Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

On Flexible Neural Networks: Some System-Theoretic Properties and a New Class

Yazdan Bavafa-Toosi and Hiromitsu Ohmori
School of Integrated Design Engineering
Keio University, Yokohama
Kanagawa 223-8522, Japan

Abstract—**Although flexible neural networks (FNNs) have been used more successfully than classical neural networks (CNNs), nothing is rigorously known about their properties. In fact, they are not even well known to the systems and control community. In this paper, theoretical evidence is given for their superiority over CNNs. Following an overview of flexible bipolar sigmoid functions (FBSFs), several fundamental properties of feedforward and recurrent FNNs are established. For the feedforward case, it is proven that similar to CNNs, FNNs with as few as a single hidden layer (SHL) are universal approximators. It is also proven that unlike irreducible SHL CBSNNs, irreducible SHL FBSNNs are nonuniquely determined by their input-output (I-O) maps, up to a finite group of symmetries. Then, recurrent FNNs are introduced. It is observed that they can be interpreted as a generalization of the conventional state-space framework. For the recurrent case, it is substantiated that similar to CBSNNs, FBSNNs are universal approximators. Necessary and sufficient conditions for the controllability and observability of a generic class of them are established. For a subclass of this class, it is proven that unlike CBSNNs, FBSNNs are nonuniquely determined by their I-O maps, up to a finite group of symmetries, and that every system inside this subclass is minimal. Finally, a new class of FNNs, namely, flexible bipolar radial basis neural networks (FBRBNNs) is introduced. It is proven that as in the case of classical radial basis neural networks (CRBNNs), feedforward SHL FBRBNNs are universal approximators.**

I. Introduction

ARTIFICIAL neural networks (ANNs) have been a topic of extensive research in the last few decades. Several types of ANNs have found their ways to control systems: Hopfield network, Kohonen's self-organizing map, Boltzmann machine, support vector machine, recurrent network, multilayer perception, and FNNs, to mention among others [1]-[7].

FBSNNs and FUSNNs (flexible unipolar sigmoid neural networks) were introduced in [1]-[4] where it was shown that they have superior learning capacity over CBSNNs and CUSNNs, respectively. Despite the successful applications of FNNs, they are not yet well known to the systems and control community. Besides, nothing is rigorously known about their properties, while several fundamental properties of CNNs have already been established.

The universal approximation capabilities of feedforward CNNs, which comprise *uni*variate function units, have been established in a number of works, see e.g. [8]-[10] and the references therein[1].

In this work, the same results are proven for feedforward FNNs, which consist of *bi*variate function units. More precisely, it is substantiated that feedforward FNNs with as few as an SHL are universal approximators.

The interpolation capabilities of feedforward SHL CNNs have been extensively studied in e.g. [13]-[15] and the references therein. The question of the number of hidden nodes needed to achieve a specific objective was first addressed in [13]. In [14] it was proven that two irreducible nets with the same I-O maps are equivalent, i.e., related by a transformation in a finite group of symmetries (permutations of the hidden nodes, and changing the sign of all the weights associated with a particular hidden node).

In this work, the same property is investigated for feedforward SHL FNNs. More exactly, the following question is addressed: To what extent is a feedforward SHL FBSNN uniquely determined by its I-O map? (Note that this is also referred to as the identifiability problem.) To this end, in line with [14], we first derive conditions under which an FBSNN is irreducible. By irreducibility we mean that the number of hidden nodes cannot be decreased without changing the I-O map of the net. These conditions can easily be checked. Then, we define a minimal net as one whose I-O map cannot be obtained from a net with fewer hidden nodes. This is not an easy condition to check on its own. Based on the above definitions, it is clear that a minimal net is irreducible. We prove that the converse is also true, as the corollary of our theorem on the equivalence of irreducible nets.

On the other hand, recurrent CNNs are teeming in the literature. They can be interpreted as a representation of the evolution of a collection of parallel processing neurons. They have been used in sequence extrapolation for time series

This work was supported in part by a Grant in Aid for the 21st Century Center of Excellence for "System Design: Paradigm Shift from Intelligence to Life" from The Ministry of Education, Culture, Sport and Technology of Japan.

1: The mathematical tools for the derivation of these results typically fall in two main categories: a) algebras of functions (leading to Stone-Weierstrass theorem [11]) and b) translation invariant subspaces (leading to Tauberian theorems [12]).

prediction, signal processing, control, and associative memories (Tank-Hopfield networks), to mention among others [5]-[7]. To date, several fundamental properties, or rather system-theoretic properties, of recurrent CNNs have been discovered. For example, their approximation capability, controllability, observability, identifiability, and minimality have been extensively studied in [16]-[21].

In this work, recurrent CBSNNs are extended to recurrent FBSNNs through the use of flexible neurons. In line with [16]-[19], the aforementioned properties are then addressed for a generic class of them. As will be observed, recurrent FBSNNs are more powerful than recurrent CBSNNs, since e.g. the flexible parameters may be used to make the system controllable/observable and/or to satisfy some other design objectives. Due to lack of space the results are presented only for continuous-time systems.

Finally, a new class of FNNs namely FBRBNNs is introduced. It is proven that similar to CRBNNs, feedforward SHL FBRBNNs are universal approximators.

The organization of this work is as follows. First an overview of the FBSF is given in the following. Then, feedforward and recurrent FNNs are studied in Sections II and III, respectively. FBRBNNs are introduced in Section IV while conclusions are presented in Section V. The paper is wrapped up in Section VI by some hints for future work.

A. *An Overview of FBSFs*

An FBSNN is basically composed of some FBSFs as its parallel processing units (or nodes) in an SHL, where the output of the *i*th unit is given by the bivariate function,

$$f(x_i, z_i) = \frac{1 - \exp(-2z_i x_i)}{z_i(1 + \exp(-2z_i x_i))}, \tag{1}$$

in which x_i is the input to the *i*th node and z_i is its parameter which must be trained. The FBSF is depicted in Figure 1.

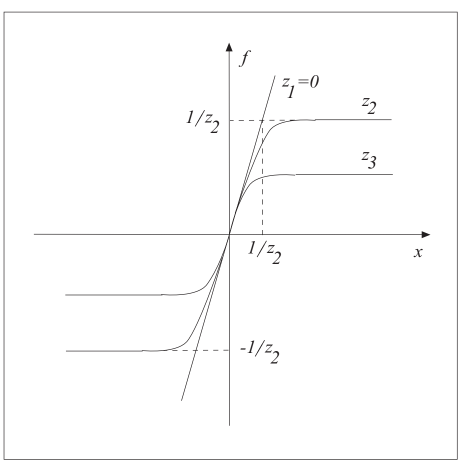

Fig. 1. Shapes of the FBSF

It is well known that the CBSF (i.e., (1) with $z = 1$) is linear well inside the input interval $(-1,1)$ and saturates outside that. However, it is easy to verify that the FBSF exhibits both linear ($f \to x$ as $z \to 0$) and nonlinear (otherwise, unless well inside the input interval $(-1/z, 1/z)$) behavior over the whole range of the input. This accounts for its superiority over CBSF which has been verified by simulation and experimental results of various case studies like connected tanks and robot manipulators [1]-[4].

In this work we further reveal this superiority in a theoretical framework by e.g. characterizing the structure of the flexibility in both feedforward and recurrent FBSNNs. We also prove that by the proper choice of the flexible parameters a recurrent FBSNN may become controllable/observable while its classical counterpart is not.

II. FEEDFORWARD FNNS

The approximation capabilities of FNNs are studied in the following development.

A. *Universal Approximation Property*

Since the extension of our results to the multi-output case is trivial, only single-output FNNs are considered. Given an FNN with n units in its SHL and a summation node in its output, the input to the *j*th hidden node is given by,

$$v_j = y_{0j} + \sum_{i=1}^{m} y_{ij} x_i, \tag{2}$$

where $x = [x_1, \ldots, x_m]^T \in R^m$ is the input vector, $Y = [y_1, \ldots, y_n]$ where $y_j = [y_{0j}, y_{1j}, \ldots, y_{mj}]^T \in R^{m+1}$ in which y_{0j} is the threshold (or the bias term) of the *j*th hidden node and the rest of the entries are the connection weights from the inputs to that node. The output of the network is thus,

$$F(x, Y, z, w) = w_0 + \sum_{j=1}^{n} w_j f(v_j, z_j), \tag{3}$$

where $z = [z_1, \ldots, z_n]^T \in R^n$ in which z_i is the parameter of the *j*th hidden node, $w = [w_0, w_1, \ldots, w_n]^T \in R^{n+1}$ where w_0 is the bias term of the output node and w_j is the output weight of the *j*th hidden node.

Let I_m be the m-dimensional unit cube, $[0,1]^m$. The following theorems can be presented.

Theorem 2.1: Finite sums of the form (3) where f is given by (1) are dense in the space of continuous functions on the m-dimensional cube, $C(I_m)$. In other words, for any

$h(x) \in C(I_m)$ and $\varepsilon > 0$ there exists a finite sum $F(x, Y, z, w)$ for which $|F(x, Y, z, w) - h(x)| < \varepsilon$, $\forall x \in I_m$.

Theorem 2.2: Finite sums of the form (3) where f is given by (1) are dense in the space of integrable functions on the m-dimensional cube, $L^1(I_m)$. In other words, for any $h(x) \in L^1(I_m)$ and $\varepsilon > 0$ there exists a finite sum $F(x, Y, z, w)$ for which $\int_{I_m} |F(x, Y, z, w) - h(x)|\, dx < \varepsilon$, $\forall x \in I_m$.

Proofs: It is easily seen that (3) can be rewritten as or rather reduced to,

$$F(x, Y, z, w) = w_0 + \Sigma_{j=1}^{n} \frac{w_j}{z_j} g(v_j z_j), \tag{4}$$

where $g(.) = \frac{1 - \exp(-2\ .)}{1 + \exp(-2\ .)}$ is a univariate function and (4) fits the class of systems considered in [8, theorem 1] and [8, theorem 4], respectively.

Δ.

Remark 2.1: It is clear that the above theorems also apply to FUSNNs in which f is given by $f(x, z) = 2|z|/(1 + \exp(-2|z|x))$.

In the sequel subsection, the identification of the parameters and weights of an SHL FBSNN from its I-O data is studied.

B. *Identifiability and Minimality*

Let a) $N_{m,n,z}$ denote the set of all feedforward SHL FBSNNs with m inputs (labeled $1,\ldots,m$), n hidden nodes (labeled $1,\ldots,n$) with parameters $z = \{z_1, \ldots, z_n\}$ ($z_i \neq 0,\ \forall i$), and a single output given by a summation node, as in Subsection II.A, and b) a *net* be a member of $N_{m,n,z}$ for some m, n and z.

Therefore, the output of the network is given by (4) where g is the hyperbolic tangent function. The function $x \to F(x, Y, z, w)$ is the *I-O map* of the net. For a given m, we call two nets *I-O equivalent* if their I-O maps are the same. Our main theorem states that two I-O equivalent nets are the same up to a finite group of symmetries. It follows from this theorem that an irreducible net is minimal.

Let the two functionals $z_{j_1} v_{j_1}$ and $z_{j_2} v_{j_2}$ be called *sign-equivalent* if $|z_{j_1} v_{j_1}| = |z_{j_2} v_{j_2}|$. We call a net *reducible* if one of the following conditions holds:

a) one of the output weights w_j is zero;

b) there exist two different hidden nodes j_1 and j_2 such that $|z_{j_1} v_{j_1}| = |z_{j_2} v_{j_2}|$;

c) one of the functionals v_j is a constant.

It is clear that if (a) holds, the corresponding hidden node makes no contribution to the output and thus can be excluded form the net. As for (b) note that f is odd with respect to v, even with respect to z, and odd with respect to v and z. Equally, g is odd with respect to zv. Therefore, if (b) holds, i.e., $z_{j_1} v_{j_1} = \rho z_{j_2} v_{j_2}$ where $\rho = 1$ or $\rho = -1$, the contribution of these hidden nodes to the output will be $\frac{w_{j_1}}{z_{j_1}} g(z_{j_1} v_{j_1}) + \frac{w_{j_2}}{z_{j_2}} g(z_{j_2} v_{j_2}) = (\rho \frac{w_{j_1}}{z_{j_1}} + \frac{w_{j_2}}{z_{j_2}}) g(z_{j_2} v_{j_2})$ and thus the number of the hidden nodes can be reduced. Finally, if (c) holds, the corresponding node can be excluded provided the bias of the output node is changed accordingly.

The aforementioned finite group of *symmetries* $\mathcal{G}$ is introduced in the following. There are some obvious transformations which do not alter the I-O map of the net. These are: 1) changing the sign of any z_j, 2) changing the sign of the output and all the input weights of any hidden node (note that the output weight of the jth hidden node is w_j / z_j based on (4) which we are using, and its input weights are v_j), and 3) interchanging (i.e., relabeling) two hidden nodes, as well as their associated weights and parameters (note that this also results in relabeling their $z_j v_j$'s). It is clear that any combination of the above transformations is also acceptable. The above-defined transformations form a finite group of transformations, denoted by $\mathcal{G}_{m,n,z}$, on the set $N_{m,n,z}$. We call two nets in this set *equivalent* if they are related by a transformation in this group. It thus follows that two equivalent nets are I-O equivalent. The following theorem states that the converse is true for irreducible nets.

Theorem 2.3: If N_1 and N_2 are two irreducible I-O equivalent nets in the sets N_{m,n^1,z^1} and N_{m,n^2,z^2}, respectively, then (a) $n^1 = n^2$ and (b) N_1 and N_2 are equivalent.

Proof: The assumption that N_1 and N_2 are I-O equivalent means $w_0^1 + \sum_{j=1}^{n^1} w_j^1 f(v_j^1, z_j^1) = w_0^2 + \sum_{j=1}^{n^2} w_j^2 f(v_j^2, z_j^2)$ where the involved terms have obvious meanings. Define $J = \{1, \ldots, n^1 + n^2\}$ and $b_0 = w_0^1 - w_0^2$. Also, for

$1 \le j \le n^1$ define $\varphi_j = v_j^1$, $b_j = w_j^1$, $d_j = z_j^1$, and for $n^1 + 1 \le j \le n^1 + n^2$ define $\varphi_j = v_{j-n^1}^2$, $b_j = -w_{j-n^1}^2$, $d_j = z_{j-n^1}^2$. Then the assumption becomes $b_0 + \sum_{j \in J} b_j f(\varphi_j, d_j) = 0$ or equally $b_0 + \sum_{j \in J} \frac{b_j}{d_j} g(d_j \varphi_j) = 0$. This last equation is in the form [14, (2)] and thus [14, lemma 1] applies.

Lemma 1: Let J be a finite set and let $\{d_j \varphi_j\}_{j \in J}$ be a family of non-constant linear affine functions on R^m, no two of which are sign-equivalent. Then the functions $g(d_j \varphi_j)$, $j \in J$, and the constant function 1 are linearly independent.

Coming back to the proof of the theorem, with a similar argument as that in [14, page 593] we conclude that $n^1 = n^2$, $w_0^1 = w_0^2$, $z_j^1 v_j^1 \equiv \pm z_j^2 v_j^2$ and $\frac{w_j^1}{z_j^1} = \pm \frac{w_j^2}{z_j^2}$ with the same choice of sign in both. In other words, N_1 and N_2 are equivalent.

Δ.

Corollary 2.1: An irreducible net is minimal.

Proof: The proof is the same as that of [14, corollary 1].

Δ.

Remark 2.2: It is worth mentioning that our result contains that of [14] as a special case ($d_j = 1,\ j \in J$) and that the structure of the inherent flexibility has been completely characterized. This freedom (or flexibility) in the choice of parameters and weights further explains the superior learning capabilities of FBSNNs over CBSNNs, as previously stated and verified in [1]-[4].

Remark 2.3: An interpretation of the result is that there is no mechanism other than the three mentioned in the definition of reducibility to reduce the number of hidden nodes without changing the I-O map of the net.

III. Recurrent FBSNNs

To see how the dynamics of the general system is derived, the following illustrative example is considered, see Fig. 2.

In this figure, the system is described by $\dot{x}_1 = f(-3x_1 + u_1 - 2u_2, z_1)$, $\dot{x}_2 = f(x_1 + 4x_2 + 5u_2, z_2)$, $y = x_2$ where f is given by (1), i.e., $f(r,z) = \frac{1}{z}\tanh(zr) = \frac{1}{z} g(s)$ and thus $\dot{x} = Z^{-1}\vec{g}(ZAx + ZBu)$, $y = Cx$ in which $x = [x_1, x_2]^{\mathrm{T}}$ $Z = [z_1, 0; 0, z_2]$, $A = [-3, 0; 1, 4]$, $B = [1, -2; 0, 5]$, $C = [0, 1]$, and $\vec{g}(s) = [g(s), 0; 0, g(s)]$.

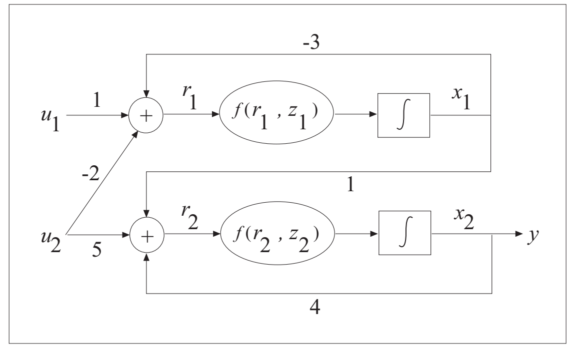

Fig. 2. A 2-dimensional, 2-input, 2-output recurrent FBSNN

Hence, given the quadruple (Z, A, B, C) where $Z = diag\{z_1, \ldots, z_n\}$, $z_i \ne 0$ $(i = 1, \ldots, n)$, $A \in R^{n \times n}$, $B \in R^{n \times m}$, and $C \in R^{p \times n}$ are the parameter, state, input and output matrices, respectively, the continuous-time recurrent FBSNN is described by the state-space equations,

$$\dot{x} = Z^{-1}\vec{g}(ZAx + ZBu), \quad y = Cx, \tag{5}$$

in which $\vec{g}$ is the diagonal mapping,

$$\vec{g} : R^n \to R^n := \begin{pmatrix} g & & 0 \\ & \ddots & \\ 0 & & g \end{pmatrix}, \tag{6}$$

and $g(.) = \tanh(.)$. To further reveal the structure of the flexibility, the transformation $v = Zx$ is applied by which (5) is reduced to,

$$\dot{v} = \vec{g}(ZAZ^{-1}v + ZBu), \quad y = CZ^{-1}v, \tag{7}$$

denoted by G, which is reminiscent of similarity transformations and equivalence in the conventional state-space framework. The following two points are noteworthy: i) It is possible to derive some slightly different models from (5),(7) (see e.g. [16]-[19]), but this work is restricted to (5),(7) only. The main advantage is that this model can be viewed as a generalization of the conventional model considered in the context of linear systems theory in which $\vec{g}$ is the identity function. Therefore, their theoretical properties may be studied in parallel, as done inhere; ii) In the above procedure the bivariate function f has been reduced to the univariate function g, and that (5),(7) contain the representation of recurrent CBSNNs as the special case that Z is the identity matrix.

Equations of the form (5),(7) can be interpreted as follows. The vector x represents the evolution of n flexible neurons where each coordinate x_i of x is the internal state of the ith flexible neuron. Each flexible neuron has its own adaptable/ flexible parameter z_i. The elements of matrices A, B, and C denote the weights (or synaptic strengths) of different connections. The function $g: R \to R$ is the activation function.

In the following subsections some properties of recurrent FBSNNs are presented. Prior to this, it should be mentioned that $\vec{g}$ in (7) is exactly the same as that considered in [16]-[19] and thus satisfies the following conditions:

a) $\forall (\alpha_i, \beta_i) \in R^2$, $\alpha_i \neq 0$, $i = 1,\ldots,n$, and $(\alpha_i, \beta_i) \neq (\alpha_j, \beta_j)$ $i \neq j$, the functions $g(\alpha_i x + \beta_i)$, $i = 1,\ldots,n$, and the constant function 1 are linearly independent;
b) g is odd and locally Lipschitz;
c) $g_\infty := \lim_{t \to +\infty} g(t)$ exists and is positive;
d) $\forall r \in R: g_r < g_\infty$;
e) $\forall \alpha, \beta \in R$, $\alpha > 1$: $\lim_{t \to +\infty} \dfrac{g_\infty - g(\alpha t + \beta)}{g_\infty - g(t)} = 0$.

Remark 3.1: The sequel properties which are presented for the system G in (7) imply the same for the original system in (5).

Remark 3.2: The following results can be obtained for more general mappings than the hyperbolic tangent embedded in FBSNNs, see [16]-[19] for more details.

A. Universal Approximation Property

Recurrent FBSNNs of the form (7) are universal approximators in the following sense. Given any system $\hat{G}$: $\dot{\hat{v}} = f(\hat{v}, u)$, $\hat{y} = h(\hat{v})$ with dimension $\hat{n}$, m inputs and p outputs, where $f: R^{\hat{n}} \times R^m \to R^{\hat{n}}$ is continuously differentiable, f and $f_{\hat{v}}$ are continuous on $\hat{v}$ and u, and $h: R^{\hat{n}} \to R^p$ is continuous. We assume that the solution $\hat{v}(t, \hat{v}(0), u)$, $t \in [0,T]$ exists for every $\hat{v}(0) \in R^{\hat{n}}$ and every measurable, essentially bounded $u: [0,T] \to R^m$.

Consider the two compact subsets $K_1 \subseteq R^{\hat{n}}$, $K_2 \subseteq R^m$, an $\varepsilon > 0$ and a $T > 0$. We say that system G (with dimension n, m inputs, and p outputs) in the form of (7) *approximates* or *simulates* $\hat{G}$ on the sets K_1, K_2 in time T and up to accuracy ε if for every $\hat{v}(0) \in K_1$ and every (measurable, essentially bounded) input $u: [0,T] \to K_2$ there exist two differentiable maps $\alpha: R^n \to R^{\hat{n}}$, $\beta: R^{\hat{n}} \to R^n$ for which,

$$\| \hat{v}(t, \hat{v}(0), u) - \alpha(v(t, \beta(\hat{v}(0)), u) \| < \varepsilon,$$
$$\| h(\hat{v}(t, \hat{v}(0), u)) - CZ^{-1}(v(t, \beta(\hat{v}(0)), u)) \| < \varepsilon, \qquad (8)$$

for all $t \in [0,T]$, where $v(t, v(0), u)$, in general, denotes the unique solution $v: [0,T] \to R^n$ of G with the initial condition $v(0)$ and the (measurable, essentially bounded input) $u: [0,T] \to R^m$, and similarly for $\hat{v}(t, \hat{v}(0), u)$.

The sequel theorem can be offered whose proof is in line with that of [16, theorem 1].

Theorem 3.1: For every system $\hat{G}$ and every K_1, K_2, ε, T as above, there is a system G in the form of (7) which simulates $\hat{G}$ on the sets K_1, K_2 in time T up to accuracy ε.

Δ.

Remark 3.3: The above result can be seen as a counterpart for the convergence property of some classes of recurrent NNs studied in [20],[21] and the references therein.

In the rest of this section, as in the case of the literature on recurrent CBSNNs, the sequel *generic* set $\mathcal{B}$ will play a crucial role,

$$\mathcal{B}_{n,m} := \left\{ B \in R^{n \times m} \mid \forall i: B_i \neq 0 \text{ and } \forall i \neq j: B_i \neq \pm B_j \right\}, \qquad (9)$$

where B_i denotes the ith row of B. It is clear that in the case of single-input systems, i.e., $m = 1$, $ZB \in \mathcal{B}$ iff all the elements of the single-column matrix ZB are nonzero and have different absolute values.

The results of the following subsections are in a geometric framework, see [22] for an introduction.

B. Controllability

The definition of controllability for system (7) is the same as that for conventional linear systems. More precisely, the system is said to be controllable if every initial state can be steered (or controlled) to every state in some finite time. The subsequent results can be proven similar to those of [17].

Theorem 3.2: If $ZB \in \mathcal{B}$, then system (7) is controllable (for all ZAZ^{-1}).

Δ.

Note that it is possible that $ZB \notin \mathcal{B}$ but the system is controllable for *some* ZAZ^{-1}. A trivial converse of the above theorem is as follows.

Theorem 3.3: If system (7) is controllable for all ZAZ^{-1}, then $ZB \in \mathcal{B}$.

Δ.

It is remarkable that the above results are in contrast to those of conventional linear systems, in which the pair (ZAZ^{-1}, ZB) is controllable independent of ZAZ^{-1} iff ZB is of full rank n. Besides, the effect of Z (on the controllability of the system) is partly characterized by the fact that if $B \notin \mathcal{B}$ because $B_i = \pm B_j$, then $ZB \in \mathcal{B}$ if $z_i \neq \pm z_j$.

C. Observability

The definition of observability for system (7) is also the same as that for conventional linear systems. More exactly, the system is called observable if there is some control that gives different outputs when the system is initialized at any two different initial states.

Let $M \in R^{n \times n}$ and $\mathcal{V} \subseteq R^n$ be a subspace. It is said that $\mathcal{V}$ is *M*-invariant if $M\mathcal{V} \subseteq \mathcal{V}$. Also, let e_i, $i = 1,...,n$ denote the canonical basis elements in R^n. A subspace spanned by $e_{i_1},...,e_{i_j}$, $j > 0$, is called a coordinate subspace. The following results can be proven along those of [18].

Theorem 3.4: If $ZB \in \mathcal{B}$, then system (7) is observable iff $\ker ZAZ^{-1} \cap \ker CZ^{-1} = \mathcal{O}_c(ZAZ^{-1}, CZ^{-1}) = 0$, where the unique largest ZAZ^{-1}-invariant coordinate subspace included in $\ker CZ^{-1}$ is denoted by $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1})$.

Δ.

Let the largest ZAZ^{-1}-invariant subspace included in $\ker CZ^{-1}$ be denoted by $O(ZAZ^{-1}, CZ^{-1})$. Thus, in the realm of conventional linear systems observability of the pair (ZAZ^{-1}, CZ^{-1}) is equivalent to $O(ZAZ^{-1}, CZ^{-1}) = 0$. Since $O(ZAZ^{-1}, CZ^{-1})$ includes both $\ker CZ^{-1}$ and $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1})$, the above theorem results in:

Corollary 3.1: If $ZB \in \mathcal{B}$ and the pair (ZAZ^{-1}, CZ^{-1}) is observable, then system (7) is observable.

Δ.

There are different ways to compute $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1})$ and $O(ZAZ^{-1}, CZ^{-1})$. In particular, a recursive algorithm can easily be designed for the former similar to that of [18].

On the other hand, if A is invertible, then $\ker ZAZ^{-1} \cap \ker CZ^{-1} = 0$. Hence:

Corollary 3.2: If $ZB \in \mathcal{B}$ and A is invertible, then system (7) is observable iff $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1}) = 0$.

Δ.

Also, it can easily be shown that if C has no zero column, then $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1}) = 0$. Consequently:

Corollary 3.3: If $ZB \in \mathcal{B}$, $\ker ZAZ^{-1} \cap \ker CZ^{-1} = 0$ and there is no zero column in C, then system (7) is observable.

Δ.

It should be mentioned that if $ZB \notin \mathcal{B}$, then a) the observability of the pair (ZAZ^{-1}, CZ^{-1}) is no longer sufficient for the observability of system (7), and b) the conditions of theorem 3.4 are still necessary (but not sufficient) for the observability of system (7). These are examples of the well-known fact that for nonlinear systems – in contrast to conventional linear systems – observability may depend on the input matrix. Moreover, it is not difficult to show that $\ker ZAZ^{-1} = Z \ker A$ and $\ker CZ^{-1} = Z \ker C$, but $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1}) \neq K\mathcal{O}_c(A, C)$. Thus, Z affects the observability of the system (independently of its effect on B).

D. Identifiability and Minimality

Identifiability refers to the possibility of identifying (or recovering) the elements of the quadruple (Z,A,B,C) from the *I-O map* of the system *G*, given by the function $u \to y$ in (7).

Given the quadruples (Z_1, A_1, B_1, C_1) and (Z_1, A_2, B_2, C_2) denoting the systems G_1 and G_2, respectively, with *arbitrary* initial states. We call the systems *equivalent* if $n_1 = n_2$ and there exists a matrix $T \in \Phi$ such that,

$$
\begin{aligned}
v_1 &= Tv_2, \\
Z_1 A_1 Z_1^{-1} &= TZ_2 A_2 Z_2^{-1} T^{-1}, \\
Z_1 B_1 &= TZ_2 B_2, \\
C_1 Z_1^{-1} &= C_2 Z_2^{-1} T^{-1},
\end{aligned} \tag{10}
$$

where $\Phi = \{T \mid T = PQ\}$ in which P is a permutation matrix and $\Phi = diag\{\rho\}$, $\rho = 1$ or $\rho = -1$. The following results hold similar to those of [19].

Theorem 3.5: Let the systems G_1 and G_2 in the form of (7) be observable and satisfy the generic condition (9). Then, they are I-O equivalent iff they are equivalent.

Δ.

It is worth mentioning that conditions (10) completely characterize the structure of the inherent flexibility in observable recurrent FBSNNs satisfying (9).

Remark 3.4: If $ZB \in \mathcal{B}$, observability is a necessary condition for I-O equivalence to result in equivalence.

Remark 3.5: If $ZB \in \mathcal{B}$ and the system is not observable because $\mathcal{O}_c(ZAZ^{-1}, CZ^{-1}) \neq 0$, it can be reduced to an I-O equivalent observable system of a lower dimension.

A recurrent FBSNN is said to be minimal if its I-O map cannot be obtained from a recurrent FBSNN with a lower dimension, i.e., with fewer flexible units. An important interpretation of theorem 3.5 is that inside the class of observable recurrent FBSNNs satisfying (9), there is no way to reduce the number of the nodes of a system without altering its I-O map. In other words, inside this class a system is always minimal. This is in contrast to that of linear systems, in which it is well known that the order of an uncontrollable (unobservable) system can be reduced without changing its I-O map.

Remark 3.6: By the appropriate selection of Z it may be possible to satisfy some of the conditions in this section. Thus, the use of flexible neurons makes recurrent FBSNNs more powerful than recurrent CBSNNs.

IV. FBRBNNs: A New Class of FNNs

The proposed FBRBNN basically consists of a collection of parallel processing FBRBF units in an SHL, where the output of the *i*th node is given by the bivariate function,

$$f(x_i, z_i) = z_i \exp(-|x_i - c_i|^2 / z_i^2) \tag{11}$$

in which x_i is the input to the *i*th node, c_i is its center, and z_i is its size of influence which must be trained. The FBRBF is depicted in Figure 3.

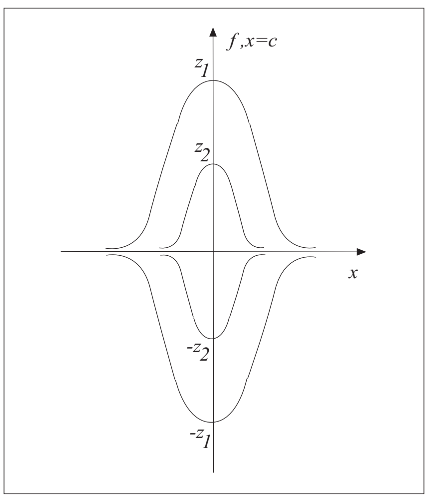

Fig.3. Shapes of the FBRBF

Similar to [1]-[4], FBRBNNs can easily be used in both supervised and unsupervised learning modes, and both connection weights and z_i's can be adjusted. FBRBNNs have superior learning capabilities over CRBNNs *since the size of influence of each unit is trained*. Simulation results and a detailed theoretical treatment are reported elsewhere. (Here we suffice to note that one can get either excitory or inhibitory units in FBRBNNs by training their size of influence only, leaving output/connection weights untrained. More precisely, for a given x_i and c_i, the FBRBF is onto with respect to z_i. This is in sharp contrast with CRBNNs.)

As in the case of FBSNNs/FUSNNs, a single-output network is considered in the following. The formulation is as that in Subsection II.A.

Theorem 4.1: Finite sums of the form (3) where f is given by (11) are dense in the space of continuous functions on the *m*-dimensional cube, $C(I_m)$. In other words, for any $h(x) \in C(I_m)$ and $\varepsilon > 0$ there exists a finite sum $F(x, Y, z, w)$ for which $|F(x, Y, z, w) - h(x)| < \varepsilon$, $\forall x \in I_m$.

Theorem 4.2: Finite sums of the form (3) where f is given by (11) are dense in the space of integrable functions on the *m*-dimensional cube, $L^1(I_m)$. In other words, for any $h(x) \in L^1(I_m)$ and $\varepsilon > 0$ there exists a finite sum $F(x, Y, z, w)$ for which $\int_{I_m} |F(x, Y, z, w) - h(x)|\,dx < \varepsilon$, $\forall x \in I_m$.

Proof: It is easily seen that (3) can be rewritten as,

$$F(x, Y, z, w) = w_0 + \sum_{j=1}^{n} w_j z_j f(v_j, z_j), \tag{12}$$

where f is the class of systems considered in [10, page 255]. Δ.

It should be noted that our proposed function differs from that of [10, page 255] in that the function considered therein has no coefficient, i.e., it is simply $\exp(-|x_i - c_i|^2 / z_i^2)$ where z_i can vary. (Note that this makes no difference in the proof.) Moreover, [10], giving reference to some old papers, does not consider this class as practically/theoretically important and is restricted to the case of a fixed influence size. This *wrong* idea has been the prevailing stereotype since the early introduction of NNs and unfortunately still to some extent continues to be.

V. Conclusions

The superiority of FNNs over CNNs has been shown and explained in a theoretical framework. Several fundamental properties of feedforward and recurrent FNNs have been established. It has been substantiated that: a) feedforward SHL FNNs are universal approximators, b) irreducible feedforward SHL FBSNNs are nonuniquely determined by their I-O maps, up to a finite group of symmetries, and c) recurrent FBSNNs are universal approximators. For a generic class of recurrent FBSNNs: d) necessary and sufficient conditions for the controllability and observability have been presented. For a subclass of this class: e) it has been proven that they are nonuniquely determined by their I-O maps, up to a finite group of symmetries, and e) every system in this subclass is minimal. It has also been shown that recurrent FNNs represent a generalization of the state-space framework. Then, a new class of FNNs, i.e., FBRBNNs has been proposed. It has been proven that feedforward SHL FBRBNNs are also universal approximators. It is hoped that this paper helps inspire further investment in a more powerful and versatile class of NNs, namely, FNNs.

VI. Future Work

Among the topics for future research on FNNs are the computational power and sample complexity for learning in both feedforward and recurrent FNNs, the effect of noise/disturbance on identifiability, and the complete characterization of the effect of Z (and in general, similarity transformation-like transformations, a special case being T in (10)) on the approximation error (and convergence rate), controllability and observability of recurrent FNNs. In this direction, the other system-theoretic properties like decoupling, sensitivity, failure tolerance, and stability [23]-[26] can also be investigated since recurrent FBSNNs are a generalization of the conventional state-space framework. It is clear that the results will contain those on recurrent CBSNNs – yet not fully established, see e.g. [16]-[19], [27]-[31] – as the special case that $Z = I$.

On the other hand, the properties of discrete-time recurrent FNNs should be studied. Although in general similar results for the properties considered in this paper are valid for them, in particular their controllability is still an (almost) open problem [16]-[19],[32].

The application domain of FNNs is quite wide. Of particular interest is the use of FNNs as controllers [1]-[4]. A typical field would be decentralized (adaptive) control of large-scale systems [24],[26],[33],[34] where CNNs, in particular CRBNNs, are already at issue, see e.g. [35],[36] and the references therein.

References

[1] M. Teshnehlab and K. Watanabe, "Self tuning of computed torque gains by using neural networks with flexible structures," *IEE Proceedings- Control Theory and Applications,* vol. 141, no. 4, pp. 235-242, 1994.

[2] M. Teshnehlab and K. Watanabe, "Flexible structural learning control of a robotic manipulator using artificial neural networks," *JSME Int. Jr.,* vol. 13, pp. 1-21, 1995.

[3] M. Teshnehlab and K. Watanabe, "Neural network controller with flexible structure based on feedback-error-learning," *Journal of Intelligent and Robotic Systems*, vol. 15, no. 4, pp. 367-387, 1996.

[4] M. Teshnehlab and K. Watanabe, *Intelligent control based on flexible neural networks*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.

[5] S. Haykin, *Neural networks: A comprehensive foundation, 2nd ed*. NJ: Prentice-Hall, 1998.

[6] Y. Bengio, *Neural networks for speech and sequence recognition*. Boston: Thompson Computer Press, 1996.

[7] R. Zbikowski and K. J. Hunt, Eds. *Neural adaptive control technology*. World Scientific Publishing, 1996.

[8] G. V. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303-314, 1989.

[9] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Neworks*, vol. 2, pp. 359-366, 1989.

[10] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, pp. 246-257, 1991.

[11] S. Lang, *Real and functional analysis, 3rd ed.* NY: Springer-Verlag, 1993.

[12] W. Rudin, *Functional analysis.* NY: McGraw-Hill, 1973.

[13] E. D. Sontag, "Remarks on interpolation and recognition using neural nets," *Advances in Neural Information Processing Systems*. vol. 3, CA: Morgan Kaufmann, pp. 939-945, 1991.

[14] H. J. Sussmann, "Uniqesness of weights for minimal feedforward nets with a given input-output map," *Neural Networks*, vol. 5, pp. 589-593, 1992.

[15] J. A. Leonard, M. A. Kramer, and L. H. Ungar, "Using radial basis functions to approximate a function and its error bounds," *IEEE Trans. Neural Networks*, vol. 3, no. 4, pp. 624-627, 1992.

[16] E. D. Sontag, "Recurrent neural networks: Some system-theoretic aspects," *Dealing with complexity: A neural network approach*, M. Karny, K. Warwick, and V. Kurkova, eds., London: Springer-Verlag, pp. 1-12, 1997.

[17] E. D. Sontag and H. J. Sussmann, "Complete controllability of continuous-time recurrent neural networks," *Systems & Control Letters*, vol. 30, pp. 177-183, 1997.

[18] F. Albertini and E. D. Sontag, "State observability in recurrent neural networks," *Systems & Control Letters*, vol. 22, pp. 235-244, 1994.

[19] F. Albertini and E. D. Sontag, "Uniqueness of weights for recurrent nets," *Mathematical Theory of Networks and Systems*, vol. 2, Regensburg : Akad Verlag, pp. 599-602, 1993.

[20] M. P. Joy and V. Tavsanoglu, "An equilibrium analysis of CNN's," *IEEE Trans. Circuits and Systems–I*, vol. 45, no.1, pp. 94-98, 1998.

[21] M. Forti, S. Manetti, M. Marini, "Necessary and sufficient condition for absolute stability of neural network," *IEEE Trans. Circuits and Systems–I*, vol. 41, no.7, pp. 491-494, 1994.

[22] W. M. Wonham, *Linear multivariable control: A geometric approach, 2nd ed.* NY: Springer-Verlag, 1985.

[23] Y. Bavafa-Toosi, "On multivariable linear output feedback," *MEng Thesis (Control)*, Department of Electrical Engineering, K.N. Toosi University of Technology, Tehran, Iran, Mar. 2000.

[24] Y. Bavafa-Toosi, H. Ohmori, and B. Labibi, "A generic approach to the design of decentralized linear output-feedback controllers," *Systems & Control Letters*, 2005 (in press).

[25] Y. Bavafa-Toosi and A. Khaki-Sedigh, "Stability is a fuzzy concept," *The 6th IEEE-IMACS Conf. Circuits, Systems, Communications and Computers*, Athens, Greece, July 2002.

[26] Y. Bavafa-Toosi, H. Ohmori, and B. Labibi, "Failure-tolerant performance stabilization of the generic large-scale system by decentralized linear output feedback," *ISA Trans.*, 2005 (in press).

[27] J. Peng, Z.-B. Xu, H. Qiao, and B. Zhang, "A critical analysis on global convergence of Hopfield-type neural networks," *IEEE Trans. Circuits and Systems–I*, vol. 52, no. 4, pp. 804-814, 2005.

[28] J. Cao and J. Wang, "Global asymptotic and robust stability of recurrent neural networks with time delays," *IEEE Trans. Circuits and Systems–I*, vol. 52, no. 2, pp. 417-426, 2005.

[29] X. Zeng and D. S. Yeung, "Sensitivity analysis of multilayer perceptron to input and weight perturbations," *IEEE Trans. Neural Networks*, vol. 12, no. 6, pp. 1358-1366, 2001.

[30] J.-Y. Choi and C.-H. Choi, "Sensitivity analysis of multilayer perceptron with differentiable activation functions," *IEEE Trans. Neural Networks*, vol. 3, no. 1, pp. 101-107, 1992.

[31] T. Petsche and B. W. Dickinson, "Trellis codes, receptive fields, and fault tolerant, self-repairing neural networks," *IEEE Trans. Neural Networks*, vol. 1, no. 2, pp. 154-166, 1990.

[32] F. Albertini and P. Dai Pra, "Forward accessability for recurrent neural networks," *IEEE Trans. Automtic Control*, vol. 40, no. 11, pp.1962-1968, 1995.

[33] H. Ito, H. Ohmori, and A. Sano, "Robust performance of decentralized control systems by expanding sequential designs," *Int. Jr. Control*, vol. 61, no. 6, pp. 1297-1311, 1995.

[34] H. Ito, H. Ohmori, and A. Sano, "A subsystem design approach to continuous-time performance of decentralized multirate sampled-data systems," *Int. Jr. Systems Science*, vol. 26, no. 6, pp. 1263-1287, 1995.

[35] J. T. Spooner and K. M. Passino, "Decentralized adaptive control of nonlinear systems using radial basis neural networks," *IEEE Trans. Automatic Control*, vol. 44, no. 11, pp. 2050-2057, 1999.

[36] N. Hovakimyan, F. Nardi, A. Calise, and N. Kim, "Adaptive output feedback control of uncertain systems using single hidden layer neural networks," *IEEE Trans. Neural Networks*, vol.13, No.6, pp.1420-1431, 2002.