JOURNAL OF APPROXIMATION THEORY 75, 295–311 (1993)

Fundamentality of Ridge Functions

VLADIMIR YA. LIN AND ALLAN PINKUS

Department of Mathematics, Technion, Israel Institute of Technology, Haifa 32000, Israel

Communicated by Carl de Boor

Received January 21, 1992; accepted in revised form July 18, 1992

For a given integer d, $1 \leqslant d \leqslant n-1$, let Ω be a subset of the set of all $d \times n$ real matrices. Define the subspace $\mathcal{M}(\Omega) = \text{span}\{g(A\mathbf{x}): A \in \Omega, g \in C(\mathbb{R}^d, \mathbb{R})\}$. We give necessary and sufficient conditions on Ω so that $\mathcal{M}(\Omega)$ is dense in $C(\mathbb{R}^n, \mathbb{R})$ in the topology of uniform convergence on compact subsets. This generalizes work of Vostrecov and Kreines. We also consider some related problems.

1. INTRODUCTION

Ridge functions on $\mathbb{R}^n$, in their simplest case, are functions F of the form

$$F(\mathbf{x}) = f(\mathbf{a} \cdot \mathbf{x}),$$

where $f: \mathbb{R} \to \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ is a fixed vector, $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$, and $\mathbf{a} \cdot \mathbf{x} = \sum_{i=1}^n a_i x_i$. Such functions, and both generalizations and linear combinations thereof, arise in various contexts. They arise in problems of tomography (see, e.g., [8, 12, 13] and references therein), projection pursuit in statistics (see, e.g., [6, 9]), neural networks (see [2, 3, 11] and references therein), partial differential equations [10] (where they are called "plane waves"), and approximation theory (see, e.g., [1, 2, 4, 5, 14]).

We consider, for given d, $1 \leqslant d \leqslant n-1$, functions G of the form

$$G(\mathbf{x}) = g(A\mathbf{x}),$$

where A is a fixed $d \times n$ real matrix, and $g: \mathbb{R}^d \to \mathbb{R}$. For $d = 1$, this reduces to ridge functions. In this paper we let Ω be a subset of all $d \times n$ real matrices. Set

$$\mathcal{M}(\Omega) = \text{span}\{g(A\mathbf{x}): A \in \Omega, g \in C(\mathbb{R}^d, \mathbb{R})\}.$$

(We run over all $A \in \Omega$ and *all* $g \in C(\mathbb{R}^d, \mathbb{R})$.) In Section 2 we determine necessary and sufficient conditions on Ω such that the subspace $\mathcal{M}(\Omega)$ is

0021-9045/93 $5.00

dense in $C(\mathbb{R}^n, \mathbb{R})$, i.e., $\overline{\mathcal{M}(\Omega)} = C(\mathbb{R}^n, \mathbb{R})$, in the topology of uniform convergence on compact subsets. In Section 3 we highlight various consequences of this result. In Section 4 we discuss the question of characterizing $\overline{\mathcal{M}(\Omega)}$ in general, and its relationship to kernels of differential operators. Finally, in Section 5 we ask a slightly different question. We fix a natural number k and ask whether it is possible to approximate functions in $C(\mathbb{R}^n, \mathbb{R})$ (in the above sense) by linear combinations of k functions of the form $g(A\mathbf{x})$, where we are free to choose the k "directions" A, as well as the functions g. We prove the answer is no.

We learned from Professor Brudnyi, only after we completed the work on this paper, of work of Vostrecov and Kreines [15, 16]. Two main results of our paper, namely Theorem 2.1 and part of Theorem 4.1, were proven for the case $d = 1$ in these two papers from the early 60's. These papers were unfortunately overlooked. We hope that they now receive the attention which is their due. (For example, in [8, 12] can be found a version of Theorem 2.1 in the case $d = 1$ and $n = 2$.) Other papers of Vostrecov (see especially [17]) are very much related to more recent work in [5].

A paper dealing with the above topics should note, and possibly use, the interrelationship between the results of this paper and Radon transform theory, polynomial ideals, exponential solutions, Zariski topology, Zariski closure, and other related matters. It was our desire to write this paper in as elementary a fashion as possible. We hope that the reader will make the appropriate connections.

2. Main Result and Proof

Let Ω and $\mathcal{M}(\Omega)$ be as defined above. For each $A \in \Omega$, we let $L(A)$ denote the span of the d rows of A. In what follows $A, B \in \Omega$ are considered distinct if $L(A) \neq L(B)$, because

$$\operatorname{span}\{g(A\mathbf{x}) : g \in C(\mathbb{R}^d, \mathbb{R})\} = \operatorname{span}\{g(B\mathbf{x}) : g \in C(\mathbb{R}^d, \mathbb{R})\}$$

if and only if $L(A) = L(B)$. Set

$$L(\Omega) = \bigcup_{A \in \Omega} L(A).$$

Let H_k^n denote the set of homogeneous polynomials of n variables of total degree k, i.e.,

$$H_k^n = \left\{ \sum_{|\mathbf{m}| = k} c_{\mathbf{m}} \mathbf{s}^{\mathbf{m}} \right\},$$

and H^n the set of all homogeneous polynomials of n variables, i.e.,

$$H^n = \bigcup_{k=0}^{\infty} H_k^n.$$

We use the standard notation $\mathbf{m} = (m_1, ..., m_n) \in \mathbb{Z}_+^n$, $|\mathbf{m}| = m_1 + \cdots + m_n$, and $\mathbf{s}^{\mathbf{m}} = s_1^{m_1} \cdots s_n^{m_n}$. Note that $\dim H_k^n = \binom{n-1+k}{k}$.

THEOREM 2.1. *The linear space $\mathcal{M}(\Omega)$ is dense in $C(\mathbb{R}^n, \mathbb{R})$ in the topology of uniform convergence on compact subsets if and only if the only polynomial in H^n which vanishes identically on $L(\Omega)$ is the zero polynomial.*

Proof. $(\Rightarrow)$. Assume that for some $k \in \mathbb{N}$ there exists a $p \in H_k^n \setminus \{0\}$ such that $p(\xi) = 0$ for all $\xi \in L(\Omega)$. Let

$$p(\xi) = \sum_{|\mathbf{m}|=k} b_{\mathbf{m}} \xi^{\mathbf{m}}.$$

Choose any $\phi \in C_0^\infty(\mathbb{R}^n)$, $\phi \neq 0$, i.e., ϕ is a nontrivial C^∞ function with compact support. For each $\mathbf{m} \in \mathbb{Z}_+^n$, $|\mathbf{m}| = k$, set

$$D^{\mathbf{m}} = \frac{\partial^k}{\partial x_1^{m_1} \cdots \partial x_n^{m_n}}.$$

We define

$$\psi(\mathbf{x}) = \sum_{|\mathbf{m}|=k} b_{\mathbf{m}} D^{\mathbf{m}} \phi(\mathbf{x}).$$

Note that $\psi \in C_0^\infty(\mathbb{R}^n)$, $\psi \neq 0$, $(\operatorname{supp} \psi \subseteq \operatorname{supp} \phi)$, and

$$\hat{\psi} = i^k \hat{\phi} p,$$

where $\hat{}$ denotes the Fourier transform.

We claim that

$$\int_{\mathbb{R}^n} g(A\mathbf{x})\, \psi(\mathbf{x})\, d\mathbf{x} = 0 \tag{2.1}$$

for all $A \in \Omega$ and $g \in C(\mathbb{R}^d, \mathbb{R})$; i.e., the nontrivial linear functional defined by integrating against ψ annihilates $\mathcal{M}(\Omega)$. This implies the desired result.

We prove (2.1) as follows. For given $A \in \Omega$, let $m = \dim L(A) \leqslant d$. Write $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$, where $(\mathbf{x}', \mathbf{0})$ and $(\mathbf{0}, \mathbf{x}'')$ are the orthogonal projections of $\mathbf{x}$ onto $L(A)$ and its orthogonal complement, respectively. Then for any $\rho \in C_0^\infty(\mathbb{R}^n)$

$$\int_{\mathbb{R}^n} \rho(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^{n-m}} \rho(\mathbf{x}', \mathbf{x}'')\, d\mathbf{x}'' \right] d\mathbf{x}'.$$

Every $\mathbf{c}' \in \mathbb{R}^m$ is such that $\mathbf{c} = (\mathbf{c}', \mathbf{0}) \in L(A)$. Thus

$$0 = i^k \hat{\phi}(\mathbf{c})\, p(\mathbf{c}) = \hat{\psi}(\mathbf{c}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \psi(\mathbf{x})\, e^{-i\mathbf{c}\cdot\mathbf{x}}\, d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^{n-m}} \psi(\mathbf{x}', \mathbf{x}'')\, d\mathbf{x}'' \right] e^{-i\mathbf{c}'\cdot\mathbf{x}'}\, d\mathbf{x}'.$$

Set

$$H(\mathbf{x}') = \int_{\mathbb{R}^{n-m}} \psi(\mathbf{x}', \mathbf{x}'')\, d\mathbf{x}''.$$

Then $H \in C_0^\infty(\mathbb{R}^m)$, and the previous equation can be rewritten as

$$0 = \frac{1}{(2\pi)^{m/2}} \int_{\mathbb{R}^k} H(\mathbf{x}')\, e^{i\mathbf{c}'\cdot\mathbf{x}'}\, d\mathbf{x}' = \hat{H}(\mathbf{c}')$$

for all $\mathbf{c}' \in \mathbb{R}^m$. Thus $H = 0$.

Set $\tilde{\mathbf{x}}' = (\mathbf{x}', 0)$. Since $(\mathbf{0}, \mathbf{x}'')$ is orthogonal to $L(A)$, it is clear that

$$A\mathbf{x} = A\tilde{\mathbf{x}}'$$

for all $\mathbf{x} = (\mathbf{x}', \mathbf{x}'') \in \mathbb{R}^n$. Thus for any $g \in C(\mathbb{R}^d, \mathbb{R})$,

$$\int_{\mathbb{R}^n} g(A\mathbf{x})\, \psi(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^{n-m}} \psi(\mathbf{x}', \mathbf{x}'')\, d\mathbf{x}'' \right] g(A\tilde{\mathbf{x}}')\, d\mathbf{x}'$$

$$= \int_{\mathbb{R}^m} H(\mathbf{x}')\, g(A\tilde{\mathbf{x}}')\, d\mathbf{x}' = 0.$$

($\Leftarrow$). Assume that for a given $k \in \mathbb{N}$ no non-trivial $p \in H_k^n$ vanishes identically on $L(\Omega)$. We will prove that $H_k^n \subseteq \mathcal{M}(\Omega)$. If the above holds for all $k \in \mathbb{Z}_+$, it then follows that $\mathcal{M}(\Omega)$ contains all polynomials, and from the Weierstrass theorem, $\overline{\mathcal{M}(\Omega)} = C(\mathbb{R}^n, \mathbb{R})$.

Let $\mathbf{d} \in L(\Omega)$. Then there exists a $\mathbf{y} \in \mathbb{R}^d$ and an $A \in \Omega$ such that $\mathbf{d} = \mathbf{y}A$. Set $g(A\mathbf{x}) = (\mathbf{y} \cdot A\mathbf{x})^k = (\mathbf{d} \cdot \mathbf{x})^k$. Thus $(\mathbf{d} \cdot \mathbf{x})^k \in \mathcal{M}(\Omega)$ for all $\mathbf{d} \in L(\Omega)$.

Since $D^{\mathbf{m}_1}\mathbf{x}^{\mathbf{m}_2} = \delta_{\mathbf{m}_1, \mathbf{m}_2} k!$, for $\mathbf{m}_1, \mathbf{m}_2 \in \mathbb{Z}_+^n$, $|\mathbf{m}_1| = |\mathbf{m}_2| = k$, it easily follows that every linear functional l on the finite dimensional linear space H_k^n may be represented by some $q \in H_k^n$ via

$$l(p) = q(D)p$$

for each $p \in H_k^n$.

For any given $q \in H_k^n$,

$$q(D)(\mathbf{d} \cdot \mathbf{x})^k = k!\, q(\mathbf{d}).$$

If the linear functional l annihilates $(\mathbf{d}\cdot\mathbf{x})^k$ for all $\mathbf{d}\in L(\Omega)$, then its representor $q\in H_k^n$ vanishes on $L(\Omega)$. By assumption this implies that $q=0$. Thus $H_k^n=\operatorname{span}\{(\mathbf{d}\cdot\mathbf{x})^k : \mathbf{d}\in L(\Omega)\}\subseteq \mathcal{M}(\Omega)$. ∎

Remark 2.1. The proof of this theorem in the case $d=1$ in Vostrecov and Kreines [15] is much the same, although we like to think that our proof is somewhat more elegant.

Remark 2.2. In Section 4 we provide a different proof of sufficiency. In fact that proof will be more general. The above proof is given because it is elementary and highlights an important fact. Namely, $\mathcal{M}(\Omega)$ is dense in $C(\mathbb{R}^n, \mathbb{R})$ in the topology of uniform convergence on compact subsets if and only if $\mathcal{M}(\Omega)$ explicitly contains the polynomials.

Remark 2.3. Our choice of the topology of uniform convergence on compact subsets is rather arbitrary. Theorem 2.1 will hold for many other linear spaces defined on $\mathbb{R}^n$ or subsets thereof. It is sufficient that the polynomials are dense therein (the Weierstrass theorem holds) and we can choose ψ, as in the first part of the proof of Theorem 2.1, in such a way that it defines a continuous functional on the space which annihilates $\mathcal{M}(\Omega)$.

3. Consequences and Remarks

In this section we note some simple consequences of Theorem 2.1. But first we remark that the property that no nontrivial polynomial in H^n identically vanishes on $L(\Omega)$ is equivalent to

$$\dim H_k^n|_{L(\Omega)}=\dim H_k^n,$$

for each $k\in\mathbb{N}$. For notational ease, we sometimes use this latter form.

Proposition 3.1. *If* $\dim H_k^n|_{L(\Omega)}=\dim H_k^n$ *for some* $k\in\mathbb{N}$, *then* $\dim H_l^n|_{L(\Omega)}=\dim H_l^n$ *for all* $l<k$.

Proof. If $\dim H_l^n|_{L(\Omega)}<\dim H_l^n$, then there exists a $p\in H_l^n\backslash\{0\}$ vanishing on $L(\Omega)$. Let $q\in H_{k-l}^n\backslash\{0\}$. Then $pq\in H_k^n\backslash\{0\}$ and vanishes on $L(\Omega)$. A contradiction. ∎

Proposition 3.2. *If* $\Omega=\Omega_1\cup\Omega_2$ *then* $\overline{\mathcal{M}(\Omega)}=C(\mathbb{R}^n,\mathbb{R})$ *if and only if* $\overline{\mathcal{M}(\Omega_j)}=C(\mathbb{R}^n,\mathbb{R})$ *for* $j=1$ *or* $j=2$.

Proof. The proof in one direction is simple. To prove the other direction, assume $\overline{\mathcal{M}(\Omega_j)}\neq C(\mathbb{R}^n,\mathbb{R})$ for $j=1$ and $j=2$. As such there exist $p_j \in H^n \backslash \{0\}$ such that p_j vanishes on $L(\Omega_j)$, $j=1,2$. Then $p = p_1 p_2 \in H^n \backslash \{0\}$ and vanishes on $L(\Omega)$. Thus $\overline{\mathcal{M}(\Omega)} \neq C(\mathbb{R}^n, \mathbb{R})$. ▮

As a consequence of Proposition 3.2, we have:

COROLLARY 3.3. *If Ω contains only a finite number of distinct elements, then $\overline{\mathcal{M}(\Omega)} \neq C(\mathbb{R}^n, \mathbb{R})$.*

Proof. Given a $d \times n$ matrix A, $d \leqslant n-1$, there exists a vector $\mathbf{c} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ such that $A\mathbf{c} = \mathbf{0}$. Set $p(\xi) = \mathbf{c} \cdot \xi$. p vanishes on $L(A)$. Apply Proposition 3.2. ▮

The case $d=1$ $(n=2)$ of this next result was proved in [15] (see also [12, 8]) by a much different method.

PROPOSITION 3.4. *If $d = n-1$, and Ω contains an infinite number of distinct A of rank $n-1$, then $\mathcal{M}(\Omega) = C(\mathbb{R}^n, \mathbb{R})$.*

Proof. Recall that $A, B \in \Omega$ are distinct if $L(A) \neq L(B)$. For each $A \in \Omega$ of rank $n-1$, there exists a unique (up to multiplication by a constant) $\mathbf{c}_A \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ (the normal to $L(A)$) such that $A\mathbf{c}_A = \mathbf{0}$. Let

$$p_A(\xi) = \mathbf{c}_A \cdot \xi.$$

p_A vanishes exactly on $L(A)$ and is irreducible.

Assume $\overline{\mathcal{M}(\Omega)} \neq C(\mathbb{R}^n, \mathbb{R})$. Thus for some $k \in \mathbb{N}$, there exists a $p \in H_k^n \backslash \{0\}$ which vanishes on $L(\Omega)$. For each $A \in \Omega$, p vanishes on $L(A)$. Thus, by the Hilbert Nullstellensatz, the polynomial p_A must be a divisor of p. This is true for an infinite number of irreducible distinct polynomials, which is a contradiction. ▮

Let Π_k^n denote the space of algebraic polynomials of n variables and total degree k, i.e., $\Pi_k^n = \bigoplus_{l=0}^k H_l^n$. Note that $\dim \Pi_k^n = \dim H_k^{n+1}$. We also let Π^n denote the set of all algebraic polynomials of n variables. For each $\mathbf{a} \in \mathbb{R}^n$ with $a_n \neq 0$, set

$$\tau(\mathbf{a}) = \left(\frac{a_1}{a_n}, ..., \frac{a_{n-1}}{a_n} \right).$$

$(\tau(\lambda\mathbf{a}) = \tau(\mathbf{a})$ for all $\lambda \in \mathbb{R} \backslash \{0\}$.) If $a_n = 0$, $\tau(\mathbf{a})$ is not defined.

PROPOSITION 3.5. *The following are equivalent:*

(a) *The only polynomial in H^n which vanishes identically on $L(\Omega)$ is the zero polynomial.*

(b) *The only polynomial in Π^n which vanishes identically on $L(\Omega)$ is the zero polynomial.*

(c) *The only polynomial in* Π^{n-1} *which vanishes identically on* $\tau(L(\Omega))$ *is the zero polynomial.*

Proof. We first prove the equivalence of (a) and (b). Obviously (b) implies (a). To prove that (a) implies (b), we first note that $L(\Omega)$ is a "balanced cone"; i.e., if $\mathbf{a} \in L(\Omega)$ then $\lambda\mathbf{a} \in L(\Omega)$ for all $\lambda \in \mathbb{R}$. Thus if $p \in \Pi^n \setminus \{0\}$ vanishes on $L(\Omega)$, then for each $\mathbf{a} \in L(\Omega)$ and every $\lambda \in \mathbb{R}$

$$0 = p(\lambda\mathbf{a}) = \sum_{j=0}^{k} q_j(\lambda\mathbf{a}) = \sum_{j=0}^{k} \lambda^j q_j(\mathbf{a}),$$

where $q_j \in H_j^n$ in the expansion of p. But this implies that

$$q_j(\mathbf{a}) = 0, \qquad j = 0, 1, ..., k.$$

Thus there exists a nontrivial homogeneous polynomial vanishing on $L(\Omega)$.

We now prove the equivalence of (a) and (c). Assume (c) does not hold. Then, for some $k \in \mathbb{N}$, there exists a $q \in \Pi_k^{n-1} \setminus \{0\}$ which vanishes on $\tau(L(\Omega))$. Let

$$q(\mathbf{x}) = \sum_{|\mathbf{m}| \leq k} d_{\mathbf{m}} \mathbf{x}^{\mathbf{m}},$$

where $\mathbf{x} = (x_1, ..., x_{n-1})$. Set

$$p(\mathbf{x}) = \sum_{|\mathbf{m}| \leq k} d_{\mathbf{m}} \mathbf{x}^{\mathbf{m}} x_n^{k+1-|\mathbf{m}|}.$$

Then $p \in H_{k+1}^n \setminus \{0\}$, and as is easily checked, p vanishes on $L(\Omega)$. Thus (a) does not hold.

Now assume that (a) does not hold. For some $k \in \mathbb{N}$ there exists a $p \in H_k^n \setminus \{0\}$ which vanishes on $L(\Omega)$. Let

$$p(\mathbf{x}) = \sum_{|\mathbf{m}| = k} c_{\mathbf{m}} \mathbf{x}^{\mathbf{m}},$$

$\mathbf{x} = (x_1, ..., x_n)$. Set

$$q(\mathbf{x}) = \sum_{|\mathbf{m}| = k} c_{\mathbf{m}} x_1^{m_1} \cdots x_{n-1}^{m_{n-1}}.$$

Then $q \in \Pi_k^{n-1} \setminus \{0\}$ and q vanishes on $\tau(L(\Omega))$. Thus (c) does not hold. ∎

Let $U_1, ..., U_n \subseteq \mathbb{R}$. By

$$\mathcal{U} = U_1 \times \cdots \times U_n$$

we mean the set of *all* vectors $\mathbf{a}=(a_1, ..., a_n) \in \mathbb{R}^n$ with $a_i \in U_i$, $i=1, ..., n$. Given d, $1 \leqslant d \leqslant n-1$, let $\Omega(\mathscr{U})$ denote the subset of the set of $d \times n$ matrices, the rows of which are all possible vectors in $\mathscr{U}$.

PROPOSITION 3.6. $\overline{\mathscr{M}(\Omega(\mathscr{U}))} = C(\mathbb{R}^n, \mathbb{R})$ *if and only if*

(a) *at least $n-d$ of the $U_1, ..., U_n$ have an infinite number of distinct elements;*

(b) *at most one of the $U_1, ..., U_n$ has only one element, and none has only the zero element.*

Proof. ($\Rightarrow$) (a) Assume $U_1, ..., U_{d+1}$ each have a finite number of elements. For each set of distinct d vectors $\tilde{A} = \{\tilde{\mathbf{a}}_i\}_{i=1}^d$ in $U_1 \times \cdots \times U_{d+1} \subseteq \mathbb{R}^{d+1}$, let $\tilde{\mathbf{c}}_{\tilde{A}} \in \mathbb{R}^{d+1}$ satisfy $\tilde{\mathbf{c}}_{\tilde{A}} \cdot \tilde{\mathbf{a}}_i = 0$, $i=1, ..., d$. Let $\mathbf{c}_{\tilde{A}} = (\tilde{\mathbf{c}}_{\tilde{A}}, 0, ..., 0) \in \mathbb{R}^n$. Take the product of all linear polynomials $\mathbf{c}_{\tilde{A}} \cdot \mathbf{x}$, for different A. (There are a finite number of such polynomials.) This product is a homogeneous polynomial which vanishes on $L(\Omega(\mathscr{U}))$.

(b) If $U_1 = \{0\}$, let $p(\mathbf{x}) = x_1$. If $U_1 = \{u_1\}$ and $U_2 = \{u_2\}$, where $u_1, u_2 \neq 0$, let $p(\mathbf{x}) = (u_2 x_1 - u_1 x_2)$. These homogeneous polynomials vanish on their respective $L(\Omega(\mathscr{U}))$.

($\Leftarrow$) Let us assume that both (a) and (b) hold. Our argument is via induction on n (with d fixed). As such we first prove the result for $d=n-1$.

Assume $d=n-1$, U_n has an infinite number of elements, and (b) holds. Thus assume U_i has at least two distinct elements a_i, b_i, $i=1, ..., n-2$, and $b_{n-1} \in U_{n-1}$ with $b_{n-1} \neq 0$. Let B denote the $n-1 \times n-1$ matrix

$$B = \begin{pmatrix} a_1 & b_2 & \cdots & b_{n-2} & b_{n-1} \\ b_1 & a_2 & \cdots & b_{n-2} & b_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_1 & b_2 & \cdots & a_{n-2} & b_{n-1} \\ b_1 & b_2 & \cdots & b_{n-2} & b_{n-1} \end{pmatrix}.$$

This matrix has rank $n-1$. Let $u_n^1, ..., u_n^{n-1}$ be distinct elements of U_n. For any $x \in U_n$, consider the $n-1 \times n$ matrix

$$U_x = \begin{pmatrix} & u_n^1 \\ B & \vdots \\ & u_n^{n-1} \\ & x \end{pmatrix}.$$

Each U_x is in Ω, and it is easily checked that for distinct $x \in U_n$, the associated U_x are distinct in the sense of Proposition 3.4. Thus, by the result therein, $\overline{\mathscr{M}(\Omega(\mathscr{U}))} = C(\mathbb{R}^n, \mathbb{R})$.

We now continue the induction argument. Assume that $1 \leqslant d < n-1$ and U_n contains an infinite number of distinct elements. Let $p \in H_k^n$ vanish on $L(\Omega(\mathscr{U}))$. Then

$$p(\mathbf{x}) = q_0(\tilde{\mathbf{x}})\, x_n^k + q_1(\tilde{\mathbf{x}})\, x_n^{k-1} + \cdots + q_k(\tilde{\mathbf{x}}),$$

where $\tilde{\mathbf{x}} = (x_1, ..., x_{n-1})$, and $q_j(\tilde{\mathbf{x}})$ is a homogeneous polynomial of degree j in the $n-1$ variables $x_1, ..., x_{n-1}$.

For $\tilde{\mathbf{u}} \in \tilde{\mathscr{U}} = U_1 \times \cdots \times U_{n-1}$,

$$p(\tilde{\mathbf{u}}, x_n) = q_0(\tilde{\mathbf{u}})\, x_n^k + q_1(\tilde{\mathbf{u}})\, x_n^{k-1} + \cdots + q_k(\tilde{\mathbf{u}})$$

vanishes at each element of U_n. Since there are an infinite number of such elements, we have

$$p(\tilde{\mathbf{u}}, x_n) \equiv 0$$

for each $\tilde{\mathbf{u}} \in \tilde{\mathscr{U}}$. Thus

$$q_0(\tilde{\mathbf{u}}) = \cdots = q_k(\tilde{\mathbf{u}}) = 0$$

for each $\tilde{\mathbf{u}} \in \tilde{\mathscr{U}}$. We apply the induction hypothesis to obtain

$$q_0 \equiv \cdots \equiv q_k \equiv 0.$$

Thus $p = 0$. ∎

For $d = 1$, this result was proved in [14].

The condition that no nontrivial polynomial in H^n (or Π^n) vanishes identically on $L(\Omega)$ is not one which is easily checked, unless $d = n-1$. The fact is that no simple condition seems possible because of the complicated nature of the zero set of multivariate polynomials.

4. The Closure of $\mathscr{M}(\Omega)$

Assume $\overline{\mathscr{M}(\Omega)} \neq C(\mathbb{R}^n, \mathbb{R})$; i.e., $\mathscr{M}(\Omega)$ does not satisfy the conditions of Theorem 2.1. Can we then identify in some way the closure of $\mathscr{M}(\Omega)$? Before stating the result of this section, we need some additional notation.

For Ω as previously defined, let $\mathscr{P}_\Omega$ denote the set of those polynomials which vanish on $L(\Omega)$, i.e.,

$$\mathscr{P}_\Omega = \{p : p \in \Pi^n, p|_{L(\Omega)} = 0\}.$$

$\mathscr{P}_\Omega$ is a polynomial ideal. Set

$$N = \ker \mathscr{P}_\Omega = \bigcap_{p \in \mathscr{P}_\Omega} \ker p = \{\mathbf{s} : p(\mathbf{s}) = 0, \text{ all } p \in \mathscr{P}_\Omega\}.$$

Note that $L(\Omega) \subseteq N$, and in general N may be much the larger set.

It is somewhat of a problem to simultaneously work with both Ω, which is a subset of $d \times n$ matrices, and N, a subset of $\mathbb{R}^n$ ($1 \times n$ matrices). As such, let us note that in Theorem 2.1 the case where $d > 1$ is essentially equivalent to the case $d = 1$ in the following sense: Given Ω, let

$$\mathscr{M}(L(\Omega)) = \operatorname{span}\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in L(\Omega), f \in C(\mathbb{R}, \mathbb{R})\}.$$

Then $\overline{\mathscr{M}(\Omega)} = \overline{\mathscr{M}(L(\Omega))}$. To see this simply note that as an application of Theorem 2.1, for each $d \times n$ matrix A

$$\overline{\operatorname{span}}\{g(A\mathbf{x}) : g \in C(\mathbb{R}^d, \mathbb{R})\} = \overline{\operatorname{span}}\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in L(A), f \in C(\mathbb{R}, \mathbb{R})\}.$$

THEOREM 4.1. *In the topology of uniform convergence on compact subsets, the following three sets are equal*:

$$\begin{aligned} &(1) \quad \mathscr{A} = \overline{\operatorname{span}}\{g(A\mathbf{x}) : A \in \Omega, g \in C(\mathbb{R}^d, \mathbb{R})\}, \\ &(2) \quad \mathscr{B} = \overline{\operatorname{span}}\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in N, f \in C(\mathbb{R}, \mathbb{R})\}, \\ &(3) \quad \mathscr{C} = \overline{\operatorname{span}}\{q(\mathbf{x}) : q \in \Pi^n, p(D)q = 0 \text{ for all } p \in \mathscr{P}_\Omega\}. \end{aligned}$$

Proof. By definition, and from the remark previous to the statement of Theorem 4.1,

$$\mathscr{A} = \overline{\mathscr{M}(\Omega)} = \overline{\mathscr{M}(L(\Omega))} = \overline{\operatorname{span}}\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in L(\Omega), f \in C(\mathbb{R}, \mathbb{R})\}.$$

Since $L(\Omega) \subseteq N$, it follows that $\overline{\mathscr{M}(L(\Omega))} \subseteq \mathscr{B}$. Thus $\mathscr{A} \subseteq \mathscr{B}$.

The set $L(\Omega)$ is a "balanced cone" (see Proposition 3.5). Let $p \in \mathscr{P}_\Omega$, and write p in the form

$$p = \sum_{k=0}^{r} p_k,$$

where $r \in \mathbb{N}$, and each p_k is a homogeneous polynomial of total degree k. Since

$$p(\lambda \mathbf{a}) = \sum_{k=0}^{r} \lambda^k p_k(\mathbf{a})$$

for each $\mathbf{a} \in \mathbb{R}^n$, it follows from the balanced cone property of $L(\Omega)$ that $p_k|_{L(\Omega)} = 0$ for each $k = 0, 1, ..., r$. Thus $p_k \in \mathscr{P}_\Omega$ for each k, and from the definition of N we see that N is also a balanced cone.

We now prove that $\bar{\mathscr{B}} \subseteq \mathscr{C}$. For each $\mathbf{a} \in \mathbb{R}^n$, $l \in \mathbb{Z}_+$, and homogeneous polynomial p_k of total degree k

$$p_k(D)(\mathbf{a} \cdot \mathbf{x})^l = \begin{cases} 0, & k > l \\ \dfrac{l!}{(l-k)!} p_k(\mathbf{a})(\mathbf{a} \cdot \mathbf{x})^{l-k}, & k \leqslant l. \end{cases}$$

Thus for each $\mathbf{a} \in N$, and every $l \in \mathbb{Z}_+$, $p(D)(\mathbf{a} \cdot \mathbf{x})^l = 0$ for all $p \in \mathscr{P}_\Omega$ since each $p \in \mathscr{P}_\Omega$ vanishes on N. Therefore $(\mathbf{a} \cdot \mathbf{x})^l \in \mathscr{C}$ for each $\mathbf{a} \in N$ and $l \in \mathbb{Z}_+$. The Weierstrass theorem implies that $\bar{\mathscr{B}} \subseteq \mathscr{C}$.

It remains to prove that $\mathscr{C} \subseteq \bar{\mathscr{A}}$. Since $\bar{\mathscr{A}}$ is a closed linear subspace of $\mathscr{C}$, it suffices to prove that each continuous linear functional on $C(\mathbb{R}^n, \mathbb{R})$ which annihilates $\bar{\mathscr{A}}$ also annihilates $\mathscr{C}$.

Every continuous linear functional m on $C(\mathbb{R}^n, \mathbb{R})$ (in the topology of uniform convergence on compact subsets) may be represented in the form

$$m(h) = \int_{\mathbb{R}^n} h(\mathbf{x})\, d\mu(\mathbf{x}),$$

where μ is a Borel measure of finite total variation and compact support; see, e.g., [7, p. 203]. Set

$$\hat{\mu}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\xi \cdot \mathbf{x}}\, d\mu(\mathbf{x});$$

i.e., $\hat{\mu}$ is the Fourier transform of μ. As is well known (see, e.g., [7, p. 389]), $\hat{\mu}$ is an entire analytic function on $\mathbb{C}^n$. Furthermore, assuming that μ annihilates $\bar{\mathscr{A}}$ (i.e., $\int_{\mathbb{R}^n} g(A\mathbf{x})\, d\mu(\mathbf{x}) = 0$ for all $A \in \Omega$, $g \in C(\mathbb{R}^d, \mathbb{R})$), we have that $\hat{\mu}$ vanishes on $L(\Omega)$. Set

$$\hat{\mu}(\xi) = \sum_{k=0}^{\infty} u_k(\xi),$$

where u_k is the homogeneous polynomial of total degree k in the power series expansion of $\hat{\mu}$. Since $L(\Omega)$ is a balanced cone it follows, as previously shown, that each u_k vanishes on $L(\Omega)$. That is, $u_k \in \mathscr{P}_\Omega$ for each $k \in \mathbb{Z}_+$.

Now

$$u_k(\xi) = \sum_{|\mathbf{m}| = k} a_{\mathbf{m}} \xi^{\mathbf{m}},$$

and

$$m_1! \cdots m_n!\, a_{\mathbf{m}} = D^{\mathbf{m}} \hat{\mu}(\xi)|_{\xi = \mathbf{0}} = \frac{(-i)^k}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathbf{x}^{\mathbf{m}}\, d\mu(\mathbf{x}).$$

Therefore for any homogeneous polynomial q_k of total degree k of the form

$$q_k(\mathbf{x}) = \sum_{|\mathbf{m}|=k} c_{\mathbf{m}} \mathbf{x}^{\mathbf{m}},$$

we have

$$\begin{aligned}\int_{\mathbb{R}^n} q_k(\mathbf{x})\, d\mu(\mathbf{x}) &= (2\pi)^{n/2} (-i)^{-k} \sum_{|\mathbf{m}|=k} m_1! \cdots m_n!\, a_{\mathbf{m}} c_{\mathbf{m}} \\ &= (2\pi)^{n/2} (-i)^{-k} u_k(D)\, q_k(\mathbf{x}).\end{aligned}$$

Furthermore, as is easily checked,

$$u_k(D)\, q_l(\mathbf{x})|_{\mathbf{x}=\mathbf{0}} = 0$$

if $l \neq k$. Thus if $q \in \Pi^n$, and

$$q = \sum_{k=0}^{N} q_k,$$

where each q_k is a homogeneous polynomial of total degree k, then

$$\int_{\mathbb{R}^n} q(\mathbf{x})\, d\mu(\mathbf{x}) = (2\pi)^{n/2} \sum_{k=0}^{N} (-i)^{-k} u_k(D)\, q(\mathbf{x})|_{\mathbf{x}=\mathbf{0}}.$$

This formula together with the fact that $u_k \in \mathscr{P}_\Omega$ for all k implies that for each $q \in \Pi^n$ satisfying $p(D)q = 0$ for all $p \in \mathscr{P}_\Omega$, we have

$$\int_{\mathbb{R}^n} q(\mathbf{x})\, d\mu(\mathbf{x}) = 0.$$

This proves that $\bar{\mathscr{C}} \subseteq \bar{\mathscr{A}}$. ∎

Remark 4.1. If no nontrivial polynomial vanishes on $L(\Omega)$, then Theorem 4.1 implies that

$$\begin{aligned}\overline{\mathscr{M}(\Omega)} &= \overline{\operatorname{span}}\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in \mathbb{R}^n, f \in C(\mathbb{R}, \mathbb{R})\} \\ &= \overline{\operatorname{span}}\{q : q \in \Pi^n\} = C(\mathbb{R}^n, \mathbb{R}).\end{aligned}$$

This is the different proof of sufficiency in Theorem 2.1 alluded to in Remark 2.2.

Remark 4.2. In [16], Vostrecov and Kreines prove the equality $\bar{\mathscr{A}} = \bar{\mathscr{B}}$ in the case $d = 1$. That is, an arbitrary function $f(\mathbf{b} \cdot \mathbf{x})$ ($f \in C(\mathbb{R}, \mathbb{R})$) can be uniformly approximated on compact subsets by functions from

$$\operatorname{span}\{g(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in L(\Omega), g \in C(\mathbb{R}, \mathbb{R})\}$$

(Ω a subset of $\mathbb{R}^n$) if and only if all (homogeneous) polynomials which vanish on $L(\Omega)$ also vanish at $\mathbf{b}$.

5. Variable "Directions"

If we are given a finite number of $d \times n$ matrices ($d < n$ is fixed throughout this section) $A_1, ..., A_k$, then we know that

$$\mathcal{M}(A_1, ..., A_k) = \left\{ \sum_{i=1}^{k} g_i(A_i \mathbf{x}) : g_i \in C(\mathbb{R}^d, \mathbb{R}) \right\}$$

is not dense in $C(\mathbb{R}^n, \mathbb{R})$ (Corollary 3.3). However, it does not follow from this fact that if we are permitted to vary the $\{A_i\}_{i=1}^k$, while keeping k fixed (k may be very large), we do not get all of $C(\mathbb{R}^n, \mathbb{R})$. This is the problem we address in this section. We set

$$\mathcal{M}_k = \bigcup_{A_1, ..., A_k} \mathcal{M}(A_1, ..., A_k)$$

and ask whether to each $f \in C(\mathbb{R}^n, \mathbb{R})$ and compact K in $\mathbb{R}^n$

$$\inf_{g \in \mathcal{M}_k} \|f - g\|_{L^\infty(K)} = 0.$$

The answer is no. However, this is a natural question to ask as one of the objects and advantages of working with ridge functions is in "choosing" the directions $A_1, ..., A_k$ depending upon the function f.

We prove the following result.

Theorem 5.1. *Given any $k \in \mathbb{N}$, there exist an $f \in C(\mathbb{R}^n, \mathbb{R})$ and a $K \subset \mathbb{R}^n$, compact, such that*

$$\inf_{g \in \mathcal{M}_k} \|f - g\|_{L^\infty(K)} > 0.$$

Our proof of Theorem 5.1 is elementary, but not short. We start with some preliminary material.

In the proof of Theorem 2.1 we constructed a linear functional which vanished on $\mathcal{M}(\Omega)$. For Ω with only a finite number of terms $A_1, ..., A_k$, there exist simpler linear functionals annihilating $\mathcal{M}(A_1, ..., A_k)$. One such set of linear functionals was given in [1] and in this more general setting may be defined as follows.

For each $i \in \{1, ..., k\}$, let $\mathbf{b}^i \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ satisfy $A_i \mathbf{b}^i = \mathbf{0}$. (Assume for convenience that the A_i are distinct, $L(A_i) \neq L(A_j)$ for $i \neq j$, and the vectors

$\mathbf{b}^i$ are distinct (and even pairwise linearly independent).) Consider the 2^k points (not necessarily all distinct) of the form

$$\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j,$$

where $\varepsilon_j \in \{0, 1\}$, $j=1, ..., k$, and $\mathbf{y}$ is any fixed vector in $\mathbb{R}^n$. To each such point we associate the weight $(-1)^{|\varepsilon|}$, where $|\varepsilon|=\sum_{j=1}^{k} \varepsilon_j$.

LEMMA 5.1. *The nontrivial linear functional on* $C(\mathbb{R}^n, \mathbb{R})$ *given by*

$$l(h)=\sum_{\substack{\varepsilon_j \in \{0,1\} \\ j=1,...,k}} (-1)^{|\varepsilon|} h\left(\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j\right)$$

annihilates $\mathcal{M}(A_1, ..., A_k)$.

Proof. It suffices to prove that

$$\sum_{\substack{\varepsilon_j \in \{0,1\} \\ j=1,...,k}} (-1)^{|\varepsilon|} g\left(A_i\left(\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j\right)\right)=0$$

for every $g \in C(\mathbb{R}^d, \mathbb{R})$ and each $i \in \{1, ..., k\}$. Since $A_i\mathbf{b}^i=\mathbf{0}$, it follows that the sum of the coefficients $(-1)^{|\varepsilon|}$ of each of the *distinct* points $A_i(\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j)$ in the above sum is zero. That is, for each given $\{\varepsilon_j\}$, $j \neq i$, $A_i(\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j)$ is a constant vector independent of $\varepsilon_i \in \{0, 1\}$, and thus both $+1$ and -1 are the coefficients of $g(A_i(\mathbf{y}+\sum_{j=1}^{k} \varepsilon_j \mathbf{b}^j))$ in the above sum. ∎

We also use this next result.

LEMMA 5.2. *There exists a constant* $c(k, n)$ *depending only on* k *and* n, *such that for any given unit vectors* $\mathbf{c}^1, ..., \mathbf{c}^k \in \mathbb{R}^n$ *there exists a unit* $\mathbf{v} \in \mathbb{R}^n$ *satisfying*

$$|\mathbf{c}^i \cdot \mathbf{v}| \geqslant c(k, n),$$

$i=1, ..., k$.

Proof. Let α_n be the surface area of the unit ball in $\mathbb{R}^n$. For each $t \in (0, 1]$, let $\gamma(t)$ denote the surface area of the unit ball covered by the set

$$\{\mathbf{w}: \|\mathbf{w}\|=1, |\mathbf{c} \cdot \mathbf{w}|<t\},$$

where $\mathbf{c}$ is any fixed unit vector in $\mathbb{R}^n$. Obviously $\gamma(t)$ is a continuous function of t ($\gamma(1)=\alpha_n$), and $\lim_{t \to 0^+} \gamma(t)=0$. Let t_k satisfy $k\gamma(t_k)=\alpha_n$.

Let $\mathbf{c}^1, \ldots, \mathbf{c}^k$ be any k unit vectors. By construction there must exist a unit vector $\mathbf{v}$ such that

$$|\mathbf{c}^i \cdot \mathbf{v}| \geqslant t_k, \qquad i=1, \ldots, k.$$

Set $c(k, n) = t_k$. ▮

We are now prepared to prove Theorem 5.1.

Proof of Theorem 5.1. Let $A_1, \ldots, A_k$ be any k $d \times n$ matrices. With no loss of generality, assume that $L(A_i) \neq L(A_j)$ for $i \neq j$. Let K be any compact subset of $\mathbb{R}^n$ with interior. Without loss of generality we will also assume that K contains the ball, centered at the origin, with radius σk, some $\sigma > 0$. Let $f \in C(\mathbb{R}^n, \mathbb{R})$ vanish on K outside the ball with center $\mathbf{0}$ and radius $\sigma c(k, n)$, and $f(\mathbf{0}) = 1$. To prove the theorem we show that

$$\inf_{g \in \mathcal{M}_k} \|f - g\|_{L^\infty(K)} \geqslant \frac{1}{2^k}.$$

We recall that if L is a subspace of a normed linear space X, and $l \in X^*$ annihilates L, then

$$\inf_{g \in L} \|f - g\|_X \geqslant \frac{|l(f)|}{\|l\|_{X^*}}.$$

(This follows very easily.) We apply this inequality to prove our result.

For each $A_1, \ldots, A_k$ as above, choose unit vectors $\mathbf{c}^1, \ldots, \mathbf{c}^k$ such that $A_i \mathbf{c}^i = \mathbf{0}$, $i = 1, \ldots, k$. Let $\mathbf{v}$ be a unit vector satisfying

$$|\mathbf{c}^i \cdot \mathbf{v}| \geqslant c(k, n), \qquad i = 1, \ldots, k,$$

as given by Lemma 5.2. Choose $\delta_i \in \{-1, 1\}$ so that

$$\delta_i \mathbf{c}^i \cdot \mathbf{v} \geqslant c(k, n), \qquad i = 1, \ldots, k,$$

and set $\mathbf{b}^i = \sigma \delta_i \mathbf{c}^i$, $i = 1, \ldots, k$.

Define

$$l(g) = \sum_{\substack{\varepsilon_j \in \{0, 1\} \\ j = 1, \ldots, k}} (-1)^{|\varepsilon|} g\left(\sum_{j=1}^k \varepsilon_j \mathbf{b}^j\right).$$

l is a nontrivial linear functional on $C(K)$ (since each of the $\sum_{j=1}^k \varepsilon_j \mathbf{b}^j$ lies in K) of norm at most 2^k. By Lemma 5.1, l annihilates $\mathcal{M}(A_1, \ldots, A_k)$. Now if any of the $\{\varepsilon_j\}_{j=1}^k$ are not zero, then

$$\left\|\sum_{j=1}^k \varepsilon_j \mathbf{b}^j\right\| \geqslant \left(\sum_{j=1}^k \varepsilon_j \mathbf{b}^j \cdot \mathbf{v}\right) \geqslant \sum_{j=1}^k \varepsilon_j (\mathbf{b}^j \cdot \mathbf{v}) \geqslant \sigma c(k, n).$$

Thus

$$l(f)=\sum_{\substack{\varepsilon_j\in\{0,1\}\\ j=1,\ldots,k}}(-1)^{|\varepsilon|}f\left(\sum_{j=1}^{k}\varepsilon_j\mathbf{b}^j\right)=f(\mathbf{0})=1.$$

This proves the theorem. ∎

Remark 5.1. $\mathcal{M}_k$ is not a closed set ($k>1$). That is, there exist $g\in\overline{\mathcal{M}_k}$ which are not in $\mathcal{M}(A_1, \ldots, A_k)$ for any choice of distinct $A_1, \ldots, A_k$. For example, if $k=2$, $n=2$ (and thus $d=1$), then $x_1x_2^2\in\overline{\mathcal{M}_2}$, but $x_1x_2^2\notin\mathcal{M}(\mathbf{a}^1, \mathbf{a}^2)$ for any two vectors $\mathbf{a}^1$, $\mathbf{a}^2\in\mathbb{R}^2$. The mechanisms here are understood (see, e.g., [17]) but are technical. We will not go into detail.

Acknowledgments

We thank Professors Ron Aharoni, Yuri Brudnyi, Amos Ron, and Vladimir Tikhomirov for their comments, suggestions, and contributions.

References

1. D. Braess and A. Pinkus, Interpolation by ridge functions, *J. Approx. Theory* **73** (1993), 218–236.
2. C. K. Chui and Xin Li, Approximation by ridge functions and neural networks with one hidden layer, *J. Approx. Theory* **70** (1992), 131–141.
3. G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control, Signals, Systems* **2** (1989), 303–314.
4. W. Dahmen and C. A. Micchelli, Some remarks on ridge functions, *Approx. Theory Appl.* **3** (1987), 139–143.
5. P. Diaconis and M. Shahshahani, On nonlinear functions of linear combinations, *SIAM J. Sci. Statist. Comput. Appl.* **5** (1984), 175–191.
6. D. L. Donoho and I. M. Johnstone, Projection-based approximation and a duality method with kernel methods, *Ann. Statist.* **17** (1989), 58–106.
7. R. E. Edwards, "Functional Analysis, Theory and Applications," Holt, Rinehart & Winston, New York, 1965.
8. C. Hamaker and D. C. Solomon, The angles between the null spaces of X-rays, *J. Math. Anal. Appl.* **62** (1978), 1–23.
9. P. J. Huber, Projection pursuit, *Ann. Statist.* **13** (1985), 435–475.
10. F. John, "Plane Waves and Spherical Means Applied to Partial Differential Equations," Interscience, New York, 1955.
11. L. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken, Multilayer feedforward networks with a non-polynomial activation function can approximate any function, *Neural Networks*, to appear.
12. B. F. Logan and L. A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math. J.* **42** (1975), 645–659.
13. F. Natterer, "The Mathematics of Computerized Tomography," Wiley, New York, 1986.
14. X. Sun and E. W. Cheney, The fundamentality of sets of ridge functions, preprint.

15. B. A. Vostrecov and M. A. Kreines, Approximation of continuous functions by superpositions of plane waves, *Dokl. Akad. Nauk SSSR* **140** (1961), 1237–1240 (*Soviet Math. Dokl.* **2** (1961), 1326–1329).
16. B. A. Vostrecov and M. A. Kreines, Approximation of a plane wave by superpositions of plane waves of given directions, *Dokl. Akad. Nauk SSSR* **144** (1962), 1212–1214 (*Soviet Math. Dokl.* **3** (1962), 875–877).
17. B. A. Vostrecov, Conditions for a function of many variables to be representable as a sum of a finite number of plane waves traveling in given directions, *Dokl. Akad. Nauk SSSR* **153** (1963), 16–19 (*Soviet Math. Dokl.* **4** (1963), 1588–1591).