Available online at www.sciencedirect.com

ScienceDirect

Indagationes Mathematicae 24 (2013) 725–738

indagationes mathematicae

www.elsevier.com/locate/indag

Smoothness and uniqueness in ridge function representation

A. Pinkus

Department of Mathematics, Technion, Haifa, Israel

Abstract

In this note we consider problems of uniqueness, smoothness and representation of linear combinations of a finite number of ridge functions with fixed directions.

Keywords: Ridge functions; Smoothness; Uniqueness

1. Introduction

A *ridge function*, in its simplest format, is a multivariate function of the form

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

defined for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, where $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ is a fixed non-zero vector, called a *direction*, $\mathbf{a} \cdot \mathbf{x} = \sum_{j=1}^{n} a_j x_j$ is the usual inner product, and f is a real-valued function defined on $\mathbb{R}$. Note that

$$f(\mathbf{a} \cdot \mathbf{x})$$

is constant on the hyperplanes $\{\mathbf{x} : \mathbf{a} \cdot \mathbf{x} = c\}$. Ridge functions are relatively simple multivariate functions. Ridge functions (formerly known as *plane waves*) were so-named in 1975 by Logan and Shepp [11]. They appear in various areas and under numerous guises.

In this note we consider problems of uniqueness, smoothness and representation of linear combinations of a finite number of ridge functions. That is, assume we are given a function F of

E-mail address: pinkus@technion.ac.il.

doi:10.1016/j.indag.2012.10.004

the form

$$F(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{a}^i \cdot \mathbf{x}), \tag{1.1}$$

where m is finite, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. We ask and answer the following questions. If F is of a certain smoothness class, what can we say about the smoothness of the f_i? How many different ways can we write F as a linear combination of a finite number of ridge functions, i.e., to what extent is a representation of F in the form (1.1) unique? And, finally, which other ridge functions $f(\mathbf{a} \cdot \mathbf{x})$ can be written in the form (1.1) with $\mathbf{a} \neq \alpha \mathbf{a}^i$, for any $\alpha \in \mathbb{R}$ and $i = 1, \ldots, m$?

In Section 4 we generalize the main results of this paper to finite linear combinations of functions of the form

$$f(A\mathbf{x})$$

where A is a fixed $d \times n$ matrix, $1 \leq d < n$, and f is a real-valued function defined on $\mathbb{R}^d$. For $d = 1$, this reduces to a ridge function.

2. Smoothness

Let $C^k(\mathbb{R}^n)$, $k \in \mathbb{Z}_+$, denote the usual set of real-valued functions with all derivatives of order up to and including k being continuous. Assume $F \in C^k(\mathbb{R}^n)$ is of the form (1.1). What does this imply, if anything, about the smoothness of the f_i? In the case $m = 1$ there is nothing to prove. That is, if

$$F(\mathbf{x}) = f_1(\mathbf{a}^1 \cdot \mathbf{x})$$

is in $C^k(\mathbb{R}^n)$ for some $\mathbf{a}^1 \neq \mathbf{0}$, then obviously $f_1 \in C^k(\mathbb{R})$. This same result holds when $m = 2$. As the $\mathbf{a}^1$ and $\mathbf{a}^2$ are linearly independent, there exists a vector $\mathbf{c} \in \mathbb{R}^n$ satisfying $\mathbf{a}^1 \cdot \mathbf{c} = 0$ and $\mathbf{a}^2 \cdot \mathbf{c} = 1$. Thus

$$F(t\mathbf{c}) = f_1(\mathbf{a}^1 \cdot t\mathbf{c}) + f_2(\mathbf{a}^2 \cdot t\mathbf{c}) = f_1(0) + f_2(t).$$

As $F(t\mathbf{c})$ is in $C^k(\mathbb{R})$, as a function of t, so is f_2. The same result holds for f_1.

However this result is no longer valid when $m \geq 3$, without some assumptions on the f_i. To see this, let us recall that the Cauchy Functional Equation

$$g(x + y) = g(x) + g(y) \tag{2.1}$$

has, as proved by Hamel [8] in 1905, very badly behaved solutions; see e.g., Aczél [1] for a discussion of the solutions of this equation. As such, setting $f_1 = f_2 = -f_3 = g$, we have very badly behaved (and certainly not in $C^k(\mathbb{R})$) f_i, $i = 1, 2, 3$, that satisfy

$$0 = f_1(x_1) + f_2(x_2) + f_3(x_1 + x_2)$$

for all $(x_1, x_2) \in \mathbb{R}^2$. That is, the very smooth function on the left-side of this equation is a sum of three unruly ridge functions. As shall shortly become evident, this Cauchy Functional Equation is critical in the analysis of our problem for all $m \geq 3$.

It was proved by Buhmann and Pinkus [2] that if $F \in C^k(\mathbb{R}^n)$, and if $f_i \in L^1_{\text{loc}}(\mathbb{R})$ for each i, then $f_i \in C^k(\mathbb{R})$ for each i, if $k \geq m - 1$. The method of proof therein used smoothing and generalized functions. In this note we remove the restriction $k \geq m - 1$, have different assumptions on the f_i, and provide an alternative, and we believe, much more natural and elementary approach to this problem.

For ease of exposition, let us denote by $\mathcal{B}$ any class of real-valued functions f defined on $\mathbb{R}$ such that if there is a function $r \in C(\mathbb{R})$ such that $f - r$ satisfies the Cauchy Functional Equation (2.1), then $f - r$ is necessarily linear, i.e., $(f - r)(x) = cx$ for some constant c, and all $x \in \mathbb{R}$. $\mathcal{B}$ includes, for example, the set of all functions that are continuous at a point, or monotonic on an interval, or bounded on one side on a set of positive measure, or Lebesgue measurable; again see e.g., Aczél [1].

Theorem 2.1. *Assume $F \in C^k(\mathbb{R}^n)$ is of the form (1.1), i.e.,*

$$F(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{a}^i \cdot \mathbf{x}),$$

where m is finite, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. Assume, in addition, that each $f_i \in \mathcal{B}$. Then, necessarily, $f_i \in C^k(\mathbb{R})$ for $i = 1, \ldots, m$.

Proof. The proof will be by induction on m. As we have seen, this result is valid when $m = 1$. Let $\mathbf{c} \in \mathbb{R}^n$ satisfy $(\mathbf{c} \cdot \mathbf{a}^m) = 0$ and $(\mathbf{c} \cdot \mathbf{a}^i) = b_i \neq 0$ for $i = 1, \ldots, m-1$. Such a $\mathbf{c}$ exists. Now

$$F(\mathbf{x} + t\mathbf{c}) - F(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{a}^i \cdot \mathbf{x} + t\mathbf{a}^i \cdot \mathbf{c}) - f_i(\mathbf{a}^i \cdot \mathbf{x}).$$

By construction we have $f_m(\mathbf{a}^m \cdot \mathbf{x} + t\mathbf{a}^m \cdot \mathbf{c}) - f_m(\mathbf{a}^m \cdot \mathbf{x}) = f_m(\mathbf{a}^m \cdot \mathbf{x}) - f_m(\mathbf{a}^m \cdot \mathbf{x}) = 0$, while $f_i(\mathbf{a}^i \cdot \mathbf{x} + t\mathbf{a}^i \cdot \mathbf{c}) - f_i(\mathbf{a}^i \cdot \mathbf{x}) = f_i(\mathbf{a}^i \cdot \mathbf{x} + tb_i) - f_i(\mathbf{a}^i \cdot \mathbf{x})$ for $i = 1, \ldots, m-1$. Thus

$$H(\mathbf{x}) := F(\mathbf{x} + t\mathbf{c}) - F(\mathbf{x}) = \sum_{i=1}^{m-1} h_i(\mathbf{a}^i \cdot \mathbf{x})$$

where $h_i(y) = f_i(y + tb_i) - f_i(y)$. Since $H \in C^k(\mathbb{R}^n)$, it follows by our induction assumption that $h_i \in C^k(\mathbb{R})$. Note that this is valid for each and every $t \in \mathbb{R}$.

We have therefore reduced our problem to the following. Assume $b \neq 0$, and for each $t \in \mathbb{R}$ the function h, defined by

$$h(y) = f(y + tb) - f(y),$$

is in $C^k(\mathbb{R})$. When does this imply that $f \in C^k(\mathbb{R})$? A detailed answer is contained in the paper by de Bruijn [4]. What is proved therein is that if $h \in C^k(\mathbb{R})$, then f is necessarily of the form $f = r + s$ where $r \in C^k(\mathbb{R})$ and s satisfies the Cauchy Functional Equation (2.1). Thus each f_i is of the form $f_i = r_i + s_i$, with r_i and s_i as above. By our assumption, each f_i is in $\mathcal{B}$, and from the definition of $\mathcal{B}$ it follows that $f_i - r_i = s_i$ is a linear function, i.e., $s_i(t) = c_i t$ for some constant c_i. Thus $f_i = r_i + s_i$, where both $r_i, s_i \in C^k(\mathbb{R})$, implying that $f_i \in C^k(\mathbb{R})$. This is valid for $i = 1, \ldots, m-1$, and hence also for $i = m$. $\square$

Remark 2.1. In Theorem 2.1 it actually suffices to only assume that $m - 2$ of the functions $\{f_i\}_{i=1}^m$ are in $\mathcal{B}$. To see this, assume $f_1, \ldots, f_{m-2}$ are in $\mathcal{B}$. From the above proof it follows that $f_1, \ldots, f_{m-2} \in C^k(\mathbb{R})$. Thus

$$G(\mathbf{x}) := F(\mathbf{x}) - \sum_{i=1}^{m-2} f_i(\mathbf{a}^i \cdot \mathbf{x}) = f_{m-1}(\mathbf{a}^{m-1} \cdot \mathbf{x}) + f_{m-2}(\mathbf{a}^{m-2} \cdot \mathbf{x})$$

is a function in $C^k(\mathbb{R}^n)$. We now apply the reasoning in the case $m = 2$ that appeared prior to the statement of Theorem 2.1.

Remark 2.2. In de Bruijn [4,5], there are delineated various classes of real-valued functions $\mathcal{D}$ with the property that if

$$h_t := \Delta_t f = f(\cdot + t) - f(\cdot) \in \mathcal{D}$$

for all $t \in \mathbb{R}$, then f is necessarily of the form $f = r + s$, where $r \in \mathcal{D}$ and s satisfies the Cauchy Functional Equation. Some of these classes $\mathcal{D}$ are: $C^k(\mathbb{R})$, functions with k continuous derivatives; $\widetilde{C}^k(\mathbb{R})$, functions that are k times differentiable (but their kth derivative need not be continuous); $C^\infty(\mathbb{R})$ functions; analytic functions; functions which are absolutely continuous on any finite interval; functions having bounded variation over any finite interval; algebraic polynomials; trigonometric polynomials; and Riemann integrable functions. Theorem 2.1 can be suitably restated for any of these classes $\mathcal{D}$.

3. Uniqueness and representation

In this section we discuss the question of the uniqueness of the representation (1.1). We ask when, and for which functions $\{g_i\}_{i=1}^k$ and $\{h_i\}_{i=1}^\ell$, we can have

$$F(\mathbf{x}) = \sum_{i=1}^{k} g_i(\mathbf{b}^i \cdot \mathbf{x}) = \sum_{j=1}^{\ell} h_i(\mathbf{c}^i \cdot \mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^n$, where k and ℓ are finite, and the $\mathbf{b}^1, \ldots, \mathbf{b}^k, \mathbf{c}^1, \ldots, \mathbf{c}^\ell$ are $k + \ell$ pairwise linearly independent vectors in $\mathbb{R}^n$? From linearity this is, of course, equivalent to the following. Assume

$$\sum_{i=1}^{m} f_i(\mathbf{a}^i \cdot \mathbf{x}) = 0 \tag{3.1}$$

for all $\mathbf{x} \in \mathbb{R}^n$, where m is finite, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. What does this imply regarding the f_i? We first prove that with minimal requirements the f_i must be polynomials of degree $\leq m - 2$. This generalizes a result of Buhmann and Pinkus [2]; see also Falconer [7] and Petersen et al. [12]. We will later extend this result.

Let Π_r^n denote the set of polynomials of *total degree* at most r in n variables. That is,

$$\Pi_r^n = \left\{ \sum_{|\mathbf{k}| \leq r} b_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} \right\}$$

where we use standard multi-index notation, i.e., $b_{\mathbf{k}} \in \mathbb{R}$, $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{Z}_+^n$, $|\mathbf{k}| = k_1 + \cdots + k_n$, and $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} \cdots x_n^{k_n}$.

Proposition 3.1. *Assume (3.1) holds where m is finite, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. Assume, in addition, that $f_i \in \mathcal{B}$, for $i = 1, \ldots, m$. Then $f_i \in \Pi_{m-2}^1$, $i = 1, \ldots, m$.*

For each $\mathbf{c} \in \mathbb{R}^n$ let

$$D_{\mathbf{c}} = \sum_{i=1}^{n} c_i \frac{\partial}{\partial x_i}$$

denote directional differentiation in the direction $\mathbf{c}$. Assume $f \in C^1(\mathbb{R})$. When considering ridge functions, the following simple formula is fundamental

$$D_{\mathbf{c}} f(\mathbf{a} \cdot \mathbf{x}) = (\mathbf{a} \cdot \mathbf{c}) f'(\mathbf{a} \cdot \mathbf{x}).$$

Using this formula we easily prove Proposition 3.1.

Proof. For $m = 1$, the result is obviously true, where we define $\Pi^1_{-1} = \{0\}$.

From de Bruijn [4] and the method of proof of Theorem 2.1, it follows that each f_i is a polynomial (see Remark 2.2). In fact we only need the sufficient smoothness of each f_i which is a direct consequence of Theorem 2.1. We now apply an elementary argument using directional derivatives as may be found, for example, in Diaconis and Shahshahani [6]; see also Buhmann and Pinkus [2].

Fix $t \in \{1, \ldots, m\}$. For each $j \in \{1, \ldots, m\}$, $j \neq t$, let $\mathbf{c}^j \in \mathbb{R}^n$ satisfy

$$\mathbf{c}^j \cdot \mathbf{a}^j = 0 \quad \text{and} \quad \mathbf{c}^j \cdot \mathbf{a}^t \neq 0.$$

This is possible since the $\mathbf{a}^i$ are pairwise linearly independent. Now, as each f_i is sufficiently smooth,

$$\begin{aligned} 0 &= \prod_{\substack{j=1 \\ j \neq t}}^m D_{\mathbf{c}^j} \sum_{i=1}^m f_i(\mathbf{a}^i \cdot \mathbf{x}) \\ &= \sum_{i=1}^m \left(\prod_{\substack{j=1 \\ j \neq t}}^m (\mathbf{c}^j \cdot \mathbf{a}^i) \right) f_i^{(m-1)}(\mathbf{a}^i \cdot \mathbf{x}) \\ &= \prod_{\substack{j=1 \\ j \neq t}}^m (\mathbf{c}^j \cdot \mathbf{a}^t) f_t^{(m-1)}(\mathbf{a}^t \cdot \mathbf{x}). \end{aligned}$$

Note that $\prod_{\substack{j=1 \\ j \neq t}}^m (\mathbf{c}^j \cdot \mathbf{a}^t) \neq 0$. Thus

$$f_t^{(m-1)}(\mathbf{a}^t \cdot \mathbf{x}) = 0$$

for all $\mathbf{x} \in \mathbb{R}^n$. Therefore

$$f_t^{(m-1)}(y) = 0$$

for all $y \in \mathbb{R}$ and f_t is a polynomial of degree at most $m - 2$. □

By the same method of proof we in fact have the following.

Corollary 3.2. *Assume $F \in \Pi^n_r$ has the form*

$$F(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{a}^i \cdot \mathbf{x}),$$

where m is finite, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. Assume, in addition, that $f_i \in \mathcal{B}$, for $i = 1, \ldots, m$. Then $f_i \in \Pi^1_s$, $i = 1, \ldots, m$, where $s = \max\{r, m - 2\}$.

One immediate consequence of Proposition 3.1 is the following which easily follows by taking $f_i = f$ for $i = 1, \ldots, m$.

Proposition 3.3. *Assume $f \in \mathcal{B}$ and f is not a polynomial. Then for any finite m, and pairwise linearly independent vectors $\mathbf{a}^1, \ldots, \mathbf{a}^m$ in $\mathbb{R}^n \setminus \{\mathbf{0}\}$, the functions*

$$\{f(\mathbf{a}^1 \cdot \mathbf{x}), \ldots, f(\mathbf{a}^m \cdot \mathbf{x})\}$$

are linearly independent.

Proposition 3.3 is a generalization of a result by Dahmen and Micchelli [3], where they prove, by different methods, that if the dimension of the span of the space $\{f(\mathbf{a} \cdot \mathbf{x}) : \mathbf{a} \in \mathbb{R}^n\}$ is finite, and f is Lebesgue measurable, then f is a polynomial.

Remark 3.1. It is often the case that the result of Proposition 3.1 can be obtained with $f_i \in \Pi_k^1$, where k is significantly less than $m-2$. Recall that we took, for each $t \in \{1, \ldots, m\}$, a collection of $m-1$ vectors $\mathbf{c}^j \in \mathbb{R}^n$, $j \in \{1, \ldots, m\}$, $j \neq t$, such that $\mathbf{c}^j \cdot \mathbf{a}^j = 0$, $\mathbf{c}^j \cdot \mathbf{a}^t \neq 0$, for $j \neq t$, so that

$$\prod_{\substack{j=1 \\ j \neq t}}^{m} (\mathbf{c}^j \cdot \mathbf{a}^i) = 0$$

for all $i \neq t$. This then implied that $f_t^{(m-1)} = 0$, whence f_t is a polynomial of degree at most $m-2$. If the $\mathbf{a}^j$ are in generic position, i.e., any n of them are linearly independent, then we can take $\mathbf{c}$ orthogonal to any $n-1$ of the $\mathbf{a}^j$, $j \neq t$, satisfying $\mathbf{c} \cdot \mathbf{a}^t \neq 0$. In this case we will only need $[(m-2)/(n-1)] + 1$ vectors $\mathbf{c}$ to obtain the same desired result, and thus each f_t must be a polynomial of degree at most $[(m-2)/(n-1)]$. However as the $\mathbf{a}^j$ are only pairwise linearly independent, they can all lie in a subspace of dimension 2, and if this is the case (which is the same as taking $n = 2$) then we do need $m-1$ $\mathbf{c}^j$'s in the above proof. Moreover this is not just an artifact of the method of proof. For each m there exist pairwise distinct $\mathbf{a}^i \in \mathbb{R}^n$, $i = 1, \ldots, m$, and polynomials f_i of exact degree $m-2$ such that $\sum_{i=1}^m f_i(\mathbf{a}^i \cdot \mathbf{x}) = 0$. To see that this holds, simply consider pairwise linearly independent $\mathbf{a}^i$ of the form $\mathbf{a}^i = (a_1^i, a_2^i, 0, \ldots, 0)$, $i = 1, \ldots, m$. The polynomials $(\mathbf{a}^i \cdot \mathbf{x})^{m-2}$, $i = 1, \ldots, m$, are homogeneous of degree $m-2$. The space of homogeneous polynomials of degree $m-2$ in two variables has dimension $m-1$. Thus some non-trivial linear combination of these $(\mathbf{a}^i \cdot \mathbf{x})^{m-2}$, $i = 1, \ldots, m$, vanishes identically.

We will consider this uniqueness result in more detail. What more can we say regarding the polynomials f_i satisfying (3.1)?

To this end, let H_r^n denote the set of homogeneous polynomials of degree r in n variables, i.e.,

$$H_r^n = \left\{ \sum_{|\mathbf{k}|=r} b_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} \right\}.$$

Then we have the following.

Proposition 3.4. *Assume m is finite, $f, f_i \in \mathcal{B}$, $i = 1, \ldots, m-1$, the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$, and $\mathbf{a} \neq \alpha \mathbf{a}^i$ for any $\alpha \in \mathbb{R}$ and $i \in \{1, \ldots, m-1\}$. Then we have*

$$f(\mathbf{a} \cdot \mathbf{x}) = \sum_{i=1}^{m-1} f_i(\mathbf{a}^i \cdot \mathbf{x}) \tag{3.2}$$

if and only if f is a polynomial of exact degree r and for every $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$, $i = 1, \ldots, m-1$, we have $q(\mathbf{a}) = 0$.

Remark 3.2. Eq. (3.2) is, of course, a rewrite of (3.1) where $f(\mathbf{a}\cdot\mathbf{x}) = -f_m(\mathbf{a}^m\cdot\mathbf{x})$. Thus we necessarily have $r \le m-2$. However this automatically follows from the statement of Proposition 3.4 since, for $r \ge m-1$, there always exists a $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$, $i = 1,\ldots,m-1$, and $q(\mathbf{a}) \neq 0$. Namely, choose $\mathbf{c}^i \in \mathbb{R}^n$, $i = 1,\ldots,m-1$, satisfying $\mathbf{c}^i\cdot\mathbf{a}^i = 0$ and $\mathbf{c}^i\cdot\mathbf{a} \neq 0$, and set $q(\mathbf{x}) = p(\mathbf{x})\prod_{i=1}^{m-1}(\mathbf{c}^i\cdot\mathbf{x})$, where $p \in H_{r-m+1}^n$ satisfies $p(\mathbf{a}) \neq 0$.

Proof. Assume (3.2) holds. Then from Proposition 3.1 it follows that $f, f_i \in \Pi_{m-2}^1$, $i = 1,\ldots,m-1$. Let

$$f(t) = \sum_{j=0}^{r} d_j t^j, \quad d_r \neq 0,$$

and

$$f_i(t) = \sum_{j=0}^{r_i} d_{ij} t^j, \quad i = 1,\ldots,m-1,$$

where $r, r_i \le m-2$. We rewrite (3.2) as

$$\sum_{j=0}^{r} d_j(\mathbf{a}\cdot\mathbf{x})^j = \sum_{i=1}^{m-1}\sum_{j=0}^{r_i} d_{ij}(\mathbf{a}^i\cdot\mathbf{x})^j.$$

A polynomial is identically zero if and only if each of its homogeneous components is zero. Thus

$$d_r(\mathbf{a}\cdot\mathbf{x})^r = \sum_{i=1}^{m-1} d_{ir}(\mathbf{a}^i\cdot\mathbf{x})^r, \tag{3.3}$$

where we set $d_{ir} = 0$ if $r > r_i$.

We claim that (3.3) can hold if and only if for every polynomial $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$, for i such that $d_{ir} \neq 0$, we have $q(\mathbf{a}) = 0$. To prove this fact we use a variant of an argument in Lin and Pinkus [10].

For $\mathbf{k} \in \mathbb{Z}_+^n$, set

$$D^{\mathbf{k}} = \frac{\partial^{|\mathbf{k}|}}{\partial x_1^{k_1}\cdots\partial x_n^{k_n}}.$$

Given $q \in H_r^n$,

$$q(\mathbf{x}) = \sum_{|\mathbf{k}|=r} b_{\mathbf{k}}\mathbf{x}^{\mathbf{k}},$$

set

$$q(D) = \sum_{|\mathbf{k}|=r} b_{\mathbf{k}} D^{\mathbf{k}}.$$

When $\mathbf{k} \in \mathbb{Z}_+^n$, $|\mathbf{k}| = r$, a simple calculation shows that

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

Thus, for $q \in H_r^n$ we have

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

Given $\mathbf{k}^1, \mathbf{k}^2 \in \mathbb{Z}_+^n$, $|\mathbf{k}^1| = |\mathbf{k}^2| = r$, we have

$$D^{\mathbf{k}^1}\mathbf{x}^{\mathbf{k}^2} = \delta_{\mathbf{k}^1,\mathbf{k}^2} k_1^1! \cdots k_n^1!.$$

This implies that every non-trivial linear functional ℓ on the finite-dimensional linear space H_r^n may be represented by some $q \in H_r^n$ via

$$\ell(p) = q(D)p$$

for each $p \in H_r^n$. Now

$$(\mathbf{a} \cdot \mathbf{x})^r \in \operatorname{span}\{(\mathbf{a}^i \cdot \mathbf{x})^r : d_{ir} \neq 0\}$$

if and only if every linear functional on H_r^n that annihilates the $(\mathbf{a}^i \cdot \mathbf{x})^r$, where $d_{ir} \neq 0$, i.e., every $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$ for all i with $d_{ir} \neq 0$, also annihilates $(\mathbf{a} \cdot \mathbf{x})^r$, i.e., satisfies $q(\mathbf{a}) = 0$. Thus, if (3.2) holds, then for every $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$, $i = 1, \ldots, m$, we have $q(\mathbf{a}) = 0$.

Assume the converse, i.e., f is a polynomial of degree r and for every $q \in H_r^n$ satisfying $q(\mathbf{a}^i) = 0$, $i = 1, \ldots, m-1$, we have $q(\mathbf{a}) = 0$. Let

$$f(t) = \sum_{j=0}^{r} d_j t^j, \quad d_r \neq 0.$$

By the above argument we have that

$$d_r(\mathbf{a} \cdot \mathbf{x})^r = \sum_{i=1}^{m-1} d_{ir}(\mathbf{a}^i \cdot \mathbf{x})^r,$$

for some set of $\{d_{ir}\}_{i=1}^{m-1}$. We claim that a similar formula holds for all other (lower) powers. This follows by suitable differentiation. Alternatively, based on the above, assume that for some $s < r$ we have a $\widetilde{q} \in H_s^n$ satisfying $\widetilde{q}(\mathbf{a}^i) = 0$, $i = 1, \ldots, m-1$, and $\widetilde{q}(\mathbf{a}) \neq 0$. Choose any $p \in H_{r-s}^n$ such that $p(\mathbf{a}) \neq 0$. Then $q = p\widetilde{q} \in H_r^n$ satisfies $q(\mathbf{a}^i) = 0$, $i = 1, \ldots, m-1$, and $q(\mathbf{a}) \neq 0$, contradicting our assumptions. Thus, for each $j = 0, 1, \ldots, r$, we have

$$d_j(\mathbf{a} \cdot \mathbf{x})^j = \sum_{i=1}^{m-1} d_{ij}(\mathbf{a}^i \cdot \mathbf{x})^j,$$

for some set of $\{d_{ij}\}$, proving that (3.2) holds. $\square$

Based on Proposition 3.4 we can now present a strengthened version of Proposition 3.1.

Corollary 3.5. *Assume that m is finite, $f_i \in \mathcal{B}$, $i = 1, \ldots, m$, and the $\mathbf{a}^i$ are pairwise linearly independent vectors in $\mathbb{R}^n$. Then we have*

$$\sum_{i=1}^{m} f_i(\mathbf{a}^i \cdot \mathbf{x}) = 0$$

if and only if for each i, f_i is a polynomial of exact degree r_i and if $q_i \in H_{r_i}^n$ satisfies $q_i(\mathbf{a}^j) = 0$, $j \in \{1, \ldots, m\} \setminus \{i\}$, then $q_i(\mathbf{a}^i) = 0$.

4. Smoothness and uniqueness in the multivariate form

One possible generalization of a ridge function is to a multivariate function of the form

$$f(A\mathbf{x})$$

defined for $\mathbf{x} \in \mathbb{R}^n$, where A is a fixed $d \times n$ matrix, $1 \le d < n$, and f is a real-valued function defined on $\mathbb{R}^d$. For $d = 1$, this reduces to a ridge function.

As previously, assume we are given a function F of the form

$$F(\mathbf{x}) = \sum_{i=1}^{m} f_i(A^i\mathbf{x}), \tag{4.1}$$

where m is finite, the A^i are $d \times n$ matrices, for some fixed d, $1 \le d < n$, and each $f_i : \mathbb{R}^d \to \mathbb{R}$. (In fact we could also consider A^i with different numbers of rows. The analysis would be much the same.) We again ask what the smoothness of F implies regarding the smoothness of the f_i.

The situation here is slightly more problematic, as redundancies can easily occur. Consider, for example, when $n = 3$, $m = 2$, $d = 2$, and

$$A^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad A^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus

$$F(x_1, x_2, x_3) = f_1(x_1, x_2) + f_2(x_2, x_3).$$

Setting $f_1(x_1, x_2) = g(x_2)$ and $f_2(x_2, x_3) = -g(x_2)$ for any arbitrary univariate function g, we have

$$0 = f_1(x_1, x_2) + f_2(x_2, x_3),$$

and yet f_1 and f_2 do not exhibit any of the smoothness or polynomial properties of the left-hand-side of this equation.

This simple example generalizes as follows. For convenience we will, in what follows, always assume that the A^i are of full rank d.

Proposition 4.1. *Assume there exist $i, j \in \{1, \ldots, m\}$, $i \ne j$, such that the $2d$ rows of A^i and A^j are linearly dependent. Then there exist non-smooth functions f_i and f_j such that*

$$f_i(A^i\mathbf{x}) + f_j(A^j\mathbf{x}) = 0$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. Since the $2d$ rows of A^i and A^j are linearly dependent and, in addition, A^i, A^j are of full rank d, there exist $\mathbf{c}^i, \mathbf{c}^j \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ such that

$$\mathbf{c}^i A^i = \mathbf{c}^j A^j \ne \mathbf{0}.$$

Thus for all $\mathbf{x} \in \mathbb{R}^n$, and any arbitrary non-smooth univariate function g we have

$$g(\mathbf{c}^i A^i\mathbf{x}) = g(\mathbf{c}^j A^j\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^n$. Set

$$f_i(A^i\mathbf{x}) = g(\mathbf{c}^i A^i\mathbf{x}),$$

and

$$f_j(A^j\mathbf{x}) = -g(\mathbf{c}^j A^j\mathbf{x}).$$

Thus, as above,

$$f_i(A^i\mathbf{x}) + f_j(A^j\mathbf{x}) = 0$$

and yet f_i and f_j do not exhibit any of the smoothness or polynomial properties of the right-hand-side of this equation. □

Note that the condition that the $2d$ rows of A^i and A^j be linearly independent implies that $d \leq n/2$. Thus for $d > n/2$ we can never make any smoothness claims on the f_i based on the smoothness of F. This is unfortunate, as functions of the form (4.1) where $d = n - 1$ are of interest.

When considering ridge functions, i.e., when $d = 1$, we very naturally demanded that the $\mathbf{a}^i$ be pairwise linearly independent. That is, we exactly claimed the linear independence of the $2d$ rows of A^i and A^j, for all $i \neq j$, for $d = 1$.

What if we assume the linear independence of the $2d$ rows of A^i and A^j for all $i \neq j$? Do the f_i of (4.1) then inherit, under some weak assumptions, smoothness properties from F? The answer is yes. Here we utilize a generalization of the one-dimensional results of de Bruijn [4,5]; see de Bruijn [4] and Kemperman [9]. Parallel to $\mathcal{B}$ of Section 2, let us define $\mathcal{B}_d$ to be any class of real-valued functions f defined on $\mathbb{R}^d$ such that if there is a function $r \in C(\mathbb{R}^d)$ such that $f - r$ satisfies the multivariate Cauchy Functional Equation

$$g(\mathbf{s}+\mathbf{t}) = g(\mathbf{s}) + g(\mathbf{t}) \tag{4.2}$$

for all $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, then $f - r$ is necessarily linear, i.e., $(f - r)(\mathbf{s}) = \mathbf{c}\cdot\mathbf{s}$ for some constant vector $\mathbf{c} \in \mathbb{R}^d$, and all $\mathbf{s} \in \mathbb{R}^d$. This holds, for example, if f is continuous at a point, or bounded on a set of positive measure, etc. We now prove the multivariate analogue of Theorem 2.1.

Theorem 4.2. *Assume $F \in C^k(\mathbb{R}^n)$ is of the form (4.1), where the $2d$ rows of A^i and A^j are linearly independent, for all $i \neq j$. Assume, in addition, that each $f_i \in \mathcal{B}_d$. Then, necessarily, $f_i \in C^k(\mathbb{R}^d)$ for $i = 1, \ldots, m$.*

Remark 4.1. It is readily verified that the $2d$ rows of A^i and A^j are linearly independent if and only if

$$\ker A^i + \ker A^j = \mathbb{R}^n.$$

Proof. The proof is much the same as the proof of Theorem 2.1, with slight modifications. As previously, our proof will be by induction. The result is obviously valid for $m = 1$.

For given A^1 and A^m, let $\mathbf{d}^1, \ldots, \mathbf{d}^d \in \mathbb{R}^n$ satisfy

$$A^m\mathbf{d}^j = \mathbf{0}, \quad j = 1, \ldots, d, \tag{4.3}$$

and

$$A^1\mathbf{d}^j = \mathbf{e}^j, \quad j = 1, \ldots, d, \tag{4.4}$$

where $\mathbf{e}^j$ denotes the jth unit vector in $\mathbb{R}^d$. Such $\mathbf{d}^j$ exist by our assumption that the $2d$ rows of A^1 and A^m are linearly independent.

For each choice of $p_1, \ldots, p_d$, consider

$$H(\mathbf{x}) := F\left(\mathbf{x} + \sum_{j=1}^{d} p_j \mathbf{d}^j\right) - F(\mathbf{x})$$
$$= \sum_{i=1}^{m} f_i\left(A^i\mathbf{x} + A^i\left(\sum_{j=1}^{d} p_j \mathbf{d}^j\right)\right) - f_i(A^i\mathbf{x}).$$

Set

$$h_i(\mathbf{y}) = f_i\left(\mathbf{y} + \sum_{j=1}^{d} p_j A^i \mathbf{d}^j\right) - f_i(\mathbf{y}), \quad i = 1, \ldots, m,$$

for $\mathbf{y} \in \mathbb{R}^d$. From (4.3),

$$h_m(\mathbf{y}) = 0,$$

and from (4.4),

$$h_1(\mathbf{y}) = f_1(\mathbf{y} + \mathbf{p}) - f_1(\mathbf{y})$$

where $\mathbf{p} = (p_1, \ldots, p_d)$.

Thus,

$$H(\mathbf{x}) = \sum_{i=1}^{m-1} h_i(A^i\mathbf{x}),$$

and by the induction hypothesis we may infer that $h_i \in C^k(\mathbb{R}^d)$ for each $i = 1, \ldots, m-1$. In particular, we have that for each and every $\mathbf{p} \in \mathbb{R}^d$, the function

$$h_1(\mathbf{y}) = f_1(\mathbf{y} + \mathbf{p}) - f_1(\mathbf{y})$$

is in $C^k(\mathbb{R}^d)$. From Kemperman [9, Section 5], see also de Bruijn [4, p. 196], it follows that $f_1 = r_1 + g_1$, where $r_1 \in C^k(\mathbb{R}^d)$ and g_1 satisfies the Cauchy Functional Equation (4.2). Since $f_1 \in \mathcal{B}_d$ we have $g_1(\mathbf{s}) = \mathbf{c} \cdot \mathbf{s}$ for some constant vector $\mathbf{c} \in \mathbb{R}^d$, and therefore $f_1 \in C^k(\mathbb{R}^d)$. Thus

$$F(\mathbf{x}) - f_1(A^1\mathbf{x}) = \sum_{i=2}^{m} f_i(A^i\mathbf{x})$$

is in $C^k(\mathbb{R}^n)$, and again by our induction assumption we have that $f_i \in C^k(\mathbb{R}^d)$ for $i = 2, \ldots, m$. $\square$

Using Theorem 4.2 and the ideas from Proposition 3.1, we obtain an analogue of this latter result, namely:

Proposition 4.3. *Assume*

$$0 = \sum_{i=1}^{m} f_i(A^i\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^n$, *where* m *is finite, and the* $2d$ *rows of* A^i *and* A^j *are linearly independent, for all* $i \neq j$. *Assume, in addition, that each* $f_i \in \mathcal{B}_d$. *Then* $f_i \in \Pi^d_{m-2}$, $i = 1, \ldots, m$.

Proof. From Theorem 4.2 it follows that each of the f_i is infinitely smooth. For $m = 1$ the result is obviously true, where we define $\Pi^d_{-1} = \{0\}$. Recall that for $\mathbf{d} \in \mathbb{R}^n$

$$D_{\mathbf{d}} = \sum_{i=1}^{n} d_i \frac{\partial}{\partial x_i}.$$

In addition, if A is a $d \times n$ matrix with row vectors $\mathbf{a}^1, \ldots, \mathbf{a}^d$, then

$$D_{\mathbf{d}} f(A\mathbf{x}) = \sum_{k=1}^{d} (\mathbf{a}^k \cdot \mathbf{d}) \frac{\partial f}{\partial y_k}(A\mathbf{x})$$

where by $\frac{\partial f}{\partial y_k}$ we mean the derivative of f with respect to its kth argument.

The proof is notationally challenging, so let us first detail the case $m = 2$. Let $\mathbf{d}^j \in \mathbb{R}^n$, $j = 1, \ldots, d$, satisfy

$$A^1 \mathbf{d}^j = \mathbf{0}, \quad j = 1, \ldots, d,$$

and

$$A^2 \mathbf{d}^j = \mathbf{e}^j, \quad j = 1, \ldots, d,$$

where $\mathbf{e}^j$ denotes the jth unit vector in $\mathbb{R}^d$. Such $\mathbf{d}^j$ exist since the $2d$ rows of A^1 and A^2 are linearly independent. Thus

$$0 = D_{\mathbf{d}^j}[f_1(A^1\mathbf{x}) + f_2(A^2\mathbf{x})] = \frac{\partial f_2}{\partial y_j}(A^2\mathbf{x}), \quad j = 1, \ldots, d.$$

As A^2 is of full rank this implies that

$$\frac{\partial f_2}{\partial y_j} = 0, \quad j = 1, \ldots, d,$$

and $f_2 \in \Pi^d_0$ (a constant function). The same result holds for f_1, proving the case $m = 2$.

For general m, let $j_1, \ldots, j_{m-1}$ be arbitrary values in $\{1, \ldots, d\}$. We will prove that for all such $j_1, \ldots, j_{m-1}$ we have

$$\frac{\partial^{m-1} f_i}{\partial y_{j_1} \cdots \partial y_{j_{m-1}}} = 0, \quad i = 1, \ldots, m.$$

This implies that $f_i \in \Pi^d_{m-2}$, $i = 1, \ldots, m$. We prove this result for $i = m$.

For each $k = 1, \ldots, m-1$, and $j_1, \ldots, j_{m-1} \in \{1, \ldots, d\}$, let $\mathbf{d}^{j_k,k} \in \mathbb{R}^n$ satisfy

$$A^k \mathbf{d}^{j_k,k} = \mathbf{0},$$

and

$$A^m \mathbf{d}^{j_k,k} = \mathbf{e}^{j_k}.$$

Such vectors exist since the $2d$ rows of A^k and A^m are linearly independent. From above we have that

$$D_{\mathbf{d}^{j_k,k}} g_k(A^k\mathbf{x}) = 0$$

for every choice of sufficiently smooth g_k. Since the differential operators $D_{\mathbf{d}^{j_k},k}$ commute, it therefore follows that

$$\prod_{k=1}^{m-1} D_{\mathbf{d}^{j_k},k} f_i(A^i\mathbf{x}) = 0, \quad i = 1, \ldots, m-1.$$

Furthermore

$$D_{\mathbf{d}^{j_k},k} f_m(A^m\mathbf{x}) = \frac{\partial f_m}{\partial y_{j_k}}(A^m\mathbf{x}).$$

Thus

$$\begin{aligned} 0 &= \prod_{k=1}^{m-1} D_{\mathbf{d}^{j_k},k} \sum_{i=1}^{m} f_i(A^i\mathbf{x}) \\ &= \prod_{k=1}^{m-1} D_{\mathbf{d}^{j_k},k} f_m(A^m\mathbf{x}) \\ &= \frac{\partial^{m-1} f_m}{\partial y_{j_1} \cdots \partial y_{j_{m-1}}}(A^m\mathbf{x}). \end{aligned}$$

As A^m is of full rank, and the above holds for all $j_1, \ldots, j_{m-1} \in \{1, \ldots, d\}$ this implies that $f_m \in \Pi^d_{m-2}$. □

Recall that Proposition 4.3 is, in fact, a result concerning the uniqueness, up to polynomials of some order, of the representation of these multivariate ridge functions.

Remark 4.2. For the sake of convenience we stated the results of this paper over $\mathbb{R}^n$. In fact they hold, mutatis mutandis, over any open set in $\mathbb{R}^n$.

Acknowledgment

The author thanks Professor Zoltán Daróczy for bringing the paper by de Bruijn [4] to his attention.

References

[1] J. Aczél, Functional Equations and their Applications, Academic Press, New York, 1966.
[2] M.D. Buhmann, A. Pinkus, Identifying linear combinations of ridge functions, Adv. Appl. Math. 22 (1999) 103–118.
[3] W. Dahmen, C.A. Micchelli, Some remarks on ridge functions, Approx. Theory Appl. 3 (1987) 139–143.
[4] N.G. de Bruijn, Functions whose differences belong to a given class, Nieuw Arch. Wiskd. 23 (1951) 194–218.
[5] N.G. de Bruijn, A difference property for Riemann integrable functions and for some similar classes of functions, Nederl. Akad. Wetensch. Proc. Ser. A. 55= Indag. Math. 14 (1952) 145–151.
[6] P. Diaconis, M. Shahshahani, On nonlinear functions of linear combinations, SIAM J. Sci. Stat. Comput. Appl. 5 (1984) 175–191.
[7] K.J. Falconer, Consistency conditions for a finite set of projections of a function, Math. Proc. Cambridge Philos. Soc. 85 (1979) 61–68.
[8] G. Hamel, Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x+y) = f(x) + f(y)$, Math. Ann. 60 (1905) 459–462.
[9] J.H.B. Kemperman, A general functional equation, Trans. Amer. Math. Soc. 86 (1957) 28–56.
[10] V.Ya. Lin, A. Pinkus, Fundamentality of ridge functions, J. Approx. Theory 75 (1993) 295–311.

[11] B.F. Logan, L.A. Shepp, Optimal reconstruction of a function from its projections, Duke Math. J. 42 (1975) 645–659.

[12] B.E. Petersen, K.T. Smith, D.C. Solmon, Sums of plane waves, and the range of the Radon transform, Math. Ann. 243 (1979) 153–161.