Representations of Dynamical Systems by Recurrent Neural Networks

MASAHIRO KIMURA

Abstract

Neural networks were orginally mathematical models made to elucidate the computational principles of networks of biological neurons, and it is natural to pursue the possibility of neural network computation from the point of view of computer science.

This paper considers continuous time recurrent neural networks (RNNs) called additive neural networks or Hopfield neural networks. Recently, there has been much interest in learning trajectories by RNNs. The goal of neural network training is not to learn the given data, but rather to build a model of the process which generates the data, that is, to realize the neural network model which exhibits good generalization. However, the previous works did not deal with the problem of generalization for trajectory learning by RNNs. In order to allow us to discuss this problem, this paper investigates the problem of approximating a dynamical system by an RNN as one extension of the problem of approximating a trajectory by an RNN.

The role of hidden units is crucial for the approximation capability, and an RNN with hidden units cannot produce a dynamical system on the visible state space unless a map is successfully specified from the visible state space to the hidden state space. We define affine neural dynamical systems (A-NDSs) as dynamical systems produced by RNNs, and propose A-NDS-based learning for approximating dynamical systems by RNNs. We also verify the validity of the A-NDS-based learning. Toward developing effective learning algorithms, we construct a unique parametric representation of n-dimensional A-NDSs, and concretely construct a non-redundant search set for learning dynamical systems by RNNs based on A-NDSs.

Contents

1 Introduction

Neural networks were orginally mathematical models made to elucidate the computational principles of networks of biological neurons, and it is natural to pursue the possibility of neural network computation from the point of view of computer science. Recurrent neural networks (RNNs) have connections that are allowed both ways between a pair of units and even from a unit to itself, while feed-forward neural networks (FNNs) have layered feed-forward structures. Recently, there has been much interest in potential abilities of RNNs (e.g., [10], [60], [19], [42]). On the one hand, attention has been devoted to exploiting RNNs without hidden units as associative memory models (e.g., [25], [26], [19]) or as machines to solve combinatorial optimization problems (e.g., [27], [28], [19]), and their convergent dynamics have been investigated (e.g., [8], [26], [20], [21], [7], [62], [22], [12], [58]). On the other hand, several learning algorithms of RNNs with hidden units have been proposed for approximating predetermined trajectories ([46], [10], [16], [40], [60], [43], [47], [59], [44], [52], [53], [17], [50], [55], [2], [42], [45]), and successfully used to memorize and regenerate various temporal sequences of patterns (e.g., [10], [11], [40], [41], [60], [48], [49], [57], [19], [45]). By a trajectory we mean a continuous curve in a Euclidean space, that is, a continuous map from an interval to the Euclidean space.

We consider the following continuous-time RNNs, called additive neural networks or Hopfield neural networks, since they are among the most common RNNs and their electrical circuit implementations are also investigated (e.g., [26], [21], [19], [22]): An RNN with n visible units consists of $n+r$ units (the units 1 to n are visible and the units $n+1$ to $n+r$ are hidden), and the state $u_i(t)$ of unit i at time t is governed by the system of ordinary differential equations:

$$\frac{du_i}{dt}(t) = -\frac{1}{\tau}u_i(t) + \sum_{j=1}^{n+r} w_{ij}g(u_j(t)) + w_{i0}, \qquad i = 1, \cdots, n+r, \tag{1.1}$$

where τ is a fixed positive constant called the time constant, r is a non-negative integral-valued parameter expressing the number of hidden units, w_{ij} $(i, j = 1, \cdots, n+r)$ is a real-valued parameter called the connection weight from unit j to unit i, w_{i0}

($i = 1, \cdots, n+r$) is also a real-valued parameter called the bias to unit i, and g is a fixed non-constant bounded C^∞-function called the the activation function. As the activation function $g(u)$, $\tanh(u)$ and $1/(1+e^{-u})$ are commonly used from the biological motivation. We mainly use $\tanh(u)$.

Learning temporal sequences of patterns widely occurs in biological systems, and the problem of constructing an RNN capable of generating a desired temporal sequence of patterns is fundamentally related with the applications of RNNs to spatiotemporal information processing such as robot arm control and speech recognition (e.g., [1], [10], [48], [19], [42]). Therefore, trajectory learning algorithms for RNNs have been extensively investigated as described above, and also applied to path planning and temporal pattern recognition ([5], [51]). However, the goal of network training is not to learn the given data itself, but rather to build a model of the process which generates the data (e.g., [19], [6]). This is because it is important whether the network can exhibit good *generalization*, that is, can make good predictions for new inputs. Since there exist almost infinitely many possible generalizations, mathematical studies of learning require the mathematical model of the process.

For the problem of learning trajectories on $\mathbf{R}^n$, it is fundamental to take a dynamical system on $\mathbf{R}^n$, that is, a one-parameter group of transformations of the C^∞ manifold $\mathbf{R}^n$, as the mathematical model of the process which generates the trajectories. This is because dynamical systems have been used to model many mechanical, chemical, biological, ecological and social systems, and their mathematical theory has been constructed (e.g., [23], [24]). For example, consider the behavior of a robot arm such that each movement of the robot arm is determined by a trajectory on the configuration space $\mathbf{R}^n$, and the trajectories corresponding to the desired behavior are generated from a dynamical system on $\mathbf{R}^n$ by specifying initial states.

In this paper, as one extension of the problem of approximating a predetermined trajectory on $\mathbf{R}^n$ by an RNN, we investigate the problem of approximating a dynamical system on $\mathbf{R}^n$ by an RNN. This allows us to discuss the problem of generalization for the trajectory learning by an RNN, and leads to investigating the problem of modeling a dynamical system on $\mathbf{R}^n$ using an RNN with n visible units. Hence, within the framework of dynamical system learning for RNNs, we should say that an RNN with n visible units approximates a dynamical system on $\mathbf{R}^n$ if the RNN produces on the visible state space $\mathbf{R}^n$ such a dynamical system that approximates the target dynamical system.

In trajectory learning, an RNN with n visible units and r hidden units is regarded as such a machine that generates the visible state trajectory $\mathbf{R} \ni t \mapsto (u_1(t), \cdots, u_n(t)) \in \mathbf{R}^n$ for a given initial state $p = (p_1, \cdots, p_{n+r}) \in \mathbf{R}^{n+r}$. Note that the role of hidden units is crucial for the approximation capability. Funahashi and Nakamura ([15]) proved that for any dynamical system ψ on $\mathbf{R}^n$, given constants $\varepsilon > 0$, $T > 0$, and a compact set $K \subset \mathbf{R}^n$, there exist an integer $r > 0$, a constant $\tau > 0$, and an RNN of time constant τ equipped with n visible units and r hidden units such that the trajectory of ψ starting at a point of K can be approximated within error ε by a visible state trajectory of the RNN in time interval $[0, T]$. However, the visible state trajectories of the RNN used by them to approximate the trajectories of the target dynamical system were not generated from one dynamical system on $\mathbf{R}^n$. For example, those visible state trajectories might have self-intersections and intersect each other with nonzero angles.

An RNN with n visible units and $r(> 0)$ hidden units cannot produce a dynamical system on $\mathbf{R}^n$ unless a map $h : \mathbf{R}^n \to \mathbf{R}^r$ is successfully specified to determine the initial states of the hidden units for initial states of the visible units. Therefore, toward developing dynamical system learning algorithms, it is necessary to build the framework of how such an RNN produces a dynamical system on $\mathbf{R}^n$ under map $h : \mathbf{R}^n \to \mathbf{R}^r$ to approximate a given dynamical system on $\mathbf{R}^n$. In this paper, we propose the notion of n-dimensional *affine neural dynamical systems* (A-NDSs) to this framework, that is, propose A-NDS-based learning.

An n-dimensional A-NDS is the dynamical system on $\mathbf{R}^n$ that an RNN with n

visible units and r hidden units can produce on the visible state space $\mathbf{R}^n$ under an affine map $h_A : \mathbf{R}^n \to \mathbf{R}^r$, where the case $r = 0$ is naturally included. As far as we know, an n-dimensional A-NDS is the only universal example of the dynamical system that such an RNN produces on $\mathbf{R}^n$ under a C^∞ map $h : \mathbf{R}^n \to \mathbf{R}^r$. We prove that any dynamical system on $\mathbf{R}^n$ can be approximated well by an n-dimensional A-NDS in a given finite region. These facts support the validity of the A-NDS-based learning for approximating dynamical systems by RNNs (see Theorem 3.6).

An A-NDS is represented by a suitable pair of an RNN and an affine map. However, this representation is not unique. The aim of approximating a dynamical system ψ by an RNN is to acquire an RNN model of ψ by learning the observed trajectory data of ψ, and learning trajectories by an RNN is basically performed under the gradient descent method for the error function which measures the discrepancy between the desired trajectories and the visible state trajectories of the RNN (e.g., [2], [42]). Thus, the local minima problem is associated with the learning algorithms and makes the performance dependent on the initial values of the learning parameters. From the point of view of developing effective learning algorithms, it is important to investigate a non-redundant search set for learning dynamical systems by RNNs based on A-NDSs, that is, to investigate the non-redundant representations of A-NDSs by the pairs of RNNs and affine maps. For example, such a non-redundant search set helps restrict the initial values of the learning parameters.

In this paper, we construct a unique parametric representation of n-dimensional A-NDSs by extending Sussmann's work ([54]) for the redundancy in the function representations of FNNs (see Theorem 4.7). Moreover, we construct the non-redundant representations of n-dimensional A-NDSs by the pairs of RNNs with n visible units and affine maps (see Theorem 5.1).

In general, simpler models are preferred to more complicated models to make analysis and control easier. In order to obtain a simpler RNN model of a target dynamical system, we consider simplifying the RNN model that has learned the dynamical system based on an A-NDS. As a solution, we give a method of obtaining all the minimal RNN models of an A-NDS from a given RNN model of the A-NDS. The method is reached by the straightforward application of the unique parametric representation of A-NDSs.

This paper is organized as follows: In Section 2, we describe the definitions and notation for RNNs and dynamical systems, and also discuss the issues of learning a dynamical system by an RNN from the given trajectory data. In Section 3, we define A-NDSs, and verify the validity of the A-NDS-based learning. In Section 4, we construct a unique parametric representation of n-dimensional A-NDSs. In Section 5, as the application of the unique parametric representation, we concretely construct a non-redundant search set for learning dynamical systems by RNNs based on A-NDSs, and give a method of obtaining all the minimal RNN models of an A-NDS from a given RNN model of the A-NDS. In Section 6, as a first step to a study of generalization for trajectory learning by RNNs, we investigate whether or not an A-NDS can be identified from its given trajectories. Finally, Section 7 concludes the paper.

We have obtained the results of Section 3 in our papers ([30], [31], [34], [35]). Also, we have obtained the results of Section 4 in our papers ([32], [36]). The results of Section 5 include those of our paper ([33]). We have obtained the results of Section 6 in our paper ([34]).

2 Preliminaries

2.1 Recurrent neural networks

Let us prepare notation together with some definitions.

An RNN with m units is parametrized by a matrix

$$W = (w_{ij})_{i=1,\cdots,m;\, j=0,1,\cdots,m} \in M_{m,m+1}(\mathbf{R})$$

of connection weights and biases. Here, $M_{k,\ell}(\mathbf{R})$ is the set of $k \times \ell$ real matrices. Hence, we denote by $\mathcal{N}_r^n(W)$ the RNN with n visible units and r hidden units such that the matrix of connection weights and biases is W.

Consider an RNN $\mathcal{N}_r^n(W)$, where $m = n + r$. The space of possible states of all the units is referred to as the *state space* of the RNN, which is the Euclidean space $\mathbf{R}^m$. The space of possible states of the visible units is referred to as the *visible state space* of the RNN, which is the Euclidean space $\mathbf{R}^n$. The space of possible states of the hidden units is referred to as the *hidden state space* of the RNN, which is the Euclidean space $\mathbf{R}^r$. The state space $\mathbf{R}^m$, the visible state space $\mathbf{R}^n$ and the hidden state space $\mathbf{R}^r$ have the following relation:

$$\mathbf{R}^m = \mathbf{R}^n \times \mathbf{R}^r.$$

When initial state $p = (p_1, \cdots, p_m) \in \mathbf{R}^m$ is given to the RNN, it outputs the trajectory

$$u^{\mathrm{v}}(t) = (u_1(t), \cdots, u_n(t))$$

called the *visible state trajectory* of the RNN with initial state p. Here, the trajectory

$$u(t) = (u_1(t), \cdots, u_m(t))$$

is the solution curve of system (1.1) of ordinary differential equations under initial condition $u(0) = p$, and is called the *state trajectory* of the RNN with initial state p. The trajectory

$$u^{\mathrm{h}}(t) = (u_{n+1}(t), \cdots, u_{n+r}(t))$$

is called the *hidden state trajectory* of the RNN with initial state p.

In order to simplify expression, we often use vector expression. Let us introduce notation. For an arbitrary positive integer ℓ, we regard a function f on $\mathbf{R}$ as the C^∞ map from $\mathbf{R}^\ell$ to $\mathbf{R}^\ell$ defined by

$$f(y) = (f(y_1), \cdots, f(y_\ell)), \qquad y = (y_1, \cdots, y_\ell) \in \mathbf{R}^\ell,$$

and regard an element of $\mathbf{R}^\ell$ as a column vector whenever we use vector expression and matrix operation. For $W = (w_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 1, \cdots, m; j = 0, 1, \cdots, m)$, we define

$$\begin{aligned} \overline{W} &= (w_{ij})_{i,j=1,\cdots,m} \in M_{m,m}(\mathbf{R}), \\ w_0 &= (w_{10}, \cdots, w_{m0}) \in \mathbf{R}^m, \end{aligned}$$

that is,

$$W = \begin{pmatrix} w_0 & \overline{W} \end{pmatrix},$$

and define the vector field F_W on $\mathbf{R}^m$ by

$$F_W(u) = -\frac{1}{\tau}u + \overline{W}g(u) + w_0, \qquad u \in \mathbf{R}^m. \tag{2.1}$$

Consider the RNN $\mathcal{N}_r^n(W)$ with $m = n+r$ units. Then, dynamics (1.1) of the RNN $\mathcal{N}_r^n(W)$ is described by

$$\frac{du}{dt} = F_W(u).$$

2.2 Dynamical systems

We recall some basic notions and facts for dynamical systems on a Euclidean space $\mathbf{R}^\ell$ (e.g., [38], [23]).

A dynamical system ψ on $\mathbf{R}^\ell$ is a C^∞ map $\psi : \mathbf{R} \times \mathbf{R}^\ell \to \mathbf{R}^\ell$ satisfying the following conditions: Define $\psi^t : \mathbf{R}^\ell \to \mathbf{R}^\ell$, $(t \in \mathbf{R})$ by $\psi^t(p) = \psi(t,p)$, $(p \in \mathbf{R}^\ell)$, then ψ^0 is the identity map on $\mathbf{R}^\ell$, and $\psi^s \circ \psi^t = \psi^{s+t}$, $(s, t \in \mathbf{R})$. The dynamical system ψ on $\mathbf{R}^\ell$ defines a C^∞ vector field Y on $\mathbf{R}^\ell$; i.e., for each $p \in \mathbf{R}^\ell$ define $Y(p) \in \mathbf{R}^\ell$ by $Y(p) = (d/dt)|_{t=0}\psi^t(p)$. Note that the curve $\mathbf{R} \ni t \mapsto \psi^t(p) \in \mathbf{R}^\ell$ is the solution curve of the ordiniary differential equation

$$\frac{dy}{dt} = Y(y) \tag{2.2}$$

under initial condition $y(0) = p$. The curve $\psi^t(p)$ is referred to as the *trajectory* of the dynamical system ψ through p.

Conversely, let Y be a C^∞ vector field on $\mathbf{R}^\ell$, and consider ordinary differential equation (2.2). For each $p \in \mathbf{R}^\ell$, let $\theta_p(t)$ be the solution curve of ordinary differential equation (2.2) under initial condition $\theta_p(0) = p$. The vector field Y is said to be *complete* if for each $p \in \mathbf{R}^\ell$, the curve $\theta_p(t)$ is well-defined for any $t \in \mathbf{R}$. When Y is complete, we can define the dynamical system ψ on $\mathbf{R}^\ell$ by $\psi^t(p) = \theta_p(t)$ for any $t \in \mathbf{R}$ and any $p \in \mathbf{R}^\ell$.

Consequently, the notion of dynamical systems on $\mathbf{R}^\ell$ is equivalent to the notion of complete C^∞ vector fields on $\mathbf{R}^\ell$. The following well-known result is used to prove our approximation theorem in Section 3.

Lemma 2.1. *Let ψ_1, ψ_2 be dynamical systems on $\mathbf{R}^\ell$, and Y_1, Y_2 the vector fields on $\mathbf{R}^\ell$ corresponding to ψ_1, ψ_2, respectively. Suppose that Y_1 satisfies a Lipshitz condition with constant L_1, and there exists $\delta > 0$ such that*

$$||Y_1(p) - Y_2(p)|| \leq \delta, \qquad p \in \mathbf{R}^\ell.$$

Then, we have

$$||\psi_1^t(p) - \psi_2^t(p)|| \leq \frac{\delta}{L_1}\left(e^{L_1|t|} - 1\right), \qquad t \in \mathbf{R},\ p \in \mathbf{R}^\ell,$$

where $||\cdot||$ denotes the Euclidean norm on $\mathbf{R}^\ell$.

2.3 Issues of dynamical system learning

Let us consider approximating a dynamical system by an RNN. There exist many dynamical systems that RNNs without hidden units cannot approximate well. In fact, it is known ([4]) that the numbers of equilibrium solutions of ordinary differential equation (1.1) for $n + r = 1$ and $n + r = 2$ are not more than 3 and 9, respectively. For example, no RNN consisting of only 1 unit can approximate well the dynamical system on $\mathbf{R}^1$ defined by the ordinary differential equation $dx/dt = (x+1)(x-1)(x+2)(x-2)$. Therefore, we must consider RNNs with hidden units.

Let us consider approximating a dynamical system on $\mathbf{R}^n$ by an RNN $\mathcal{N}_r^n(W)$ with hidden units. Note first that an RNN with hidden units does not define a dynamical system on the visible state space $\mathbf{R}^n$ although such RNNs have a greater potential for representing dynamical systems than RNNs without hidden units. In fact, when an element of $\mathbf{R}^n$ is input to an RNN $\mathcal{N}_r^n(W)$ as initial state of a dynamical system, the initial states of the hidden units cannot be specified, while the initial states of the visible units are specified. Namely, the RNN $\mathcal{N}_r^n(W)$ cannot output a trajectory. To make such an RNN produce a dynamical system on $\mathbf{R}^n$, we must successfully determine the initial states of the hidden units for initial states of the visible units. In what follows, we will see that this is challenging work. Therefore, there is the essential difference between the problem of approximating a dynamical system by an RNN and that of approximating some trajectories by an RNN.

Consider approximating a dynamical system ψ on $\mathbf{R}^n$ by an RNN $\mathcal{N}_r^n(W)$ using the given trajectories of ψ as training data. The following example shows that straightforward application of the existing trajectory learning algorithms is not successful in approximating ψ by $\mathcal{N}_r^n(W)$.

Example 1. Let ψ be the dynamical system on $\mathbf{R}^2$ defined by the system of ordinary differential equations

$$\begin{aligned}\frac{dx_1}{dt} &= -x_1 - \tanh(2x_1 + x_2)\\ \frac{dx_2}{dt} &= -x_2 + \tanh(x_1).\end{aligned}$$

Let $\xi(t)$ be the trajectory of the dynamical system ψ through $p = (0, 1) \in \mathbf{R}^2$; i.e., $\xi(t) = \psi^t(p)$ (see Figure 1). Consider the problem of approximating the trajectory $\xi(t)$ by an RNN using the existing trajectory learning algorithms. It turns out that the RNN $\mathcal{N}_1^2(W)$ with $\tau = 1$ and $g(s) = \tanh(s)$ is one of the solutions of this problem, where

$$W = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 \end{pmatrix},$$

that is, the dynamics of the RNN $\mathcal{N}_1^2(W)$ is described by

$$\begin{aligned}\frac{du_1}{dt} &= -u_1 - \tanh(u_3)\\ \frac{du_2}{dt} &= -u_2 + \tanh(u_1)\\ \frac{du_3}{dt} &= -u_3 + \tanh(u_1) - 2\tanh(u_3).\end{aligned}$$

In fact, if the initial state of the hidden unit 3 for the initial state p of the visible units is set to the value 1, then it is shown that the visible state trajectory $u^{\mathrm{v}}(t)$ of $\mathcal{N}_1^2(W)$ for initial state $(0, 1, 1)$ coincides with the given trajectory $\xi(t)$. Based on this fact, let us assume that the initial state of the hidden unit for initial states of the visible units is always the value 1. Then, the RNN $\mathcal{N}_1^2(W)$ does not produce a dynamical system on $\mathbf{R}^2$. In fact, for $q = (1/10, -1/20) \in \mathbf{R}^2$, the visible state trajectory of the RNN $\mathcal{N}_1^2(W)$ with initial state $(1/10, -1/20, 1)$ is never a trajectory of one dynamical system on $\mathbf{R}^2$ since it has a self-intersection (see Figure 1). Hence, we cannot make an RNN approximate the dynamical system ψ in this manner.

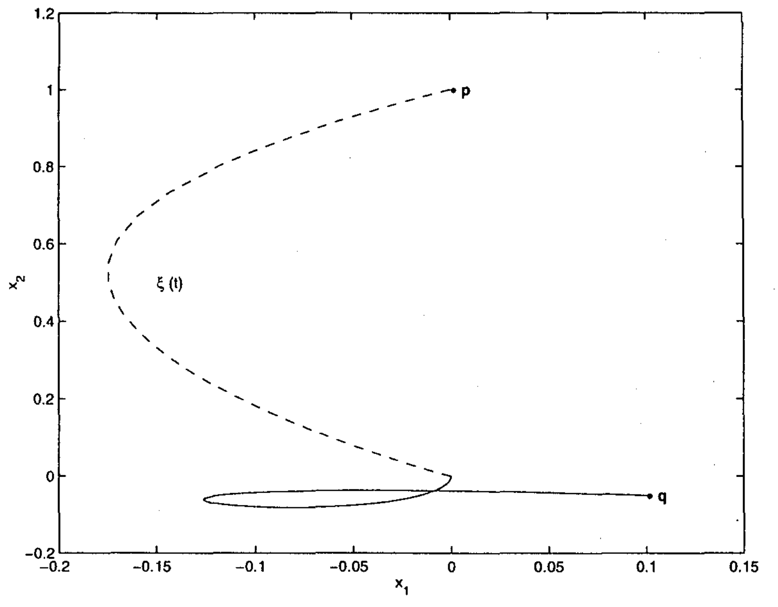


Figure 1. An unsuccesful example of learning dynamical system ψ by an RNN.

Dashed line: the visible state trajectory of the RNN with initial state $(0, 1, 1)$.

Solid line: the visible state trajectory of the RNN with initial state $(1/10, -1/20, 1)$.

3 Neural Dynamical Systems

In this section, we begin with a systematic investigation of the dynamical system produced by an RNN $\mathcal{N}_r^n(W)$ with $m = n + r$ units, and define the notion of *affine neural dynamical systems (A-NDSs)*. For approximating dynamical systems by RNNs, we propose A-NDS-based learning and verify its validity.

Definition. A dynamical system φ on $\mathbf{R}^n$ is said to be *produced* by RNN $\mathcal{N}_r^n(W)$ if for an arbitrary $x \in \mathbf{R}^n$, the trajectory $\varphi^t(x)$ of φ through x coincides with some visible state trajectory $u^{\mathrm{v}}(t)$ of $\mathcal{N}_r^n(W)$ with initial state $p \in \mathbf{R}^m$.

3.1 RNNs without hidden units

First, we consider RNNs with no hidden units, that is, we consider the RNNs with m units written in the form $\mathcal{N}_0^m(W)$, $W \in M_{m,m+1}(\mathbf{R})$. In this case, $n = m$, $r = 0$, and the visible state space coincides with the state space $\mathbf{R}^m$, We will see that these RNNs produce dynamical systems on $\mathbf{R}^m$ uniquely.

Proposition 3.1. *For each* $W = (w_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 1, \cdots, m;\ j = 0, 1, \cdots, m)$, *the vector field* F_W *on* $\mathbf{R}^m$ *is complete.*

Proof. Let $p = (p_1, \cdots, p_m) \in \mathbf{R}^m$. Consider the following initial value problem of ordinary differential equation on $\mathbf{R}^m$ with unkown functions $u(t) = (u_1(t), \cdots, u_m(t))$:

$$\begin{aligned} \frac{du}{dt} &= F_W(u), \\ u(0) &= p. \end{aligned}$$

It is sufficient to prove that the solution curve to this initial value problem can be extended to $(-\infty, \infty)$. Let $\xi(t) = (\xi_1(t), \cdots, \xi_m(t))$, $t \in (a, b)$, be the maximal solution curve of this initial value problem. Because it is known ([15]) that $b = \infty$, it suffices to prove $a = -\infty$.

Suppose $a > -\infty$. There exists a positive constant c such that

$$\left| \sum_{j=1}^{m} w_{ij}\, g(u_j) + w_{i0} \right| \leq c, \qquad u_1, \cdots, u_m \in \mathbf{R},\ i = 1, \cdots, m \tag{3.1}$$

since the activation function g is bounded. For $i = 1, \cdots, m$, we define the C^∞ functions v_i^+ and v_i^- on $\mathbf{R}$ by

$$
\begin{aligned}
v_i^+(t) &= (p_i + \tau c)e^{-t/\tau} - \tau c, \qquad t \in \mathbf{R}, \\
v_i^-(t) &= (p_i - \tau c)e^{-t/\tau} + \tau c, \qquad t \in \mathbf{R}.
\end{aligned}
$$

We will prove that

$$
v_i^-(t) \leq \xi_i(t) \leq v_i^+(t), \qquad t \in (a, 0],\ i = 1, \cdots, m. \tag{3.2}
$$

Hence, there exists a compact subset K of $\mathbf{R}^m$ such that

$$
\xi(t) \in K, \qquad t \in (a, 0].
$$

This contradicts $a > -\infty$ (e.g., [23]). This result proves the proposition.

Let us prove inequalities (3.2). First, we put

$$
z_i(t) = v_i^+(t) - \xi_i(t), \qquad t \in (a, 0].
$$

It is easily seen that the C^∞ curve $z(t) = (z_1(t), \cdots, z_m(t))$, $t \in (a, 0]$ is the solution curve of the following initial value problem of ordinary differential equation on $\mathbf{R}^m$:

$$
\begin{aligned}
\frac{du_i}{dt}(t) &= f_i(t, u_1(t), \cdots, u_m(t)), \qquad i = 1, \cdots, m, \\
u_i(0) &= 0, \qquad i = 1, \cdots, m,
\end{aligned} \tag{3.3}
$$

where each f_i is the C^∞ function on $(a, 0] \times \mathbf{R}^m$ defined by

$$
\begin{aligned}
f_i(t, u_1, \cdots, u_m) &= -\frac{u_i}{\tau} - q_i(t), \qquad t \in (a, 0],\ u_1, \cdots, u_m \in \mathbf{R}, \\
q_i(t) &= c + \sum_{j=1}^{m} w_{ij}\, g(\xi_j(t)) + w_{i0}, \qquad t \in (a, 0].
\end{aligned}
$$

For non-negative integer k, we inductively define the C^∞ curve on $\mathbf{R}^m$, $z^{(k)}(t) = (z_1^{(k)}(t), \cdots, z_m^{(k)}(t))$, $t \in (a, 0]$ as follows: For $i = 1, \cdots, m$ and $t \in (a, 0]$,

$$
\begin{aligned}
z_i^{(0)}(t) &= 0, \\
&\cdots, \\
z_i^{(k)}(t) &= \int_0^t f_i(s, z^{(k-1)}(s))\, ds, \\
&\cdots .
\end{aligned}
$$

Then, for $i = 1, \cdots, m$ and $t \in (a, 0]$,

$$\begin{aligned}
\left|z_i^{(1)}(t) - z_i^{(0)}(t)\right| &\le \int_t^0 |q_i(s)|\, ds \le 2c\,|t|, \\
\left|z_i^{(k)}(t) - z_i^{(k-1)}(t)\right| &\le \int_t^0 \left|f_i(s, z^{(k-1)}(s)) - f_i(s, z^{(k-2)}(s))\right| ds \\
&\le \int_t^0 \left|z_i^{(k-1)}(s) - z_i^{(k-2)}(s)\right| ds, \qquad 2 \le k \in \mathbf{N},
\end{aligned}$$

where $\mathbf{N}$ denotes the set of all natural numbers. By induction on k, we can prove

$$\left|z_i^{(k)}(t) - z_i^{(k-1)}(t)\right| \le 2c\frac{|t|^k}{k!}, \qquad t \in (a, 0],\ i = 1, \cdots, m,\ k \in \mathbf{N}.$$

Thus, we have

$$\left|z_i^{(k)}(t) - z_i^{(k-1)}(t)\right| \le 2c\frac{|a|^k}{k!}, \qquad t \in (a, 0],\ i = 1, \cdots, m,\ k \in \mathbf{N}.$$

Hence, there exists a curve on $\mathbf{R}^m$, $\tilde{z}(t) = (\tilde{z}_1(t), \cdots, \tilde{z}_m(t))$, $t \in (a, 0]$ such that $z^{(k)}(t)$ uniformly converges to $\tilde{z}(t)$ on $(a, 0]$. Note that $\tilde{z}(t)$ is a continuous curve since each $z^{(k)}(t)$ is continuous curve. From

$$\lim_{k\to\infty} z_i^{(k)}(t) = \int_0^t \lim_{k\to\infty} f_i(s, z^{(k-1)}(s))\, ds$$

for $t \in (a, 0]$ and $i = 1, \cdots, m$, we obtain

$$\tilde{z}_i(t) = \int_0^t f_i(s, \tilde{z}(s))\, ds, \qquad t \in (a, 0],\ i = 1, \cdots, m.$$

It is easily seen that

$$\tilde{z}_i(0) = 0, \quad \frac{d\tilde{z}_i}{dt}(t) = f_i(t, \tilde{z}(t)), \qquad t \in (a, 0],\ i = 1, \cdots, m.$$

By the uniqueness of the solution to initial value problem (3.3) of ordinary differential equation, we obtain

$$\tilde{z}(t) = z(t), \qquad t \in (a, 0].$$

Hence, $z^{(k)}(t)$ converges to $z(t)$ for all $t \in (a, 0]$. From inequality (3.1) we have

$$\begin{aligned}
z_i^{(k)}(t) &= \int_0^t \left[-z_i^{(k-1)}(s) - q_i(s)\right] ds, \qquad t \in (a, 0],\ i = 1, \cdots, m,\ k \in \mathbf{N}, \\
q_i(t) &\ge 0, \qquad t \in (a, 0],\ i = 1, \cdots, m.
\end{aligned}$$

Using induction on k it is derived that $z_i^{(k)}(t) \geq 0$ for all $t \in (a, 0]$, $i = 1, \cdots, m$, and $k \in \mathbf{N}$. Hence,

$$z_i(t) \geq 0; \text{ i.e., } \xi_i(t) \leq v_i^+(t), \qquad t \in (a, 0], \ i = 1, \cdots, m.$$

In the same way, the following inequality is proved:

$$v_i^-(t) \leq \xi_i(t), \qquad t \in (a, 0], \ i = 1, \cdots, m. \ \blacksquare$$

Proposition 3.1 implies that the vector field F_W defines a dynamical system on $\mathbf{R}^m$. We denote this dynamical system by

$$\Phi_W \ : \ \mathbf{R} \times \mathbf{R}^m \to \mathbf{R}^m.$$

Note that the trajecotry $\Phi_W^t(p)$ of the dynamical system Φ_W through p coincides with the state trajectory $u(t)$ of the RNN $\mathcal{N}_0^m(W)$ with initial state p, and is also the visible state trajectory. Hence, the RNN $\mathcal{N}_0^m(W)$ without hidden units always produces the dynamical system Φ_W on $\mathbf{R}^m$.

The following proposition shows that the RNNs consisting of m units without hidden units represent uniquely a certain class of dynamical systems on $\mathbf{R}^m$.

Proposition 3.2. *Let* $W = (w_{ij})$, $W' = (w'_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 0, 1, \cdots, m;$ $j = 1, \cdots, m)$. *Then,*

$$\Phi_W = \Phi_{W'} \quad \Rightarrow \quad W = W'.$$

Proof. Since a dynamical system and a complete vector field are equivalent notions, it suffices to prove that

$$F_W = F_{W'} \quad \Rightarrow \quad W = W'.$$

Suppose $F_W(u) = F_{W'}(u)$, $(u \in \mathbf{R}^m)$. Then, for each i, we have

$$\sum_{j=1}^{m} (w_{ij} - w'_{ij})\, g(u_j) + (w_{i0} - w'_{i0}) \ = \ 0, \qquad u = (u_1, \cdots, u_m) \in \mathbf{R}^m.$$

Differentiating this equation with respect to u_j, we obtain

$$(w_{ij} - w'_{ij}) \frac{dg}{ds}(u_j) = 0, \qquad u_j \in \mathbf{R},\ j = 1, \cdots, m.$$

Since the activaton function g is not a constant function, we obtain

$$w_{ij} \;=\; w'_{ij}, \qquad i, j = 1, \cdots, m.$$

Then it is straightforward to show that $w_{i0} = w'_{i0}$, $(i = 1, \cdots, m)$. Hence, we obtain $W = W'$. ∎

3.2 RNNs with hidden units

Next, we consider RNNs with r hidden units; i.e., $m = n + r$, $r > 0$. As seen in Section 2.3, such an RNN does not define a dynamical system on the visible state space $\mathbf{R}^n$ unless we successfully determine the initial values of the hidden units for inital values of the visible units. We will propose *neural dynamical systems* (NDSs) as dynamical systems produced by RNNs with hidden units.

Consider an RNN $\mathcal{N}_r^n(W)$ with $m = n + r$ units, where $W \in M_{m,m+1}(\mathbf{R})$. Let φ be a dynamical system on $\mathbf{R}^n$ produced by the RNN $\mathcal{N}_r^n(W)$. We begin by observing how φ is described. We fix any $x \in \mathbf{R}^n$ and consider the trajectory $\varphi^t(x)$ of the dynamical system φ through x. By definition, there exists some $p \in \mathbf{R}^m$ such that the trajectory $\varphi^t(x)$ coincides with the visible state trajectory $u^{\mathrm{v}}(t)$ of the RNN $\mathcal{N}_r^n(W)$ with initial state p. Let $u(t)$ be the state trajectory of the RNN $\mathcal{N}_r^n(W)$ with initial state p. We denote by

$$\pi \;:\; \mathbf{R}^m \to \mathbf{R}^n$$

the projection from $\mathbf{R}^m = \mathbf{R}^n \times \mathbf{R}^r$ to $\mathbf{R}^n$. Since $x = \varphi^0(x) = u^{\mathrm{v}}(0) = \pi(u(0)) = \pi(p)$, there exists some $h(x) \in \mathbf{R}^r$ such that

$$p \;=\; (x, h(x)).$$

By definition, we have

$$\varphi^t(x) \;=\; u^{\mathrm{v}}(t) \;=\; \pi(u(t)), \qquad t \in \mathbf{R}.$$

Since $u(t) = \Phi^t_W(p)$ $(t \in \mathbf{R})$, and x is any point of $\mathbf{R}^n$, we obtain

$$\varphi^t(x) = \pi\left(\Phi^t_W(x, h(x))\right), \qquad t \in \mathbf{R},\ x \in \mathbf{R}^n.$$

Note that h is a map from the visible state space $\mathbf{R}^n$ to the hidden space $\mathbf{R}^r$, but it is not necessarily continuous.

Based on the above observation, we define in the following way an NDS as one dynamical system produced by the RNN $\mathcal{N}^n_r(W)$ with hidden units.

Definition. Let $W \in M_{m,m+1}(\mathbf{R})$ and $h : \mathbf{R}^n \to \mathbf{R}^r$ a C^∞ map from the visible state space $\mathbf{R}^n$ to the hidden space $\mathbf{R}^r$. We define the C^∞ map

$$\varphi_{W,h} : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^n$$

by

$$\varphi^t_{W,h}(x) = \pi\left(\Phi^t_W(x, h(x))\right), \qquad t \in \mathbf{R},\ x \in \mathbf{R}^n. \tag{3.4}$$

We call $\varphi_{W,h}$ the *n-dimensional neural dynamical system* (NDS) produced by the RNN $\mathcal{N}^n_r(W)$ under C^∞ map h if it forms a dynamical system on $\mathbf{R}^n$ (see Figure 2).

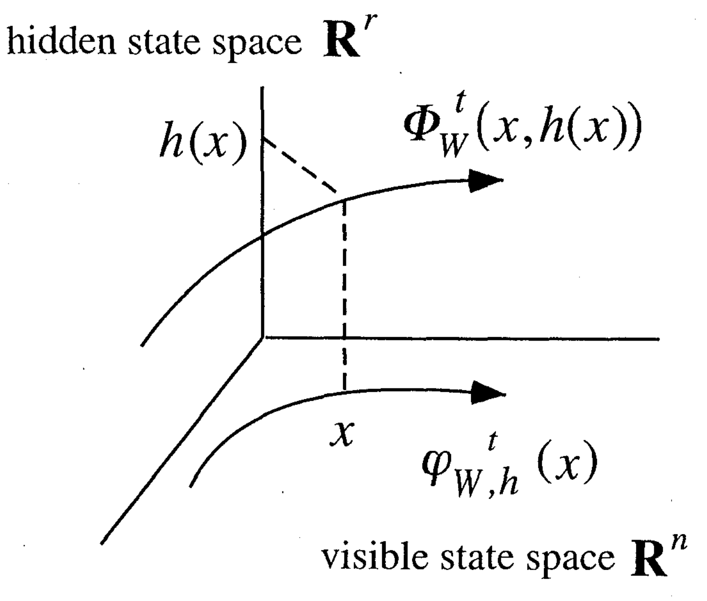


Figure 2. The definition of NDS $\varphi_{W,h}$.

We denote by $C^\infty(\mathbf{R}^n; \mathbf{R}^r)$ the set of all C^∞ maps from $\mathbf{R}^n$ to $\mathbf{R}^r$. Note that for any $W \in M_{m,m+1}(\mathbf{R})$ and any $h \in C^\infty(\mathbf{R}^n; \mathbf{R}^r)$, the C^∞ map $\varphi_{W,h}$ does not necessarily form a dynamical system on $\mathbf{R}^n$. In fact, there may exist a trajectory $\varphi^t_{W,h}(p)$ having a self-inetersection (see Figure 3 and Example 1), or there may exist distinct trajectories $\varphi^t_{W,h}(p)$ and $\varphi^t_{W,h}(q)$ intersecting with some nonzero angle (see Figure 4 and the following example).

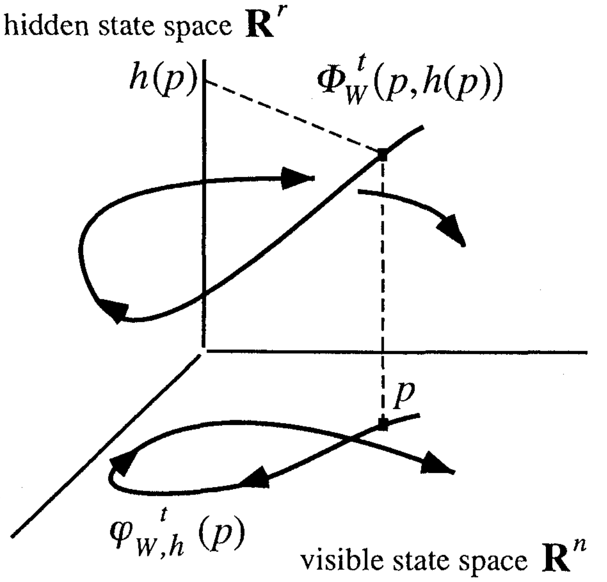


Figure 3. Case where C^∞ map $\varphi_{W,h}$ does not form a dynamical system: Self-intersection.

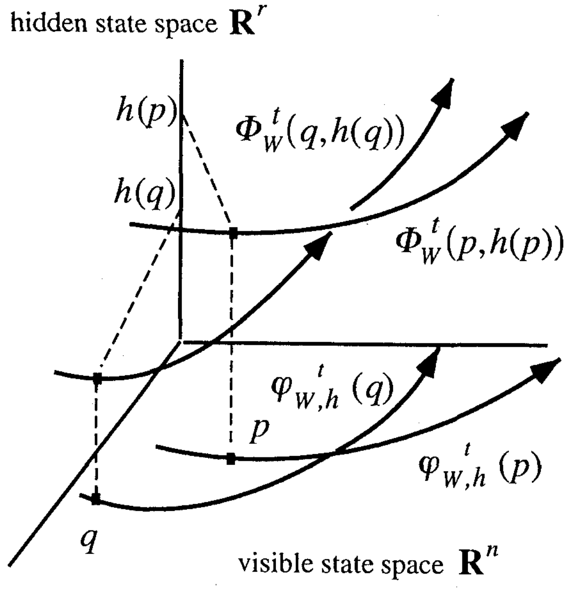


Figure 4. Case where C^∞ map $\varphi_{W,h}$ does not form a dynamical system: Intersection with a non-zero angle.

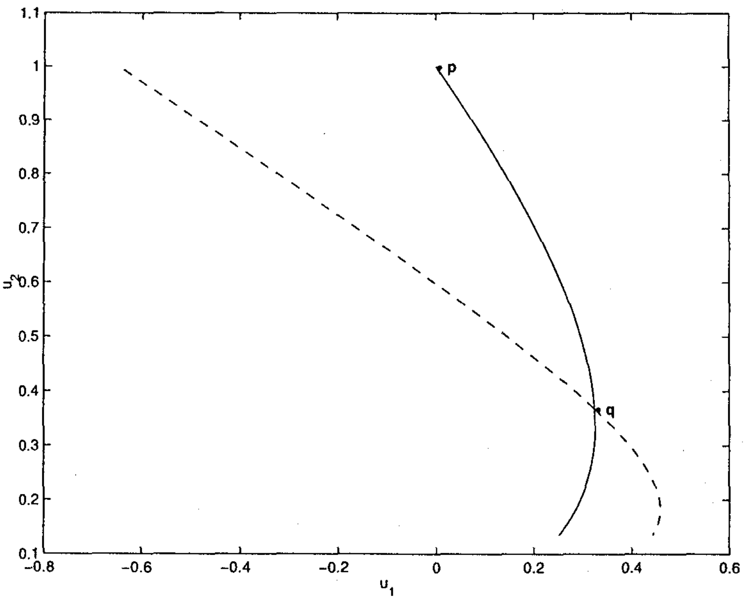
Example 2. Let $m = 3$, $n = 2$, $r = 1$, and let $W = (w_{ij}) \in M_{3,4}(\mathbf{R})$ be

$$w_{ij} = \begin{cases} 1, & (i = j = 1), \\ 0, & (\text{otherwise}). \end{cases}$$

We consider the RNN $\mathcal{N}_1^2(W)$ with $\tau = 1$ and $g(s) = \tanh(s)$. The dynamics of $\mathcal{N}_1^2(W)$ is written as follows:

$$\begin{aligned} \frac{du_1}{dt} &= -u_1 + \tanh(u_3) \\ \frac{du_2}{dt} &= -u_2 \\ \frac{du_3}{dt} &= -u_3. \end{aligned}$$

Let $h : \mathbf{R}^2 \to \mathbf{R}$ be the constant function 1. We put $p = (0, 1) \in \mathbf{R}^2$ and $q = \varphi_{W,h}^1(0, 1) \in \mathbf{R}^2$. Then, we can show that the two trajectories $\varphi_{W,h}^t(p)$ and $\varphi_{W,h}^t(q)$ intersect at the point q with some nonzero angle (see Figure 5). Hence, the C^∞ map $\varphi_{W,h}$ does not form a dynamical system on $\mathbf{R}^2$.



Fugure 5. Example where C^∞ map $\varphi_{W,h}$ does not form a dynamical system.

Solid line: the trajectory $\varphi^t_{W,h}(p)$.

Dashed line: the trajectory $\varphi^t_{W,h}(q)$.

The following proposition gives a necessary and sufficient condition to produce NDSs and the general form of NDSs. We use the following notation: For any $W = (w_{ij})_{i=1,\cdots,m;\, j=0,1,\cdots,m} \in M_{m,m+1}(\mathbf{R})$ with $m = n + r$, we set

$$\begin{aligned}
w_0^{\mathrm{v}} &= (w_{1\,0}, \cdots, w_{n\,0}) \in \mathbf{R}^n, \\
w_0^{\mathrm{h}} &= (w_{n+1\,0}, \cdots, w_{m\,0}) \in \mathbf{R}^r, \\
W^{\mathrm{v,v}} &= (w_{\alpha\beta})_{\alpha,\beta=1,\cdots,n} \in M_{n,n}(\mathbf{R}), \\
W^{\mathrm{v,h}} &= (w_{\alpha\, n+k})_{\alpha=1,\cdots,n;\, k=1,\cdots,r} \in M_{n,r}(\mathbf{R}), \\
W^{\mathrm{h,v}} &= (w_{n+k\,\alpha})_{k=1,\cdots,r;\, \alpha=1,\cdots,n} \in M_{r,n}(\mathbf{R}), \\
W^{\mathrm{h,h}} &= (w_{n+k\, n+\ell})_{k,\ell=1,\cdots,r} \in M_{r,r}(\mathbf{R}), \\
W^{\mathrm{v}} &= (w_{\alpha j})_{\alpha=1,\cdots,n;\, j=0,1,\cdots,m} \in M_{n,m+1}(\mathbf{R}),
\end{aligned}$$

that is,

$$W = \begin{pmatrix} w_0^{\mathrm{v}} & W^{\mathrm{v,v}} & W^{\mathrm{v,h}} \\ w_0^{\mathrm{h}} & W^{\mathrm{h,v}} & W^{\mathrm{h,h}} \end{pmatrix},$$

$$
\begin{aligned}
w_0 &= \begin{pmatrix} w_0^{\mathrm{v}} \\ w_0^{\mathrm{h}} \end{pmatrix}, \\
\overline{W} &= \begin{pmatrix} W^{\mathrm{v,v}} & W^{\mathrm{v,h}} \\ W^{\mathrm{h,v}} & W^{\mathrm{h,h}} \end{pmatrix}, \\
W^{\mathrm{v}} &= \begin{pmatrix} w_0^{\mathrm{v}} & W^{\mathrm{v,v}} & W^{\mathrm{v,h}} \end{pmatrix}.
\end{aligned}
$$

Proposition 3.3. *Let $W \in M_{m,m+1}(\mathbf{R})$ and $h \in C^{\infty}(\mathbf{R}^n; \mathbf{R}^r)$. Consider the RNN $\mathcal{N}_r^n(W)$ with $m = n+r$ units, $r > 0$. The RNN $\mathcal{N}_r^n(W)$ produces the n-dimensional NDS $\varphi_{W,h}$ under C^{∞} map h if and only if for any $x^0 \in \mathbf{R}^n$, the visible state trajectory $u^{\mathrm{v}}(t)$ and the hidden state trajectory $u^{\mathrm{h}}(t)$ of $\mathcal{N}_r^n(W)$ with initial state $(x^0, h(x^0)) \in \mathbf{R}^m$ satisfies*

$$g(h(u^{\mathrm{v}}(t))) - g(u^{\mathrm{h}}(t)) \in \operatorname{Ker} W^{\mathrm{v,h}}, \qquad t \in \mathbf{R},$$

that is,

$$W^{\mathrm{v,h}}\left(g(h(u^{\mathrm{v}}(t))) - g(u^{\mathrm{h}}(t))\right) = \{0\}, \qquad t \in \mathbf{R}.$$

Moreover, for the NDS $\varphi_{W,h}$ on $\mathbf{R}^n$, the corresponding vector field $X_{W,h}$ on $\mathbf{R}^n$ is given by

$$X_{W,h}(x) = -\frac{1}{\tau}x + W^{\mathrm{v,v}}g(x) + W^{\mathrm{v,h}}g(h(x)) + w_0^{\mathrm{v}}, \qquad x \in \mathbf{R}^n. \tag{3.5}$$

Proof. Define the vector field $X_{W,h}$ on $\mathbf{R}^n$ by equation (3.5). Since $\Phi_{W,h}^0$ is the identity map on $\mathbf{R}^m$ and

$$\left.\frac{d}{dt}\right|_{t=0} \Phi_{W,h}^t(x, h(x)) = F_W(x, h(x)), \qquad x \in \mathbf{R}^n,$$

we obtain by equation (3.4) that $\varphi_{W,h}^0$ is the identity mapping on $\mathbf{R}^n$ and

$$\left.\frac{d}{dt}\right|_{t=0} \varphi_{W,h}^t(x) = X_{W,h}(x), \qquad x \in \mathbf{R}^n.$$

As is well-known (e.g., [38]),

$$\varphi_{W,h}^s \circ \varphi_{W,h}^t = \varphi_{W,h}^{s+t}, \qquad s, t \in \mathbf{R},$$

if and only if for any $x^0 \in \mathbf{R}^n$,

$$\frac{d}{dt}\varphi_{W,h}^t(x^0) = X_{W,h}(\varphi_{W,h}^t(x^0)), \qquad t \in \mathbf{R}. \tag{3.6}$$

Hence, the C^∞ map $\varphi_{W,h}$ forms a dynamical system on $\mathbf{R}^n$ if and only if equation (3.6) is satisfied. We fix any $x^0 \in \mathbf{R}^n$, and consider the state trajectory $u(t)$, the visible state trajectory $u^{\mathrm{v}}(t)$ and the hidden state trajectory $u^{\mathrm{h}}(t)$ of the RNN $\mathcal{N}_r^n(W)$ with inital state $(x^0, h(x^0))$. Observe first that

$$\varphi^t_{W,h}(x^0) \;=\; u^{\mathrm{v}}(t), \qquad t \in \mathbf{R}.$$

Since $u(t) = (u^{\mathrm{v}}(t), u^{\mathrm{h}}(t))$ and $(du/dt)(t) = F_W(u(t))$ for any $t \in \mathbf{R}$, we have

$$\frac{du^{\mathrm{v}}}{dt}(t) \;=\; \pi\left(F_W(u^{\mathrm{v}}(t), u^{\mathrm{h}}(t)\right), \qquad t \in \mathbf{R}.$$

Hence, equation (3.6) is equivalent to the following equation:

$$\pi\left(F_W(u^{\mathrm{v}}(t), u^{\mathrm{h}}(t)\right) \;=\; X_{W,h}(u^{\mathrm{v}}(t)), \qquad t \in \mathbf{R}.$$

From these results, we can easily prove the proposition. ∎

Remark. If $\mathrm{rank} W^{\mathrm{v,h}} = r$, the condition $g(h(u^{\mathrm{v}}(t))) - g(u^{\mathrm{h}}(t)) \in \mathrm{Ker}\, W^{\mathrm{v,h}}$ for any $t \in \mathbf{R}$ implies that

$$u^{\mathrm{h}}(t) \;=\; h(u^{\mathrm{v}}(t)), \qquad t \in \mathbf{R}$$

in the case where the activation function g is a monotone function.

From Proposition 3.3, it turns out that for any $W \in M_{m,m+1}(\mathbf{R})$ with $W^{\mathrm{v,h}} = 0$ and any $h \in C^\infty(\mathbf{R}^n; \mathbf{R}^r)$, the RNN $\mathcal{N}_r^n(W)$ consisting of m units with $r(> 0)$ hidden units always produce the NDS $\varphi_{W,h}$ under h, which coincides with the dynamical sysytem $\Phi_{W^{\mathrm{v}}}$ produced by the RNN $\mathcal{N}_0^n(W^{\mathrm{v}})$ consisting of n units without hidden units. This gives trivial examples of NDSs. Thus, the notion of NDSs includes the dynamical systems produced by RNNs without hidden units.

Next, let us construct a non-trivial and universal example of NDSs. We restrict attention to the case where h is an affine map. We denote by

$$h_A : \mathbf{R}^n \to \mathbf{R}^r$$

the affine map corresponding to

$$A = (a_{kj})_{k=1,\cdots,r;\, j=0,1,\cdots,n} \in M_{r,n+1}(\mathbf{R}),$$

that is,

$$h_A(x) \;=\; \bar{A}x + a_0, \qquad x \in \mathbf{R}^n, \tag{3.7}$$

where $\bar{A} = (a_{k\alpha}) \in M_{r,n}(\mathbf{R})$ $(k = 1, \cdots, r;\ \alpha = 1, \cdots, n)$ and $a_0 = (a_{10}, \cdots, a_{r0}) \in \mathbf{R}^r$, that is,

$$A \;=\; \begin{pmatrix} a_0 & \bar{A} \end{pmatrix}.$$

Consider the direct product of the set of RNNs with n visible units and $r(>0)$ hidden units, and the set of affine maps from the visible state space $\mathbf{R}^n$ to the hidden state space $\mathbf{R}^r$. Note that the product set is identified with the set $M_{m,m+1}(\mathbf{R}) \times M_{r,n+1}(\mathbf{R})$, where $m = n + r$. Define its subset $\mathcal{M}_r^n$ by

$$\mathcal{M}_r^n = \left\{ (W, A) \in M_{m,m+1}(\mathbf{R}) \times M_{r,n+1}(\mathbf{R}) \;\middle|\; \begin{array}{l} \bar{A}W^{\mathrm{v,v}} = W^{\mathrm{h,v}} \\ \bar{A}W^{\mathrm{v,h}} = W^{\mathrm{h,h}} \\ \bar{A}w_0^{\mathrm{v}} + (1/\tau)a_0 = w_0^{\mathrm{h}} \end{array} \right\}. \tag{3.8}$$

Corollary 3.4. *For any $(W, A) \in \mathcal{M}_r^n$, the RNN $\mathcal{N}_r^n(W)$ produces the n-dimensional NDS φ_{W,h_A} under affine map h_A, and the corresponding vector field X_{W,h_A} on $\mathbf{R}^n$ is given by*

$$X_{W,h_A}(x) = -\frac{1}{\tau}x + W^{\mathrm{v,v}}g(x) + W^{\mathrm{v,h}}g(\bar{A}x + a_0) + w_0^{\mathrm{v}}, \qquad x \in \mathbf{R}^n. \tag{3.9}$$

Proof. Let $x^0 \in \mathbf{R}^n$, and let $u^{\mathrm{v}}(t)$ and $u^{\mathrm{h}}(t)$ be the visible and hidden state trajectories of $\mathcal{N}_r^n(W)$ with initial state $(x^0, h_A(x^0))$, respectively. Then, it is easily seen from definition (3.8) of $\mathcal{M}_r^n$ that the trajectory $(u^{\mathrm{v}}(t), \bar{A}u^{\mathrm{v}}(t) + a_0)$ on $\mathbf{R}^m$ is the state trajectory of the RNN $\mathcal{N}_r^n(W)$ with initial state $(x^0, \bar{A}x^0 + a_0)$. Hence,

$$u^{\mathrm{h}}(t) \;=\; \bar{A}u^{\mathrm{v}}(t) + a_0, \qquad t \in \mathbf{R}.$$

This means that

$$g\left(h_A(u^{\mathrm{v}}(t))\right) \;-\; g\left(u^{\mathrm{h}}(t)\right) \;=\; \{0\}, \qquad t \in \mathbf{R}.$$

Hence, we get the corollary from Proposition 3.3. ∎

Corollary 3.4 supplies universal examples of NDSs.

3.3 Affine neural dynamical systems

Now, we give the framework of approximating dynamical systems by RNNs. Based on Proposition 3.1 and Corollay 3.4, we define the notion of *affine neural dynamical systems (A-NDSs)*, and propose to use an n-dimensional A-NDS as a dynamical system that an RNN with n visible units actually produces on the visible state space $\mathbf{R}^n$ to approximate a target dynamical system on $\mathbf{R}^n$. Namely, we propose the A-NDS-based learning for approximating dynamical systems by RNNs.

Definition. First, for any $\mu = W \in \mathcal{M}_0^n$, we call the dynamical system Φ_W the n-dimensional *affine neural dynamical system (A-NDS)* produced by the RNN $\mathcal{N}_0^n(W)$, and define

$$\varphi_\mu = \Phi_W,$$

where,

$$\mathcal{M}_0^n = M_{n,n+1}(\mathbf{R}).$$

Next, let r be a positive intger and $\mu = (W, A) \in \mathcal{M}_r^n$. We call the NDS φ_{W,h_A} the n-dimensional *affine neural dynamical system (A-NDS)* produced by the RNN $\mathcal{N}_r^n(W)$ under affine map h_A, and define

$$\varphi_\mu = \varphi_{W,h_A}.$$

Note that A-NDSs are the only univeral example of NDSs as far as we know, and include the concept of *pseudo-neural system* defined by Funahashi ([14]), which was introduced as the extension of usual RNNs for approximating trajectories.

Let us consider the approximation capability of A-NDSs to verify the validitiy of the A-NDS-based learning for approximating dynamical systems by RNNs. We prove that any dynamical system on $\mathbf{R}^n$ can be approximated well by an n-dimensional A-NDS in a given finite region.

Let ψ be a dynamical system on $\mathbf{R}^n$ and Y the corresponding vector field on $\mathbf{R}^n$. We fix a positive constant ε, a bounded closed interval J containing $0 \in \mathbf{R}$, and a compact set $K \subset \mathbf{R}^n$. Our purpose is to construct such an n-dimensional A-NDS φ_μ that approximates within error ε the dynamical system ψ in $J \times K$, that is,

$$\left\|\varphi_\mu^t(x) - \psi^t(x)\right\| \;<\; \varepsilon, \qquad t \in J,\; x \in K, \tag{3.10}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^n$.

The following well-known result, due to Cybenko([9]), Funahashi ([14]), and Hornik, Stinchcombe and White ([29]), is essentially used to prove our approximation theorem. Note that this result has been variously extended (e.g, [3], [39]).

Lemma 3.5. (FNN approximation theorem) *Let K be a compact subset of $\mathbf{R}^n$ and f a continuous function on K. Then, for an arbitrary $\varepsilon > 0$, there exist a non-negative integer r, $b_k \in \mathbf{R}$, $(k = 1, \cdots, r)$, $c_{k\alpha} \in \mathbf{R}$, $(k = 1, \cdots, r;\ \alpha = 1, \cdots, n)$, and $\theta_k \in \mathbf{R}$, $(k = 1, \cdots, r)$ such that*

$$\left\|\sum_{k=1}^{r} b_k\, g\left(\sum_{\alpha=1}^{n} c_{k\alpha}\, x_\alpha \,+\, \theta_k\right) - f(x)\right\| \;<\; \varepsilon, \qquad x \in K.$$

Here, in the case of $r = 0$ the second term of the above expression is regarded as

$$\sum_{\alpha=1}^{n} c_\alpha\, x_\alpha \,+\, \theta,$$

where $c_\alpha \in \mathbf{R}$, $(\alpha = 1, \cdots, n)$ and $\theta \in \mathbf{R}$.

We fix $p \in K$. For $R > 0$, let $B_p(R)$ denote the n-dimensional closed ball with center p and radius R. Since $\psi : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^n$ is continuous and $J \times K$ is compact, there exists a positive constant ρ such that

$$\psi(J \times K) \;\subset\; B_p(\rho).$$

Since $Y : \mathbf{R}^n \to \mathbf{R}^n$ is C^∞, there exists a positive constant L such that

$$\max\{\|DY(x)\| \mid x \in B_p(\rho + \varepsilon)\} \;\leq\; L,$$

where $DY(x)$ is the Jacobian matrix of Y at x and $\|\cdot\|$ is the Euclidean norm on $M_{n,n}(\mathbf{R})$ $(\simeq \mathbf{R}^{n^2})$. Observe that Y is a Lipshitz map from $B_p(\rho + \varepsilon)$ to $\mathbf{R}^n$ with constant L. Choose a positve constant δ such that

$$\delta \;<\; \frac{L\varepsilon}{e^{LT-1}},$$

where

$$T = \max\{|t| \mid t \in J\}.$$

By virtue of Lemma 3.5, there exist a non-negative integer r, $P \in M_{n,n}(\mathbf{R})$ $Q \in M_{n,r}(\mathbf{R})$, $\theta \in \mathbf{R}^n$, $\bar{A} \in M_{r,n}(\mathbf{R})$, and $a_0 \in \mathbf{R}^r$ such that

$$\left\| Y(x) + \frac{1}{\tau}x - Pg(x) - \theta - Qg(\bar{A}x + a_0) \right\| \leq \delta, \qquad x \in B_p(\rho + \varepsilon).$$

We define $\mu = (W, A) \in \mathcal{M}_r^n$ by

$$\begin{aligned} W &= \begin{pmatrix} \theta & P & Q \\ \bar{A}\theta + \frac{1}{\tau}a_0 & \bar{A}P & \bar{A}Q \end{pmatrix} \in M_{n+r,n+r+1}(\mathbf{R}), \\ A &= \begin{pmatrix} a_0 & \bar{A} \end{pmatrix} \in M_{r,n+1}(\mathbf{R}). \end{aligned}$$

Consider the A-NDS φ_μ on $\mathbf{R}^n$ produced by the RNN $\mathcal{N}_r^n(W)$ under the affine map $h_A : \mathbf{R}^n \to \mathbf{R}^r$. Then, by equations (2.1) and (3.9), the vector field X_μ corresponding to the A-NDS φ_μ satifies the following condition:

$$\|X_\mu(x) - Y(x)\| \leq \delta, \qquad x \in B_p(\rho + \varepsilon).$$

Here, observe that

$$\varphi_\mu^t(x) \in B_p(\rho + \varepsilon), \qquad t \in J,\ x \in K.$$

By using Lemma 2.1, we can easily show that the A-NDS φ_μ satisfies condition (3.10). Hence, our purpose is accomplished.

From the above arguments, we obtain the following theorem.

Theorem 3.6.

(1) *For any $W \in \mathcal{M}_0^n$, the RNN $\mathcal{N}_0^n(W)$ without hidden units produces the dynamical system Φ_W on the visible state space $\mathbf{R}^n$.*

(2) *Let r be a positive integer. For any $\mu = (W, A) \in \mathcal{M}_r^n$, the RNN $\mathcal{N}_r^n(W)$ with r hidden units produces the dynamical system φ_μ on the visible state space $\mathbf{R}^n$ under affine map $h_A : \mathbf{R}^n \to \mathbf{R}^r$.*

(3) *Let ψ be a dynamical system on $\mathbf{R}^n$. Given a constant $\varepsilon > 0$, a bounded closed interval $J \subset \mathbf{R}$, and a compact set $K \subset \mathbf{R}^n$, there exist a non-negative integer r and $\mu \in \mathcal{M}_r^n$ such that*

$$\|{\varphi_\mu}^t(x) - \psi^t(x)\| < \varepsilon, \qquad t \in J,\ x \in K.$$

Theorem 3.6 supports the validity of the A-NDS-based learning for approximating dynamical systems by RNNs.

4 Representations of A-NDSs

For the problem of approximating a dynamical system by an RNN, we proposed A-NDS-based learning in the previous section. This section deals with a unique parametric representation of n-dimensional A-NDSs. In this and next sections, we focus on the case where the activation function $g(u)$ is $\tanh(u)$.

Let $\mathcal{D}^n$ denote the set of n-dimensional A-NDSs, and let

$$\mathcal{F} \; : \; \bigcup_{0 \leq r \in \mathbf{Z}} \mathcal{M}_r^n \; \rightarrow \; \mathcal{D}^n$$

denote the parametric representation of n-dimensional A-NDSs by RNNs and affine maps, that is,

$$\mathcal{F}(\mu) \;=\; \varphi_\mu$$

Note that for $\mu = W \in \mathcal{M}_0^n$, $\mathcal{F}(\mu)$ is the n-dimensional A-NDS produced by the RNN $\mathcal{N}_0^n(W)$ without hidden units, and for $\mu = (W, A) \in \mathcal{M}_r^n$ $(r > 0)$, $\mathcal{F}(\mu)$ is the n-dimensional A-NDS produced by the RNN $\mathcal{N}_r^n(W)$ under affine map h_A. For $\varphi \in \mathcal{D}^n$, $\mu \in \mathcal{M}_r^n$ is called a *representation* of φ if $\mathcal{F}(\mu) = \varphi$.

By definition, the map $\mathcal{F}$ is surjective. We also know by Proposition 3.2 that the restriction

$$\mathcal{F}|_{\mathcal{M}_0^n} : \mathcal{M}_0^n \rightarrow \mathcal{D}^n$$

is injective. However, as seen below, the parametric representation $\mathcal{F}$ has redundancy. By extending Sussmann's work ([54]) for the redundancy in the function representations of FNNs, we construct a unique parametric representation of $\mathcal{D}^n$ toward developing efficient learning algorithms of dynamical systems for RNNs based on A-NDSs.

4.1 Parametric representation

Suppose that r is a positive integer. By defintion (3.8), the set $\mathcal{M}_r^n$ can be identified with the set $M_{n,n+r+1}(\mathbf{R}) \times M_{r,n+1}(\mathbf{R})$ in the following way:

$$\mathcal{M}_r^n \ni (W, A) \;\leftrightarrow\; (W^{\vee}, A) \in M_{n,n+r+1}(\mathbf{R}) \times M_{r,n+1}(\mathbf{R}).$$

Thus, $\mathcal{M}_r^n$ can be regarded as the $(n^2+2nr+n+r)$-dimensional Euclidean space. From now on, we often use this identification.

Let $\mu = (W, A) \in \mathcal{M}_r^n$ $(r > 0)$, $W = (w_{ij}) \in M_{n+r,n+r+1}(\mathbf{R})$ $(i = 1, \cdots, n+r;\ j = 0, 1, \cdots, n+r)$, and $A = (a_{kj}) \in M_{r,n+1}(\mathbf{R})$ $(k = 1, \cdots, r;\ j = 1, \cdots, n)$. Then, we can describe the vector field X_μ on $\mathbf{R}^n$ corresponding to the n-dimensional A-NDS φ_μ as follows:

$$\begin{aligned}
X_\mu(x) &= \left((X_\mu)_1(x), \cdots, (X_\mu)_n(x)\right), \qquad x \in \mathbf{R}^n; \\
(X_\mu)_i(x) &= -\frac{1}{\tau} x_i + \sum_{j=1}^{n} w_{ij} \tanh(x_j) \\
&\quad + \sum_{k=1}^{r} w_{i\,n+k} \tanh\left(\sum_{j=1}^{n} a_{kj}\, x_j \;+\; a_{k0}\right) \;+\; w_{i0}, \\
&\qquad\qquad x = (x_1, \cdots, x_n) \in \mathbf{R}^n,\ i = 1, \cdots, n.
\end{aligned} \tag{4.1}$$

Example 3. Let $n = 1$ and $r = 2$. We define $\mu = (W, A) \in \mathcal{M}_2^1$ and $\tilde{\mu} = (\widetilde{W}, \tilde{A}) \in \mathcal{M}_2^1$ as follows:

$$W^{\vee} = (\ 1\quad 3\quad -1\quad 2\ ), \qquad A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix},$$

$$\widetilde{W}^{\vee} = (\ 1\quad 3\quad -2\quad -1\ ), \qquad \tilde{A} = \begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix}.$$

Note that the matrices W and $\widetilde{W}$ of connection weights and biases are respectively

$$W = \begin{pmatrix} 1 & 3 & -1 & 2 \\ 2+1/\tau & 6 & -2 & 4 \\ 3-1/\tau & 9 & -3 & 6 \end{pmatrix},$$

$$\widetilde{W} = \begin{pmatrix} 1 & 3 & -2 & -1 \\ -3+1/\tau & -9 & 6 & 3 \\ 2+1/\tau & 6 & -4 & -2 \end{pmatrix},$$

and the affine maps h_A and $h_{\tilde{A}}$ from the visible state sapce $\mathbf{R}^1$ to the hidden state space $\mathbf{R}^2$ are respectively

$$h_A(x) = (2x+1, 3x-1), \qquad x \in \mathbf{R}^1,$$

$$h_{\tilde{A}}(x) = (-3x+1, 2x+1), \qquad x \in \mathbf{R}^1.$$

By equation (4.1), we have

$$X_\mu(x) = X_{\tilde{\mu}}(x) = -\frac{1}{\tau}x + 3\tanh(x) - \tanh(2x+1) + 2\tanh(3x-1) + 1, \qquad x \in \mathbf{R}^1.$$

Hence, we obtain the following:

$$\exists\, \mu, \tilde{\mu} \in \mathcal{M}_2^1 \quad \text{s.t.} \quad \mu \neq \tilde{\mu}, \ \ \mathcal{F}(\mu) \ = \ \mathcal{F}(\tilde{\mu}).$$

Example 3 shows that the parametric representation $\mathcal{F}$ has redundancy. In particular, it implies that there exist distinct RNNs with the same number of hidden units such that they produce the same A-NDS. In the next example, we see that there exist distinct RNNs such that they produce the same A-NDS while their numbers of hidden units are different.

Example 4. For $n = 1$ and $r = 2$, we define $\mu = (W, A) \in \mathcal{M}_2^1$ by

$$W^{\mathrm{v}} = \begin{pmatrix} 5 & 1 & 2 & 3 \end{pmatrix}, \qquad A = \begin{pmatrix} -1 & 2 \\ 1 & -2 \end{pmatrix}.$$

Note that the matrix W of connection weights and biases is

$$W = \begin{pmatrix} 5 & 1 & 2 & 3 \\ 10 - 1/\tau & 2 & 4 & 6 \\ -10 + 1/\tau & -2 & -4 & -6 \end{pmatrix},$$

and the affine map $h_A : \mathbf{R}^1 \to \mathbf{R}^2$ is

$$h_A(x) = (2x - 1, -2x + 1), \qquad x \in \mathbf{R}^1.$$

Next, for $n = 1$ and $r = 1$, we define $\tilde{\mu} = (\widetilde{W}, \tilde{A}) \in \mathcal{M}_1^1$ by

$$\widetilde{W}^{\mathrm{v}} = \begin{pmatrix} 5 & 1 & -1 \end{pmatrix}, \qquad \tilde{A} = \begin{pmatrix} -1 & 2 \end{pmatrix}.$$

Note that the matrix $\widetilde{W}$ of connection weights and biases is

$$\widetilde{W} = \begin{pmatrix} 5 & 1 & -1 \\ 10 - 1/\tau & 2 & -2 \end{pmatrix},$$

and the affine map $h_{\tilde{A}} : \mathbf{R}^1 \to \mathbf{R}^1$ is

$$h_{\tilde{A}}(x) = 2x - 1, \qquad x \in \mathbf{R}^1.$$

By equation (4.1), we have

$$X_\mu(x) = X_{\tilde{\mu}}(x) = -\frac{1}{\tau}x + \tanh(x) - \tanh(2x-1) + 5, \qquad x \in \mathbf{R}^1.$$

Hence, we obtain the following:

$$\exists\, \mu \in \mathcal{M}_2^1,\ \exists\, \tilde{\mu} \in \mathcal{M}_1^1 \quad \text{s.t.} \quad \mathcal{F}(\mu) \ = \ \mathcal{F}(\tilde{\mu}).$$

4.2 Symmetric transformations

Let us consider the case where distinct RNNs with the same number of hidden units produce the same A-NDS. Assume that r is a positive integer. A C^∞ diffeomorphism of $\mathcal{M}_r^n$ is referred to as a *transformation* of $\mathcal{M}_r^n$. We introduce the notion of *symmetric transformations* of $\mathcal{M}_r^n$.

Definition. A transformation γ of $\mathcal{M}_r^n$ is said to be *symmetric* if for each $\mu \in \mathcal{M}_r^n$, μ and $\gamma(\mu)$ represents the same A-NDS, that is,

$$\mathcal{F}(\mu) \ = \ \mathcal{F}(\gamma(\mu)), \qquad \mu \in \mathcal{M}_r^n.$$

Let us construct some symmetric transformations of $\mathcal{M}_r^n$.

First, for $\ell = 1, \cdots, r$, We define the transformation θ_ℓ of $\mathcal{M}_r^n$ by

$$\begin{aligned}
\theta_\ell(W, A) &= \left(\widetilde{W}, \tilde{A}\right), \qquad (W, A) \in \mathcal{M}_r^n; \\
\tilde{w}_{ij} &= w_{ij}, \qquad i = 1, \cdots, n;\ j = 0, 1, \cdots, n, \\
\tilde{w}_{i\,n+k} &= \begin{cases} -w_{i\,n+k}, & (k = \ell), \\ w_{i\,n+k}, & (k \neq \ell), \end{cases} \qquad i = 1, \cdots, n;\ k = 1, \cdots, r \\
\tilde{a}_{kj} &= \begin{cases} -a_{kj}, & (k = \ell), \\ a_{kj}, & (k \neq \ell), \end{cases} \qquad k = 1, \cdots, r;\ j = 0, 1, \cdots, n,
\end{aligned}$$

where

$$W = (w_{ij})_{i=1,\cdots,n+r;\ j=0,1,\cdots,n+r}, \quad A = (a_{kj})_{k=1,\cdots,r;\ j=0,1,\cdots,n},$$
$$\widetilde{W} = (\tilde{w}_{ij})_{i=1,\cdots,n+r;\ j=0,1,\cdots,n+r}, \quad \tilde{A} = (\tilde{a}_{kj})_{k=1,\cdots,r;\ j=0,1,\cdots,n}.$$

Note that the trasnformation θ_ℓ means changing the signs of all the weights associated with the hidden unit $n + \ell$. By equation (4.1), it is easily proved that θ_ℓ is a symmetric transformation of $\mathcal{M}_r^n$ for $\ell = 1, \cdots, r$.

Next, let $\mathcal{S}_r$ denote the set of permutations on the set $\{1, \cdots, r\}$. For $\sigma \in \mathcal{S}_r$, define the transformation $\tilde{\sigma}$ of $\mathcal{M}_r^n$ by

$$\begin{aligned}
\tilde{\sigma}(W, A) &= (\widetilde{W}, \tilde{A}), \qquad (W, A) \in \mathcal{M}_r^n; \\
\tilde{w}_{ij} &= w_{ij}, \qquad i = 1, \cdots, n;\ j = 0, 1, \cdots, n, \\
\tilde{w}_{i\ n+\sigma(k)} &= w_{i\ n+k}, \qquad i = 1, \cdots, n;\ k = 1, \cdots, r, \\
\tilde{a}_{\sigma(k)\ j} &= a_{kj}, \qquad k = 1, \cdots, r;\ j = 0, 1, \cdots, n,
\end{aligned}$$

where

$$W = (w_{ij})_{i=1,\cdots,n+r;\ j=0,1,\cdots,n+r}, \quad A = (a_{kj})_{k=1,\cdots,r;\ j=0,1,\cdots,n},$$

$$\widetilde{W} = (\tilde{w}_{ij})_{i=1,\cdots,n+r;\ j=0,1,\cdots,n+r}, \quad \tilde{A} = (\tilde{a}_{kj})_{k=1,\cdots,r;\ j=0,1,\cdots,n}.$$

Note that the transformation $\tilde{\sigma}$ means relabeling the hidden units according to the permutation σ. By equation (4.1), it is easily proved that $\tilde{\sigma}$ is a symmetric transformation of $\mathcal{M}_r^n$ for any $\sigma \in \mathcal{S}_r$.

Let $\mathcal{G}_r^n$ be the transformation group of $\mathcal{M}_r^n$ generated by the symmetric transformations θ_ℓ ($\ell = 1, \cdots, r$) and $\tilde{\sigma}$ ($\sigma \in \mathcal{S}_r$). We can easily prove the following proposition.

Proposition 4.1. *$\mathcal{G}_r^n$ is a non-commutative finite group of degree $2^r r!$. In particular,*

(1) $\theta_k \circ \theta_k = \mathrm{id}$ $(k = 1, \cdots, r)$, *where* id *is the identity transformation of* $\mathcal{M}_r^n$,

(2) $\theta_k \circ \theta_\ell = \theta_\ell \circ \theta_k$ $\quad (k, \ell = 1, \cdots, r)$,

(3) $\tilde{\sigma} \circ \theta_\ell = \theta_{\sigma(\ell)} \circ \tilde{\sigma}$ $\quad (\sigma \in \mathcal{S}_r,\ \ell = 1, \cdots, r)$,

(4) $\mathcal{G}_r^n = \left\{ \theta_1^{\lambda_1} \circ \cdots \circ \theta_r^{\lambda_r} \circ \tilde{\sigma} \,\middle|\, \lambda_1, \cdots, \lambda_r = 0, 1;\ \sigma \in \mathcal{S}_r \right\}$, *where* $\theta_\ell^{\,0}$ *expresses* id *and* $\theta_\ell^{\,1}$ *expresses* θ_ℓ.

For the case $r = 0$, we define $\mathcal{G}_0^n$ as the trivial group consisting of the identity transformation of $\mathcal{M}_0^n$.

Example 5. Let $n = 1$ and $r = 1$. Then, we have

$$\begin{aligned} \mathcal{S}_r &= \{\mathrm{id}\}, \\ \mathcal{G}_1^1 &= \{\mathrm{id},\ \theta_1\}. \end{aligned}$$

Let us identify $\mathcal{M}_1^1$ with $M_{1,3}(\mathbf{R}) \times M_{1,2}(\mathbf{R})$. Then, the symmetric transformation θ_1 is described as follows: For each $(W^{\mathrm{v}}, A) \in \mathcal{M}_1^1 = M_{1,3}(\mathbf{R}) \times M_{1,2}(\mathbf{R})$,

$$\theta_1(W^{\mathrm{v}}, A) = \Big(\begin{pmatrix} w_{10} & w_{11} & -w_{12} \end{pmatrix},\ \begin{pmatrix} -a_{10} & -a_{11} \end{pmatrix} \Big),$$

where

$$W^{\mathrm{v}} = (w_{10}\ w_{11}\ w_{12}), \quad A = (a_{10}\ a_{11}).$$

Example 6. Let $n = 1$ and $r = 2$. Then, we have

$$\begin{aligned} \mathcal{S}_2 &= \{\mathrm{id},\ \sigma\}, \\ \mathcal{G}_2^1 &= \{\mathrm{id},\ \theta_1,\ \theta_2,\ \tilde{\sigma},\ \theta_1 \circ \theta_2,\ \theta_1 \circ \tilde{\sigma},\ \theta_2 \circ \tilde{\sigma},\ \theta_1 \circ \theta_2 \circ \tilde{\sigma}\}, \end{aligned}$$

where σ is the transposition of 1 and 2. Let us identify $\mathcal{M}_2^1$ with $M_{1,4}(\mathbf{R}) \times M_{2,2}(\mathbf{R})$. Then, the symmetric transformations θ_1, θ_2 and $\tilde{\sigma}$ are described as follows: For each $(W^{\mathrm{v}}, A) \in \mathcal{M}_2^1 = M_{1,4}(\mathbf{R}) \times M_{2,2}(\mathbf{R})$,

$$\begin{aligned} \theta_1(W^{\mathrm{v}}, A) &= \left(\begin{pmatrix} w_{10} & w_{11} & -w_{12} & w_{13} \end{pmatrix},\ \begin{pmatrix} -a_{10} & -a_{11} \\ a_{20} & a_{21} \end{pmatrix} \right), \\ \theta_2(W^{\mathrm{v}}, A) &= \left(\begin{pmatrix} w_{10} & w_{11} & w_{12} & -w_{13} \end{pmatrix},\ \begin{pmatrix} a_{10} & a_{11} \\ -a_{20} & -a_{21} \end{pmatrix} \right), \\ \tilde{\sigma}(W^{\mathrm{v}}, A) &= \left(\begin{pmatrix} w_{10} & w_{11} & w_{13} & w_{12} \end{pmatrix},\ \begin{pmatrix} a_{20} & a_{21} \\ a_{10} & a_{11} \end{pmatrix} \right), \end{aligned}$$

where

$$W^{\mathrm{v}} = (w_{10}\ w_{11}\ w_{12}\ w_{13}), \quad A = \begin{pmatrix} a_{10} & a_{11} \\ a_{20} & a_{21} \end{pmatrix}.$$

4.3 Irreducible representations

Let us consider the case where distinct RNNs with different numbers of hidden units produce the same A-NDS. Let r be a non-negative intger. First, we introduce the notion of *minimal representations* of A-NDSs.

Definition. Let $\mu \in \mathcal{M}_r^n$ be a representation of an n-dimensional A-NDS φ. Then, μ is called a minimal representation of φ if there is no other representatoin of φ using an RNN with fewer hidden units than r, that is, if

$$\exists\, \tilde{\mu} \in \mathcal{M}_{\tilde{r}}^n \quad \text{s.t.} \quad \mathcal{F}(\tilde{\mu}) = \varphi \quad \Rightarrow \quad r \leq \tilde{r}.$$

In order to state the minimality condition of $\mu \in \mathcal{M}_r^n$ in terms of μ itself, we introduce the notion of *irreducible representations* of A-NDSs.

Definition. An element $\mu = (W, A)$ of $\mathcal{M}_r^n$ is said to be *irreducible* if $r = 0$ or the following conditions hold:

A: $(w_{1\ n+k}, \cdots, w_{n\ n+k}) \neq (0, \cdots, 0), \qquad k = 1, \cdots, r.$

B: $(a_{k1}, \cdots, a_{kn}) \neq (0, \cdots, 0), \qquad k = 1, \cdots, r.$

C: $(a_{k0}, a_{k1}, \cdots, a_{kn}) \neq \pm e_1, \cdots, \pm e_n \quad (k = 1, \cdots, r),$

where $e_i = (0, \cdots, 0, \underset{\smile}{\overset{i+1}{1}}, 0, \cdots, 0) \in \mathbf{R}^{n+1}$ for $i = 1, ..., n$

D: $(a_{k0}, a_{k1}, \cdots, a_{kn}) \neq \pm(a_{\ell 0}, a_{\ell 1}, \cdots, a_{\ell n}), \qquad k, \ell = 1, \cdots, r;\ k \neq \ell,$

where $W = (w_{ij}) \in M_{n+r,n+r+1}(\mathbf{R})$ $(i = 1, \cdots, n+r;\ j = 0, 1, \cdots, n+r)$, and $A = (a_{kj}) \in M_{r,n+1}(\mathbf{R})$ $(k = 1, \cdots, r;\ j = 1, \cdots, n)$.

We call $\mu \in \mathcal{M}_r^n$ an *irreducible representation* of an n-dimensional A-NDS φ if $\mathcal{F}(\mu) = \varphi$ and μ is irreducible. Let $\widehat{\mathcal{M}_r^n}$ denote the set of irreducible elements of $\mathcal{M}_r^n$. Note that $\widehat{\mathcal{M}_0^n} = \mathcal{M}_0^n$. We call the set $\widehat{\mathcal{M}_r^n}$ the *irreducible set* of $\mathcal{M}_r^n$. By equation (4.1), we can easily prove the followin proposition.

Proposition 4.2. *If $\mu \in \mathcal{M}_r^n$ is not an irreducible representation of an n-dimensional A-NDS φ, then it is not a minimal representation of φ.*

Example 7. Let $n = 1$ and $r = 1$. We identify $\mathcal{M}_1^1$ with $M_{1,3}(\mathbf{R}) \times M_{1,2}(\mathbf{R})$. Then the irreducible set $\widehat{\mathcal{M}_1^1}$ of $\mathcal{M}_1^1$ is identified as follows:

$$\widehat{\mathcal{M}_1^1} = \left\{ (W^{\mathrm{v}}, A) \in M_{1,3}(\mathbf{R}) \times M_{1,2}(\mathbf{R}) \;\middle|\; \begin{array}{l} W^{\mathrm{v}} = (w_{1j})_{j=0,1,2} \\ A = (a_{1j})_{j=0,1} \\ w_{12} \neq 0, \;\; a_{11} \neq 0, \;\; (a_{10}, a_{11}) \neq (0, \pm 1) \end{array} \right\}.$$

Example 8. Let $n = 1$ and $r = 2$. We identify $\mathcal{M}_2^1$ with $M_{1,4}(\mathbf{R}) \times M_{2,2}(\mathbf{R})$. Then the irreducible set $\widehat{\mathcal{M}_2^1}$ of $\mathcal{M}_2^1$ is identified as follows:

$$\widehat{\mathcal{M}_2^1} = \left\{ (W^{\mathrm{v}}, A) \in M_{1,4}(\mathbf{R}) \times M_{2,2}(\mathbf{R}) \;\middle|\; \begin{array}{l} W^{\mathrm{v}} = (w_{1j})_{j=0,1,2,3} \\ A = (a_{kj})_{k=1,2;\, j=0,1} \\ w_{12} \neq 0, w_{13} \neq 0, \\ a_{11} \neq 0, a_{21} \neq 0, \\ (a_{10}, a_{11}) \neq (0, \pm 1), \\ (a_{20}, a_{21}) \neq (0, \pm 1), \\ (a_{10}, a_{11}) \neq \pm (a_{20}, a_{21}) \end{array} \right\}.$$

Here, we explain our irreducibility conditions in terms of RNNs and affine maps according to Sussmann's view ([54]) on irreducible FNNs. We begin with the following definition:

Definition. Two functions f_1 andf_2 on $\mathbf{R}^n$ are said to be *sign-equivalent* if

$$|f_1(x)| = |f_2(x)|, \qquad x \in \mathbf{R}^n.$$

Let $r > 0$, $\mu = (W, A) \in \mathcal{M}_r^n$, and let $(x_1, ..., x_n)$ denote the canonical coordinate system in $\mathbf{R}^n$. Consider the affine map h_A from the visible state space $\mathbf{R}^n$ to the hidden state space $\mathbf{R}^r$, which is written as follows:

$$\begin{aligned} h_A(x) &= \left((h_A)_1(x), \cdots, (h_A)_r(x)\right), \qquad x \in \mathbf{R}^n; \\ (h_A)_k(x) &= \sum_{j=1}^{n} a_{kj} x_j + a_{k0}, \qquad x = (x_1, \cdots, x_n), \; k = 1, \cdots, r. \end{aligned}$$

Now, we can state the conditions **A**, **B**, **C** and **D** for μ to be irreducible as follows: Condition **A** states that there exists at least one connection from each hidden unit $n + k$ to the visible units. Condition **B** states that each affine function $(h_A)_k(x)$ is not a constant function. Condition **C** states that each affine function $(h_A)_k(x)$ is not sign-equivalent to any of the affine functions $x_1, \cdots, x_n$ on $\mathbf{R}^n$. Condition **D** states that no two of the affine functions $(h_A)_1(x), \cdots, (h_A)_r(x)$ are sign-equivalent.

4.4 Unique parametric representations

Now we investigate a unique parametric representation of $\mathcal{D}^n$. First, we state the following result by Sussmann which is used to prove our proposition.

Lemma 4.3. (Sussmann [54]) *Let $f_1, \cdots, f_d$ be non-constant affine functions on $\mathbf{R}^n$ such that no two of them are sign-equivalent, where d is a positive integer. Then, the C^∞ functions* $\tanh \circ f_1, \cdots, \tanh \circ f_d$ *and the constant function* 1 *on* $\mathbf{R}^n$ *are linearly independent in the real vector space of C^∞ functions on $\mathbf{R}^n$.*

Let us prove the following proposition.

Proposition 4.4. *Suppose that Let r and $\tilde{r}$ be non-negative integers. $\mu = (W, A)$ $\in \widehat{\mathcal{M}}^n_r$ and $\tilde{\mu} = (\widetilde{W}, \tilde{A}) \in \widehat{\mathcal{M}}^n_{\tilde{r}}$ satisfy $\mathcal{F}(\mu) = \mathcal{F}(\tilde{\mu})$. Then, $r = \tilde{r}$, and there exists a $\gamma \in \mathcal{G}^n_r$ such that $\gamma(\mu) = \tilde{\mu}$.*

Proof. Put

$$\begin{aligned} W &= (w_{ij})_{i=1,\cdots,n+r;\ j=0,1,\cdots,n+r}, \\ \widetilde{W} &= (\tilde{w}_{ij})_{i=1,\cdots,n+\tilde{r};\ j=0,1,\cdots,n+\tilde{r}}, \\ A &= (a_{kj})_{k=1,\cdots,r;\, j=0,1,\cdots,n}, \\ \tilde{A} &= (\tilde{a}_{kj})_{i=1,\cdots,\tilde{r};\ j=0,1,\cdots,n}. \end{aligned}$$

Let $(x_1, \cdots, x_n)$ denote the canonical coordinate system in $\mathbf{R}^n$. Observe that the affine functions $(h_A)_1, \cdots, (h_A)_r$ and $(h_{\tilde{A}})_1, \cdots, (h_{\tilde{A}})_{\tilde{r}}$ on $\mathbf{R}^n$ are written as follows:

$$\begin{aligned} (h_A)_k(x) &= \sum_{j=1}^{n} a_{kj}x_j + a_{k0}, \qquad x = (x_1, \cdots, x_n),\ k = 1, \cdots, r, \\ (h_{\tilde{A}})_\ell(x) &= \sum_{j=1}^{n} \tilde{a}_{\ell j}x_j + \tilde{a}_{\ell 0}, \qquad x = (x_1, \cdots, x_n),\ \ell = 1, \cdots, \tilde{r}. \end{aligned}$$

By assumption, we have $\varphi_\mu = \varphi_{\tilde{\mu}}$. Thus, by equation (4.1), it follows that

$$w_{i0} - \tilde{w}_{i0} + \sum_{j=1}^{n}(w_{ij} - \tilde{w}_{ij}) \tanh(x_j) + \sum_{k=1}^{r} w_{i\, n+k} \tanh(h_k(x))$$

$$-\sum_{\ell=1}^{\tilde{r}} \tilde{w}_{i\, n+\ell} \tanh(\tilde{h}_\ell(x)) = 0, \qquad x = (x_1, \cdots, x_n),\ i = 1, \cdots, n. \quad (4.2)$$

Since μ and $\tilde{\mu}$ are irreducible, each $(h_A)_k(x)$ and each $(h_{\tilde{A}})_\ell(x)$ are non-constant affine functions, and are not sign-equivalent to any of the affine functions 1, x_1, $\cdots$, x_n on $\mathbf{R}^n$. Moreover, no two of $(h_A)_1(x)$, $\cdots$,, $(h_A)_r(x)$ are sign-equivalent, and no two of $(h_{\tilde{A}})_1(x)$, $\cdots$, $(h_{\tilde{A}})_{\tilde{r}}(x)$ are also sign-equivalent. Suppose that for any $k = 1, \cdots, r$ and any $\ell = 1, \cdots, \tilde{r}$, the affine functions $(h_A)_k(x)$ and $(h_{\tilde{A}})_\ell(x)$ are not sign-equivalent. Then, by equation (4.2) and Lemma 4.3 we have

$$w_{i\, n+k} = 0, \qquad i = 1, \cdots, n;\ k = 1, \cdots, r,$$

$$\tilde{w}_{i\, n+\ell} = 0, \qquad i = 1, \cdots, n;\ \ell = 1, \cdots, \tilde{r}$$

However, this contradicts the fact that μ and $\tilde{\mu}$ are irreducible. Hence, there exists a pair of $(h_A)_k(x)$ and $(h_{\tilde{A}})_\ell(x)$ such that they are sign-equivalent. Let $\{(k_1, \ell_1), ...,$ $(k_d, \ell_d)\}$ be all of the pairs (k, ℓ) such that $(h_A)_k(x)$ and $(h_{A'})_\ell(x)$ are sign-equivalent, and put

$$(h_{\tilde{A}})_{\ell_\alpha}(x) = \rho_{k_\alpha} (h_A)_{k_\alpha}(x), \qquad \alpha = 1, \cdots, d,$$

where $\rho_{k_\alpha} = 1$ or -1, $(\alpha = 1, \cdots, d)$. Then equation (4.2) is equivalent to the following equation:

$$w_{i0} - \tilde{w}_{i0} + \sum_{j=1}^{n}(w_{ij} - \tilde{w}_{ij}) \tanh(x_j)$$

$$+\sum_{\alpha=1}^{d}(w_{i\, n+k_\alpha} - \rho_{k_\alpha}\tilde{w}_{i\, n+\ell_\alpha}) \tanh\left((h_A)_{k_\alpha}(x)\right)$$

$$+ \sum_{k \neq k_1, \cdots, k_d} w_{i\, n+k} \tanh\left((h_A)_k(x)\right)$$

$$- \sum_{\ell \neq \ell_1, \cdots, \ell_d} \tilde{w}_{i\, n+\ell} \tanh\left((h_{\tilde{A}})_\ell(x)\right) = 0, \quad x = (x_1, \cdots, x_n),\ i = 1, \cdots, n. (4.3)$$

By Lemma 4.3 and the irreducibility of μ and $\tilde{\mu}$, the following C^∞ functions on $\mathbf{R}^n$ are linearly independent:

$$1,\ \tanh(x_1), \cdots, \tanh(x_n),$$

$$\tanh\left((h_A)_1(x)\right), \cdots, \tanh\left((h_A)_r(x)\right),\ \tanh\left((h_{\tilde{A}})_\ell(x)\right),\ (\ell \neq \ell_1, \cdots, \ell_d)$$

Hence, by equation (4.3) and the irreducibility of μ and $\tilde{\mu}$, we have

$$\begin{aligned} r &= \tilde{r} = d, \\ w_{ij} &= \tilde{w}_{ij}, \qquad i = 1, \cdots, n;\ j = 0, 1, \cdots, n, \\ \tilde{w}_{i\ n+\sigma(k)} &= \rho_k\, w_{i\ n+k}, \qquad i = 1, \cdots, n;\ k = 1, \cdots, r, \end{aligned}$$

where σ is the element of $\mathcal{S}_r$ defined by $\sigma(k_\alpha) = \ell_\alpha$, $(\alpha = 1, \cdots, r)$. Observe that

$$(h_{\tilde{A}})_{\sigma(k)}(x) = \rho_k\ (h_A)_k(x), \qquad x \in \mathbf{R}^n,\ k = 1, \cdots, r.$$

Now we define $\gamma \in \mathcal{G}_r^n$ by

$$\gamma = {\theta_1}^{(1-\rho_1)/2} \circ \cdots \circ {\theta_r}^{(1-\rho_r)/2} \circ \tilde{\sigma}.$$

Then, it can be easily verified that $\gamma(\mu) = \tilde{\mu}$. We have completed the proof of Proposition 4.4. ∎

Consider an n-dimensional A-NDS φ and an element $\hat{\mu}$ of $\mathcal{M}_{\hat{r}}^n$ with $\mathcal{F}(\hat{\mu}) = \varphi$. In general, it is difficult to decide whether $\hat{\mu}$ is a minimal representation of φ. Indeed, in order to verify the minimality of $\hat{\mu}$, we need to examine all elements of the sets $\mathcal{M}_r^n$, $(r \in \mathbf{Z};\ 0 \leq r \leq \hat{r})$. By definition, it is much easier to decide whether $\hat{\mu}$ is irreducible. We have seen in Proposition 4.2 that a minimal representation of an A-NDS is an irreducible representation of the A-NDS. The following corollary shows that the converse is also true.

Corollary 4.5. *Let φ be an n-dimensional A-NDS and μ an element of $\mathcal{M}_r^n$ such that $\mathcal{F}(\mu) = \varphi$. Then, μ is a minimal representation of φ if and only if μ is irreducible, that is, $\mu \in \widehat{\mathcal{M}}_r^n$.*

Proof. It is sufficient to prove that if μ is irreducible, then it is a minimal representation of φ. Suppose that μ is irreducible but not a minimal representation of

φ. Then, there exist a non-negative integer $\tilde{r}$ and an element $\tilde{\mu}$ of $\mathcal{M}^n_{\tilde{r}}$ such that $\tilde{r} < r$ and $\mathcal{F}(\mu) = \mathcal{F}(\tilde{\mu})$. Taking a minimal representation $\hat{\mu} \in \mathcal{M}^n_{\hat{r}}$ of φ, we have

$$\hat{r} \leq \tilde{r} < r. \tag{4.4}$$

Since minimal representations are irreducible, $\hat{\mu}$ is irreducible. Hence, we obtain $r = \hat{r}$ by Proposition 4.4. This contradict inequality (4.4). We have thus proved the corollary. ∎

Corollary 4.5 states that the notion of minimal representations of A-NDSs are equivalent to the notion of irreducible representations of A-NDSs. Let φ be an n-dimensional A-NDS and $\hat{\mu}$ an irreducible representation of φ. Proposition 4.4 implies that the set of all irreducible representations of φ is given by

$$\{\gamma(\hat{\mu}) \mid \gamma \in \mathcal{G}^n_r\}. \tag{4.5}$$

Hence, the set of all minimal representations of φ is obtained by the set (4.5). In the following examples, we explicitly express all the minimal representatations of the A-NDSs in Examples 4 and 3.

Example 9. Let φ be the 1-dimensional A-NDS in Example 4. Then, the vector field X on $\mathbf{R}^1$ corresponding to φ is of the form

$$X(x) = -\frac{1}{\tau}x + \tanh(x) - \tanh(2x-1) + 5, \qquad x \in \mathbf{R}^1.$$

Example 7 implies that $(\widetilde{W}, \tilde{A}) \in \mathcal{M}^1_1$ in Example 4 is an irreducible representation of φ. Let us identify $\mathcal{M}^1_1$ with $M_{1,3}(\mathbf{R}) \times M_{1,2}(\mathbf{R})$. From expression (4.5) and Example 5, we obtain that all the minimal representations of φ are given by

$$\Big(\begin{pmatrix} 5 & 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 2 \end{pmatrix} \Big), \quad \Big(\begin{pmatrix} 5 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -2 \end{pmatrix} \Big).$$

Example 10. Let φ be the 1-dimensional A-NDS in Example 3. Then, the vector field X on $\mathbf{R}^1$ corresponding to φ is of the form

$$X(x) = -\frac{1}{\tau}x + 3\tanh(x) - \tanh(2x+1) + 2\tanh(3x-1) + 1, \qquad x \in \mathbf{R}^1.$$

Example 8 implies that $(W, A) \in \mathcal{M}_2^1$ in Example 3 is an irreducible representation of φ. Let us identify $\mathcal{M}_2^1$ with $M_{1,4}(\mathbf{R}) \times M_{2,2}(\mathbf{R})$. From expression (4.5) and Example 6, we obtain that all the minimal representations of φ are given by

$$\left(\begin{pmatrix} 1 & 3 & -1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}\right), \quad \left(\begin{pmatrix} 1 & 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ -1 & 3 \end{pmatrix}\right),$$
$$\left(\begin{pmatrix} 1 & 3 & -1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & -3 \end{pmatrix}\right), \quad \left(\begin{pmatrix} 1 & 3 & 2 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 3 \\ 1 & 2 \end{pmatrix}\right),$$
$$\left(\begin{pmatrix} 1 & 3 & 1 & -2 \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ 1 & -3 \end{pmatrix}\right), \quad \left(\begin{pmatrix} 1 & 3 & -2 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix}\right),$$
$$\left(\begin{pmatrix} 1 & 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 3 \\ -1 & -2 \end{pmatrix}\right), \quad \left(\begin{pmatrix} 1 & 3 & -2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -3 \\ -1 & -2 \end{pmatrix}\right).$$

Let r be a non-negative integer. Note that the irreducible set $\widehat{\mathcal{M}_r^n}$ is an open set of $\mathcal{M}_r^n$ and

$$\gamma\left(\widehat{\mathcal{M}_r^n}\right) = \widehat{\mathcal{M}_r^n}, \qquad \gamma \in \mathcal{G}_r^n.$$

Consider the quotient space $\widehat{\mathcal{M}_r^n}/\mathcal{G}_r^n$. Let

$$\pi_r \; : \; \widehat{\mathcal{M}_r^n} \; \to \; \widehat{\mathcal{M}_r^n}/\mathcal{G}_r^n$$

be the canonical projection. Then, we have the following.

Proposition 4.6. *The quotient space $\widehat{\mathcal{M}_r^n}/\mathcal{G}_r^n$ has a unique structure of C^∞ manifold such that the canonical projection π_r is a covering map.*

Proof. It is sufficient to prove that $\mathcal{G}_r^n$ acts freely and properly discontinuously on $\widehat{\mathcal{M}_r^n}$. Since $\mathcal{G}_r^n$ is a finite group, it is sufficient to show that $\mathcal{G}_r^n$ acts freely on $\widehat{\mathcal{M}_r^n}$. Thus, our task is to prove that if $\mu = (W, A) \in \widehat{\mathcal{M}_r^n}$ and

$$\mathcal{G}_r^n \ni \gamma = {\theta_1}^{\lambda_1} \circ \cdots {\theta_r}^{\lambda_r} \circ \tilde{\sigma} \neq \mathrm{id} \qquad (\lambda_1, \cdots, \lambda_r \in \{0, 1\},\ \sigma \in \mathcal{S}_r),$$

then $\gamma(\mu) \neq \mu$.

First, consider the case $\sigma \neq \mathrm{id}$. Then, there exists an integer $k \in \{1, \cdots, r\}$ such that $\sigma(k) \neq k$. If $\gamma(\mu) = \mu$, then we have

$$(a_{k0},\ a_{k1},\ \cdots,\ a_{kr}) \ = \ \pm\left(a_{\sigma(k)\,0},\ a_{\sigma(k)\,1},\ \cdots,\ a_{\sigma(k)\,r}\right).$$

This contradicts $\mu \in \widehat{\mathcal{M}}_r^n$.

Next, consider the case $\sigma = \mathrm{id}$. Then, there exists an integer $k \in \{1, \cdots, r\}$ such that $\lambda_k = 1$. If $\gamma(\mu) = \mu$, then we have

$$(w_{1\ n+k},\ \cdots,\ w_{n\ n+k}) = (0,\ \cdots,\ 0).$$

This contradicts $\mu \in \widehat{\mathcal{M}}_r^n$.

Hence, we have completed the proof of Proposition 4.6. ∎

Now we construct a unique parametric representation of $\mathcal{D}^n$. We define the parameter space $\mathcal{A}^n$ by

$$\mathcal{A}^n = \bigcup_{0 \le r \in \mathbf{Z}} \left(\widehat{\mathcal{M}}_r^n / \mathcal{G}_r^n \right) \qquad \text{(disjoint union)}.$$

Note that $\mathcal{A}^n$ is a countable disjoint union of C^∞ manifolds with distinct dimensions. We also define the map $\tilde{\mathcal{F}} : \mathcal{A}^n \to \mathcal{D}^n$ by

$$\tilde{\mathcal{F}}(\pi_r(\mu)) = \mathcal{F}(\mu), \qquad \mu \in \widehat{\mathcal{M}}_r^n,\ 0 \le r \in \mathbf{Z}.$$

Note that this definition is well-defined since

$$\mathcal{F}(\mu) = \mathcal{F}(\gamma(\mu)) \qquad (\mu \in \widehat{\mathcal{M}}_r^n,\ \gamma \in \mathcal{G}_r^n,\ 0 \le r \in \mathbf{Z}).$$

Proposition 4.4 implies that the map $\tilde{\mathcal{F}}$ is injective. Corollary 4.5 implies that the map $\tilde{\mathcal{F}}$ is surjective. Hence, we have proved the following theorem.

Theorem 4.7. *Let $\mathcal{A}^n$ and $\tilde{\mathcal{F}} : \mathcal{A}^n \to \mathcal{D}^n$ be as above. Then, the set $\mathcal{A}^n$ is a countable disjoint union of C^∞ manifolds with distinct dimensions, and the map $\tilde{\mathcal{F}}$ is bijective. Namely, the map $\tilde{\mathcal{F}} : \mathcal{A}^n \to \mathcal{D}^n$ gives a unique parametric repesentation of the set $\mathcal{D}^n$ of n-dimensional A-NDSs.*

5 Application of Unique Representations

In this section, we describe two applications of the unique parametric repesentation of n-dimensional A-NDSs constructed in the previous section. One is concerned with non-redundant search sets for learning dynamical systems. The other is concerned with simplification of RNN models of dynamical systems.

5.1 Construction of a non-redundant seach set

From the point of view of developing effective learning algorithms, it is important to investigate a non-redundant search set for learning dynamical systems by RNNs based on A-NDSs. For example, such a non-redundant search set helps restrict the initial values of the learning parameters.

Let us construct the set $\mathcal{B}^n \subset \bigcup_{0 \le r \in \mathbf{Z}} \mathcal{M}_r^n$ of non-redundant representations of n-dimensional A-NDSs by the pairs of RNNs and affine maps.

In order to simplify expression, we use the following notations: First, for any positive integer r and any $(W, A) \in \mathcal{M}_r^n$, we denote by w_{ij} the (i,j)-component of W, and denote by a_{kj} the (k,j)-component of A, and define

$$\begin{aligned} w_j &= (w_{1j}, w_{2j}, \cdots, w_{nj}) \in \mathbf{R}^n, \qquad j = 0, 1, \cdots, n+r, \\ a_k &= (a_{k0}, a_{k1}, \cdots, a_{kn}) \in \mathbf{R}^{n+1}, \qquad k = 1, \cdots, r, \end{aligned}$$

Next, for any positive integer d, and $x = (x_1, \cdots, x_d)$, $y = (y_1, \cdots, y_d) \in \mathbf{R}^d$ such that $x \neq y$, we write

$$x \preceq y$$

if $x_1 < y_1$ or $x_1 = y_1$, $x_2 < y_2$, or $\cdots$, or $x_1 = y_1$, $\cdots$, $x_{d-1} = y_{d-1}$, $x_d < y_d$.

Note that for any positive integer r, any element (W, A) of $\mathcal{M}_r^n$ is determined by specifying $w_j \in \mathbf{R}^n$ $(j = 0, 1, ..., n+r)$ and $a_k \in \mathbf{R}^{n+1}$ $(k = 1, ..., r)$.

Remark. Let $\gamma = {\theta_1}^{\lambda_1} \circ \cdots \circ {\theta_r}^{\lambda_r} \circ \tilde{\sigma} \in \mathcal{G}_r^n$ $(\lambda_1, \cdots, \lambda_r \in \{0, 1\}$, $\sigma \in \mathcal{S}_r)$ and $\mu = (W, A) \in \mathcal{M}_r^n$. Then, $\gamma(\mu) = (\widetilde{W}, \tilde{A})$ can be expressed as follows:

$$\tilde{w}_j = w_j \quad (j = 0, 1, \cdots, n),$$

$$\tilde{w}_{n+k} = (-1)^{\lambda_k} w_{n+\sigma^{-1}(k)} \quad (k = 1, \cdots, r),$$
$$\tilde{a}_k = (-1)^{\lambda_k} a_{\sigma^{-1}(k)} \quad (k = 1, \cdots, r).$$

Now, we define the subset $\mathcal{B}^n$ of $\bigcup_{0 \leq r \in \mathbf{Z}} \mathcal{M}_r^n$ by

$$\mathcal{B}^n = \bigcup_{0 \leq r \in \mathbf{Z}} \mathcal{B}_r^n,$$

where

$$\mathcal{B}_r^n = \left\{ (W, A) \in \widehat{\mathcal{M}}_r^n \;\middle|\; \begin{array}{l} 0 \preceq w_{n+1}, \cdots, 0 \preceq w_{n+r} \\ a_1 \preceq \cdots \preceq a_r \end{array} \right\}.$$

The following theorem shows that the set $\mathcal{B}^n$ gives a non-redundant search set for learning dynamical systems on $\mathbf{R}^n$ by RNNs based on n-dimensional A-NDSs.

Theorem 5.1. *Let $\mathcal{B}^n$ be as above. Then, the restriction*

$$\mathcal{F}|_{\mathcal{B}^n} : \mathcal{B}^n \to \mathcal{D}^n$$

of the parametric representation $\mathcal{F}$ of $\mathcal{D}^n$ is bijective. Namely, $\mathcal{B}^n$ gives the set of non-redundant representations of n-dimensional A-NDSs by the pairs of RNNs with n visible units and affine maps.

Proof of Theorem 5.1. It is sufficient to prove that the map

$$\pi_r|_{\mathcal{B}_r^n} : \mathcal{B}_r^n \to \widehat{\mathcal{M}}_r^n / \mathcal{G}_r^n$$

is bijective for any positive integer r.

Surjectivity: First, we prove the surjectivity. It is sufficient to prove that for any $\mu \in \widehat{\mathcal{M}}_r^n$, there exsits a $\gamma \in \mathcal{G}_r^n$ such that $\gamma(\mu) \in \mathcal{B}_r^n$. The following lemma is needed to prove the surjectivity.

Lemma 5.2. *Let $\mu = (W, A) \in \mathcal{C}_r$, where*

$$\mathcal{C}_r = \left\{ (W, A) \in \widehat{\mathcal{M}}_r^n \;\middle|\; 0 \preceq w_{n+1}, \cdots, 0 \preceq w_{n+r} \right\}.$$

Then, there exists a $\gamma \in \mathcal{G}_r^n$ such that $\gamma(\mu) \in \mathcal{B}_r^n$.

Proof of Lemma 5.2. Since $\mu = (W, A) \in \mathcal{C}_r \subset \widehat{\mathcal{M}}_r^n$, $a_k \neq a_\ell$ if $k \neq \ell$. Thus, there exists a $\sigma \in \mathcal{S}_r$ such that

$$a_{\sigma^{-1}(1)} \preceq \cdots \preceq a_{\sigma^{-1}(r)}.$$

Then, it is easily verified that $\tilde{\sigma}(\mu) \in \mathcal{B}_r^n$. This completes the proof of Lemma 5.2.

Now, let us prove the surjectivity. Let $\mu = (W, A) \in \widehat{\mathcal{M}}_r^n$. Note that $w_{n+1} \neq 0, \cdots$, $w_{n+r} \neq 0$. We define the constants $\lambda_1, \cdots, \lambda_r$ by

$$\lambda_k = \begin{cases} 0 & \text{if } 0 \preceq w_{n+k}, \\ 1 & \text{if } w_{n+k} \preceq 0, \end{cases} \qquad (k = 1, \cdots, r).$$

Then, it is easily verified that

$$\theta_1{}^{\lambda_1} \circ \cdots \circ \theta_r{}^{\lambda_r}(\mu) \in \mathcal{C}_r.$$

Hence, by Lemma 5.2, we can prove the surjectivity.

Injectivity: Next, we prove the injectivity. It is sufficient to prove that if $\mu \in \mathcal{B}_r^n$ and $\gamma \in \mathcal{G}_r^n$ satisfy $\gamma(\mu) \in \mathcal{B}_r^n$, then $\gamma(\mu) = \mu$. In particular, it is sufficient to prove that

$$\mathrm{id} \neq \gamma \in \mathcal{G}_r^n, \ \mu \in \mathcal{B}_r^n \quad \Rightarrow \quad \gamma(\mu) \notin \mathcal{B}_r^n.$$

Hence, the proof of the injectivity is achieved by the following two lemmas:

Lemma 5.3. *Let $\sigma \in \mathcal{S}_r$ and $\lambda_1, \cdots, \lambda_r \in \{0, 1\}$ such that $\lambda_1 + \cdots + \lambda_r \neq 0$. If $\gamma = \theta_1{}^{\lambda_1} \circ \cdots \circ \theta_r{}^{\lambda_r} \circ \tilde{\sigma}$, then $\gamma(\mu) \notin \mathcal{B}_r^n$ for any $\mu \in \mathcal{B}_r^n$.*

Proof of Lemma 5.3. Let $\mu = (W, A) \in \mathcal{B}_r^n$ and $\gamma(\mu) = (\widetilde{W}, \tilde{A}) \in \mathcal{B}_r^n$. Suppose that $\lambda_k = 1$, where k is an integer such that $1 \leq k \leq r$. Since $0 \preceq w_{n+1}, \cdots, 0 \preceq w_{n+r}$, we can easily show that $\tilde{w}_{n+k} \preceq 0$. Hence, $\gamma(\mu) \notin \mathcal{B}_r^n$. This completes the proof of Lemma 5.3.

Lemma 5.4. *Let $\sigma \in \mathcal{S}_r$ such that $\sigma \neq \mathrm{id}$. Then, $\tilde{\sigma}(\mu) \notin \mathcal{B}_r^n$ for any $\mu \in \mathcal{B}_r^n$.*

Proof of Lemma 5.4. Let $\mu = (W, A) \in \mathcal{B}_r^n$ and $\tilde{\sigma}(\mu) = (\widetilde{W}, \tilde{A}) \in \mathcal{B}_r^n$. Since $\sigma \neq \mathrm{id}$, there exist integers k and ℓ such that $1 \leq k < \ell \leq r$ and $\sigma^{-1}(k) > \sigma^{-1}(\ell)$. Since

$a_1 \preceq \cdots \preceq a_r$, we can easily show that $\tilde{a}_\ell \preceq \tilde{a}_k$. Hence, $\tilde{\sigma}(\mu) \notin \mathcal{B}_r^n$ This completes the proof of Lemma 5.4.

We have completed the proof of Theorem 5.1. ∎

5.2 Simplifying RNN models of A-NDSs

In general, simpler models are preferred to more complicated models to make analysis and control easier. In order to obtain a simpler RNN model of a target dynamical system, we consider simplifying the RNN model that has learned the dynamical system, that is, consider decreasing the number of hidden units.

Let us develop a method of simplifying the RNN model $\langle \mathcal{N}_{r^0}^n(W^0); h_{A^0} \rangle$ that has learned a dynamical system ψ on $\mathbf{R}^n$ based on an n-dimensional A-NDS φ. As a solution, we give the method of obtaining all the minimal RNN model of the A-NDS φ from the learned RNN model $\langle \mathcal{N}_{r^0}^n(W^0); h_{A^0} \rangle$.

We put $\mu^0 = (W^0, A^0)$. A minimal RNN model of the A-NDS φ is equivalent to a minimal representation of φ. We have shown in the previous section that the notion of minimal representations are equivalent to the notion of irreduible representations. Thus, our goal is no more and no less than to obtain all the irreducible representations of φ from the given representation μ^0.

Using the criterion of irreducibility, we can easily prove that obtaining one minimal RNN model of the A-NDS φ from the RNN model $\langle \mathcal{N}_{r^0}^n(W^0); h_{A^0} \rangle$ is achieved by successively performing the following four opereations: Each operation derives an RNN model $\langle \mathcal{N}_{r-1}^n(W'); h_{A'} \rangle$ of φ from the prime model $\langle \mathcal{N}_r^n(W); h_A \rangle$ via elimination of a hidden unit $n+\ell$, where $W = (w_{ij})$, $A = (a_{k\beta})$, $W = (w'_{ij})$, $A' = (a'_{k\beta})$, $\ell = 1, \cdots, r$.

1. Eliminate the hidden unit $n + \ell$ such that

$$w_{1\,n+\ell} = \cdots = w_{n\,n+\ell} = 0,$$

and let W' be the matrix obtained by eliminating from the matix W the elements W_{ij}'s with $i = n + \ell$ or $j = n + \ell$, and let A' be the matrix obtained by eliminating from the matix A the elements $A_{\ell\beta}$'s.

2. Eliminate the hidden unit $n+\ell$ such that

$$a_{\ell 1} = \cdots = a_{\ell n} = 0,$$

and let A' be the matrix obtained by eliminating from the matix A the elements $a_{\ell\beta}$'s, and let W' be the matrix obtained by eliminating from the matix W the elements w_{ij}'s with $i = n+\ell$ or $j = n+\ell$, and

$$\begin{aligned} w'_{\alpha 0} &= w_{\alpha 0} + w_{\alpha\, n+\ell} \tanh(a_{\ell 0}) \qquad (\alpha = 1, \cdots, n), \\ w'_{n+k\, 0} &= \sum_{\alpha=1}^{n} a'_{k\alpha} w'_{\alpha 0} + (a'_{k0}/\tau) \qquad (k = 1, \cdots, r-1). \end{aligned}$$

3. Eliminate the hidden unit $n+\ell$ such that there exists one and only one β_0 $(1 \leq \beta_0 \leq n)$ such that

$$a_{\ell\, \beta_0} = \rho\ (= 1 \text{ or } -1), \qquad a_{\ell\beta} = 0\ (\beta \neq \beta_0),$$

and let A' be the matrix obtained by eliminating from the matix A the elements $a_{\ell\beta}$'s, and let W' be the matrix obtained by eliminating from the matix W the elements w_{ij}'s with $i = n+\ell$ or $j = n+\ell$, and

$$\begin{aligned} w'_{\alpha\, \beta_0} &= w_{\alpha\, \beta_0} + \rho\, w_{\alpha\, n+\ell} \qquad (\alpha = 1, \cdots, n), \\ w'_{n+k\, \beta_0} &= \sum_{\alpha=1}^{n} a'_{k\alpha} w'_{\alpha\, \beta_0} \qquad (k = 1, \cdots, r-1). \end{aligned}$$

4. Eliminate the hidden unit $n+\ell$ such that there exists some k_0 $(1 \leq k_0 \leq r)$ such that

$$k_0 < \ell, \qquad a_{k_0\, \beta} = \rho\, a_{\ell\beta} \quad (\beta = 0, 1, \cdots, n),$$

where $\rho = 1$ or -1, and let A' be the matrix obtained by eliminating from the matix A the elements $a_{\ell\beta}$'s, and let W' be the matrix obtained by eliminating from the matix W the elements w_{ij}'s with $i = n+\ell$ or $j = n+\ell$, and

$$\begin{aligned} w'_{\alpha\, n+k_0} &= w_{\alpha\, n+k_0} + \rho\, w_{\alpha\, n+\ell} \qquad (\alpha = 1, \cdots, n), \\ w'_{n+k\, n+k_0} &= \sum_{\alpha=1}^{n} a'_{k\alpha} w'_{\alpha\, n+k_0} \qquad (k = 1, \cdots, r-1). \end{aligned}$$

Consequently, we have achieved a method of obtaining the minimal RNN model $\langle \mathcal{N}_{\tilde{r}}^{n}(\widetilde{W}); \tilde{A} \rangle$ of the A-NDS φ from the given RNN model $\langle \mathcal{N}_{r^0}^{n}(W^0); A^0 \rangle$. Hence, all the minimal RNN models of the A-NDS φ is obtained by

$$\left\{ \langle \mathcal{N}_{\tilde{r}}^{n}(W); A \rangle \;\middle|\; (W, A) = \gamma(\widetilde{W}, \tilde{A}),\ \gamma \in \mathcal{G}_{\tilde{r}}^{n} \right\}.$$

6 A Study of Generalization

In this section, we deal with the problem of generalization for trajectory learning by RNNs within the framework of dynamical system learning. Namely, we would like to discuss whether or not an RNN model of a dynamical system can be acquired by learning its given trajectories. As a first step, we investigate whether or not an A-NDS can be identified from its given trajectories. First, for an A-NDS produced by an RNN without hidden units, we give a geometric criterion for whether or not the A-NDS can be identified from its given trajectories, and present those examples that the A-NDS can and cannot be identified from its given trajectories. Next, this theory is slightly extended to the case of A-NDSs produced by RNNs with hidden units.

6.1 A-NDSs produced by RNNs without hidden units

Consider the set

$$\mathcal{RN}_0^m = \{\mathcal{N}_0^m(W) \mid W \in \mathcal{M}_0^m = M_{m,m+1}(\mathbf{R})\}$$

of all RNNs consisting of m units without hidden units. Fix $W^* = (w_{ij}^*) \in M_{m,m+1}(\mathbf{R})$, $(i = 1, \cdots, m;\ j = 0, 1, \cdots, m)$, and consider the m-dimensional A-NDS Φ_{W^*} produced by the RNN $\mathcal{N}_0^m(W^*)$. Suppose that $\xi^\lambda(t)$ $(\lambda \in \Lambda)$ are the observed trajectories of the dynamical system Φ_{W^*} on $\mathbf{R}^m$ through $p^\lambda \in \mathbf{R}^m$; i.e.,

$$\xi^\lambda(t) = \Phi_{W^*}^t(p^\lambda).$$

Our purpose is to investigate whether or not the A-NDS Φ_{W^*} can be identified from the given trajectories $\xi^\lambda(t)$ $(\lambda \in \Lambda)$ in the set

$$\mathcal{D}_0^m = \{\Phi_W \mid W \in M_{m,m+1}(\mathbf{R})\}$$

of m-dimensional A-NDSs produced by those RNNs belonging to the set $\mathcal{RN}_0^m$.

By Proposition 3.2, identifying the A-NDS Φ_{W^*} in the set $\mathcal{D}_0^m$ is equivalent to identifying W^* in the set $M_{m,m+1}(\mathbf{R})$. We put

$$\mathcal{W}(\xi^\lambda; \lambda \in \Lambda) = \left\{W \in M_{m,m+1}(\mathbf{R}) \mid \Phi_W^t(p^\lambda) = \xi^\lambda(t), \quad t \in \mathbf{R},\ \lambda \in \Lambda\right\}.$$

Then,

$$\left\{\Phi_W \in \mathcal{D}_0^m \mid W \in \mathcal{W}(\xi^\lambda; \lambda \in \Lambda)\right\}$$

is the set of those A-NDSs belonging to $\mathcal{D}_0^m$ that generate the given trajectories $\xi^\lambda(t)$ $(\lambda \in \Lambda)$.

The next proposition gives a geometric interpretation for the set $\mathcal{W}(\xi^\lambda; \lambda \in \Lambda)$.

Proposition 6.1. *Let* $W = (w_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 1, \cdots, m;\ j = 0, 1, \cdots, m)$. *Then,*

$$W \in \mathcal{W}(\xi^\lambda; \lambda \in \Lambda)$$

if and only if

$$(w_{i0}, w_{i1}, \cdots, w_{im}) \in (w_{i0}^*, w_{i1}^*, \cdots, w_{im}^*) + \mathcal{V}(\xi^\lambda; \lambda \in \Lambda)^\perp, \qquad i = 1, \cdots, m,$$

where $\mathcal{V}(\xi^\lambda; \lambda \in \Lambda)$ *is the vector subspace of* $\mathbf{R}^{m+1}$ *spanned by the set*

$$\bigcup_{\lambda \in \Lambda} \left\{\left(1,\ g(\xi^\lambda(t))\right) \in \mathbf{R} \times \mathbf{R}^m \mid t \in \mathbf{R}\right\},$$

and the symbol "⊥" means the orthogonal complement with respect to the canonical inner-product $\langle\,,\,\rangle$ *on* $\mathbf{R}^{m+1}$ *(see Figure 6).*

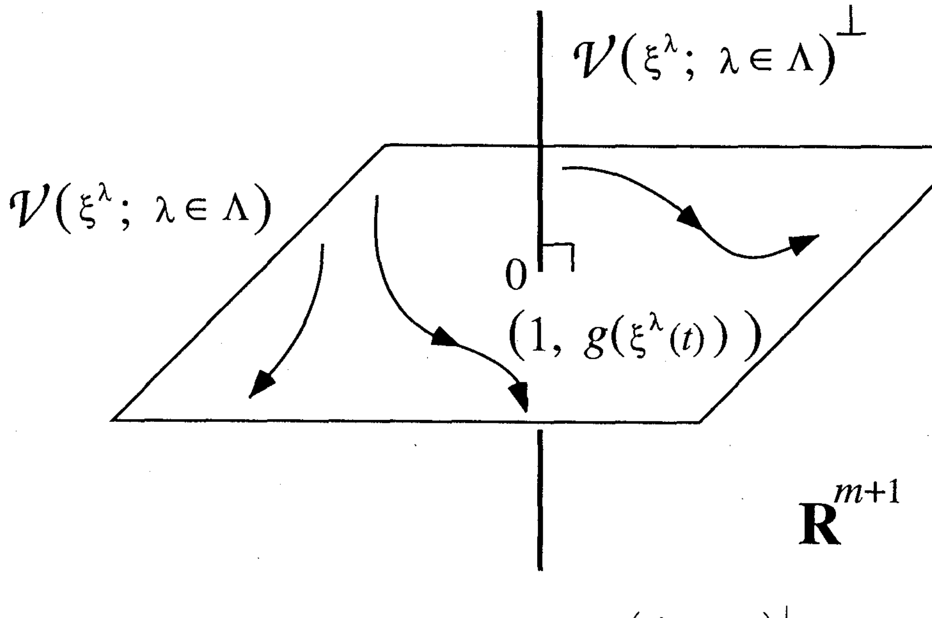


Figure 6. An illustration of the subspace $\mathcal{V}\left(\xi^\lambda; \lambda \in \Lambda\right)^\perp$ of $\mathbf{R}^{m+1}$.

Proof. We observe that $W \in \mathcal{W}(\xi^\lambda; \lambda \in \Lambda)$ if and only if $\xi^\lambda(t)$ is the solution curve of the ordinary differential equation $du/dt = F_W(u)$ under initial condition $u(0) = p^\lambda$ for any $\lambda \in \Lambda$. Hence, a necessary and sufficient condition for $W \in \mathcal{W}(\xi^\lambda; \lambda \in \Lambda)$ is that

$$F_W(\xi^\lambda(t)) = F_{W^*}(\xi^\lambda(t)), \qquad t \in \mathbf{R},\ \lambda \in \Lambda. \tag{6.1}$$

A straightforward calculation yields that condition (6.1) is equivalent to the following condition:

$$\left(\overline{W} - \overline{W}^*\right) g\left(\xi^\lambda(t)\right) + (w_0 - w_0^*) = 0, \qquad t \in \mathbf{R},\ \lambda \in \Lambda. \tag{6.2}$$

Condition (6.2) is also equivalent to the condition

$$\left\langle (w_{i0} - w_{i0}^*, w_{i1} - w_{i1}^*, \cdots, w_{im} - w_{im}^*,)\ ,\ \left(1,\ g(\xi^\lambda(t))\right)\right\rangle = 0,$$

for any $t \in \mathbf{R}$, $\lambda \in \Lambda$, and $i = 1, \cdots, m$. From these results, we can easily prove the proposition. ∎

By Proposition 6.1, we get the following geometric criterion for the possibility of identifying the A-NDS Φ_{W^*} from the given trajectories $\xi^\lambda(t)$ $(\lambda \in \Lambda)$.

Corollary 6.2.

$$\mathcal{W}(\xi^\lambda; \lambda \in \Lambda) = \{W^*\}$$

if and only if

$$\mathcal{V}(\xi^\lambda; \lambda \in \Lambda) = \mathbf{R}^{m+1}.$$

Next, using Corollary 6.2, we present those examples that the A-NDS Φ_{W^*} can and cannot be identified from the given trajectories $\xi^\lambda(t)$ $(\lambda \in \Lambda)$. We put $\tau = 5$ and $g(s) = \tanh(s)$.

Example 11. Let $m = 2$, $W^* = 0 \in M_{2,3}(\mathbf{R})$, $\Lambda = \{\lambda\}$, $p^\lambda = (p_1^\lambda, p_2^\lambda) \in \mathbf{R}^2$ and $\xi^\lambda(t) = \Phi_{W^*}^t(p^\lambda)$. Then, the dynamics of the A-NDS Φ_{W^*} is described as follows:

$$\frac{du_1}{dt} = -\frac{1}{5}u_1$$

$$
\begin{aligned}
\frac{du_2}{dt} &= -\frac{1}{5}u_2 \\
\frac{du_2}{dt} &= -\frac{1}{5}u_3.
\end{aligned}
$$

Figure 7 illustrates the phase portrait of the A-NDS Φ_{W^*}. We can easily prove that

$$
\dim \mathcal{V}(\xi^\lambda) = \begin{cases} 1, & (p^\lambda = 0) \\ 3, & (p_1^\lambda p_2^\lambda \neq 0,\ p_1^\lambda \neq \pm p_2^\lambda) \\ 2, & (\text{otherwise}). \end{cases}
$$

Hence, from the given trajectory $\xi^\lambda(t)$ in Figure 6, the A-NDS Φ_{W^*} can be identified for $\lambda = 9, \cdots, 12$, while it cannot be identified for $\lambda = 1, \cdots, 8$.

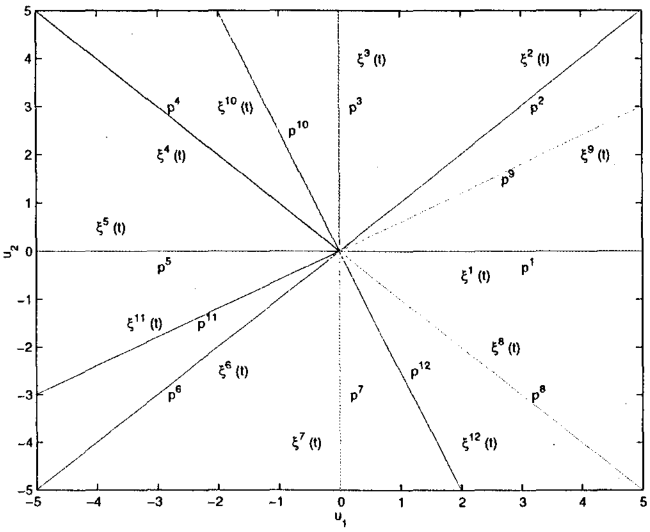


Figure 7. Phase portrait of dynamical system Φ_{W^*}.

$p^1 = (3, 0)$, $p^2 = (3, 3)$, $p^3 = (0, 3)$, $p^4 = (-3, 3)$,
$p^5 = (-3, 0)$, $p^6 = (-3, -3)$, $p^7 = (0, -3)$,
$p^8 = (3, -3)$, $p^9 = (2.5, 1.5)$, $p^{10} = (-1, 2.5)$,
$p^{11} = (-2.5, -1.5)$, and $p^{12} = (1, -2.5)$.

Example 12. Let $m = 3$,

$$
W^* = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},
$$

$\Lambda = \{\lambda\}$, $p^1 = (-3.5, 0.5, 3.2)$, $p^2 = (0.1, 0.2, 0)$ and $\xi^\lambda(t) = \Phi^t_{W^*}(p^\lambda)$, $(\lambda = 1, 2)$. Then, the dynamics of the A-NDS Φ_{W^*} is described as follows:

$$\begin{aligned} \frac{du_1}{dt} &= -\frac{1}{5}u_1 - \tanh(u_3) \\ \frac{du_2}{dt} &= -\frac{1}{5}u_2 + \tanh(u_1) \\ \frac{du_2}{dt} &= -\frac{1}{5}u_3 + \tanh(u_2). \end{aligned}$$

Using the result of Hastings, Tyson and Webster ([18]) for monotone cyclic feedback systems, which has been improved by Mallet-Paret and Smith ([37]), it has been shown ([1], [62], [61]) that the dynamical system Φ_{W^*} has a limit cycle around the origin and the two trajectories $\xi^1(t)$ and $\xi^2(t)$ converge to the limit cycle (see Figure 8). We can easily prove that

$$\dim \mathcal{V}(\xi^\lambda) = 4, \qquad \lambda = 1, 2.$$

Hence, the A-NDS Φ_{W^*} can be identified from the given trajectory $\xi^\lambda(t)$ for $\lambda = 1, 2$.

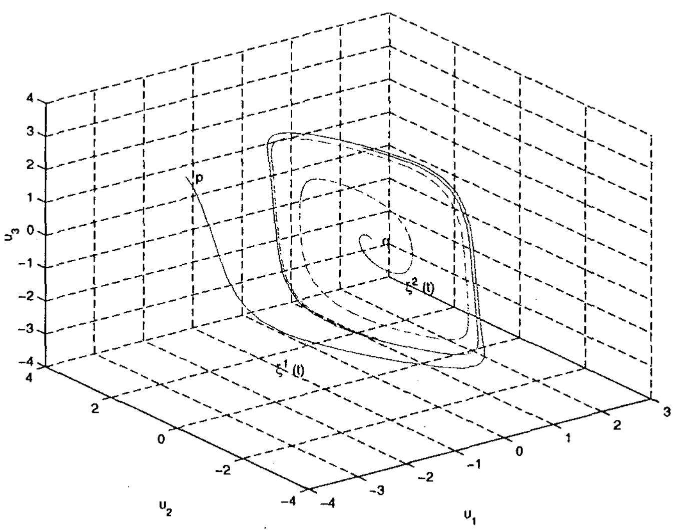


Figure 8. Phase portrait of dynamical system Φ_{W^*}.

Solid line: the trajectory $\xi^1(t)$.

Dashed line: the trajectory $\xi^2(t)$.

6.2 A-NDSs produced by RNNs with hidden units

Let us attempt to extend our theory to the case of RNNs with hidden units. Consider RNNs consisting of m units with n visible units and $r(>0)$ hidden units. Fix $A \in M_{r,n+1}(\mathbf{R})$ and $W^* = (w^*_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 1, \cdots, m;\ j = 0, 1, \cdots, m)$ such that $(W^*, A) \in \mathcal{M}^n_r$. Consider the n-dimensional A-NDS φ_{W^*,h_A} produced by the RNN $\mathcal{N}^n_r(W^*)$ under affine map h_A. Suppose that $\xi^\lambda(t)$ $(\lambda \in \Lambda)$ are the observed trajectories of the dynamical system φ_{W^*,h_A} on $\mathbf{R}^n$ through $p^\lambda \in \mathbf{R}^n$; i.e.,

$$\xi^\lambda(t) = \varphi^t_{W^*,h_A}(p^\lambda).$$

Our purpose is to investigate whether or not the A-NDS φ_{W^*,h_A} can be identified from the given trajectories $\xi^\lambda(t)$, $(\lambda \in \Lambda)$ in the set

$$\mathcal{D}^n_r(A) = \{\varphi_{W,h_A} \mid (W, A) \in \mathcal{M}^n_r\}$$

of n-dimensional A-NDSs produced by RNNs with n visible units and r hidden units under affine map h_A.

We put

$$\mathcal{W}_A(\xi^\lambda; \lambda \in \Lambda) = \left\{ W \in M_{m,m+1}(\mathbf{R}) \;\middle|\; \begin{array}{lcll} (W, A) & \in & \mathcal{M}^n_r, & \\ \varphi^t_{W,h_A}(p^\lambda) & = & \xi^\lambda(t), & t \in \mathbf{R},\ \lambda \in \Lambda \end{array} \right\}.$$

Then,

$$\left\{ \varphi_{W,h_A} \in \mathcal{D}^n_r(A) \mid W \in \mathcal{W}_A(\xi^\lambda; \lambda \in \Lambda) \right\}$$

is the set of those A-NDSs belonging to $\mathcal{D}^n_r(A)$ that generate the given trajectories $\xi^\lambda(t)$ $(\lambda \in \Lambda)$.

The following proposition gives a geometric interpretation for the set $\mathcal{W}_A(\xi^\lambda; \lambda \in \Lambda)$, which is an extension of Proposition 6.1. Using Corollary 3.4, the proposition can be proved in the same way as Proposition 6.1.

Proposition 6.3. *Let $W = (w_{ij}) \in M_{m,m+1}(\mathbf{R})$ $(i = 1, \cdots, m;\ j = 0, 1, \cdots, m)$ such that $(W, A) \in \mathcal{M}^n_r$. Then,*

$$W \in \mathcal{W}_A(\xi^\lambda; \lambda \in \Lambda)$$

if and only if

$$(w_{i0}, w_{i1}, \cdots, w_{im}) \in (w_{i0}^*, w_{i1}^*, \cdots, w_{im}^*) + \mathcal{V}_A(\xi^\lambda; \lambda \in \Lambda)^\perp, \qquad i = 1, \cdots, n,$$

where $\mathcal{V}_A(\xi^\lambda; \lambda \in \Lambda)$ *is the vector subspace of* $\mathbf{R}^{m+1}$ *spanned by the set*

$$\bigcup_{\lambda \in \Lambda} \left\{ \left(1,\ g(\xi^\lambda(t)),\ g \circ h_A(\xi^\lambda(t))\right) \in \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^r \mid t \in \mathbf{R} \right\}.$$

As an extension of Corollary 6.2, we get the following geometric criterion for the possiblity of identifying the A-NDS φ_{W^*,h_A} from the given trajectories $\xi^\lambda(t)$, $(\lambda \in \Lambda)$.

Corollary 6.4.

$$\mathcal{W}_A(\xi^\lambda; \lambda \in \Lambda) = \{W^*\}$$

if and only if

$$\mathcal{V}_A(\xi^\lambda; \lambda \in \Lambda) = \mathbf{R}^{m+1}.$$

7 Conclusion

We have considered continuous time RNNs called additive neural networks or Hopfield neural networks. As one extension of the problem of approximating a trajectory on $\mathbf{R}^n$ by an RNN, we have investigated the problem of approximating a dynamical system on $\mathbf{R}^n$ by an RNN. This has allowed us to discuss the problem of generalization for trajectory learning by RNNs, and leads to investigating the problem of acquiring an RNN model of a dynamical system on $\mathbf{R}^n$ by learning its observed trajectories.

The role of hidden units is crucial for the approximation capability. However, an RNN with n visible units and $r(>0)$ hidden units cannot produce a dynamical system on the visible state space unless a map $h : \mathbf{R}^n \to \mathbf{R}^r$ is successfully specified to determine the initial states of the hidden units for initial states of the visible units. Therefore, toward developing dynamical system learning algorithms, it is necessary to build the framework of how such an RNN produces a dynamical system on $\mathbf{R}^n$ under map $h : \mathbf{R}^n \to \mathbf{R}^r$ to approximate a given dynamical system on $\mathbf{R}^n$. By investigating systematically the dynamical systems produced by RNNs, we have defined the notion of n-dimensional A-NDSs, and proposed to use an n-dimensional A-NDS as a dynamical system that an RNN with n visible units actually produces on the visible state space to approximate a target dynamical system on $\mathbf{R}^n$, that is, proposed A-NDS-based learning for approximating dynamical systems by RNNs. By proving that any dynamical system on $\mathbf{R}^n$ can be approximated well by an n-dimensional A-NDS in a given finite region, we have verified the validity of the A-NDS-based learning (see Theorem 3.6).

An A-NDS is represented by a suitable pair of an RNN and an affine map. However, this representation is not unique. The aim of approximating a dynamical system by an RNN is to acquire an RNN model of the dynamical system by learning its given trajectories, and the existing trajectory learning alogrithms for RNNs are based on the gradient descent method for an error function. Thus, the local minima problem is associated with the learning algorithms, and makes the performance dependent on the initial values of the learning parameters. From the point of view of developing effective learning algorithms, it is significant to understand a non-redundant search set for learning dynamical systems by RNNs based on A-NDSs. For example, such a non-redundant search set helps restrict the initial values of the learning parameters. We have constructed a unique parametric representation of n-dimensional A-NDSs (see Theorem 4.7), and also constructed the non-redundant representations of n-dimensional A-NDSs by the pairs of RNNs with n visible units and affine maps, that is, concretely constructed a non-redundant search set for learning. (see Theorem 5.1).

As the straightforward application of the unique parametric representation of n-dimensional A-NDSs, we have presented a method of obtaining all the minimal RNN models of an A-NDS from a given RNN model of the A-NDS. This method helps simplify the RNN model that has learned a target dynamical system based on an A-NDS. We have also investigated whether or not an A-NDS can be identified from its given trajectories as a first step to a study of generalization for trajectory learning by RNNs.

References

[1] Atiya, A. and Baldi, P., Oscillations and synchronizations in neural networks: An exploration of the labeling hypothesis, *International Journal of Neural Systems*, **2** (1989), 103–124.

[2] Baldi, P., Gradient descent learning algorithm overview: A general dynamical systems perspective, *IEEE Transactions on Neural Networks*, **6** (1995), 182–195.

[3] A. R. Barron, Universal approximation bounds for superposition of a sigmoidal function, *IEEE Transactions on Information Theory*, **39** (1993), 930–945.

[4] Beer, R. D., On the dynamics of small continuous-time recurrent neural networks, *Adaptive Behavior*, **3** (1995), 469–509.

[5] Bersini, H., Saerens, M. and Sotelino, L. G., Hopfield net generation, encoding and classification of temporal trajectories, *IEEE Transactions on Neural Networks*, **5** (1994), 945–953.

[6] Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford Univesity Press, Oxford, 1995.

[7] Cohen, M. A., The construction of arbitrary stable dynamics in nonlinear neural networks, *Neural Networks*, **5** (1992), 83–104.

[8] Cohen M. A. and Grossberg S., Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, **13** (1983), 815–826.

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

[10] Doya, K. and Yoshizawa, S., Adaptive neural oscillator using continuous time back-propagation learning, *Neural Networks*, **2** (1989), 375–385.

[11] Doya, K. and Yoshizawa, S., Memorizing oscillatory patterns in the analog neuron network, In *Proceedings of 1989 International Joint Conference on Neural Networks, Washington D.C.*, **1**, IEEE (1989), 27–32.

[12] Fiedler, B. and Gedeon, T, A class of convergent neural network dynamics, *Physica D*, **111** (1998), 288–294.

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

[14] Funahashi, K., Approximation theory, dynamical systems, and recurrent neural networks, In *Proceedings of the 5th International Colloquium on Differential Equations, Plovdiv, Bulgaria*, **2**, SCT Publishing (1994), 51–58.

[15] Funahashi, K. and Nakamura, Y., Approximation of dynamical systems by continuous time recurrent neural networks, *Neural Networks*, **6** (1993), 801–806.

[16] Gherrity, M., A learning algorithm for analog fully recurrent neural networks, In *Proceedings of 1989 International Joint Conference on Neural Networks, Washington D.C.*, **1**, IEEE (1989), 643–644.

[17] Gouhara, K., Yokoi, K. and Uchikawa, Y., Valley searching method for recurrent neural networks, In *Proceedings of 1992 International Joint Conference on Neural Networks, Baltimore*, **3**, IEEE (1992), 972–979.

[18] Hastings, S. P., Tyson, J. and Webster, D., Existence of periodic solutions for negative feedback cellular control systems, *Journal of Differential Equations*, **25** (1977), 39–64.

[19] Hertz, J., Krogh, A. and Palmer, R. G., *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity, Lecture Notes, **1**, Addison-Wesley, Reading, MA, 1991.

[20] Hirsh, M. W., Convergent activation dynamics in continuous time networks, *Neural Networks*, **2** (1989), 331–349.

[21] Hirsh, M. W., Network dynamics: principles and problems, In F. Pasemann and H-D Doebner (Eds.), *Neurodynamics, Proceedings of the 9th Summer Workshop on Mathematical Physics, Clausthal, Germany*, World Scientific (1991), 3–29.

[22] Hirsh, M. W., Dynamical systems, In P. Smolensky, M. C. Mozer and D. E. Rumelhart (Eds.), *Mathematical Perspectives on Neural Networks*, Lawrence Erlbaum Associates, (1996), 271–323.

[23] Hirsh, M. W. and Smale, S., *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, New York, 1974.

[24] Hofbauer, J. and Sigmund, K., *The Theory of Evolution and Dynamical Systems*, London Mathematical Society Student Texts **7**, Cambridge University Press, Cambridge, 1988.

[25] Hopfield, J. J., Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences, USA*, **79** (1982), 2554–2558.

[26] Hopfield, J. J., Neurons with graded response have collective computational properties like those of two-state neurons, *Proceedings of the National Academy of Sciences, USA*, **81** (1984), 3088–3092.

[27] Hopfield, J. J. and Tank, D. W., Neural computation of decisions in optimization problems, *Biological Cybernetics*, **52** (1985), 141–152.

[28] Hopfield, J. J. and Tank, D. W., Computing with neural circuits: A model, *Science*, **233** (1986), 625–633.

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

[30] M. Kimura and R. Nakano, Learning dynamical systems from trajectories by continuous time recurrent neural networks, In *Proceedings of 1995 IEEE International Conference on Neural Networks*, Perth, Australia, IEEE (1995), 2992–2997.

[31] M. Kimura and R. Nakano, Learning dynamical systems produced by recurrent neural networks, In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff (Eds.), *Artificial Neural Networks - ICANN 96*, Bochum, Germany, Lecture Notes in Computer Science **1112**, Springer (1996), 133–138.

[32] M. Kimura and R. Nakano, Unique representations of dynamical systems produced by recurrent neural networks, In W. Gerstner, A. Germond, M. Hasler, and J. D. Nicoud (Eds.), *Artificial Neural Networks - ICANN'97*, Lausanne, Switzerland, Lecture Notes in Computer Science **1327**, Springer (1997), 403–408.

[33] M. Kimura and R. Nakano, Simplifying fully recurrent neural networks that have learned dynamical systems, In *JCIS'98 Proceedings Volume II - 3rd International Conference on Computational Intelligence and Neuroscience*, Research Triangle Park, North Carolina, USA, Association for Intelligent Machinery (1998), 143–146.

[34] M. Kimura and R. Nakano, Learning dynamical systems by recurrent neural networks from orbits, *Neural Networks*, **11** (1998), 1589–1599.

[35] M. Kimura and R. Nakano, Dynamical systems produced by recurrent neural networks, *Transactions of Institute of Electronics, Information and Communication Engineers*, **J82-D-II** (1999), 818–828, in Japanese.

[36] M. Kimura and R. Nakano, A unique representation of affine neural dynamical systems, *Transactions of Japan Society for Industrial and Applied Mathematics*, **9** (1999), 37–50, in Japanese.

[37] Mallet-Paret, J. and Smith, H. L., The Poincaré-Bendixson theorem for monotone cyclic feedback systems, *Journal of Dynamics and Differential Equations*, **2** (1990), 367–421.

[38] Matsushima, Y., *Differentiable Manifolds*, Marcel-Dekker, New York, 1972.

[39] H. N. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, *Neural Computation*, **8** (1996), 164–177.

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

[41] Pearlmutter, B. A., Learning state space trajectories in recurrent neural networks, In *Proceedings of 1989 International Joint Conference on Neural Networks, Washington D.C.*, **2**, IEEE (1989), 365–372.

[42] B. A. Pearlmutter, Gradient calculations for dynamic recurrent neural networks: A survey, *IEEE Transactions on Neural Networks*, **6** (1995), 1212–1228.

[43] Pineda, F. J., Time dependent adaptive neural networks, In D. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, **2**, Morgan Kaufmann (1990), 710–718.

[44] Reiss, M. and Taylor, J. G., Storing temporal sequences, *Neural Networks*, **4** (1991), 773–787.

[45] Ruiz, A., Owens, D. H. and Townley, S., Existence, learning, and replication of periodic motions in recurrent neural networks, *IEEE Transactions on Neural Networks*, **9** (1998), 651–661.

[46] Rumelhart, D. E., Hinton, G. E. and Williams, R. J., Learning internal representations by error propagation, In D. E. Rumelhart and J. L McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press (1986), 318–362.

[47] Sato, M., A learning algorithm to teach spatiotemporal patterns to recurrent neural networks, *Biological Cybernetics*, **62** (1990), 259–263.

[48] Sato, M., Joe, K. and Hirahara, T., APOLONN brings us to the real world, In *Proceedings of 1990 International Joint Conference on Neural Networks, San Diego*, **1**, IEEE (1990), 581–587.

[49] Sato, M., Murakami, Y. and Joe, K., Learning chaotic dynamics by recurrent neural networks, In *Proceedings of the 1990 International Conference on Fuzzy Logic and Nerual Networks*, (1990), 601–604.

[50] Schmidhuber, J., A fixed size storage $O(n^3)$ time complexity learning algorithm for fully recurrent continually running networks, *Neural Computation*, **4** (1992), 243–248.

[51] Sotelino, L. G., Saerens, M. and Bersini, H., Classification of temporal trajectories by continuous time recurrent nets, *Neural Networks*, **7** (1994), 767–776.

[52] Sun, G. Z., Chen, H. H. and Lee, Y. C., A fast on-line learning algorithm for recurrent neural networks, In *Proceedings of 1991 International Joint Conference on Neural Networks, Seattle*, **2**, IEEE (1991), 13–18.

[53] Sun, G. Z., Chen, H. H. and Lee, Y. C., Green function method for fast on-line learning algorithm of recurrent neural networks, In J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), *Advances in Neural Information Processing Systems*, **4**, Morgan Kaufmann (1992), 333–340.

[54] Sussmann, H. J., Uniqueness of the weights for minimal feedforward nets with a given input-output map, *Neural Networks*, **5** (1992), 589–593.

[55] Toomarian, N. B. and Barhen, J., Learning a trajectory using adjoint functions and teacher forcing, *Neural Networks*, **5** (1992), 473–484.

[56] Tsung, F-S. and Cottrell, G. W., Phase-space learning, In D. S. Touretzky, G. Tesauro and T. K. Leen (Eds.), *Advances in Neural Information Processing Systems*, **7**, MIT Press (1995), 481–488.

[57] Tsung, F-S., Cottrell, G. W. and Selverston, A. I., Some experiments on learning stable network oscillations, In *Proceedings of 1990 International Joint Conference on Neural Networks, San Diego*, **1**, IEEE (1990), 169–174.

[58] van den Driessche, P. and Zou, X., Global attractivity in delayed Hopfield neural network models, *SIAM Journal on Applied Mathematics*, **58** (1998), 1878–1890.

[59] Williams, R. J. and Peng, J., An efficient gradient-based algorithm for on-line training of recurrent network trajectories, *Neural Computation*, **2** (1990), 490–501.

[60] Williams, R. J. and Zipser, Z., A learning algorithm for continually running fully recurrent neural networks, *Neural Computation*, **1** (1989), 270–280.

[61] Yang, H, Intrinsic oscillations of neural networks. In *Proceedings of 1995 International Conference on Neural Networks, Perth, Australia*, **6**, IEEE (1995), 3044–3047.

[62] Yang, H. and Dillon, T. S., Exponential stability and oscillation of Hopfield graded response neural networks, *IEEE Transactions on Neural Networks*, **5** (1994), 719–729.

Paper List

Journal Papers:

J1. M. Kimura and R. Nakano, Learning dynamical systems by recurrent neural networks from orbits, *Neural Networks*, **11** (1998), 1589–1599.

J2. M. Kimura and R. Nakano, Dynamical systems produced by recurrent neural networks, *Transactions of Institute of Electronics, Information and Communication Engineers*, **J82-D-II** (1999), 818–828, in Japanese.

J3. M. Kimura and R. Nakano, A unique representation of affine neural dynamical systems, *Transactions of Japan Society for Industrial and Applied Mathematics*, **9** (1999), 37–50, in Japanese.

International Conference Papers:

C1. M. Kimura and R. Nakano, Learning dynamical systems from trajectories by continuous time recurrent neural networks, In *Proceedings of 1995 IEEE International Conference on Neural Networks*, Perth, Australia, IEEE (1995), 2992–2997.

C2. M. Kimura and R. Nakano, Learning dynamical systems produced by recurrent neural networks, In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff (Eds.), *Artificial Neural Networks - ICANN 96*, Bochum, Germany, Lecture Notes in Computer Science **1112**, Springer (1996), 133–138.

C3. M. Kimura and R. Nakano, Unique representations of dynamical systems produced by recurrent neural networks, In W. Gerstner, A. Germond, M. Hasler, and J. D. Nicoud (Eds.), *Artificial Neural Networks - ICANN'97*, Lausanne, Switzerland, Lecture Notes in Computer Science **1327**, Springer (1997), 403–408.

C4. M. Kimura and R. Nakano, Simplifying fully recurrent neural networks that have learned dynamical systems, In *JCIS'98 Proceedings Volume II - 3rd International Conference on Computational Intelligence and Neuroscience*, Research Triangle Park, North Carolina, USA, Association for Intelligent Machinery (1998), 143–146.