Neural Networks, Vol. 6, pp. 801–806, 1993
Printed in the USA.

0893-6080/93 $6.00 + .00

ORIGINAL CONTRIBUTION

Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks

KEN-ICHI FUNAHASHI AND YUICHI NAKAMURA

Toyohashi University of Technology

(*Received* 16 *March* 1992; *revised and accepted* 10 *November* 1992)

Abstract—*In this paper, we prove that any finite time trajectory of a given n-dimensional dynamical system can be approximately realized by the internal state of the output units of a continuous time recurrent neural network with n output units, some hidden units, and an appropriate initial condition. The essential idea of the proof is to embed the n-dimensional dynamical system into a higher dimensional one which defines a recurrent neural network. As a corollary, we also show that any continuous curve can be approximated by the output of a recurrent neural network.*

Keywords—Approximation, Continuous time recurrent neural network, Dynamical system, Autonomous system, Trajectory, Internal state, Hidden unit, Continuous curve.

1. INTRODUCTION

There are two types of connections in neural networks. Neural networks with only feedforward connections are called feedforward networks, and neural networks with arbitrary connections are often called recurrent networks. In the case of feedforward networks, ever since the back propagation learning algorithm was proposed by Rumelhart, Hinton, and Williams (1986), several applications have been made, mainly to static information processing, such as pattern recognition. On the theoretical capability of these networks, Cybenko (1989), Funahashi (1989), and Hornik, Stinchcombe, and White (1989) proved mathematically that a given continuous mapping on a compact set can be approximately realized by three-layer feedforward neural networks with any precision. We call this "the fundamental approximation theorem" in the following.

The nonlinear dynamical behavior of recurrent networks is suitable for spatio-temporal information processing. Theoretical studies on recurrent networks have been mainly concerned with the stability of convergence of the trajectory to the equilibria (e.g., Hirsch, 1989). The Hopfield network with symmetrical weights of connection has been applied to the content addressable memory and combinatorial optimization.

Learning algorithms which employ the steepest descent method for the modification of recurrent network weights have been proposed both by Williams and Zipser (1990) in the case of discrete time systems, and by Pearlmutter (1989), Pineda (1987), and Sato (1990), etc., for continuous time systems. Sato and Murakami (1991) proposed both the modified recurrent network in order to approximate the dynamical system and its learning algorithm by the use of the fundamental approximation theorem. Furthermore, they applied the algorithm to the approximation of nonlinear dynamical systems. Since their network is far from the ordinary recurrent networks, the theoretical capability of the recurrent network has still not been clarified. Seidl and Lorenz (1991) proved the approximation theorem for the trajectory of the discrete dynamical system by the use of the fundamental approximation theorem.

The main goal of this study is to elucidate the theoretical capability of continuous time recurrent networks. In this paper we will prove that the internal state of output units of the continuous time recurrent network approximates the finite time trajectory of the given dynamical system with any precision. The proof is given by the use of the fundamental approximation theorem and the fundamental theorems on dynamical systems.

2. CONTINUOUS TIME RECURRENT NEURAL NETWORKS

There are two types of recurrent neural networks: discrete time recurrent neural networks and continuous time ones. In this paper, we study the latter.

Acknowledgement: We would like to thank Dr. M. Sakakibara for critical reading of the manuscript.

Requests for reprints should be sent to K. Funahashi, Department of Information and Computer Sciences, Toyohashi University of Technology, Hibarigaoka, Tenpaku, Toyohashi 441, Japan.

The dynamics of the continuous time recurrent neural network with m units, which is discussed in this paper, is described by the following system of ordinary differential equations:

$$\frac{du_i(t)}{dt} = -\frac{u_i(t)}{\tau_i} + \sum_{j=1}^{m} w_{ij}\sigma(u_j(t)) + I_i(t) \qquad (i = 1, \ldots, m), \quad (1)$$

where $u_i(t)$ is the internal state of the i-th unit, τ_i is the time constant of the i-th unit, w_{ij} are connection weights, $I_i(t)$ is the input to the i-th unit, and $\sigma(u_i(t))$ is the output of the i-th unit. Here, σ is called the output function, and C^1-sigmoid functions (nonconstant, bounded, and monotone increasing C^1-functions) are used. As σ,

$$\sigma(x) = 1/(1 + \exp(-x)) \quad (2)$$

is usually used.

In the following, we deal with recurrent neural networks with the same time constant τ (i.e., $\tau_i = \tau$) and without input (i.e., $I_i(t) = 0$). We set $u(t) = {}^t(u_1(t), \ldots, u_m(t))$ and use $W = (w_{ij})$ as an $m \times m$ weight matrix. Let $\sigma : \mathbf{R}^m \rightarrow \mathbf{R}^m$ be denoted by a sigmoid mapping

$$\sigma({}^t(u_1, \ldots, u_m)) = {}^t(\sigma(u_1), \ldots, \sigma(u_m)); \quad (3)$$

then the vector expression of eqn (1) is

$$\dot{u}(t) = -\frac{1}{\tau} u(t) + W\sigma(u(t)). \quad (4)$$

3. APPROXIMATE REALIZATION THEOREMS OF DYNAMICAL SYSTEM TRAJECTORIES

Let the point of the n-dimensional Euclidian space $\mathbf{R}^n$ be denoted by $x = {}^t(x_1, \ldots, x_n)$ and the Euclidian norm of x be defined by $|x|$.

The dynamical system on an open set of $\mathbf{R}^n$ means a system defined by an autonomous ordinary differential equation which has global solutions in the open set. Let the output function σ of recurrent neural networks be a C^1-sigmoid function. As recurrent neural networks studied here have no input, some units are called output units and the others are called hidden units.

In this paper, we prove the following theorems.

THEOREM 1. *Let D be an open subset of $\mathbf{R}^n$, $F : D \rightarrow \mathbf{R}^n$ be a C^1-mapping, and $\tilde{K}$ be a compact subset of D. Suppose that there is a subset $K \subset \tilde{K}$ such that any solution $x(t)$ with initial value $x(0)$ in K of an ordinary differential equation*

$$\dot{x} = F(x), \quad x(0) \in K \quad (5)$$

is defined on $I = [0, T](0 < T < \infty)$ and $x(t)$ is included in $\tilde{K}$ for any $t \in I$. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N and a recurrent neural network with n output units and N hidden units such that for a solution $x(t)$ satisfying eqn (5) and an appropriate initial state of the network,

$$\max_{t \in I} |x(t) - u(t)| < \varepsilon \quad (6)$$

holds, where $u(t) = {}^t(u_1(t), \ldots u_n(t))$ is the internal state of output units of the network.

For dynamical systems, the condition for $\tilde{K}$ is satisfied a priori by definition, so we can restate Theorem 1 as follows.

THEOREM 2. *Let $D \subset \mathbf{R}^n$ and $F : D \rightarrow \mathbf{R}^n$ be the same as above, and suppose that $\dot{x} = F(x)$ defines a dynamical system on D. Let K be a compact subset of D and we consider trajectories of the system on the interval $I = [0, T](0 < T < \infty)$. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N and a recurrent neural network with n output units and N hidden units such that for any trajectory $\{x(t); 0 \leqq t \leqq T\}$ of the system with initial value $x(0) \in K$ and an appropriate initial state of the network,*

$$\max_{t \in I} |x(t) - u(t)| < \varepsilon$$

holds, where $u(t) = {}^t(u_1(t), \ldots u_n(t))$ is the internal state of output units of the network.

As a corollary of Theorem 2, we obtain the following:

COROLLARY 1. *Let σ be a strictly increasing C^1-sigmoid function such that $\sigma(\mathbf{R}) = (0, 1)$. Let D be an open subset of $(0, 1)^n$, $F : D \rightarrow \mathbf{R}^n$ be a C^1-mapping, and suppose that $\dot{x} = F(x)$ defines a dynamical system on D. Let K be a compact subset of D and we consider trajectories of the system on interval $I = [0, T](0 < T < \infty)$. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N and a recurrent neural network with n output units and N hidden units such that for any trajectory $\{x(t); 0 \leqq t \leqq T\}$ of the system with initial value $x(0) \in K$ and an appropriate initial state of the network,*

$$\max_{t \in I} |x(t) - y(t)| < \varepsilon \quad (7)$$

holds, where $y(t) = {}^t(y_1(t), \ldots, y_n(t))$ is the output of the recurrent network with the sigmoid output function σ.

By the use of Theorem 1, we also obtain the following:

THEOREM 3. *Let $f : I = [0, T] \rightarrow \mathbf{R}^n$ be a continuous curve, where $0 < T < \infty$. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N and a recurrent network with n output units and N hidden units such that*

$$\max_{t \in I} |f(t) - u(t)| < \varepsilon, \quad (8)$$

where $u(t) = {}^t(u_1(t), \ldots u_n(t))$ is the internal state of output units of the network.

In the same way as Corollary 1, the following Corollary can be obtained from Theorem 3.

COROLLARY 2. *Let σ be a strictly increasing C^1-sigmoid function such that $\sigma(\mathbf{R}) = (0, 1)$, and $f: I = [0, T] \to (0, 1)^n$ be a continuous curve, where $0 < T < \infty$. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N and a recurrent neural network with n output units and N hidden units such that*

$$\max_{t \in I} |f(t) - y(t)| < \varepsilon, \tag{9}$$

where $y(t) = {}^t(y_1(t), \ldots, y_n(t))$ is the output of the recurrent network with the sigmoid output function σ.

4. PRELIMINARY 1

The following Theorem by Funahashi (1989) is the base for the proofs of our theorems in Section 3.

THEOREM. *Let $\sigma(x)$ be a sigmoid function (i.e., a nonconstant, increasing, and bounded continuous function on $\mathbf{R}$). Let K be a compact subset of $\mathbf{R}^n$, and $f(x_1, \ldots, x_n)$ be a continuous function on K. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N, real constants c_i, θ_i $(i = 1, \ldots, N)$ and w_{ij} $(i = 1, \ldots, N; j = 1, \ldots, n)$ such that*

$$\max_{x \in K} \left| f(x_1, \ldots, x_n) - \sum_{i=1}^{N} c_i \sigma\left(\sum_{j=1}^{n} w_{ij} x_j - \theta_i \right) \right| < \varepsilon \tag{10}$$

holds.

This Theorem shows that three-layer feedforward neural networks whose output layer has linear units can approximate any continuous mapping $f: \mathbf{R}^n \to \mathbf{R}^m$ uniformly on an arbitrary compact set.

THEOREM (*The fundamental approximation theorem*). *Let K be a compact subet of $\mathbf{R}^n$, and $f: K \to \mathbf{R}^m$ be a continuous mapping. Then, for an arbitrary $\varepsilon > 0$, there exist an integer N, an $m \times N$ matrix A, an $N \times n$ matrix B, and an N dimensional vector θ such that*

$$\max_{x \in K} | F(x) - A\sigma(Bx + \theta)| < \varepsilon \tag{11}$$

holds, where $\sigma: \mathbf{R}^N \to \mathbf{R}^N$ is a sigmoid mapping defined by

$$\sigma({}^t(u_1, \ldots, u_N)) = {}^t(\sigma(u_1), \ldots, \sigma(u_N)).$$

Similar results have been obtained by Cybenko (1989) and Hornik, Stinchcombe, and White (1989).

5. PRELIMINARY 2

In the following, we restate the basic facts of the theory of dynamical systems which are used in the proofs of our theorems (see, e.g., Hirsch & Smale, 1974).

Let D be an open subset of $\mathbf{R}^n$. A mapping $F: D \to \mathbf{R}^n$ is said to be Lipschitz on D if there exists a constant L such that

$$| F(x) - F(y)| \le L|x - y| \tag{12}$$

for all $x, y \in D$. We call L a Lipschitz constant for F. We call F locally Lipschitz if each point of D has a neighborhood D_0 in D such that the restriction $F \mid D_0$ is Lipschitz.

LEMMA 1. *Let a mapping $F: D \to \mathbf{R}^n$ be C^1. Then F is locally Lipschitz. Moreover, if $A \subset D$ is compact, then the restriction $F \mid A$ is Lipschitz. (For proof, see Hirsch & Smale, 1974, chap. 8, §3. Lemma and §6. Lemma).*

LEMMA 2. *Let $F: D \to \mathbf{R}^n$ be a C^1-mapping and $x_0 \in D$. Then there is some $a > 0$ and a unique solution x: $(-a, a) \to D$ of the differential equation*

$$\dot{x} = F(x) \tag{13}$$

satisfying the initial condition $x(0) = x_0$. (For proof, see Hirsch & Smale, 1974, chap. 8, §2. Theorem 1).

LEMMA 3. *Let D be an open subset of $\mathbf{R}^n$ and $F: D \to \mathbf{R}^n$ be a C^1-mapping. Let $x(t)$ be a solution on a maximal open interval $J = (\alpha, \beta) \subset \mathbf{R}$ with $\beta < \infty$. Then for any given compact subset $K \subset D$, there is some $t \in (\alpha, \beta)$ with $x(t) \notin K$. (For proof, see Hirsch & Smale, 1974, chap. 8, §5. Theorem).*

LEMMA 4. *Let $F: \mathbf{R}^n \to \mathbf{R}^n$ be a bounded C^1-mapping. Then, the differential equation*

$$\dot{x} = -\frac{1}{\tau} x + F(x), \tag{14}$$

where $\tau > 0$ has a unique solution on $[0, \infty)$.

Proof. From the assumption, we can take a constant $M > 0$ such that

$$| F_i(x)| \le M \quad (\forall i = 1, \ldots, n) \tag{15}$$

for all $x \in \mathbf{R}^n$. By comparing the solution $x(t)$ with solutions of the following equations

$$\dot{y} = -\frac{1}{\tau} y + M, \tag{16}$$

$$\dot{y} = -\frac{1}{\tau} y - M, \tag{17}$$

we can easily show that

$$| x_i(t)| \le \max\{| x_i(0)|, \tau M\} = C_i. \tag{18}$$

If we set $C = \max\{C_i\}$, then the solution $x(t)$ satisfies

$$| x(t)| \le \sqrt{n} C \tag{19}$$

on the existing interval of the solution. From Lemma 3 and Lemma 2, the solution $x(t)$ exists uniquely on the interval $[0, \infty)$. Q.E.D.

Lemma 4 guarantees that eqn (4) describing a recurrent neural network has a unique solution on $[0, \infty)$ because the output function σ is bounded and C^1.

LEMMA 5. *Let $F, \tilde{F}: D \to \mathbf{R}^n$ be Lipschitz continuous mappings and L be a Lipschitz constant of F. Suppose that for all $x \in D$,*

$$| F(x) - \tilde{F}(x) | < \varepsilon. \tag{20}$$

If $x(t)$, $y(t)$ are solutions to

$$\dot{x} = F(x) \tag{21}$$

$$\dot{y} = \tilde{F}(y), \tag{22}$$

respectively, on some interval J, such that $x(t_0) = y(t_0)$; then

$$| x(t) - y(t) | \le \frac{\varepsilon}{L}(\exp L | t - t_0 | - 1) \tag{23}$$

holds, for all $t \in J$. (For proof, see Hirsch & Smale, 1974, chap. 15, §1. Theorem 3).

6. PROOF OF THE THEOREMS

Using the above preliminaries, we will prove the theorems stated in Section 3.

Proof of Theorem 1.

Step 1.

For given $\varepsilon > 0$, we choose η so that $0 < \eta < \min\{\varepsilon, \lambda\}$, where λ is the distance between $\tilde{K}$ and the boundary ∂D of D. We set

$$K_\eta = \{x \in \mathbf{R}^n; \exists z \in \tilde{K}, | x - z | \le \eta\}; \tag{24}$$

then K_η is a compact subset of D, because $\tilde{K}$ is compact. Therefore, by Lemma 1, F is Lipschitz on K_η. We also choose $\varepsilon_1 > 0$ so that

$$\varepsilon_1 < \frac{\eta L_F}{2(\exp L_F T - 1)}, \tag{25}$$

where L_F is a Lipschitz constant of $F | K_\eta$.

By the fundamental approximation theorem, there exist an integer N, an $n \times N$ matrix A, an $N \times n$ matrix B and an N-dimensional vector θ such that

$$\max_{x \in K_\eta} | F(x) - A\sigma(Bx + \theta) | < \frac{\varepsilon_1}{2}. \tag{26}$$

We define a C^1-mapping $\tilde{F} : \mathbf{R}^n \to \mathbf{R}^n$ by

$$\tilde{F}(x) = -\frac{1}{\tau}x + A\sigma(Bx + \theta), \tag{27}$$

where τ is chosen large enough so that the following conditions are satisfied:

(a) $\forall x \in K_\eta$; $\left|\frac{x}{\tau}\right| < \frac{\varepsilon_1}{2}$

(b) $\left|\frac{\theta}{\tau}\right| < \frac{\eta L_{\tilde{G}}}{2(\exp L_{\tilde{G}} T - 1)}$ and $\left|\frac{1}{\tau}\right| < \frac{L_{\tilde{G}}}{2}$,

where $L_{\tilde{G}}$ is a constant and $L_{\tilde{G}}/2$ is a Lipschitz constant for the mapping $W\sigma : \mathbf{R}^{n+N} \to \mathbf{R}^{n+N}$, which will be defined later (W is defined by A and B).

Then, by eqns (26) and (27),

$$\max_{x \in K_\eta} | F(x) - \tilde{F}(x) | < \varepsilon_1 \tag{28}$$

holds. We set $x(t)$ and $\tilde{x}(t)$ as the solutions of the following equations:

$$\dot{x} = F(x), \tag{29}$$

$$\dot{\tilde{x}} = \tilde{F}(\tilde{x}), \tag{30}$$

with initial condition $x(0) = \tilde{x}(0) = x_0 \in K$, respectively. Then, by Lemma 5, for any $t \in I$,

$$| x(t) - \tilde{x}(t) | \le \frac{\varepsilon_1}{L_F}(\exp L_F t - 1)$$

$$\le \frac{\varepsilon_1}{L_F}(\exp L_F T - 1). \tag{31}$$

Therefore, by the condition of ε,

$$\max_{t \in I} | x(t) - \tilde{x}(t) | < \frac{\eta}{2} \tag{32}$$

holds.

Step 2.

We consider the following dynamical system defined by $\tilde{F}$ stated in Step 1.

$$\dot{\tilde{x}} = -\frac{1}{\tau}\tilde{x} + A\sigma(B\tilde{x} + \theta). \tag{33}$$

If we set $\tilde{y} = B\tilde{x} + \theta$, then

$$\dot{\tilde{y}} = B\dot{\tilde{x}} = -\frac{1}{\tau}\tilde{y} + C\sigma(\tilde{y}) + \frac{1}{\tau}\theta, \tag{34}$$

where $C = BA$ and C is an $N \times N$ matrix. We set

$$\tilde{z} = {}^t(\tilde{x}_1, \ldots, \tilde{x}_n, \tilde{y}_1, \ldots, \tilde{y}_N) \tag{35}$$

and we define a mapping $\tilde{G} : \mathbf{R}^{n+N} \to \mathbf{R}^{n+N}$ by

$$\tilde{G}(\tilde{z}) = -\frac{1}{\tau}\tilde{z} + W\sigma(\tilde{z}) + \frac{1}{\tau}\theta_1, \tag{36}$$

where W is an $(n + N) \times (n + N)$ matrix and θ_1 is an $(n + N)$ vector defined by

$$W = \begin{pmatrix} 0 & A \\ 0 & C \end{pmatrix}, \quad \theta_1 = \begin{pmatrix} 0 \\ \theta \end{pmatrix}, \tag{37}$$

respectively. Then, by Lemma 2, the first n components of the solution of the equation of

$$\dot{\tilde{z}} = \tilde{G}(\tilde{z}), \quad \tilde{y}(0) = B\tilde{x}(0) + \theta \tag{38}$$

are equivalent to the solution of the system (33).

Now, we define a mapping $G : \mathbf{R}^{n+N} \to \mathbf{R}^{n+N}$ by the use of τ and W stated above, as the following:

$$G(z) = -\frac{1}{\tau}z + W\sigma(z), \tag{39}$$

where

$$z = {}^t(u_1, \ldots, u_n, h_1, \ldots, h_N). \tag{40}$$

Then the dynamical system defined by G.

$$\dot{z} = -\frac{1}{\tau}z + W\sigma(z), \tag{41}$$

is realized by a recurrent neural network, if we set $u(t)$

$= {}^t(u_1(t), \ldots, u_n(t))$ as the internal state of n output units and $h(t) = {}^t(h_1(t), \ldots, h_N(t))$ as the internal state of N hidden units. As $\tilde{G}$ and G are C^1-mappings, and $\sigma'(x)$ is a bounded function, so the mapping $\tilde{z} \to W\sigma(\tilde{z})$ is Lipschitz on $\mathbf{R}^{n+N}$ and we set $L_{\tilde{G}}/2$ as its Lipschitz constant. Then, $L_{\tilde{G}}$ is a Lipschitz constant of $\tilde{G}$ because $L_{\tilde{G}}/2$ is a Lipschitz constant of $-\tilde{z}/\tau$ by the condition (b) of τ.

Using eqn (36), eqn (39) and the condition (b) of τ, we see that for any $z \in \mathbf{R}^{n+N}$,

$$|\tilde{G}(z) - G(z)| = \left|\frac{\theta}{\tau}\right| < \frac{\eta L_{\tilde{G}}}{2(\exp L_{\tilde{G}}T - 1)} \tag{42}$$

holds. Therefore, we set $\tilde{z}(t)$ and $z(t)$ as the solutions of the following equations, respectively:

$$\dot{\tilde{z}} = \tilde{G}(\tilde{z}), \quad \begin{cases} \tilde{x}(0) = x_0 \in K \\ \tilde{y}(0) = Bx_0 + \theta \end{cases} \tag{43}$$

$$\dot{z} = G(z), \quad \begin{cases} u(0) = x_0 \in K \\ h(0) = Bx_0 + \theta. \end{cases} \tag{44}$$

Then, by Lemma 5 we obtain

$$\max_{t \in I} |\tilde{z}(t) - z(t)| \le \frac{\eta}{2}, \tag{45}$$

and hence

$$\max_{t \in I} |\tilde{x}(t) - u(t)| \le \frac{\eta}{2} \tag{46}$$

holds, where $\tilde{x}(t)$ is the same as $\tilde{x}(t)$ in eqn (32).

Step 3.

Using eqn (32) and (46) stated above, for a given $\varepsilon > 0$, we can construct a recurrent neural network with internal state $z(t)$ by τ and W introduced above. For $x(t)$ satisfying eqn (5), if we set the initial state of the network by

$$u(0) = x(0) \quad \text{and}$$
$$h(0) = Bx(0) + \theta, \tag{47}$$

we get

$$\max_{t \in I} |x(t) - u(t)| \le \frac{\eta}{2} + \frac{\eta}{2} = \eta < \varepsilon. \tag{48}$$

Q.E.D.

REMARKS. The recurrent network constructed in the above proof has connections between hidden units as well as connections from hidden units to output units, but has no connection from output units to hidden units. The estimate using Lemma 5 is essential in the proof of Theorem 1. Therefore, by a similar estimate using Lemma 5, it follows that we can construct a recurrent network with very small connection weights from output units to hidden units which satisfies the statement of the Theorem, and so we omit the details here.

Proof of Theorem 2. Because the flow $\phi_t(x)$ of the dynamical system is a continuous mapping $\mathbf{R} \times D \to D$ $((t, x) \to \phi_t(x))$, (see Hirsch & Smale, 1974), the set $\tilde{K}$ of trajectories on time interval I whose initial points are in the compact set K:

$$\tilde{K} = \{x(t) \in \mathbf{R}^n;\ x(0) \in K,\ 0 \le t \le T\} \tag{49}$$

is a compact subset of D. By establishing correspondence of K and $\tilde{K}$ to K and $\tilde{K}$ in Theorem 1, respectively, our Theorem is proved. Q.E.D.

Proof of Corollary 1. By continuity of $\sigma^{-1} : (0, 1) \to \mathbf{R}$, $D_1 = \sigma^{-1}(D)$ is an open subset of $\mathbf{R}^n$, and $K_1 = \sigma^{-1}(K)$ is a compact subset of D^{-1}. For $x \in (0, 1)^n$, let $u \in \mathbf{R}^n$ be denoted by

$${}^t(u_1, \ldots, u_n) = \sigma^{-1}({}^t(x_1, \ldots, x_n)). \tag{50}$$

Then, by the sigmoid mapping σ, the given dynamical system $\dot{x} = F(x)$ on D is transformed to a dynamical system defined by

$$\frac{du_i}{dt} = \frac{1}{\sigma'(u_i)} F_i(\sigma(u_1), \ldots, \sigma(u_n))\ (i = 1, \ldots, n) \tag{51}$$

on $D_1 \subset \mathbf{R}^n$. From this fact, our Corollary can be easily proved by the use of Theorem 2. Q.E.D.

Proof of Theorem 3. Using a mollifier, we can take a C^∞-curve $\tilde{f} : (-\delta, T + \delta) \to \mathbf{R}^n$ for some $\delta > 0$ such that

$$\max_{t \in I} |f(t) - \tilde{f}(t)| < \frac{\varepsilon}{2}. \tag{52}$$

If we set $g(t) = (\tilde{f}(t), t) \in \mathbf{R}^n \times \mathbf{R} = \mathbf{R}^{n+1}$ for $t \in [0, T]$, then g is an injective mapping and so there exists a one-dimensional compact C^∞-submanifold M of $\mathbf{R}^{n+1}$ such that $g([0, T]) \subset M$.

Taking a tubular neighborhood V of M in $\mathbf{R}^{n+1}$ (see Hirsch, 1976, Theorem 5.1), we can easily construct a system of ordinary differential equations $\dot{x} = F(x)$ defined on V such that $F \in C^\infty$ on V and $g([0, T])$ is a part of a trajectory of the system with $x(0) = g(0)$. Using Theorem 1, there exists a recurrent network with $n + 1$ output units such that

$$\max_{t \in I} |g(t) - \tilde{u}(t)| < \frac{\varepsilon}{2}, \tag{53}$$

where $\tilde{u}(t) = {}^t(u_1(t), \ldots, u_{n+1}(t))$ is the internal state of output units. Considering the projection $\tilde{f}(t)$ of $g(t)$ to $\mathbf{R}^n$ by $\pi : \mathbf{R}^{n+1} \to \mathbf{R}^n({}^t(x_1, \ldots, x_{n+1}) \to {}^t(x_1, \ldots, x_n))$, we obtain a recurrent network with n output units whose internal state $u(t) = {}^t(u_1(t), \ldots u_n(t))$ satisfies

$$\max_{t \in I} |\tilde{f}(t) - u(t)| < \frac{\varepsilon}{2}. \tag{54}$$

Therefore, from eqns (52) and (54) we obtain

$$\max_{t \in I} |f(t) - u(t)| < \varepsilon. \tag{55}$$

Q.E.D.

7. CONCLUSIONS

We proved that the finite time trajectories of a given n-dimensional dynamical system are approximated by the internal states of output units of a recurrent neural network with n output units, N hidden units and appropriate initial states. The important point of the proof is the use of the fundamental approximation theorem to embed the given dynamical system into a higher dimensional dynamical system which defines a recurrent neural network. We consider one of the capability problems of continuous time recurrent neural networks to be solved in the form of an existence theorem of networks which approximate trajectories of a given dynamical system. As a corollary of our theorem, we also proved that any continuous curve can be approximated by the output of a recurrent neural network.

We consider that our theorems are the first step to study the capability problems of continuous time recurrent neural networks.

REFERENCES

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, **2**, 303–314.

Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, **2**(3), 183–191.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, **2**(5), 331–349.

Hirsch, M. W. (1976). *Differential topology*. New York: Springer-Verlag.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. San Diego, CA: Academic Press, Inc.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**(5), 359–366.

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Proceedings of the IJCNN*, **2**, 365–372.

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation*, **1**, 263–269.

Pineda, F. J. (1987). Generalization of backpropagation to recurrent neural networks. *Physical Review Letters*, **18**, 2229–2232.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing* (pp. 318–362). Cambridge, MA: MIT Press.

Sato, M. (1990). A learning algorithm to teach spatio-temporal patterns to recurrent neural networks. *Biological Cybernetics*, **62**, 259–263.

Sato, M., & Murakami, Y. (1991). Learning nonlinear dynamics by recurrent neural networks. In H. Kawakami (Ed.), *Proceedings of the Symposium on Some Problems on the Theory of Dynamical Systems in Applied Sciences, Advanced Series in Dynamical Systems*. Singapore: World Scientific, **10**, 49–64.

Seidl, D. R., & Lorenz, R. D. (1991). A structure by which a recurrent neural network can approximate a nonlinear dynamic system. *Proceedings of the IJCNN, Seattle, WA*, **2**, 709–714.

Williams, R. J., & Zipser, D. (1990). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, **1**, 256–263.