JOURNAL OF APPROXIMATION THEORY **85**, 269–287 (1996)
ARTICLE NO. 0042

TDI-Subspaces of $C(\mathbb{R}^d)$ and Some Density Problems from Neural Networks

Allan Pinkus

Department of Mathematics, Technion, Haifa, 32000, Israel
E-mail: pinkus@techunix.technion.ac.il

Communicated by E. W. Cheney

Received May 11, 1994; accepted in revised form November 3, 1995

In this paper we consider translation and dilation invariant subspaces of $C(\mathbb{R}^d)$. We characterize all such subspaces and also identify those $f \in C(\mathbb{R}^d)$, the span of whose translates and dilates generate non-trivial subspaces. We apply these and related results to some mathematical models in the theory of neural networks.

1. Introduction

Some time ago Burkhard Lenze asked me the following question: For which functions $\sigma \in C(\mathbb{R})$ is

$$\operatorname{span}\left\{\sigma\left(\rho \prod_{i=1}^{d} (x_i - b_i)\right): \rho, b_1, ..., b_d \in \mathbb{R}\right\}$$

dense in $C(\mathbb{R}^d)$? Here we consider $C(\mathbb{R}^d)$ with the topology of uniform convergence on compact subsets. Lenze's interest in this question was prompted by a mathematical model from the theory of neural networks. More specifically, a type of multilayer feedforward network with a single hidden layer, see Lenze [11] and [12]. The consideration of this specific problem led us to the study of a more general question on translation and dilation invariant subspaces. In this paper we will report on some of the results, both old and new, in this area and apply them to the analysis of three different mathematical models which appear in the theory of neural networks.

The present form of this paper owes much to Aharon Atzmon of Tel-Aviv University. In particular, he brought the important references Schwartz [18] and Harasymiv [6] to our attention, and also gave willingly and patiently of his time and knowledge.

0021-9045/96 $18.00

2. Preliminaries and History

We recall that a subspace W of $C(\mathbb{R}^d)$ is *translation-invariant* if $f \in W$ implies $f(\cdot - \mathbf{b}) \in W$ for all $\mathbf{b} \in \mathbb{R}^d$. The subspace W of $C(\mathbb{R}^d)$ is *dilation-invariant* if $f \in W$ implies $f(a_1 \cdot, ..., a_d \cdot) \in W$ for all $\mathbf{a} = (a_1, ..., a_d) \in \mathbb{R}^d$. We say that a subspace is a TDI-subspace if it is closed translation and dilation invariant subspace. For each $f \in C(\mathbb{R}^d)$, we set

$$\mathcal{M}_f = \overline{\text{span}}\{f(a_1 \cdot - b_1, ..., a_d \cdot - b_d) \colon \mathbf{a}, \mathbf{b} \in \mathbb{R}^d\},$$

where $\mathbf{a} = (a_1, ..., a_d)$ and $\mathbf{b} = (b_1, ..., b_d)$. That is, $\mathcal{M}_f$ is the closure of the smallest TDI-subspace which contains (or is generated by) f.

This paper contains two main results regarding TDI-subspaces. Firstly we identify those $f \in C(\mathbb{R}^d)$ which generate non-trivial TDI-subspaces (Theorem 3), i.e., for which $\mathcal{M}_f$ is neither the empty set nor all of the space $C(\mathbb{R}^d)$. Secondly we characterize the TDI-subspaces themselves (Theorem 4). These two problems are connected, but distinct. In Section 4 we apply Theorem 3 and also the related Proposition 2, due to Schwartz, to certain problems motivated by mathematical models from the theory of neural networks.

We first recall some standard notation. For $\mathbf{r} = (r_1, ..., r_d) \in \mathbb{Z}_+^d$ we set

$$D^{\mathbf{r}} = \frac{\partial^{|\mathbf{r}|}}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}$$

where $|\mathbf{r}| = r_1 + \cdots + r_d$. We let $C_0^\infty(\mathbb{R}^d)$ denote the set of C^∞ functions defined on $\mathbb{R}^d$ with compact support. For $f \in C(\mathbb{R}^d)$ and p a polynomial of degree m of the form

$$p(\mathbf{x}) = \sum_{|\mathbf{r}| \leqslant m} a_{\mathbf{r}} \mathbf{x}^{\mathbf{r}}$$

we say that f satisfies the differential equation

$$p(D) f := \sum_{|\mathbf{r}| \leqslant m} a_{\mathbf{r}} D^{\mathbf{r}} f = 0$$

in the *weak sense* (or generalized sense) if

$$\int_{\mathbb{R}^d} f(p^*(D) \phi) = 0$$

for all $\phi \in C_0^\infty(\mathbb{R}^d)$, where $p^*(\mathbf{x}) = \sum_{|\mathbf{r}| \leqslant m} a_{\mathbf{r}} (-1)^{|\mathbf{r}|} \mathbf{x}^{\mathbf{r}}$.

Translation invariant subspaces have been much studied over the years. It is an important topic in harmonic analysis. In the space $C(\mathbb{R}^d)$, with the topology of uniform convergence on compact subsets, this subject was considered by Schwartz in [19]. He introduced the class of mean-periodic functions (functions whose linear span of translates are not dense in $C(\mathbb{R}^d)$) and was able to characterize such functions in the case $d=1$. Unfortunately for $d>1$ the theory does not seem to have a simple extension, see Gurevich [5]. TDI-subspaces are somewhat simpler. It is well-known that the only TDI-subspaces of $L^2(\mathbb{R}^d)$ are trivial ones. This result was extended to $L^p(\mathbb{R}^d)$ for all $1<p<\infty$ by Harasymiv [6]. The case $p=1$, where non-trivial TDI-subspaces do exist, is considered in Harasymiv [7]. See also Harasymiv [8] for other related results, and Alexsandrov, Kargaev [1] for examples of non-trivial TDI-subspaces in $L^p(\mathbb{R}^d)$ where $0<p<1$. In another direction, Brown, Schreiber, Taylor [3] considered translation and rotation invariant closed subspaces of $C(\mathbb{R}^d)$, while Sternfeld, Weit [20] classify the TDI-subspaces of $C(\mathbb{C})$. For some similar problems relating to the space $C_0(\mathbb{R})$, see Atzmon [2].

We will recall two results of Schwartz [18]. For a given $f\in C(\mathbb{R}^d)$, set

$$\mathscr{G}_f=\overline{\operatorname{span}}\{f(\lambda\cdot-b_1,...,\lambda\cdot-b_d):\lambda\in\mathbb{R},\mathbf{b}\in\mathbb{R}^d\}$$

where $\mathbf{b}=(b_1,...,b_d)$. Note that the same dilation is used in each variable. Then,

Theorem 1 (Schwartz [18]). *Let $f\in C(\mathbb{R}^d)$. Then $\mathscr{G}_f\neq C(\mathbb{R}^d)$ if and only if there exists a non-trivial homogeneous polynomial p such that $p(D)f=0$ in the weak sense.*

We will not prove this theorem here. We quote this theorem for two reasons. Firstly, this result is related to Theorem 3, and we use essentially the same method of proof in our proof of the first part of Theorem 3. Secondly, one consequence of Theorem 1 is the following result, also due to Schwartz, which we will apply later in Section 4.

Proposition 2 (Schwartz [18]). *Let $f\in C(\mathbb{R}^d)$ and $\mathscr{G}_f$ be as defined above. Let $\mathscr{R}_f$ denote the closed rotation invariant subspace generated by $\mathscr{G}_f$. Then $\mathscr{R}_f\neq C(\mathbb{R}^d)$ if and only if $\Delta^m f=0$, in the weak sense, for some positive integer m, where Δ is the Laplacian operator.*

3. TDI-Subspaces

We first delineate those $f\in C(\mathbb{R}^d)$ for which $\mathscr{M}_f\neq C(\mathbb{R}^d)$.

Theorem 3. *For $f\in C(\mathbb{R}^d)$ the following are equivalent:*

(a) $\mathscr{M}_f\neq C(\mathbb{R}^d)$.

(b) *For some $\mathbf{r}\in\mathbb{Z}^d_+$, $D^{\mathbf{r}}f=0$ in the weak sense.*

(c) *For some* $\mathbf{r} \in \mathbb{Z}_+^d$, *f has the form*

$$f(x_1, ..., x_d) = \sum_{i=1}^{d} \sum_{j=0}^{r_i - 1} g_{ij}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_d)\, x_i^j, \tag{1}$$

where $g_{ij} \in C(\mathbb{R}^{d-1})$ *for all i and j.*

Remark 1. For $d=1$ we understand (1) to mean that the g_{1j} are constants, i.e., f is a polynomial. For the case $d=1$ we have $\mathcal{M}_f = \mathcal{G}_f$.

Proof. For convenience we will abuse notation and write

$$(\mathbf{a} \cdot \mathbf{x} - \mathbf{b}) = (a_1 x_1 - b_1, ..., a_d x_d - b_d).$$

((a) $\Rightarrow$ (b)). Since $\mathcal{M}_f \neq C(\mathbb{R}^d)$ there exists a non-trivial Borel measure μ of finite total variation and of compact support such that

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

for all $h \in \mathcal{M}_f$. Since μ is non-trivial and polynomials are dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compact subsets, there exists an $\mathbf{r} \in \mathbb{Z}_+^d$ such that

$$\int_{\mathbb{R}^d} \mathbf{x}^{\mathbf{r}}\, d\mu(\mathbf{x}) \neq 0.$$

It is relatively simple to show that for each $\phi \in C_0^\infty(\mathbb{R}^d)$, the convolution $(f * \phi)$ is contained in $\mathcal{M}_f$. Since both f and ϕ are in $C(\mathbb{R}^d)$, and ϕ has compact support, this can be proven by taking Riemann sums of the convolution integral, see e.g. [13]. We consider taking derivatives as a limiting operation in taking differences. Since $(f * \phi) \in C^\infty(\mathbb{R}^d)$, and thus it and all its derivatives are uniformly continuous on every compact set, it follows that for each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$

$$\frac{\partial^{\mathbf{r}}}{\partial \mathbf{a}^{\mathbf{r}}} (f * \phi)(\mathbf{a} \cdot \mathbf{x} - \mathbf{b}) = \mathbf{x}^{\mathbf{r}} (D^{\mathbf{r}}(f * \phi))(\mathbf{a} \cdot \mathbf{x} - \mathbf{b}) \in \mathcal{M}_f,$$

i.e., the appropriate difference operator converges uniformly to the corresponding derivative on each compact subset of $\mathbb{R}^d$. Thus

$$\int_{\mathbb{R}^d} \mathbf{x}^{\mathbf{r}} (D^{\mathbf{r}}(f * \phi))(\mathbf{a} \cdot \mathbf{x} - \mathbf{b})\, d\mu(\mathbf{x}) = 0.$$

Letting $\mathbf{a}=\mathbf{0}$, we see that

$$(D^{\mathbf{r}}(f*\phi))(-\mathbf{b})\int_{\mathbb{R}^d}\mathbf{x}^{\mathbf{r}}\,d\mu(\mathbf{x})=0$$

for each choice of $\mathbf{b}\in\mathbb{R}^d$ and $\phi\in C_0^\infty(\mathbb{R}^d)$. This implies, since $\int_{\mathbb{R}^d}\mathbf{x}^{\mathbf{r}}\,d\mu(\mathbf{x})\neq 0$, that

$$D^{\mathbf{r}}(f*\phi)=0$$

for all $\phi\in C_0^\infty(\mathbb{R}^d)$. (b) follows readily from this fact.

((b) $\Rightarrow$ (c)). We were surprised not to have found this explicit form in the literature. We feel certain it is there somewhere. It is a folk theorem, is well-known, and is not particulary difficult to prove. It follows from the fact (found in a number of different books) that if

$$\frac{\partial}{\partial x_i}h(x_1,...,x_d)=0$$

in the weak sense, then

$$h(x_1,...,x_d)=g(x_1,...,x_{i-1},x_{i+1},...,x_d).$$

That is, h is a function independent of x_i.

((c) $\Rightarrow$ (a)). Assume f has the form (1) for some $g_{ij}\in C(\mathbb{R}^{d-1})$. Then for each $\mathbf{a},\mathbf{b}\in\mathbb{R}^d$, the function $f(\mathbf{a}\cdot\mathbf{x}-\mathbf{b})$ is also of the form

$$\sum_{i=1}^{d}\sum_{j=0}^{r_i-1}h_{ij}(x_1,...,x_{i-1},x_{i+1},...,x_d)\,x_i^j$$

for some $h_{ij}\in C(\mathbb{R}^{d-1})$. That is,

$$\mathscr{M}_f\subseteq\mathscr{A}=\left\{\sum_{i=1}^{d}\sum_{j=0}^{r_i-1}h_{ij}(x_1,...,x_{i-1},x_{i+1},...,x_d)\,x_i^j\colon h_{ij}\in C(\mathbb{R}^{d-1})\right\}.$$

There exist many appropriate linear functionals which annihilate the subspace $\mathscr{A}$. This implies that $\mathscr{A}\neq C(\mathbb{R}^d)$, and therefore $\mathscr{M}_f\neq C(\mathbb{R}^d)$. Let us construct such a continuous linear functional. If ℓ_k is any continuous linear functional with compact support which annihilates

$$\sum_{j=0}^{r_k-1}h_j(x_1,...,x_{k-1},x_{k+1},...,x_d)\,x_k^j$$

(for all $h_j\in C(\mathbb{R}^{d-1})$, $j=1,...,r_k-1$), and such that the ℓ_k commute, then $\prod_{k=1}^d\ell_k$ annihilates $\mathscr{A}$. For example, we can take ℓ_k to be a sufficiently high order difference in the variable x_k. That is, we let ℓ_k be defined as follows: For $f \in C(\mathbb{R}^d)$ and some $h > 0$, fixed,

$$\ell_k(f) = (S_{h,k})^{r_k} f(x_1, ..., x_d)$$

for any choice of $(x_1, ..., x_d) \in \mathbb{R}^d$, where $S_{h,k}$ is the linear operator

$$S_{h,k} f(x_1, ..., x_d) = f(x_1, , ..., x_k + h, ..., x_d) - f(x_1, ..., x_k, ..., x_d).$$

There are many other possible choices of ℓ_k, $k = 1, ..., d$. ▮

In the proof of (a) ⇒ (b) in the above, we obtained a result well worth highlighting. We also use this fact in the subsequent analysis.

Corollary 4. *Let $f \in C(\mathbb{R}^d)$. If $D^{\mathbf{r}} f \neq 0$ (in the weak sense), then $\mathbf{x}^{\mathbf{r}} \in \mathcal{M}_f$.*

Remark 2. Functions of the form (1) (for $d \geqslant 2$) have been termed *quasipolynomials* (see e.g., Brudnyi [4]), and *generalized polynomials* (see e.g., Vaindiner [21]).

Remark 3. It follows from Theorem 3 that if $f \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, $(f \neq 0)$, then $\mathcal{M}_f = C(\mathbb{R}^d)$. One can also obtain this result more directly by using the representation of the linear functionals on $C(\mathbb{R}^d)$ and Fourier transforms. In fact the translation invariant subspace generated by f (without dilation) is itself dense in $C(\mathbb{R}^d)$.

Theorem 3 delineates those functions f which generate non-trivial TDI-subspaces of $C(\mathbb{R}^d)$. However it does not characterize the subspaces themselves. We now do just that. We identify all TDI-subspaces, of which there are only a countable number. They have a rather simple elegant structure.

To explain, we first set

$$\mathcal{Z} = \overline{\mathbb{Z}}^d_+ = \{\mathbf{k} = (k_1, ..., k_d) \colon k_i \text{ is a nonnegative integer or infinity}\}.$$

For $\mathbf{k} \in \mathcal{Z}$, we let

$$\mathcal{A}_{\mathbf{k}} = \overline{\operatorname{span}}\{x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} \colon \mathbf{j} \in \mathcal{Z}, \mathbf{j} \leqslant \mathbf{k}\},$$

where by $\mathbf{j} \leqslant \mathbf{k}$ we mean $j_i \leqslant k_i$, $i = 1, ..., d$. (We have slightly abused notation. If, for example, $k_i = \infty$, then we mean that x_i^j appears for all nonnegative integers j.) If $k_1 = \cdots = k_\ell = \infty$, and $k_{\ell+1}, ..., k_d$ are each finite nonnegative integers, then

$$\mathcal{A}_{\mathbf{k}} = \operatorname{span}\{h(x_1, ..., x_\ell)\, x_{\ell+1}^{j_{\ell+1}} \cdots x_d^{j_d} \colon 0 \leqslant j_i \leqslant k_i, i = \ell + 1, ..., d, h \in C(\mathbb{R}^\ell)\}.$$

Note that if $\mathbf{k} = (\infty, ..., \infty)$, then $\mathscr{A}_{\mathbf{k}} = C(\mathbb{R}^d)$. It is simple to check that each $\mathscr{A}_{\mathbf{k}}$ is a TDI-subspace. The result is the following.

THEOREM 5. *Every TDI-subspace is a finite sum of spaces* $\mathscr{A}_{\mathbf{k}}$.

We divide the proof of Theorem 5 into a series of steps.

LEMMA 6. *Any sum of a countable number of the* $\mathscr{A}_{\mathbf{k}}$ *is also a sum of a finite number of the* $\mathscr{A}_{\mathbf{k}}$.

Proof. For each $\mathbf{k} \in \mathscr{Z}$, let

$$\mathscr{B}_{\mathbf{k}} = \{\mathbf{j}\colon \mathbf{j} \in \mathscr{Z}, \mathbf{j} \leqslant \mathbf{k}\}.$$

Let $\mathscr{B}$ be any set in $\mathscr{Z}$ with the property that if $\mathbf{k} \in \mathscr{B}$ then $\mathbf{j} \in \mathscr{B}$ for all $\mathbf{j} \leqslant \mathbf{k}$. We claim that $\mathscr{B}$ is the union of a finite number of $\mathscr{B}_{\mathbf{k}}$. This is equivalent to the statement of the lemma. It is important to recall that some of the k_i may be infinite, and this should be allowed for in what follows.

Let K be the set of maximal points in $\mathscr{B}$, i.e., the set of $\mathbf{k} \in \mathscr{B}$ for which if $\mathbf{j} \in \mathscr{B}$ and $\mathbf{k} \leqslant \mathbf{j}$, then $\mathbf{k} = \mathbf{j}$. Obviously $\mathscr{B}$ is the union of the $\mathscr{B}_{\mathbf{k}}$ with $\mathbf{k} \in K$. Our claim is thus equivalent to the claim that K is finite.

Let $d = 2$. If $\mathbf{k} = (k_1, k_2)$ and $\mathbf{m} = (m_1, m_2)$ are in K, then $(k_1 - m_1)(k_2 - m_2) < 0$ (with the proper understanding if some of the components equal ∞). If K has a countable number of distinct points, then ordering them so that their first component is increasing, we see that the second component thereof is a countable strictly decreasing sequence of non-negative integers. This is impossible.

Now assume $d > 2$. For each pair of distinct points $\mathbf{k} = (k_1, ..., k_d)$, $\mathbf{m} = (m_1, ..., m_d)$ in K there exists a pair of indices (i, j), where $(i, j) \in \{1, ..., d\}$, $i \neq j$, such that $(k_i - m_i)(k_j - m_j) < 0$. This is the important property we will use. Assuming K has an infinite number of points all with equal rth coordinate, for some $r \in \{1, ..., d\}$. This then implies that for such points in K the property mentioned above holds for $i, j \in \{1, ..., d\} \backslash \{r\}$, $i \neq j$. An induction argument then proves the result.

We divide the proof into two cases. Firstly assume that for each $\mathbf{k} \in K$ we have $\max\{k_i\colon i = 1, ..., d\} = \infty$. Since K has an infinite number of points there must exist an $r \in \{1, ..., d\}$ and an infinite subset of K such that $k_r = \infty$ for all points in this subset. This proves the case.

We now assume the existence of a $\mathbf{k} \in K$ for which $k_i < \infty$, $i = 1, ..., d$. For each $\mathbf{m} \in K$, $\mathbf{m} \neq \mathbf{k}$, there exists a pair of indices (i, j) where $i, j \in \{1, ..., d\}$, $i \neq j$, such that $(k_i - m_i)(k_j - m_j) < 0$. Since K has an infinite number of distinct points there must exist a pair of indices (r, s), as above, such that this property holds for this particular pair and an infinite number of points in K. Thus for an infinite number of points in K either $m_r < k_r$ or $m_s < k_s$

for fixed r and s. Assume $m_r < k_r$ for an infinite number of points $\mathbf{m}$ in K. Since each $m_r \in \{0, 1, ..., k_r - 1\}$ and k_r is finite, there must exist an infinite number of points in K with equal rth coordinate. ∎

Let $f \in C(\mathbb{R}^d)$ be such that $\mathcal{M}_f \neq C(\mathbb{R}^d)$. Thus f has the form

$$f(x_1, ..., x_d) = \sum_{i=1}^{d} \sum_{j=0}^{r_i - 1} g_{ij}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_d)\, x_i^j. \tag{2}$$

Consider each $g_{ij} \in C(\mathbb{R}^{d-1})$. If $D^{\mathbf{s}} g_{ij} \neq 0$ (in the weak sense) for each $\mathbf{s} \in \mathbb{Z}_+^{d-1}$, then we leave g_{ij} as it is. If, however, $D^{\mathbf{s}} g_{ij} = 0$ (in the weak sense) for some $\mathbf{s} \in \mathbb{Z}_+^{d-1}$, then we may write g_{ij} as a sum, as in (2), of functions of fewer variables times monomials in the other variables. That is, it is possible, after a finite number of steps, to bring f to the form

$$f(x_1, ..., x_d) = \sum_{\ell, \mathbf{i}, \mathbf{j}} g_{\mathbf{i}, \mathbf{j}}(x_{i_1}, ..., x_{i_\ell})\, x_{i_{\ell+1}}^{j_{\ell+1}} \cdots x_{i_d}^{j_d} \tag{3}$$

where the (finite) sum is taken over $\ell \in \{0, 1, ..., d\}$, the $\mathbf{i} = (i_1, ..., i_d)$ vary among the possible permutations of $\{1, ..., d\}$, the $\mathbf{j} = (j_{\ell+1}, ..., j_d\}$ vary over the nonnegative integers less than some constant (depending on f), and

$$D^{\mathbf{s}} g_{\mathbf{i}, \mathbf{j}}(x_{i_1}, ..., x_{i_\ell}) \neq 0, \tag{4}$$

(in the weak sense) for each choice of $\mathbf{s} \in \mathbb{Z}_+^\ell$. (If $\ell = 0$, then $g_{\mathbf{i}, \mathbf{j}}$ is a constant.) The reason for writing f in the form (3) may be seen from the next two propositions. The essential ingredient is contained in the statement of Corollary 4.

Proposition 7. *Let $\mathbf{k} \in \mathbb{Z}$ where $k_1 = \cdots = k_\ell = \infty$, and $k_{\ell+1}, ..., k_d$ are each finite nonnegative integers. Let*

$$f(x_1, ..., x_d) = g(x_1, ..., x_\ell)\, x_{\ell+1}^{k_{\ell+1}} \cdots x_d^{k_d}$$

where $g \in C(\mathbb{R}^\ell)$ is any function for which $D^{\mathbf{s}} g \neq 0$ (in the weak sense) for each $\mathbf{s} \in \mathbb{Z}_+^\ell$. Then $\mathcal{M}_f = \mathcal{A}_{\mathbf{k}}$.

Proof. Since $D^{\mathbf{s}} g \neq 0$ (in the weak sense) for each $\mathbf{s} \in \mathbb{Z}_+^\ell$ we have $D^{\mathbf{r}} f \neq 0$ where $\mathbf{r} = (s_1, ..., s_\ell, k_{\ell+1}, ..., k_d)$. It thus follows from Corollary 4 that $\mathbf{x}^{\mathbf{r}} \in \mathcal{M}_f$ for each such $\mathbf{r} \in \mathbb{Z}_+^d$. A simple argument shows that the TDI-subspaces (only translation invariance is in fact necessary) generated by the $\mathbf{x}^{\mathbf{r}}$ is exactly $\mathcal{A}_{\mathbf{r}}$. (That is, it includes all monomials of lower order power.) Since $\mathcal{A}_{\mathbf{r}} \subseteq \mathcal{M}_f$ for each choice of non-negative integers $(s_1, ..., s_\ell)$ in $\mathbf{r} = (s_1, ..., s_\ell, k_{\ell+1}, ..., k_d)$, it follows that $\mathcal{A}_{\mathbf{r}} \subseteq \mathcal{M}_f$. On the other hand $\mathcal{M}_f \subseteq \mathcal{A}_{\mathbf{k}}$ by inspection. ∎

PROPOSITION 8. *$\mathcal{M}_f$ is a finite sum of the $\mathcal{A}_{\mathbf{k}}$.*

Proof. Let f have the form (3) where each $g_{\mathbf{i},\mathbf{j}}$ from (3) satisfies (4). It suffices to show that

$$g_{\mathbf{i},\mathbf{j}}(x_{i_1}, ..., x_{i_\ell})\, x_{i_{\ell+1}}^{j_{\ell+1}} \cdots x_{i_d}^{j_d} \in \mathcal{M}_f$$

for some such $g_{\mathbf{i},\mathbf{j}}$. For if this is the case, then the associated $\mathcal{A}_{\mathbf{k}}$ is in $\mathcal{M}_f$ (by Proposition 7) and since $\mathcal{M}_f$ is a closed linear space, it also follows that

$$f(x_1, ..., x_d) - g_{\mathbf{i},\mathbf{j}}(x_{i_1}, ..., x_{i_\ell})\, x_{i_{\ell+1}}^{j_{\ell+1}} \cdots x_{i_d}^{j_d} \in \mathcal{M}_f.$$

We can then continue this process, and in this way show that each term in the expansion (3) is in $\mathcal{M}_f$, and the result then follows from Proposition 7.

Among the (non-zero) terms in the sum (3) choose $\ell \in \{0, 1, ..., d\}$ maximal for which

$$g_{\mathbf{p},\mathbf{q}}(x_{p_1}, ..., x_{p_\ell})\, x_{p_{\ell+1}}^{q_{\ell+1}} \cdots x_{p_d}^{q_d}$$

appears, and $(q_{\ell+1}, ..., q_d)$ is also maximal in the sense that there is in (3) no other non-zero term in the form

$$g_{\mathbf{p},\mathbf{j}}(x_{p_1}, ..., x_{p_\ell})\, x_{p_{\ell+1}}^{j_{\ell+1}} \cdots x_{p_d}^{j_d}$$

with $q_i \leqslant j_i$, $i = \ell+1, ..., d$, and with strict inequality for some such i.

Assume for ease of notation that $p_i = i$, $i = 1, ..., d$. Let $\mathbf{r} = (r_1, ..., r_\ell, q_{\ell+1}, ..., q_d)$. For r_i sufficiently large, $i = 1, ..., \ell$, the operator $D^{\mathbf{r}}$ will annihilate all terms in the sum (3) except

$$g_{\mathbf{p},\mathbf{q}}(x_1, ..., x_\ell)\, x_{\ell+1}^{q_{\ell+1}} \cdots x_d^{q_d}.$$

Thus $\mathbf{x}^{\mathbf{r}} \in \mathcal{M}_f$ for all such $\mathbf{r} \in \mathbb{Z}_+^d$ (Corollary 4) which implies that $\mathcal{A}_k \in \mathcal{M}_f$ where $k_i = \infty$, $i = 1, ..., \ell$, $k_i = q_i$, $i = \ell+1, ..., d$. Thus

$$g_{\mathbf{p},\mathbf{q}}(x_1, ..., x_\ell)\, x_{\ell+1}^{q_{\ell+1}} \cdots x_d^{q_d} \in \mathcal{M}_f.$$

This proves the proposition. ∎

Proof of Theorem 5. Let $\mathcal{M}$ be any TDI-subspace of $C(\mathbb{R}^d)$. Set

$$K = \{\mathbf{k}\colon \mathcal{A}_{\mathbf{k}} \subseteq \mathcal{M}\}.$$

If $\mathcal{M} \neq \sum_{\mathbf{k}\in K} \mathcal{A}_{\mathbf{k}}$, then there exists an $f \in \mathcal{M} \backslash \sum_{\mathbf{k}\in K} \mathcal{A}_{\mathbf{k}}$. Now $\mathcal{M}_f \subseteq \mathcal{M}$, but $\mathcal{M}_f \nsubseteq \sum_{\mathbf{k}\in K} \mathcal{A}_{\mathbf{k}}$. Thus from Proposition 7 $\mathcal{A}_{\mathbf{s}} \subseteq \mathcal{M}$ for some $\mathbf{s} \notin K$. This contradicts the definition of K. Thus

$$\mathcal{M} = \sum_{\mathbf{k}\in K} \mathcal{A}_{\mathbf{k}}.$$

By Lemma 6 we can rewrite this as a sum over a finite number of terms. ▮

Remark 4. It follows from the proof of Theorem 5 that every TDI-subspace may be represented in the form $\mathscr{M}_f$ for various choices of $f \in C(\mathbb{R}^d)$, i.e., is generated by a single function.

Remark 5. This same type of analysis on $C(\mathbb{C})$ is done in Sternfeld, Weit [20].

4. Applications to Neural Network Models

We were motivated to look at these problems because of specific examples which arose in neural networks models. Two of these examples are related to Theorem 3 and are special cases of functions f of the form $\sigma(p(\cdot))$, where p is a specified polynomial on $\mathbb{R}^d$ and $\sigma \in C(\mathbb{R})$. The question asked was to characterize, for a given p, those σ for which $\mathscr{M}_{\sigma(p)} = C(\mathbb{R}^d)$. Obviously if σ is itself a polynomial then $\mathscr{M}_{\sigma(p)} \neq C(\mathbb{R}^d)$. This follows from Theorem 3, but also much more simply from the observation that if σ is a polynomial, then $\mathscr{M}_{\sigma(p)}$ is a (finite-dimensional) space of polynomials. In the first example, we also consider the matrix version of the problem. The third example is related to a model based on radial-basis functions. Here we will use Proposition 2 to characterize those functions for which density holds in $C(\mathbb{R}^d)$.

Example 1. In [13] it is proved that if $\sigma \in C(\mathbb{R})$, then

$$C(\mathbb{R}^d) = \overline{\text{span}} \left\{ \sigma\left(\sum_{i=1}^{d} a_i x_i + b \right) : (a_1, ..., a_d) \in \mathbb{R}^d, b \in \mathbb{R} \right\}$$

if and only if σ is not a polynomial. For a discussion of this problem in neural network, see [13]. In fact, it follows quite easily, again see [13], that the above holds for all $d \geqslant 1$ if and only if it holds for $d = 1$. Thus this result is also a consequence of Theorem 1 due to Schwartz [18].

We might also consider this result as a special case of Theorem 3. Set

$$f(x_1, ..., x_d) = \sigma(x_1 + \cdots + x_d).$$

Thus

$$f(a_1 x_1 - b_1, ..., a_d x_d - b_d) = \sigma\left(\sum_{i=1}^{d} a_i x_i - \left(\sum_{i=1}^{d} b_i \right) \right).$$

Now

$$D^{\mathbf{r}}\sigma(x_1+\cdots+x_d)=0$$

in the weak sense is equivalent to

$$\sigma^{(|\mathbf{r}|)}=0$$

in the weak sense. As is well-known, this implies that $\sigma^{(|\mathbf{r}|)}=0$ in the usual sense. That is, σ is a polynomial of degree at most $|\mathbf{r}|-1$.

A different (and totally equivalent) approach is based on (c) of Theorem 3. If $\mathscr{M}_f \neq C(\mathbb{R}^d)$, then

$$\sigma(x_1+\cdots+x_d)=\sum_{i=1}^{d}\sum_{j=0}^{r_i-1} g_{ij}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_d)\, x_i^j. \tag{5}$$

Now

$$\prod_{k=1}^{d}(S_{h,k})^{r_k}\,\sigma(x_1+\cdots+x_d)=0$$

since this linear operator annihilates the right-hand side of (5). Moreover

$$\prod_{k=1}^{d}(S_{h,k})^{r_k}\,\sigma(x_1+\cdots+x_d)=(S_h)^{|\mathbf{r}|}\,\sigma(t)$$

where $t=x_1+\cdots+x_d$, and

$$S_h\sigma(t)=\sigma(t+h)-\sigma(t).$$

It is known (and easy to prove) that if

$$(S_h)^{|\mathbf{r}|}\,\sigma(t)=0$$

for all $t\in\mathbb{R}$ and $h>0$, then σ is a polynomial of degree at most $|\mathbf{r}|-1$.

Let us now consider the matrix version of this problem. See Mhaskar, Micchelli [15] for some partial results. We wish to determine necessary and sufficient conditions on $\sigma\in C(\mathbb{R}^m)$, $d\geqslant m\geqslant 1$, such that

$$C(\mathbb{R}^d)=\overline{\operatorname{span}}\{\sigma(A\cdot+\mathbf{b})\colon A\in\mathbb{R}^{m\times d}, \mathbf{b}\in\mathbb{R}^m\}$$

i.e., where A ranges over all $m\times d$ real matrices. We show that this holds if and only if σ is not a polynomial.

PROPOSITION 9. *Let $\sigma\in C(\mathbb{R}^m)$. Then*

$$\operatorname{span}\{\sigma(A\cdot+\mathbf{b})\colon A\in\mathbb{R}^{m\times d}, \mathbf{b}\in\mathbb{R}^m\}$$

is dense in $C(\mathbb{R}^d)$ if and only if σ is not a polynomial.

Proof. ($\Rightarrow$) If σ is a polynomial then the above span is, for each $A \in \mathbb{R}^{m \times d}$ and $\mathbf{b} \in \mathbb{R}^m$, contained in a fixed finite-dimensional space of polynomials. As such it cannot be dense in $C(\mathbb{R}^d)$.

($\Leftarrow$) We base the proof of this direction on the fact that if σ is not a polynomial, then

$$h(t) = \sigma(c_1 t + d_1, ..., c_m t + d_m)$$

is not a polynomial for some choice of $c_1, ..., c_m$ and $d_1, ..., d_m$. In fact we can take all the c_j except one to be zero. That is, from Palais [16], it follows that there exists an $i \in \{1, ..., m\}$ and constants $\{b_j\}_{j \neq i}$ such that

$$h(t) = \sigma(b_1, ..., b_{i-1}, t, b_{i+1}, ..., b_m)$$

is not a polynomial.

By the result of [13] (see above)

$$\operatorname{span}\{h(\mathbf{a} \cdot + b) \colon \mathbf{a} \in \mathbb{R}^d, b \in \mathbb{R}\}$$

is dense in $C(\mathbb{R}^d)$. Substituting it follows that

$$\operatorname{span}\{\sigma(A \cdot + \mathbf{b}) \colon A \in \mathbb{R}^{m \times d}, \mathbf{b} \in \mathbb{R}^m\}$$

is dense in $C(\mathbb{R}^d)$ for a rather small and special class of matrices and translates. ∎

Example 2. This is the example motivated by Burkhard Lenze. For discussion of this model and some of the mathematical results, see Lenze [11] and [12]. We consider the problem of characterizing the $\sigma \in C(\mathbb{R})$ for which the closure of the translates and dilates of $\sigma(x_1, ..., x_d)$ (the product of the x_i) span $C(\mathbb{R}^d)$. That is, we set

$$f(x_1, ..., x_d) = \sigma(x_1 \cdots x_d),$$

and then

$$\mathcal{M}_f = \overline{\operatorname{span}} \left\{ \sigma\left(\rho \prod_{i=1}^{d} (\cdot - b_i) \right) \colon \rho, b_1, ..., b_d \in \mathbb{R} \right\}.$$

The only sufficiently smooth functions σ for which $\mathcal{M}_f \neq C(\mathbb{R}^d)$ are polynomials. This may be proven directly from consideration of $D^{\mathbf{r}} \sigma(x_1 \cdots x_d) = 0$ which eventually translates into an Euler's equation. However there are other solutions in $C(\mathbb{R})$. We will prove this result using difference type operators, rather than via a differential equation.

THEOREM 10. *Let $d \geqslant 2$ and $\sigma \in C(\mathbb{R})$. Then*

$$f(x_1, ..., x_d) = \sigma(x_1 \cdots x_d)$$

is such that $\mathcal{M}_f \neq C(\mathbb{R}^d)$ if and only if

$$\sigma(t) = \sum_{j=0}^{r} b_{0_j} t^j + \sum_{j=1}^{r} \sum_{i=1}^{d-1} b_{ij} t^j (\ln |t|)^i \tag{6}$$

for some finite r and coefficients $\{b_{ij}\}$.

We will use the following result to prove Theorem 10.

PROPOSITION 11. *Let $g \in C(\mathbb{R} \backslash \{0\})$. For each $a \in \mathbb{R}$ and $k \in \mathbb{Z}_+$ set*

$$T_{k,a} g(t) = g(at) - a^k g(t).$$

Then

$$\prod_{k=0}^{r} (T_{k,a})^{\ell_k} g = 0$$

for all $a \in \mathbb{R} \backslash \{0\}$ if and only if

$$g(t) = \sum_{j=0}^{r} \sum_{i=0}^{\ell_{j-1}} b_{ij} t^j (\ln |t|)^i.$$

Remark 6. If $\ell_j = 0$, then $\sum_{j=0}^{\ell_j - 1}$ is understood to be empty.

Proof. $(\Leftarrow)$

$$\begin{aligned} T_{k,a}&(t^j (\ln |t|)^i) \\ &= (at)^j (\ln |at|)^i - a^k t^j (\ln |t|)^i = (at)^j (\ln |a| + \ln |t|)^i - a^k t^j (\ln |t|)^i \end{aligned}$$

$$= \begin{cases} \sum_{p=0}^{i} c_p t^j (\ln |t|)^p, & \text{if} \quad j \neq k, \\ \sum_{p=0}^{i-1} c_p t^j (\ln |t|)^p, & \text{if} \quad j = k, i \geqslant 1, \\ 0, & \text{if} \quad j = k, i = 0, \end{cases} \tag{7}$$

where $c_i \neq 0$ if $j \neq k$, and $c_{i-1} \neq 0$ if $j = k$ and $i \geqslant 1$. The coefficients c_p are easily calculated. They depend on a, i, j and k, but not on t.

From the above formula (7) it follows that

$$(T_{k,a})^{\ell_k}\left(\sum_{i=0}^{\ell_k-1} b_{ik}t^k(\ln|t|)^i\right)=0$$

for each k, and thus

$$\prod_{k=0}^{r}(T_{k,a})^{\ell_k}\left(\sum_{j=0}^{r}\sum_{i=0}^{\ell_j-1} b_{ij}t^j(\ln|t|)^i\right)=0.$$

($\Rightarrow$) This direction is most easily proved by induction. The essential idea is that of peeling off the $T_{k,a}$, one by one, and the following uniqueness result.

If $g\in C(\mathbb{R}\backslash\{0\})$, $k\in\mathbb{Z}_+$, and

$$T_{k,a}\,g=0$$

for all $a\in\mathbb{R}\backslash\{0\}$, then $g(t)=At^k$ for some constant A. To prove this let $g^*\in C(\mathbb{R}\backslash\{0\})$ satisfy

$$T_{k,a}\,g^*=0.$$

Then $\tilde{g}(t)=g^*(t)-g^*(1)\,t^k$ is in this same class, satisfies the equation

$$0=T_{k,a}\tilde{g}=\tilde{g}(at)-a^k\tilde{g}(t)$$

for all $a, t\in\mathbb{R}\backslash\{0\}$, and also $\tilde{g}(1)=0$. Setting $t=1$ in the above equation, it follows that $\tilde{g}(a)=0$ for all $a\in\mathbb{R}\backslash\{0\}$. That is $g^*(t)=At^k$.

We now prove the result by induction on $\sum_{j=1}^r \ell_j$. For $\sum_{j=1}^r \ell_j=1$ we just proved the result. Assume therefore that the result holds for $\sum_{j=1}^r \ell_j\leqslant t-1$, $(t\geqslant 2)$. Let $\sum_{j=1}^r \ell_j=t$. Choose s such that $\ell_s\geqslant 1$. Set

$$\tilde{\ell}_j=\begin{cases}\ell_j, & j\neq s\\ \ell_j-1, & j=s,\end{cases}$$

and $h=T_{s,a}\,g$.

If

$$\prod_{k=0}^{r}(T_{k,a})^{\ell_k}g=0,$$

then

$$\prod_{k=0}^{r}(T_{k,a})^{\tilde{\ell}_k}h=0.$$

From the induction hypothesis

$$h(t)=\sum_{j=0}^{r}\sum_{i=0}^{\ell_j-1} b_{ij}t^j(\ln|t|)^i.$$

That is,

$$(T_{s,a}g)(t)=\sum_{j=0}^{r}\sum_{i=0}^{\ell_j-1} b_{ij}t^j(\ln|t|)^i.$$

From (7) there exists a $\tilde{g}\in C(\mathbb{R}\backslash\{0\})$ of the form

$$\tilde{g}(t)=\sum_{j=0,j\neq s}^{r}\sum_{i=0}^{\ell_j-1} c_{ij}t^j(\ln|t|)^i+\sum_{i=0}^{\ell_s} c_{is}t^s(\ln|t|)^i,$$

where the c_{ij} depend on everything (including a) except t, such that

$$T_{s,a}g=T_{s,a}\tilde{g}.$$

Thus

$$T_{s,a}(g-\tilde{g})=0.$$

By the uniqueness result $(g-\tilde{g})(t)=At^s$. That is

$$g(t)=\tilde{g}(t)+At^s$$

which is the desired form. ∎

Remark 7. As has been pointed out by the referee, Proposition 11 also follows from the substitution $y=\log t$, $\alpha=\log a$, $G(y)=g(e^y)$ (for $a, t>0$) and solving the resulting difference equation.

We now return to the proof of Theorem 10.

Proof of Theorem 10. From (c) of Theorem 3 we have that $\mathcal{M}_f\neq C(\mathbb{R}^d)$ if and only if

$$\sigma(x_1\cdots x_d)=\sum_{i=1}^{d}\sum_{j=0}^{r} g_j(x_1,...,x_{i-1},x_{i+1},...,x_d)\,x_i^j. \tag{8}$$

(From the symmetry of the $\{x_i\}_{i=1}^d$ we may assume $r_i-1=r$ and $g_{ij}=g_j$ for each $i=1,...,d$.)

($\Leftarrow$) We will show that each of the $t^j(\ln|t|)^i$, $j=0,...,r$, $i=0,...,d-1$, may be written in the form (8), (except that the $(\ln|t|)^i$ are not in $C(\mathbb{R})$).

Let $h_{ij}(t)=t^j(\ln|t|)^i$, $0\leqslant i\leqslant d-1$. Then

$$\begin{aligned}h_{ij}(x_1\cdots x_d)&=(x_1\cdots x_d)^j\,(\ln|x_1\cdots x_d|)^i\\&=(x_1\cdots x_d)^j\,(\ln|x_1|+\cdots+\ln|x_d|)^i.\end{aligned}$$

The expression $(\ln|x_1|+\cdots+\ln|x_d|)^i$ may be written as a linear combination of terms of the form $\prod_{\ell=1}^{d}(\ln|x_\ell|)^{m_\ell}$ where the m_ℓ are non-negative integers which sum to $i\leqslant d-1$. Thus in each such term at least one of the m_ℓ equals 0, and hence each term depends on at most $d-1$ variables. It therefore follows that h_{ij} is of the form (8).

($\Rightarrow$) We claim that for σ of the form (8),

$$\prod_{k=0}^{r}(T_{k,a})^d\,\sigma=0, \tag{9}$$

for each $a\in\mathbb{R}\backslash\{0\}$. If (9) holds then the result follows from Proposition 11, where we must have $b_{i0}=0$ for $i\geqslant 1$, since $\sigma\in C(\mathbb{R})$ (while $(\ln|t|)^i$ is not continuous at $t=0$).

Note that $T_{k,a}\sigma(x_1\cdots x_d)=\sigma(ax_1\cdots x_d)-a^k\sigma(x_1\cdots x_d)$. We can consider the a in the first term on the right of the equality sign as belonging to any of the x_ℓ. That is, for $f(x_1,...,x_d)$ and $\ell\in\{1,...,d\}$, set

$$T_{k,a,\ell}f(x_1,...,x_d)=f(x_1,...,ax_\ell,...,x_d)-a^kf(x_1,...,x_d).$$

Then

$$\prod_{k=0}^{r}(T_{k,a})^d\,\sigma(x_1\cdots x_d)=\prod_{\ell=1}^{d}\prod_{k=0}^{r}(T_{k,a,\ell})\,\sigma(x_1\cdots x_d).$$

Now, as is readily checked,

$$\left(\prod_{k=0}^{r}(T_{k,a,\ell})\right)\left(\sum_{j=0}^{r}g_j(x_1,...,x_{\ell-1},x_{\ell+1},...,x_d)\,x_\ell^j\right)=0.$$

Since the operators $T_{k,a,\ell}$ commute, it therefore follows that

$$\prod_{k=0}^{r}(T_{k,a})^d\,\sigma=0,$$

which proves our theorem. ∎

It is an interesting and open question to determine, for a given polynomial p, necessary and sufficient conditions on $\sigma\in C(\mathbb{R})$ such that $\mathcal{M}_{\sigma(p)}=C(\mathbb{R}^d)$. We know that if σ is a polynomial, then $\mathcal{M}_{\sigma(p)}\neq C(\mathbb{R}^d)$, and we have seen an example of a polynomial p where $\sigma(t)=t^j(\ln|t|)^i$ is such that $\sigma(p(\cdot))$ also satisfies (c) of Theorem 3. In fact, if p_k is a polynomial of $d-1$ variables, $k=1, ..., d$, and

$$p(x_1, ..., x_d)=\prod_{k=1}^{d} p_k(x_1, ..., x_{k-1}, x_{k+1}, ..., x_d)$$

then $\sigma(t)=t^j(\ln|t|)^i$ is such that $\sigma(p(\cdot))$ again satisfies (c) of Theorem 3 for $j\geqslant 1$ and $i\in\{0, ..., d-1\}$. It is also possible to construct polynomials p for which $\sigma(t)=t^\alpha$ is such that $\sigma(p(\cdot))$ satisfies (c) of Theorem 3 for any specified rational α. It would be nice to know which possible functions $\sigma\in C(\mathbb{R})$ are such that $\mathcal{M}_{\sigma(p)}\neq C(\mathbb{R}^d)$ for some polynomial p, and also to determine conditions on a polynomial p such that $\mathcal{M}_{\sigma(p)}\neq C(\mathbb{R}^d)$ holds only if σ is a polynomial.

EXAMPLE 3. A different mathematical formulation which addresses the same basic neural network model leads to the following problem, see e.g., Park, Sandberg [17]. Let $k\in C(\mathbb{R}_+)$, and

$$\mathcal{H}_k=\overline{\text{span}}\{k(\rho\,\|\cdot-\mathbf{a}\|)\colon \mathbf{a}\in\mathbb{R}^d, \rho>0\}$$

where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^d$. What are conditions on k such that $\mathcal{H}_k=C(\mathbb{R}^d)$? If $k(\|\cdot\|)$ is also in $L^1(\mathbb{R}^d)$, then standard arguments imply that $\mathcal{H}_k=C(\mathbb{R}^d)$ (see Remark 3). However this is not in the least necessary. A different condition is that k not be an even polynomial. For if k was an even polynomial of degree $2n$, then $k(\|\cdot-\mathbf{a}\|)$ would be a polynomial (in $\mathbb{R}^d$) of degree at most $2n$ for each $\mathbf{a}$. Even polynomials are exactly all the functions for which $\mathcal{H}_k\neq C(\mathbb{R}^d)$.

THEOREM 12. *Let $k\in C(\mathbb{R}_+)$. Then $\mathcal{H}_k\neq C(\mathbb{R}^d)$ if and only if k is an even polynomial.*

Proof. We first note that $\mathcal{H}_k$ is a rotation invariant space of the form $\mathcal{G}_f$ where

$$f(x_1, ..., x_d)=k((x_1^2+\cdots+x_d^2)^{1/2}).$$

Thus it follows from Proposition 2 that $\mathcal{H}_k\neq C(\mathbb{R}^d)$ if and only if

$$\Delta^m k=0 \tag{10}$$

in the weak sense, for some positive integer m. Now, since k is itself rotation invariant, (10) reduces to

$$\left(\frac{\partial^2}{\partial r^2}+\frac{d-1}{r}\frac{\partial}{\partial r}\right)^m k(r)=0 \tag{11}$$

for $r \geqslant 0$, in the weak sense, where $r = (x_1^2 + \cdots + x_d^2)^{1/2}$. On $(0, \infty)$ every weak solution to this equation is a classical solution.

A simple calculation shows that the $2m$ linearly independent solutions of (11) on $(0, \infty)$ are dependent on m, d, and the parity of d. For d odd, they are the functions in the span of

$$1, r^2, ..., r^{2(m-1)}, \frac{1}{r^{d-2}}, ..., \frac{1}{r^{d-2m}}.$$

For d even, they are given by the span of the m functions

$$1, r^2, ..., r^{2(m-1)},$$

and the first m functions in the sequence

$$\frac{1}{r^{d-2}}, ..., \frac{1}{r^2}, \ln r, r^2 \ln r, r^4 \ln r, \ldots$$

We now return to (10). It is known that every weak solution of this equation is in fact a C^∞ function. This is sometimes called Weyl's lemma, see Hörmander [9, Theorem 4.4.3] or Hörmander [10, Chap. XI]. But the only C^∞ functions in $(x_1, ..., x_d)$ among those listed above are the even polynomials. This proves the theorem. ∎

Acknowledgment

We thank Professors Atzmon, Brudnyi, Dyn'kin, Elias, Lenze, and Lin and the referee for their help and suggestions.

References

1. A. B. Alexsandrov and P. P. Kargaev, Hardy classes of functions harmonic in a half space, *Algebra i Anal.* **5** (1993), 1–73; *St. Petersburg Math. J.* **5** (1994), 229–286.
2. A. Atzmon, Uniform approximation by linear combinations of translations and dilations of a function, *J. London Math. Soc.* **27** (1983), 51–54.
3. L. Brown, M. Schreiber, and B. A. Taylor, Spectral synthesis and the Pompeiu problem, *Ann. Inst. Fourier, Grenoble* **23** (1973), 125–154.
4. Ju. A. Brudnyi, Approximation of functions of n variables by quasipolynomials, *Izv. Akad. Nauk SSSR* **34** (1970), 564–583; *Math. USSR-Izv.* **4** (1970), 568–586.
5. D. I. Gurevich, Counterexamples to a problem of L. Schwartz, *Funct. Anal. Appl.* **197** (1975), 116–120.
6. S. R. Harasymiv, A note of dilations in L^p, *Pacific J. Math.* **21** (1967), 493–501.
7. S. R. Harasymiv, A note on approximation by automorphic images of functions in L^1 and L^∞, *J. Austral. Math. Soc.* **8** (1968), 222–230.

8. S. R. Harasymiv, Dilations of rapidly decreasing functions, *Pacific J. Math.* **31** (1969), 395–402.
9. L. Hörmander, "The Analysis of Linear Partial Differential Operators, I," Springer-Verlag, Berlin, 1983.
10. L. Hörmander, "The Analysis of Linear Partial Differential Operators, II," Springer-Verlag, Berlin, 1983.
11. B. Lenze, Constructive multivariate approximation via sigmoidal functions with application to neural networks, *in* "Numerical Methods of Approximation Theory" (D. Braess and L. L. Schumaker, Eds.), pp. 155–175, ISNM 105, Birkhäuser, Basel, 1992.
12. B. Lenze, Quantitative approximation results for sigma-pi-type neural network operators, *in* "Multivariate Approximation: From CAGD to Wavelets" (K. Jetter and F. Utreras, Eds.), pp. 193–209, World Scientific, Singapore, 1993.
13. M. Leshno, Ya. V. Lin, A. Pinkus, and S. Schocken, Multilayer feedforward networks with a non-polynomial activation function can approximate any function, *Neural Networks* **6** (1993), 861–867.
14. V. Ya. Lin and A. Pinkus, Fundamentality of ridge functions, *J. Approx. Theory* **75** (1993), 295–311.
15. H. N. Mhaskar and C. A. Micchelli, Approximation by superposition of a sigmoidal function and radial basis functions, *Adv. in Appl. Math.* **13** (1992), 350–373.
16. R. S. Palais, Some analogues of Hartog's theorem in an algebraic setting, *Amer. J. Math.* **100** (1978), 387–405.
17. J. Park and I. W. Sandberg, Universal approximation using radial-basis-function networks, *Neural Comp.* **3** (1991), 246–257.
18. L. Schwartz, Sur certaines familles non fondamentales de fonctions continues, *Bull. Soc. Math. France* **72** (1944), 141–145.
19. L. Schwartz, Théorie générale des fonctions moyenne-périodiques, *Ann. of Math.* **48** (1947), 857–928.
20. Y. Sternfeld and Y. Weit, Affine invariant subspaces of $C(\mathbb{C})$, *Proc. Amer. Math. Soc.* **107** (1989), 231–236.
21. A. I. Vaindiner, Approximation of continuous and differentiable functions of several variables by generalized polynomials (Finite linear combinations of functions of fewer variables), *Dokl. Akad. Nauk.* **192** (1970), 483–486; *Soviet. Math. Dokl.* **11** (1970), 648–652.