JOURNAL OF APPROXIMATION THEORY 70, 131–141 (1992)

Approximation by Ridge Functions and Neural Networks with One Hidden Layer*

CHARLES K. CHUI AND XIN LI

Department of Mathematics, Texas A & M University, College Station, Texas 77843, U.S.A.

Communicated by E. W. Cheney

Received October 5, 1990; revised April 16, 1991

We describe the configuration of an infinite set V of vectors in R^s, $s \geqslant 1$, for which the closure with respect to $C(K)$ of the algebraic span of $\{f(\langle \mathbf{v}, \cdot \rangle): \mathbf{v} \in V, f \in C(R)\}$ is all of $C(K)$, where K is any compact set in R^s. This configuration also guarantees that for any sigmoidal function $\sigma \in C(R)$, the span of $\{\sigma(m\langle \mathbf{v}, \cdot \rangle + k): \mathbf{v} \in V; m, k \in Z\}$ is already dense in $C(K)$. In particular, neural networks with one hidden layer of the form

$$\sum_{(i,k) \in J} c(i, k)\, \sigma(\langle \mathbf{i}, \mathbf{x} \rangle + k),$$

where $k \in Z$, $c(i, k) \in R$, and $\mathbf{i} \in Z^s$, can be designed to approximate any continuous functions in s variables. © 1992 Academic Press, Inc.

1. INTRODUCTION

Mathematical models of neural networks were studied as early as 1943 by McCulloch and Pitts (cf. Funahashi [7]). In recent years, with rapid advances in computer technologies and information sciences, neural networks have found important applications in various fields of technology (cf. [1, 8, 10, 12, 16]). However, many mathematical problems on neural networks remain unsolved. One of the most challenging is the problem of designing a neural network with only one hidden layer, by using a single but arbitrary sigmoidal function $\sigma(t)$, such that any continuous function in s variables, where $s \geqslant 1$ is usually quite large, can be uniformly approximated within a preassigned tolerance. While there is still no constructive solution to this problem, the objective of this paper is to prove

* Research supported by NSF Grant DMS 89-0-1345 and ARO Contract DAAL 03-90-G-0091.

0021-9045/92 $5.00

that such a design is always possible for any sigmoidal function $\sigma(t)$, by using only integral scalings and translations in approximating continuous functions in any number of variables. Our approach differs from that in Cybenko [5] in that instead of using arguments in harmonic analysis, it establishes a density result on ridge functions. This should provide an intermediate step for coming up with a constructive proof.

Our investigation was motivated by the important work of Cybenko [5, 6], and we thank Professor Cybenko for providing us with these two preprints and for raising several important related questions. In fact, Cybenko [5] already established such a possibility provided that all scalings $a \in \mathbb{R}^s$ and translates $b \in \mathbb{R}$ are used. In Section 2, we show that Cybenko's proof of the one-variable result can be easily modified to show that integral scalings and translations are already sufficient. In establishing the general multivariable result, we first develop a general theory of approximation by ridge functions. This is done in Section 3. The proof of the multivariate result follows from applying the one-variable setting established in Section 2 and the density result of ridge functions established in Section 3. This will be given in Section 4.

Hence, our approach is different from the derivation given by Funahashi in his recent paper [7] and from Cybenko's work [5]. In [7], since an integral formula of Irie and Miyake [9] was used, the sigmodal function $\sigma(t)$ had to be assumed to be an increasing function.

2. Approximation by a Sigmoidal Function

A continuous function $\sigma(t)$ in one real variable is called a sigmoidal function if it satisfies

$$\begin{aligned} \sigma(t) &\to 1 \quad \text{as} \quad t \to +\infty; \\ \sigma(t) &\to 0 \quad \text{as} \quad t \to -\infty. \end{aligned} \tag{2.1}$$

Let s be any positive integer and denote, as usual, the set of integers by $\mathbb{Z}$ and the set of real numbers by $\mathbb{R}$. The following result will be established in this paper.

THEOREM 2.1. *Let $\sigma \in C(\mathbb{R})$ be any sigmoidal function and K any compact set in $\mathbb{R}^s$. Then the linear span*

$$S_s(\sigma) := \text{span}\{\sigma(\langle \cdot, \mathbf{m} \rangle + k) : \mathbf{m} \in \mathbb{Z}^s, k \in \mathbb{Z}\} \tag{2.2}$$

is dense in $C(K)$.

Here and throughout, $\langle \cdot, \cdot \rangle$ denotes, as usual, the inner product in $\mathbb{R}^s$.

For completeness, we include a proof of the univariate result by following Cybenko [5], where, by considering $\mathbf{m} \in \mathbb{R}^s$ and $k \in \mathbb{R}$ in (2.2), the density result is established by using the fact that the zero measure is the only Borel measure whose Fourier transform vanishes identically.

For $s = 1$, we may assume, without loss of generality, that $K = I := [0, 1]$. Suppose, on the contrary, that the theorem is false. Then there exists a nontrivial regular Borel measure μ on I which annihilates $S_1(\sigma)$; that is,

$$\int_I \sigma(mx + k)\, d\mu(x) = 0, \tag{2.3}$$

for all m and $k \in \mathbb{Z}$, but

$$\int_I f(x)\, d\mu(x) \neq 0 \tag{2.4}$$

for some $f \in C(I)$. Let $k, p, q \in \mathbb{Z}$, with $q > 0$, be arbitrarily chosen and fixed, and consider the function

$$r(x) = \begin{cases} 1 & \text{if} \quad x > \dfrac{p}{q} \\ 0 & \text{if} \quad x < \dfrac{p}{q} \\ \sigma(k) & \text{if} \quad x = \dfrac{p}{q}. \end{cases}$$

Then it is clear from (2.1) that

$$\sigma\left(nq\left(x - \frac{p}{q}\right) + k\right) \to r(x)$$

for each $x \in \mathbb{R}$ as $n \to +\infty$. Hence, by the Lebesgue Dominated Convergence Theorem and (2.3), we have

$$\mu\left(\left\{x : x > \frac{p}{q}\right\}\right) + \sigma(k)\, \mu\left(\left\{\frac{p}{q}\right\}\right) = \int_I r(x)\, d\mu(x) = 0,$$

which implies, by noting that $k \in \mathbb{Z}$ is arbitrary, that

$$\mu\left(\left\{x : x > \frac{p}{q}\right\}\right) = \mu\left(\left\{\frac{p}{q}\right\}\right) = 0. \tag{2.5}$$

Since p/q may be any arbitrary rational number in $[0, 1]$, the contradiction of (2.5) with (2.4) establishes Theorem 1.1 for the case $s = 1$.

To establish Theorem 2.1 for $s > 1$, we note that for any regular Borel measure μ, we have

$$\mu(\{\mathbf{x}: \langle \mathbf{x}, \mathbf{y} \rangle + \alpha > 0\}) = 0, \qquad \mathbf{y} \in \mathbb{R}^s, \alpha \in \mathbb{R}, \tag{2.6}$$

provided that

$$\mu(\{\mathbf{x}: \langle \mathbf{x}, \mathbf{m} \rangle + k > 0\}) = 0, \qquad \mathbf{m} \in \mathbb{Z}^s, k \in \mathbb{Z}.$$

Hence, by using an argument in Fourier transforms as in [5], we also establish Theorem 2.1 for any $s \geqslant 1$. However, this proof does not provide any information on the construction of approximants from $S_s(\sigma)$ in (2.2). For this reason, we include a study of approximation by ridge functions in Section 3; and in Section 4, we show that even a stronger result than Theorem 2.1 is a consequence of this approach.

3. Approximation by Ridge Functions

In this section, we establish a density result on ridge functions. Let $\mathbf{v} \in \mathbb{R}^s$ be nonzero. Then corresponding to any continuous function f of one variable, we have a "ridge function" $f(\langle \cdot, \mathbf{v} \rangle)$ which is in $C(\mathbb{R}^s)$. The reason for this terminology is that $f(\langle \cdot, \mathbf{v} \rangle)$ is a constant, which happens to be $f(0)$, on the $(s-1)$-dimensional space $\mathbf{v}^\perp$, the orthogonal complement of $\mathbf{v}$ in $\mathbb{R}^s$. For any nonempty set $V \subset \mathbb{R}^s \backslash \{\mathbf{0}\}$ and any compact set $K \subset \mathbb{R}^s$, let

$$M(V, K) = \operatorname{clos}_{C(K)} \operatorname{span}\{f(\langle \cdot, \mathbf{v} \rangle): \mathbf{v} \in V, f \in C(\mathbb{R})\}. \tag{3.1}$$

The objective of this section is to identify those sets V for which $M(V, K) = C(K)$. We first note that as pointed out by Sun [13] V cannot be too small as in the following. For more details, the reader is referred to the expanded version of this paper [3].

Proposition 3.1. *Let $s \geqslant 2$. Then a necessary condition for $M(V, I^s) = C(I^s)$ is that V contains an infinite set*

$$\tilde{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$$

with different directions, in the sense that $\lambda_i \mathbf{v}_i + \lambda_j \mathbf{v}_j = \mathbf{0}$ where $\lambda_i, \lambda_j \in \mathbb{R}$ and $i \neq j$, implies that $\lambda_i = \lambda_j = 0$.

In the following, we see, however, that the necessary condition in Proposition 3.1 is not sufficient. For example, let

$$V = \{(k, m, 0): k, m \in \mathbb{Z}\} \backslash \{\mathbf{0}\}$$

in $\mathbb{R}^3$. Then for any $f \in C(\mathbb{R})$, we have

$$f(\langle \mathbf{x}, \mathbf{v} \rangle) = f(kx_1 + mx_2),$$

where $\mathbf{v} = (k, m, 0) \in V$ and $\mathbf{x} = (x_1, x_2, x_3)$, so that

$$D^{(0,0,1)} f(\langle \cdot, \mathbf{v} \rangle) = 0.$$

Here and throughout, the standard notation

$$D^{\mathbf{m}} = \left(\frac{\partial}{\partial x_1}\right)^{m_1} \cdots \left(\frac{\partial}{\partial x_s}\right)^{m_s},$$

$\mathbf{m} = (m_1, ..., m_s) \in \mathbb{Z}_+^s$, is used, and $\mathbb{Z}_+ := \{k \in \mathbb{Z} : k \geqslant 0\}$.

To find sufficient conditions on V for which $M(V, K) = C(K)$, we need the following notion of "interpolation property." First, let π_k^s denote the space of polynomials in s variables of total degree $\leqslant k$ and $\pi^s = \bigcup_{k \in \mathbb{Z}_+} \pi_k^s$.

Definition. *A set $\mathscr{U} \subset \mathbb{R}^{s-1}$ is said to have the interpolation property relative to π^{s-1} if for any positive integer k, there exist $\mathbf{u}_1, ..., \mathbf{u}_N \in \mathscr{U}$, where*

$$N = N_k^{s-1} := \binom{k+s-1}{s-1},$$

such that $\{\mathbf{u}_1, ..., \mathbf{u}_N\}$ admits unique Lagrange interpolation from π_k^{s-1}; that is, for any $y_1, ..., y_N$ in $\mathbb{R}$, a unique polynomial $p_k \in \pi_k^{s-1}$ can be found to satisfy $p_k(\mathbf{u}_j) = y_j$, $j = 1, ..., N$.

To apply this notion to study sufficient conditions on V, we first introduce a map τ from $\mathbb{R}^s$ to $\mathbb{R}^{s-1}$ defined by

$$\tau(x_1, ..., x_s) = \left(\frac{x_2}{x_1}, ..., \frac{x_s}{x_1}\right), \qquad x_1 \neq 0.$$

Hence, by setting

$$V' := \{(x_1, ..., x_s) \in V : x_1 \neq 0\},$$

we have

$$\tau V' = \left\{\left(\frac{x_2}{x_1}, ..., \frac{x_s}{x_1}\right) : (x_1, ..., x_s) \in V, x_1 \neq 0\right\};$$

and $\tau V'$ may be considered as a "projection" of V' into $\mathbb{R}^{s-1}$. Of course, to every $\mathbf{u} \in \tau V'$, there is some $\mathbf{v} \in V'$ such that $\tau \mathbf{v} = \mathbf{u}$. The following is the main result of this section.

THEOREM 3.1. *Let $V \subset \mathbb{R}^s \backslash \{\mathbf{0}\}$ be given such that $\tau V'$ has the interpolation property relative to π^{s-1}. Then $M(V, K) = C(K)$ for all compact sets K in $\mathbb{R}^s$.*

Proof. Since any compact set K necessarily lies in some $[a, b]^s$ and $C(K)$ can be extended to $C([a, b]^s)$, we may assume, without loss of generality, that $K = I^s$. In addition, since π^s is dense in $C(I^s)$, it is sufficient to prove that $\pi^s \subset M(V, I^s)$. For this purpose, let k be any positive integer, and consider the subspace H_k^s of π_k^s consisting of all homogeneous polynomials of degree k; that is, $p \in H_k^s$ if and only if $p \in \pi_k^s$ and $p(\lambda \mathbf{x}) = \lambda^k p(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^s$, for all constants λ. It is clear that

$$\bigcup_{k \in \mathbb{Z}_+} H_k^s = \pi^s \tag{3.2}$$

and the dimension of each H_k^s is

$$\dim H_k^s = \dim \pi_k^{s-1} = N = N_k^{s-1} := \binom{k+s-1}{s-1}. \tag{3.3}$$

In fact, the collection of all monomials,

$$\mathbf{x}^m := x_1^{m_1} \cdots x_s^{m_s}, \qquad |\mathbf{m}| := m_1 + \cdots + m_s = k,$$

of total degree k is a basis of H_k^s. In the following, we show that H_k^s has another basis consisting only of ridge functions (relative to the direction set V), so that in view of (3.2), it follows that $\pi^s \subset M(V, I^s)$, and this completes the proof of the theorem.

To find this basis of ridge (polynomial) functions, we first choose a set of $N(=N_k^{s-1})$ distinct elements $\mathbf{u}_1, ..., \mathbf{u}_N$ from $\tau V'$ that admits unique Lagrange interpolation from π_k^{s-1}. That this is possible follows from the hypothesis of the theorem. Next, select $\mathbf{v}_1, ..., \mathbf{v}_N$ from V' such that $\tau \mathbf{v}_i = \mathbf{u}_i$, $i = 1, ..., N$. Now, the ridge (polynomial) functions we are looking for are given by

$$g_i(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v}_i \rangle^k = (b_{i,1} x_1 + \cdots + b_{i,s} x_s)^k, \tag{3.4}$$

$i = 1, ..., N$, where $\mathbf{x} = (x_1, ..., x_s)$ and $\mathbf{v}_i = (b_{i,1}, ..., b_{i,s})$. It is clear that each g_i is in H_k^s. Since the number of g_i's agrees with the dimension of H_k^s as given in (3.3), in order to show that these ridge functions form a basis of H_k^s, it is sufficient to show that if

$$F(\mathbf{x}) := \sum_{i=1}^{N} c_i g_i(\mathbf{x}) \equiv 0, \tag{3.5}$$

then $c_1 = \cdots = c_N = 0$. Let us assume that (3.5) holds and that, on the contrary, the row vector $[c_1, ..., c_N]$ is nontrivial. Then if $\mathbf{m}_1, ..., \mathbf{m}_N$ are the N distinct elements in $\mathbb{Z}^s$ with $|\mathbf{m}_1| = \cdots = |\mathbf{m}_N| = k$, the set of identities

$$(D^{\mathbf{m}_i}F)(\mathbf{x}) = 0, \qquad i = 1, ..., N,$$

is equivalent to the system of linear equations

$$\sum_{j=1}^{N} c_j \mathbf{v}_j^{m_i} = 0, \qquad i = 1, ..., N;$$

or in matrix form,

$$[c_1 \cdots c_N] A = 0,$$

where

$$A = \begin{bmatrix} \mathbf{v}_1^{\mathbf{m}_1} \cdots \mathbf{v}_1^{\mathbf{m}_N} \\ \cdots \cdots \cdots \\ \mathbf{v}_N^{\mathbf{m}_1} \cdots \mathbf{v}_N^{\mathbf{m}_N} \end{bmatrix}.$$

Since we have assumed that $[c_1 \cdots c_N]$ is nontrivial, the matrix A must be singular, and hence, there is a nontrivial column vector $\mathbf{a} = [a_1 \cdots a_N]^T$ that satisfies $A\mathbf{a} = \mathbf{0}$. That is, we have obtained a nontrivial function $G \in H_k^s$, defined by

$$G(\mathbf{x}) = \sum_{j=1}^{N} a_j \mathbf{x}^{\mathbf{m}_j},$$

which satisfies the interpolatory condition

$$G(\mathbf{v}_i) = 0, \qquad i = 1, ..., N. \tag{3.6}$$

To return from $\mathbb{R}^s$ to R^{s-1}, we set $\mathbf{y} = \tau\mathbf{x}$ for any $\mathbf{x} = (x_1, ..., x_s)$ with $x_1 \neq 0$, and consider the nontrivial $h \in \pi_k^{s-1}$ defined by

$$\begin{aligned} h(\mathbf{y}) = h(\tau\mathbf{x}) &:= x_1^{-k} G(\mathbf{x}) \\ &= \sum_{j=1}^{N} a_j (1, y_1, ..., y_{s-1})^{\mathbf{m}_j}, \end{aligned}$$

where $\mathbf{y} = (y_1, ..., y_{s-1}) \in \mathbb{R}^{s-1}$. Therefore, recalling that $\mathbf{u}_i = \tau\mathbf{v}_i$, we have, from (3.6), that

$$h(\mathbf{u}_i) = 0, \qquad i = 1, ..., N.$$

Hence, since $\{\mathbf{u}_1, ..., \mathbf{u}_N\}$ admits unique Lagrange interpolation from π_k^{s-1}, the polynomial h must be identically zero. This contradicts $\mathbf{a} \neq \mathbf{0}$, and completes the proof of the theorem. ∎

For $s \geqslant 2$, it is clear that for $V = \mathbb{Z}^s$, we have $\tau V' \supset \mathbb{Z}^{s-1}$, which clearly has the interpolation property relative to π^{s-1}. Of course there is nothing special about the set $\mathbb{Z}^{s-1}$. In fact, it is at least intuitively obvious that $\mathbb{Z}^{s-1}$ can be generalized to

$$A = A_1 \times \cdots \times A_{s-1}, \tag{3.7}$$

where $A_1, ..., A_{s-1}$ are infinite sets in $\mathbb{R}$. That is, we have the following.

PROPOSITION 3.2. *Let $A_1, ..., A_{s-1}$ be infinite sets in $\mathbb{R}$. Then the set A defined in* (3.7) *has the interpolation property relative to π^{s-1}.*

The proof of this result depends on the following lemma which is obvious.

LEMMA 3.1. *Let $\{\mathbf{x}_1, ..., \mathbf{x}_m\} \subset \mathbb{R}^{s-1}$, $s \geqslant 2$. Suppose that the only polynomial $p \in \pi_k^{s-1}$ that satisfies $p(\mathbf{x}_1) = \cdots = p(\mathbf{x}_m) = 0$ is the zero polynomial. Then there exists a subset $\{\mathbf{x}_{m_1}, ..., \mathbf{x}_{m_N}\}$, where $N = \binom{k+s-1}{s-1}$ as in* (3.3), *that admits unique Lagrange interpolation from π_k^{s-1}.*

We now prove Proposition 3.2 by using induction on s. Since A_1 is an infinite set, the induction procedure can be initiated. Let $k \in \mathbb{Z}_+$ be fixed. By the induction hypothesis, there exists a subset $B \subset A_1 \times \cdots \times A_{s-2}$ that admits unique Lagrange interpolation from π_k^{s-2}. Let $\{t_1, ..., t_{k+1}\}$ be any set of distinct points in A_{s-1} and define

$$W = \{(\mathbf{b}, t_i) : \mathbf{b} \in B,\ i = 1, ..., k+1\}$$

which is a subset of $A_1 \times \cdots \times A_{s-1}$. It is sufficient to show that W admits unique Lagrange interpolation from π_k^{s-1}. But this easily follows from Lemma 3.1. Indeed, if $p \in \pi_k^{s-1}$ satisfies

$$p(\mathbf{b}, t_i) = 0 \qquad \text{for all} \qquad (\mathbf{b}, t_i) \in W,$$

then writing

$$p(\mathbf{x}) = p_1(x_1, ..., x_{s-2})\, x_{s-1}^k + p_2(x_1, ..., x_{s-2})\, x_{s-1}^{k-1} + \cdots + p_{k+1}(x_1, ..., x_{s-2}),$$

where $\mathbf{x} = (x_1, ..., x_{s-1})$ and $p_1, ..., p_{k+1}$ are in π_k^{s-2}, we have

$$0 = p(\mathbf{b}, t_i) = p_1(\mathbf{b})\, t_i^k + p_2(\mathbf{b})\, t_i^{k-1} + \cdots + p_{k+1}(\mathbf{b})$$

for $i=1, ..., k+1$. Since $t_1, ..., t_{k+1}$ are distinct, it follows that

$$p_j(\mathbf{b})=0, \qquad j=1, ..., k+1.$$

This holds for all $\mathbf{b} \in B$, where B admits Lagrange interpolation from π_k^{s-2}. Hence, $p_1, ..., p_{k+1}$ are identically zero, or p is the zero polynomial.

Hence, the following density result of Sun and Cheney [15] is a consequence of Proposition 3.2 and Theorem 3.1.

COROLLARY 3.1. *Let $s \geqslant 2$ and $A_1, ..., A_s$ be subsets of $\mathbb{R}$, such that at most one of these sets is finite and the finite set, if any, contains a nonzero element. Then*

$$M(A_1 \times \cdots \times A_s, K)=C(K)$$

for any compact set K in $\mathbb{R}^s$.

In fact, Sun and Cheney [15] also observed that the condition in Corollary 3.1 is necessary. Other sets that admit unique Lagrange interpolation from π_k^s can be found in the literature such as [2, 4, 11]. We close this section with an obvious observation. Throughout this section, V' and $\tau V'$ were defined by considering those vectors in V with nonzero first entries. Of course, the same results are obtained by restricting our attention to the ith entry, for any i, $1 \leqslant i \leqslant s$.

4. NEURAL NETWORKS WITH ONE HIDDEN LAYER

We are now ready to complete the proof of Theorem 2.1 for any $s \geqslant 1$ by using Theorem 3.1 on ridge functions. In fact, we establish the following somewhat stronger result which, in view of Corollary 3.1, yields Theorem 2.1 immediately.

THEOREM 4.1. *Let $\sigma \in C(\mathbb{R})$ be any sigmoidal function and $s \geqslant 1$. Suppose that $V \subset \mathbb{R}^s$ is so chosen that $\tau V'$ has the interpolation property relative to π^{s-1}. Then the linear span*

$$\operatorname{span}\{\sigma(\langle \cdot, \mathbf{v}\rangle + k) : \mathbf{v} \in V, k \in \mathbb{Z}\} \tag{4.1}$$

is dense in $C(K)$ for any compact subset K of $\mathbb{R}^s$.

Proof. As in the proof of Theorem 3.1, we assume, without loss of generality, that $K = I^s$. Let $\varepsilon > 0$ be arbitrarily given and $f \in C(I^s)$. By Theorem 3.1, there exist functions $f_j \in C(\mathbb{R})$ such that

$$\left| f(\mathbf{x}) - \sum_{j=1}^{k} f_j(\langle \mathbf{x}, \mathbf{v}_j \rangle) \right| < \frac{\varepsilon}{2}, \qquad \mathbf{x} \in I^s, \tag{4.2}$$

for some $\mathbf{v}_1, ..., \mathbf{v}_k \in V$. Set

$$[a_j, b_j] = \{ \langle \mathbf{x}, \mathbf{v}_j \rangle : \mathbf{x} \in I^s \}. \tag{4.3}$$

By the one-variable result established in Section 2, for each j, there exist $r_{ji} \in \mathbb{R}$ and $m_{ji}, n_{ji} \in \mathbb{Z}$, $i = 1, ..., k_j$, such that

$$\left| f_j(x) - \sum_{i=1}^{k_j} r_{ji} \sigma(m_{ji} x + n_{ji}) \right| < \frac{\varepsilon}{2k}, \qquad x \in [a_j, b_j]. \tag{4.4}$$

Hence, in view of (4.3), we have, by applying (4.2) and (4.4),

$$\left| f(\mathbf{x}) - \sum_{j=1}^{k} \sum_{i=1}^{k_j} r_{ji} \sigma(m_{ji} \langle \mathbf{x}, \mathbf{v}_j \rangle + n_{ji}) \right| < \varepsilon.$$

This completes the proof of Theorem 4.1. ▮

It should be remarked that a constructive proof of Theorem 2.1 is still not available. This and the related complexity problems are very important mathematical problems in neural networks. The interested reader is referred to Cybenko [6] for an overview of this important collection of problems and of their applications to neural computing.

ACKNOWLEDGMENTS

The first author is grateful to George Cybenko for providing him with several preprints and pointing out some important problems in neural networks and neuro computing. After the completion of this paper, we received copies of [14, 15] from Ward Cheney and X. Sun, and thank them for providing us with these interesting writings.

Note added in proof. At the time when this galley is corrected, two constructive proofs of a weaker form of Theorem 2.1 (with non-integer m and k) are available. The interested reader is referred to (1) C. K. Chui and X. Li, Realization of neural networks with one hidden layer, and (2) H. Mhaskar and C. A. Micchelli, Approximation by superposition of a sigmoidal function.

References

1. D. J. Burr, Experiments on neural net recognition of spoken and written text, *IEEE Trans. Acoust. Speech Signal Process.* **36** (1988), 1162–1168.
2. C. K. Chui, "Multivariate Splines," CBMS–NSF Series in Applied Mathematics, No. 54. SIAM, Philadelphia, 1988.
3. C. K. Chui and X. Li, CAT Report #222, Texas A&M University, College Station, TX, July, 1990.
4. P. G. Ciarlet and P. A. Raviart, General Lagrange and Hermite interpolation in R^n with applications to finite element methods, *Arch. Rational Mech. Anal.* **46** (1972), 177–199.
5. G. Cybenko, Approximation by superpositions of a sigmoid function, *Math. Control Signals Systems* **2** (1989), 303–314.
6. G. Cybenko, Mathematical problems in neural computing, to appear.
7. K. I. Funahashi, On the approximate realization of continuous mappings by neural networks, *Neural Networks* **2** (1989), 183–192.
8. R. P. Gorman and T. J. Sejnowski, Learned classification of sonar targets using a massively parallel networks, *IEEE Trans. Acoust. Speech Signal Process* **36** (1989), 1135–1140.
9. B. Irie and S. Miyake, Capabilities of three-layered perceptrons, *in* "IEEE International Conference on Neural Networks," Vol. 1, pp. 641–648, 1988.
10. R. P. Lippmann, An introduction to computing with neural nets, *IEEE ASSP Mag.* **4** (1987), 4–22.
11. R. A. Nicolaides, On a class of finite elements generated by Lagrange interpolation, *SIAM J. Numer. Anal.* **9** (1972), 435–445.
12. T. J. Sejnowski and C. J. Rosenberg, Parallel networks that learn to pronounce English text, *Complex Systems* **1** (1987), 145–168.
13. X. Sun, weekly seminar, Center for Approximation Theory, College Station, TX, September, 1989.
14. X. Sun, "Multivariate Interpolation Using Ridge or Related Functions," Ph.D. Dissertation, Univ. of Texas at Austin, August, 1990.
15. X. Sun and E. W. Cheney, "The Fundamentality of Sets of Ridge Functions," CAT Report #224, Texas A&M University, College Station, TX, 1990.
16. A. Wieland and R. Leighton, Geometric analysis of neural networks capabilities, *in* "IEEE First International Conference on Neural Networks," Vol. 3, pp. 385–392, 1987.