cdots \geq \lambda_p^P$. Suppose that the multiplicity of λ_1^P is $m (m \leq p)$, clearly, $\lambda_1^P = \lambda_2^P = \cdots = \lambda_m^P$. Denote by $V_{\lambda_1^P}^P$ the eigensubspace associated with the largest eigenvalue λ_1^P. Let $S_i^P \in V_{\lambda_1^P}^P (i = 1, \ldots, p)$ be the unit eigenvectors that are mutually orthogonal. These eigenvectors form a basis of $V_{\lambda_1^P}^P$. Suppose that

$$b_P = \sum_{i=1}^{p} \tilde{b}_i^P \cdot S_i^P.$$

Theorem 2: Suppose that $\lambda_1^P = 1$ and $b_P \perp V_{\lambda_1^P}^P$. If it holds that

$$\begin{cases} \sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P > 0 \\ W_{ZP} \cdot \left[\sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P \right] + b_Z < 0 \end{cases} \tag{5}$$

for any constants $c_i > 0 (i = 1, \ldots, m)$, then

$$C = \left\{ \left[\begin{array}{c} \sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P \\ W_{ZP} \cdot \left[\sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P \right] + b_Z \end{array} \right] \middle| \; c_i > 0 (i = 1, \ldots, m) \right\}$$

is a continuous attractor of (3).

Proof: Since $b_P \perp V_{\lambda_1^P}^P$, then $\tilde{b}_1^P = \cdots = \tilde{b}_m^P = 0$. Clearly, C is a connected set. Given any $x^* \in C$, there exist constants $c_i > 0 (i = 1, \ldots, m)$ such that

$$\begin{cases} x_P^* = \sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P > 0 \\ x_Z^* = W_{ZP} \cdot \left[\sum\limits_{i=1}^{m} c_i S_i^P + \sum\limits_{j=m+1}^{p} \frac{\tilde{b}_j^P}{1-\lambda_j^P} S_j^P \right] + b_Z < 0. \end{cases}$$

It is easy to show that x^* is an equilibrium point. Thus, the set C forms a connected set of equilibrium points. Next, we prove that C attracts trajectories, i.e., each $x^* \in C$ is stable.

Define two differentiable functions by

$$\begin{cases} V_P(t) = \|x_P(t) - x_P^*\|^2 \\ V_Z(t) = \|x_Z(t) - x_Z^*\|^2 \end{cases}$$

for $t \geq 0$.

Then, it follows from (4) that

$$\begin{aligned} \dot{V}_P(t) = &-2V_P(t) + 2\left[x_P(t) - x_P^*\right]^T \cdot W_P \cdot \left[\sigma\left(x_P(t)\right) - x_P^*\right] \\ &+2\left[x_P(t) - x_P^*\right]^T \cdot W_{PZ} \cdot \sigma\left(x_Z(t)\right) \end{aligned} \tag{6}$$

and

$$\begin{aligned} \dot{V}_Z(t) = &-2V_Z(t) + 2\left[x_Z(t) - x_Z^*\right]^T \cdot W_{ZP} \cdot \left[\sigma\left(x_P(t)\right) - x_P^*\right] \\ &+2\left[x_Z(t) - x_Z^*\right]^T \cdot W_Z \cdot \sigma\left(x_Z(t)\right) \end{aligned} \tag{7}$$

for $t \geq 0$.

Given a constant ϵ such that

$$0 < \epsilon \leq \min\left\{x_i\,(i \in P), -x_j\,(j \in Z)\right\}$$

define a neighborhood B_ϵ of x^* by

$$B_\epsilon = \left\{ x \in R^n \,\middle|\, \|x_P - x_P^*\| < \frac{\epsilon}{2}, \|x_Z - x_Z^*\| < \|W_{ZP}\| \cdot \epsilon \right\}.$$

It can be proven that each trajectory starting from B_ϵ will stay there forever. We will show this by counterproof. Suppose $x(0) \in B_\epsilon$, and the trajectory $x(t)$ starting from $x(0)$ cannot stay in B_ϵ for all $t \geq 0$, then two cases can happen.

Case 1: There exists a $t_1 > 0$ with $V_P(t_1) = (\epsilon/2)^2$, and there exists $\tilde{t}_1 > t_1$ such that $V_P(t)$ is strictly increasing on the interval $[t_1, \tilde{t}_1]$, and

$$\begin{cases} V_P(t) \leq \left(\frac{\epsilon}{2}\right)^2, & 0 \leq t < t_1 \\ V_Z(t) \leq \left(\|W_{ZP}\| \cdot \epsilon\right)^2, & 0 \leq t \leq \tilde{t}_1. \end{cases}$$

Thus, it must hold that $\dot{V}_P(t) > 0$ for $t \in [t_1, \tilde{t}_1]$. However, since $(I - W)_P$ is positive semidefinite, from (6), it follows for $t \in [t_1, \tilde{t}_1]$ that

$$\begin{aligned} \dot{V}_P(t) = &-2V_P(t) + 2\left[x_P(t) - x_P^*\right]^T \cdot W_P \cdot \left[\sigma\left(x_P(t)\right) - x_P^*\right] \\ &+ 2\left[x_P(t) - x_P^*\right]^T \cdot W_{PZ} \cdot \sigma\left(x_Z(t)\right) \\ = &-2V_P(t) + 2\left[x_P(t) - x_P^*\right]^T \cdot W_P \cdot \left[x_P(t) - x_P^*\right] \\ = &-2\left[x_P(t) - x_P^*\right]^T \cdot (I - W)_P \cdot \left[x_P(t) - x_P^*\right] \leq 0. \end{aligned}$$

This is a contradiction.

Case 2: There exists a $t_2 > 0$ such that

$$V_Z(t_2) = \left(\|W_{ZP}\| \cdot \epsilon\right)^2$$

and

$$\begin{cases} V_Z(t) < \left(\|W_{ZP}\| \cdot \epsilon\right)^2, & 0 \leq t < t_2 \\ V_P(t) \leq \left(\frac{\epsilon}{2}\right)^2, & 0 \leq t \leq t_2. \end{cases}$$

Thus, it must hold that $\dot{V}_Z(t_2) \geq 0$. However, from (7), it follows that

$$\begin{aligned}\dot{V}_Z(t_2) =& -2V_Z(t_2) + 2\left[x_Z(t_2) - x_Z^*\right]^T \cdot W_{ZP} \\ & \cdot \left[\sigma\left(x_P(t_2)\right) - x_P^*\right] + 2\left[x_Z(t) - x_Z^*\right]^T \cdot W_Z \cdot \sigma\left(x_Z(t)\right) \\ =& -2V_Z(t_2) + 2\left[x_Z(t_2) - x_Z^*\right]^T \cdot W_{ZP} \cdot \left[x_P(t_2) - x_P^*\right] \\ \leq& -\left(\|W_{ZP}\| \cdot \epsilon\right)^2 < 0.\end{aligned}$$

This is a contradiction.

The above implies that for any $x(0) \in B_\epsilon$, it holds that $x(t) \in B_\epsilon$ for all $t \geq 0$. Clearly, it shows that x^* is stable. Thus, C is a continuous attractor. The proof is complete. ■

Theorem 2 provides conditions for the network (3) to possess continuous attractors. These conditions are given in terms of the eigenvectors of the eigensubspace $V_{\lambda_1^P}^P$, thus, the continuous attractors can be easily constructed from the conditions of Theorem 2. The continuous attractors are explicitly represented by the eigenvectors and it provides a way to completely master the continuous attractors.

It should be noted that the condition (5) is sufficient but not necessary. For example, the network

$$\dot{x} + x = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sigma(x) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

possesses a continuous attractor as follows:

$$C = \left\{ \begin{bmatrix} c \\ -c-1 \\ 1 \end{bmatrix} \middle| c > 0 \right\}.$$

However, it is easy to see that the condition (5) is not satisfied.

V. Simulations

In this section, we will give some simulations to illustrate the continuous attractors theory established in the above sections.

Example 1: Let us first consider a 3-D linear neural network

$$\dot{x} + x = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix} x + \begin{bmatrix} -0.5774 \\ 0.5774 \\ 0.5774 \end{bmatrix}. \quad (8)$$

Denote

$$W = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix} \quad b = \begin{bmatrix} -0.5774 \\ 0.5774 \\ 0.5774 \end{bmatrix}.$$

It can be checked that W has the largest eigenvalue $\lambda_1 = 1$ with multiplicity 2, and W has another eigenvalue -2. Moreover, $b \perp V_{\lambda_1}$. By Theorem 1, the network possesses a continuous attractor. The continuous attractor is in two dimensions. Fig. 1 shows the 2-D continuous attractor of the network. The plane in the figure is the continuous attractor. It divides the 3-D space into two parts. The figure shows that 40 trajectories starting from randomly selected initial points converge to the attractor.

Example 2: Consider a 2-D LT network

$$\dot{x} + x = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \sigma(x) + \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \quad (9)$$

Denote

$$W = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Let $P = \{1\}$, then

$$W_P = 1 \quad W_{ZP} = -1 \quad S^P = 1$$

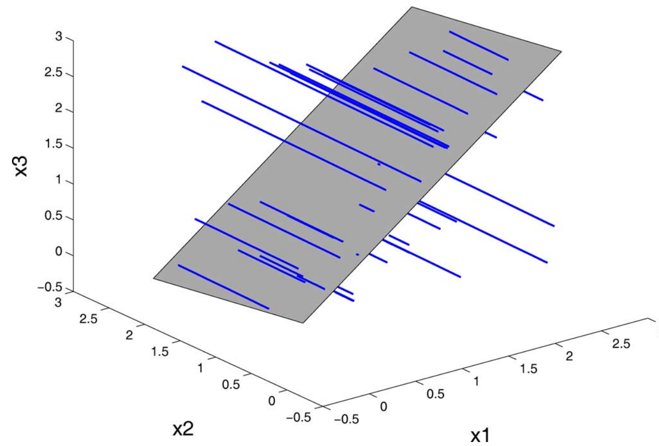

Fig. 1. Continuous attractor of linear RNN (8). It is a 2-D hyperplane. Trajectories starting from the points close to the plane converge to it.

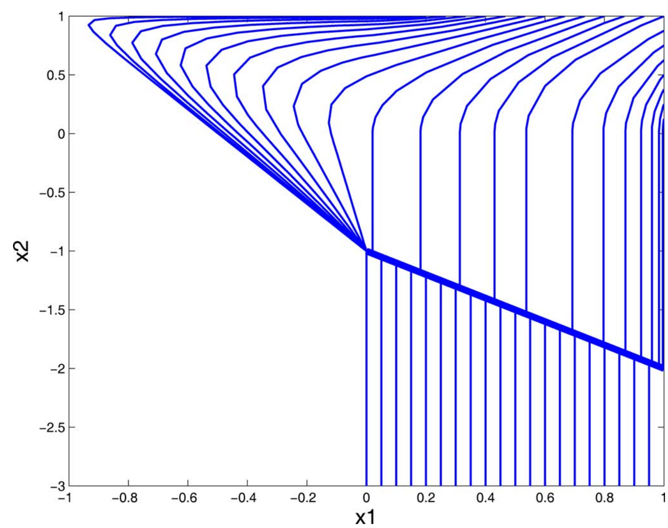

Fig. 2. Continuous attractor of LT RNN (9). The bold straight line is the continuous attractor. It is a connected set of equilibrium points and each of these points is stable.

and

$$\begin{cases} c \cdot S^P > 0 \\ W_{ZP} \cdot c \cdot S^P + b_Z = -c - 1 < 0 \end{cases}$$

for any $c > 0$. By Theorem 2

$$C = \left\{ \begin{bmatrix} c \\ -c-1 \end{bmatrix} \middle| c > 0 \right\}$$

is a continuous attractor.

Fig. 2 shows the continuous attractor of the network. The bold line in the figure is the continuous attractor. The figure shows that 50 trajectories starting from initial points close to the continuous attractor are attracted to the continuous attractor.

VI. Conclusion

In this paper, the representations of continuous attractors of linear RNNs and LT RNNs have been investigated. A continuous attractor is defined by a connected set of equilibrium points, which forms a lower dimensional manifold in the original state space. Moreover, a continuous attractor must attract trajectories in some sense. This is implemented by requiring each equilibrium point of the continuous attractor to be stable. Under some conditions, explicit representations of continuous attractors are obtained. By the representations, continuous attractors of the RNNs can be completely described. The methods in this brief could be possibly further developed to study continuous attractors of more general networks such as stochastic neural networks with time delays [16]–[18]. More research on this direction is required.

References

[1] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biol. Cybern.*, vol. 27, pp. 77–87, 1977.
[2] R. Hahnloser, "On the piecewise analysis of linear threshold neural networks," *Neural Netw.*, vol. 11, no. 2, pp. 691–697, 1998.
[3] R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, pp. 947–951, 2000.
[4] R. Hahnloser, H. S. Seung, and J. J. Slotine, "Permitted and forbidden sets in symmetric threshold-linear networks," *Neural Comput.*, vol. 15, pp. 621–638, 2003.
[5] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, no. 10, pp. 3088–3092, 1984.
[6] D. D. Lee, B. Y. Reis, H. S. Seung, and D. W. Tank, "Nonlinear network models of the oculomotor integrator," in *In Computational Neuroscience: Trends in Research*. New York: Plenum Press, 1997.
[7] C. K. Machens and C. D. Brody, "Design of continuous attractor networks with monotonic tuning using a symmetry principle," *Neural Comput.*, vol. 20, no. 2, pp. 452–485, 2008.
[8] P. Miller, "Analysis of spike statistics in neuronal systems with continuous attractors or multiple, discrete attractor states," *Neural Comput.*, vol. 18, pp. 1268–1317, 2006.
[9] A. Pouget, K. Zhang, S. Deneve, and P. E. Latham, "Statistically efficient estimation using population coding," *Neural Comput.*, vol. 10, pp. 373–401, 1998.
[10] H. S. Seung, "Continuous attractors and oculomotor control," *Neural Netw.*, vol. 11, pp. 1253–1258, 1998.
[11] H. S. Seung, "How the brain keeps the eyes still," *Proc. Nat. Acad. Sci. USA*, vol. 93, pp. 13339–13344, 1996.
[12] H. S. Seung, "Learning continuous attractors in recurrent networks," *Adv. Neural Inf. Process. Syst.*, vol. 10, pp. 654–660, 1998.
[13] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, pp. 2268–2269, 2000.
[14] S. M. Stringer, E. T. Rolls, T. P. Trappenberg, and I. E. T. de Araujo, "Self-organizing continuous attractor networks and motor function," *Neural Netw.*, vol. 16, pp. 161–182, 2003.
[15] H. J. Tang, K. C. Tan, and E. J. Teoh, "Dynamics analysis and analog associative memory with L-T neurons," *IEEE Trans. Neural Netw.*, vol. 17, no. 2, pp. 409–418, Mar. 2006.
[16] Z. D. Wang, D. W. C. Ho, and X. H. Liu, "On global asymptotic stability of neural networks with discrete and distributed delays," *Phys. Lett. A*, vol. 345, no. 4-6, pp. 299–308, Oct. 2005.
[17] Z. D. Wang, D. W. C. Ho, and X. H. Liu, "State estimation for delayed neural networks," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 279–284, Jan. 2005.
[18] Z. D. Wang, Y. R. Liu, M. Z. Li, and X. H. Liu, "Stability analysis for stochastic Cohen-Grossberg neural networks with mixed time delays," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 814–820, May 2006.
[19] H. Wersing, W. J. Beyn, and H. Ritter, "Dynamical stability conditions for recurrent neural networks with unsaturating piecewise linear transfer functions," *Neural Comput.*, vol. 13, pp. 1811–1825, 2001.
[20] S. Wu and S. Amari, "Computing with continuous attractors: Stability and online aspects," *Neural Comput.*, vol. 17, pp. 2215–2239, 2005.
[21] S. Wu, K. Hamaguchi, and S. Amari, "Dynamics and computation of continuous attractors," *Neural Comput.*, vol. 20, no. 4, pp. 994–1025, Apr. 2008.
[22] X. Xie, R. H. Hahnloser, and H. S. Seung, "Selectively grouping neurons in recurrent networks of lateral inhibition," *Neural Comput.*, vol. 14, pp. 2627–2646, 2002.
[23] Z. Yi and K. K. Tan, *Convergence Analysis of Recurrent Neural Networks*. Boston, MA: Kluwer, 2004.
[24] Z. Yi and K. K. Tan, "Multistability analysis of discrete-time recurrent neural networks with unsaturating piecewise linear activation functions," *IEEE Trans. Neural Netw.*, vol. 15, no. 2, pp. 329–336, Mar. 2004.
[25] Z. Yi, K. K. Tan, and T. H. Lee, "Multistability analysis for recurrent neural networks with unsaturating piecewise linear transfer functions," *Neural Comput.*, vol. 15, pp. 639–662, 2003.
[26] K. C. Zhang, "Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory," *J. Neurosci.*, vol. 16, no. 6, pp. 2112–2126, 1996.