SIAM J. SCI. STAT. COMPUT.
Vol. 5, No. 1, March 1984

0196-5204/84/0501-0013 $01.25/0

ON NONLINEAR FUNCTIONS OF LINEAR COMBINATIONS*

PERSI DIACONIS† AND MEHRDAD SHAHSHAHANI‡

Abstract. Projection pursuit algorithms approximate a function of p variables by a sum of nonlinear functions of linear combinations:

$$f(x_1,\cdots,x_p) \doteq \sum_{i=1}^{n} g_i(a_{i1}x_1+\cdots+a_{ip}x_p). \tag{1}$$

We develop some approximation theory, give a necessary and sufficient condition for equality in (1), and discuss nonuniqueness of the representation.

Key words. approximation theory, nonlinear high-dimensional nonparametric regression, polynomials, Schwartz distributions

1. Introduction and statement of main results.

We present some mathematical analysis for a class of curve fitting algorithms labeled "projection pursuit" algorithms by Friedman and Stuetzle (1981a, b). These algorithms approximate a general function of p variables by a sum of nonlinear functions of linear combinations:

$$f(x_1,\cdots,x_p) \doteq \sum_{i=1}^{n} g_i(a_{i1}x_1+\cdots+a_{ip}x_p). \tag{1.1}$$

In (1.1), f is a given function and univariate, nonlinear functions g_i and linear combinations $a_{i1}x_1+\cdots+a_{ip}x_p$ are sought so that a reasonable approximation is attained. Such approximation is computationally feasible and performs well in examples of nonparametric regression with noisy data, high-dimensional density estimation, and multidimensional spline approximation. In addition to the articles of Friedman and Stuetzle cited above, see Friedman and Tukey (1974), and Friedman, Grosse and Stuetzle (1983) for examples and computational details. Huber (1981a, b) begins to connect the algorithms to statistical theory. This note treats the algorithms from the point of view of approximation theory.

It is easy to show that approximation is always possible.

THEOREM 1. *Functions of the form* $\sum \alpha_i e^{\mathbf{a}^i\cdot\mathbf{x}}$, *with* α_i *real*, $\mathbf{a}^i$ *a vector of nonnegative integers, and* $\mathbf{x}=(x_1,\cdots,x_p)$ *are dense in the continuous real valued functions on* $[0,1]^p$ *under the maximum deviation norm.*

Proof. The functions $e^{\mathbf{a}\cdot\mathbf{x}}$ separate points of $[0,1]^p$ and are closed under multiplication. Finite linear combinations of such functions form a point separating algebra which is dense because of the Stone–Weierstrass theorem. □

THEOREM 2. *Functions of the form*

$$\sum \alpha_i \cos(2\pi\mathbf{a}^i\cdot x)+\beta_i \sin(2\pi b^i\cdot x)$$

are dense in $L^2[0,1]^p$.

Proof. Any function in $L^2[0,1]^p$ can be well approximated by its Fourier expansion. See Zygmund (1959, Vol. 2) and the survey article by Ash (1976) for further details and refinements. □

* Received by the editors September 26, 1981.

† Stanford University, Stanford, California 94305 and Harvard University, Cambridge, Massachusetts 02138.

‡ University of Virginia, Charlottesville, Virginia 22903 and Stanford University, Stanford, California 94305.

Sometimes equality is possible in (1.1). For example,

$$xy = \tfrac{1}{4}(x+y)^2 - \tfrac{1}{4}(x-y)^2,$$

$$\max(x, y) = \tfrac{1}{2}|x+y| + \tfrac{1}{2}|x-y|,$$

$$(xy)^2 = \tfrac{1}{4}(x+y)^4 + \frac{7}{4\cdot 3^3}(x-y)^4 - \frac{1}{2\cdot 3^3}(x+2y)^4 - \frac{2^3}{3^3}(x+\tfrac{1}{2}y)^4.$$

In what follows we will focus on conditions for equality in (1.1) as a method of determining examples to test, compare, and evaluate algorithms. Consider first a smooth function of 2 variables of the special form,

$$f(x, y) = g(ax+by).$$

Clearly,

$$\left(b\frac{\partial}{\partial x} - a\frac{\partial}{\partial y}\right) f \equiv 0.$$

If f has the form

$$f(x, y) = \sum_{i=1}^{n} g_i(a_i x + b_i y), \tag{1.2}$$

then the differential operator

$$\prod_{i=1}^{n}\left(b_i\frac{\partial}{\partial x} - a_i\frac{\partial}{\partial y}\right) = \sum_{i=0}^{n} c_i \frac{\partial^n}{\partial x^i\,\partial y^{n-i}}$$

applied to f is identically zero. The next theorem gives a converse.

THEOREM 3. *Let $f \in C^n[0, 1]^2$. Suppose that for some real numbers $c_0, \cdots, c_n$, the operator $\sum_{i=0}^{n} c_i\, \partial^n/\partial x^i\, \partial y^{n-i}$ applied to f is identically zero. If the polynomial $\sum_{i=0}^{n} c_i z^i$ has distinct real zeros then* (1.2) *holds for some (a_i, b_i). The lines $a_i x + b_i y$ are all distinct.*

Theorem 3 is proved in § 2 which also contains a discussion of techniques for finding directions (a_i, b_i) given f. Some applications of Theorem 3 are contained in the following examples.

Application 1. The functions e^{xy} and $\sin xy$ cannot be written in the form (1.1) for any finite n. Indeed, the equation $\sum c_i(\partial^n/\partial x^i\, \partial y^{n-i})\{e^{xy}\} \equiv 0$ implies $c_i \equiv 0$ and the associated polynomial has complex roots.

Application 2. Let $f(x, y)$ be a polynomial of degree m. Then

$$f(x, y) = \sum_{i=1}^{m} g_i(a_i x + b_i y),$$

where each g_i is a polynomial of degree at most m. This follows by elementary manipulations from Theorem 3. Thus, any polynomial in two variables can be represented exactly. Since polynomials are dense in $C[0, 1]^2$, this gives another proof of denseness of projection pursuit approximations. A different proof of this result is in Logan and Shepp (1975). An extension to more than two variables is in Proposition 1 of § 2.

Application 3. Representations of the form (1.1) are not necessarily unique. For example,

$$xy = c(ax+by)^2 - c(ax-by)^2$$

for any a and b satisfying $ab \neq 0$, $a^2+b^2=1$ with $c=1/4ab$. Writing $a=\cos\theta$, $b=\sin\theta$, any noncoordinate direction can be chosen for the quadratic g_1. The second direction is forced as orthogonal to this. This suggests that substantive interpretation of the linear combinations (a_i, b_i) is difficult. For a more ambitious example, consider the function $(xy)^2$. This is of 4th degree. Use of Theorem 3 as outlined in § 2, shows that $(xy)^2$ cannot be expressed as a sum of $n=3$ or fewer terms in (1.1). Four terms of 4th degree suffice:

$$(xy)^2=\alpha_1(x+b_1y)^4+\alpha_2(x+b_2y)^4+\alpha_3(x+b_3y)^4+\alpha_4(x+b_4y)^4,$$

where b_1, b_2, b_3, b_4 are chosen as distinct, and satisfying

$$b_1b_2+b_1b_3+b_1b_4+b_2b_3+b_2b_4+b_3b_4=0.$$

Then α_1, α_2, α_3, α_4 are determined by

$$\alpha_i=\frac{1}{6}\frac{\sum^* b_j}{\prod^*(b_j-b_i)},$$

where the sum and product are over $j\neq i$. This clearly defines a three-dimensional family of solutions.

In thinking about nonuniqueness, we observed that the only examples of nonunique representation we could find are polynomials. Indeed, polynomials have the following strong nonuniqueness property.

DEFINITION. A function $f(x, y)$ has *strongly nonunique representations* if there are two sets of directions $\{(a_i, b_i)\}_{i=1}^n$, $\{(\alpha_i, \beta_i)\}_{i=1}^m$, all distinct from each other, such that

$$f(x,y)=\sum_{i=1}^n g_i(a_ix+b_iy)=\sum_{i=1}^m h_i(\alpha_ix+\beta_iy)$$

for some g_i and h_i.

Polynomials have strongly nonunique representations: if $\deg P(x, y)=n$, and $(a_i, b_i)_{i=1}^{n+1}$ are any distinct directions, Theorem 1 implies that $P(x, y)$ can be represented in these directions. It turns out that *only* polynomials have this property. This is a consequence of Theorem 4.

THEOREM 4. *Let $f(x, y)\in C^{n+m}[0,1]^2$. Suppose that for some directions $\{a_i, b_i\}_{i=1}^n$ and $\{(\alpha_i, \beta_i)\}_{i=1}^m$,*

$$f(x,y)=\sum_{i=1}^n g_i(a_ix+b_iy)=\sum_{i=1}^m h_i(\alpha_ix+\beta_iy).$$

If, for some i, (α_i, β_i) is distinct from (a_j, b_j), $1\leqq j\leqq n$, then h_i is a polynomial of degree at most $n+m-2$.

Proof. Let

$$A=\prod_{j\neq i}\left(\beta_j\frac{\partial}{\partial x}-\alpha_j\frac{\partial}{\partial y}\right),$$

and

$$B=\prod_{j=i}^n\left(a_j\frac{\partial}{\partial x}-b_j\frac{\partial}{\partial y}\right).$$

Then,

$Af=\sum Ag_i$ is a sum of functions in directions (a_i, b_i);

so

$$0 = BAF = Ch_i^{(m+n-1)}(\alpha_i x + \beta_i y)$$

for

$$C = \left\{\prod_{j=1}^{n} (\alpha_i b_j - \beta_i a_j)\right\} \cdot \left\{\prod_{j \neq i} (\alpha_i \beta_j - \beta_i \alpha_j)\right\} \neq 0.$$

Thus, h_i is a polynomial of degree at most $m+n-2$. □

COROLLARY. *A function $f(x, y)$ has strongly nonunique representations if and only if f is a polynomial.*

Remark. The statement and proof of Theorem 4 carry over to functions of more than two variables in a straightforward way. See Lemma 1 of § 2.

How are the curve fitting algorithms affected by nonuniqueness? To understand this, we performed the following experiment. On each trial 200 independent, mean zero, variance 1, normal points (x_i, y_i) were generated. The algorithm of Friedman and Stuetzle was given x_i, y_i, and $x_i y_i + \varepsilon_i$, with ε_i normally distributed, mean zero, variance .1 errors. We expected the directions fit to change a great deal. In each of 100 trials the algorithm fit univariate functions in directions $(1, 1)$ and $(1, -1)$ (to two decimal places).

To understand this, it is important to consider the nature of the algorithm. At each stage, it chooses the direction which minimizes the residual sum of squares when the best fitting function in that direction is subtracted off. See Friedman and Stuetzle (1981) for a careful description. If the sample size is large, the algorithm will behave in the same way as the infinite population analogue. Thus, let X, Y be independent Gaussian variables with mean zero and variance 1. Consider approximating XY by the best linear combination of the form $aX + bY$. For fixed a and b, the L^2 norm is minimized by the function $E(XY|aX+bY)$. Which values of a and b, subject to $a^2 + b^2 = 1$, minimize

$$E\{XY - E(XY|aX+bY)\}^2?$$

Let us show that the minimum is achieved at $a = \pm b = \pm 1/\sqrt{2}$. Let $U = aX + bY$, $V = aX - bY$. Then U and V are independent standard normal and

$$XY = \frac{1}{4ab}\{U^2 - V^2\}, \qquad E(XY|aX+bY) = \frac{1}{4ab}\{U^2 - 1\}.$$

Then,

$$\begin{aligned} E\{XY - E(XY|aX+bY)\}^2 &= \frac{1}{(4ab)^2} E\{(U^2 - V^2) - (U^2 - 1)\}^2 \\ &= \frac{1}{(4ab)^2} E\{V^2 - 1\}^2. \end{aligned}$$

Since the distribution of V does not depend on a and b, the right side is minimized when $a^2 = b^2 = \frac{1}{2}$. When the best fitting function is subtracted off, the second stage of the algorithm subtracts off a quadratic in the orthogonal direction and the algorithm terminates after two steps. The same result can be shown to hold when X and Y are chosen uniformly in $[-1, 1]^2$.

Similar computations can be instructively carried out for functions other than XY. For example, consider $(XY)^2$. David Donoho has shown that if X and Y are normally distributed the algorithm chooses the four directions $(1, 1)$, $(1, -1)$, $(1, 0)$, (0, 1). Moreover, Donoho can prove that the approximation does not terminate after four steps, even though the function can be expressed as a sum of four 4th degree polynomials. Infinitely many steps are required—successive terms being cyclically added in each of the four directions.

Donoho and Ian Johnstone have independently proved that for normally distributed X and Y, the greedy approximation, which at each stage finds the a and b to minimize

$$E\{f_n(X, Y)-E\{f_n(X, Y)|aX+bY\}\}^2,$$

converges in L^2.

These results underscore a property of the projection pursuit algorithm: the directions it chooses are the directions that minimize the L^2 error. The situation is somewhat like finding the principal components of a covariance matrix. There are many possible bases, but the directions chosen have a well-defined interpretation in terms of maximum reduction of variance.

Application 4. Even if the directions (a_i, b_i) are fixed, the representation need not be unique. Suppose that n is the smallest integer such that

$$f(x, y)=\sum_{i=1}^{n} g_i(a_ix+b_iy).$$

If also

$$f(x, y)=\sum_{i=1}^{n} h_i(a_ix+b_iy),$$

then

$$f_i(t)-h_i(t)=p_i(t), \qquad 1\leqq i\leqq n,$$

with p_i a polynomial of degree at most $n-1$. The polynomials p_i can be chosen in an arbitrary way subject to the constraint $\sum p_i\equiv 0$. In particular, any $n-1$ of the p_i can be chosen arbitrarily and a final polynomial can be found to satisfy the constraint. These results all follow easily from Theorem 3; indeed the operator $L_i=\prod_{j\neq i}[b_j\,\partial/\partial x-a_j\,\partial/\partial y]$ applied to $f(x, y)$ gives

$$h_i^{(n-1)}(a_ix+b_iy)\prod_{j\neq i}(b_ja_i-a_jb_i)=g_i^{(n-1)}(a_ix+b_iy)\prod_{j\neq i}(b_ja_i-a_jb_i).$$

The products are nonvanishing because the directions are distinct. It follows that h_i differs from g_i by at most a polynomial of degree $n-1$, and that an arbitrary polynomial may be added subject to the constraint.

In the special case $n=2$, Theorem 3 was given by Dotson [4] who suggests further application to factoring probability densities and separation of variables.

The generalization to dimension greater than two is not as neat. We give a result for three-dimensions which generalizes to p-dimensions. Suppose that for n distinct directions $a^i\in\mathbb{R}^3$, a function f can be represented, for $x\in\mathbb{R}^3$, as

$$f(x)=\sum_{i=1}^{n} g_i(a^i\cdot x). \tag{1.3}$$

Let $\Pi^i=\{\rho\in\mathbb{R}^3: \rho\cdot a^i=0\}$. Let $\nabla=(\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$. Clearly,

$$\prod_{i=1}^{n}(\rho^i\cdot\nabla)f\equiv 0 \quad \text{for all } \rho^i\in\Pi^i.$$

This condition is not sufficient. To see this, consider the function $f(x_1, x_2, x_3) = x_1x_2$ and the three directions $a^1 = (1, 0, 0)$, $a^2 = (0, 1, 0)$ and $a^3 = (0, 0, 1)$. For any nonzero $\rho^i \in \Pi^i$, the operator $\Pi(\rho^i \cdot \nabla)$ applied to f is zero. Yet, f cannot be written as $f(x) = f_1(x_1) + f_2(x_2) + f_3(x_3)$.

The condition is sufficient "up to polynomials":

THEOREM 5. *Let a^r be distinct, nonzero, directions in $\mathbb{R}^3$. Let Π^r be the plane $\{\rho \in \mathbb{R}^3 : a^r \cdot \rho = 0\}$. A function $f \in C^n(\mathbb{R}^3)$ has the form*

$$f(x) = \sum_{r=1}^{n} g_r(a^r \cdot x) + P(x)$$

for a polynomial P of degree less than n, if and only if

$$\prod_{i=1}^{n} (\rho^i \cdot \nabla) f = 0 \quad \text{for all } \rho^r \in \Pi^r. \tag{1.4}$$

Remark. As noted above, condition (1.4) is not sufficient to ensure that representation (1.3) holds. If (1.3) holds, there are other obvious necessary conditions: if $\rho^{ij} \in \Pi^i \cap \Pi^j$, then

$$(\rho^{ij} \cdot \nabla) g_i(a^i \cdot x) = (\rho^{ij} \cdot \nabla) g_j(a^i \cdot x) = 0.$$

Thus, f is annihilated by differential operators of degree $[(n+1)/2], \cdots, n-1$. Unfortunately, even these conditions are not sufficient. H. Royden, in unpublished work, has determined necessary and sufficient conditions of a rather different type. These are stated at the end of this paper. In Proposition 1 of § 2 we show that any polynomial P can be written as a sum of univariate polynomials of linear combinations. If deg $P = k$, then $(k+1)(k+2)/2$ terms may be required.

Thus far we have been assuming sufficient differentiability. Versions of all theorems are valid if derivatives are interpreted in the sense of distributions. This is discussed in some detail in § 3.

Our theorems are related to Hilbert's 13th problem. In modern notation, Hilbert asked if there are genuine multivariate functions. Of course, $x + y$ is a function of two variables but $xy = e^{\log x + \log y}$ is a superposition of univariate functions and $+$. Kolmogorov and Arnold showed that, in this sense, $+$ is the only function of two variables. They constructed five monotone functions $\phi_i : [0, 1] \to \mathbb{R}$, which satisfy $|\phi_i(x) - \phi(y)| \leqq |x - y|$. These functions have the following remarkable property: for each $f \in C[0, 1]^2$ there is a $g \in C[0, 1]$ such that for all (x, y),

$$f(x, y) = \sum_{i=1}^{n} g(\phi_i(x) + \tfrac{1}{2}\phi_i(y)).$$

Thus ϕ_i are a "universal change of variables" which allows exact equality. A nice discussion of this result and its refinements can be found in Lorentz (1966), (1976) and Vitushkin (1977). While the functions ϕ_i and g are given in a constructive fashion, it does not seem that this result is used to approximate functions in an applied context. This is probably because the functions ϕ_i are fairly "wild". For example, it is known that it is not possible to choose ϕ_i to be C^1 functions, so fixed linear combinations of x and y are ruled out. It is known that $f(x, y) = \sum_{i=1}^{n} g_i(a_ix + b_iy)$ for all polynomials $f(x, y)$ is not possible with a_i, b_i fixed independent of f. In the projection pursuit approach to approximation, a_i and b_i are allowed to depend on f and Example 2 shows that now any polynomial can be written in required form. Example 1 shows that not all functions can be so expressed.

This paper has characterized functions that can be represented *exactly* as a sum of nonlinear functions of linear combinations. It is important to be able to recognize functions that can be well approximated by such a sum. Some important work on this problem is in the papers by Logan and Shepp (1975) and Logan (1975). These papers work with prespecified directions, but the main results of Logan (1975) do not depend on the directions. Roughly, Logan shows that a function on the unit disc can be well approximated, in L^2, by a sum of n univariate functions if and only if the function has bandwidth n, in the sense that its Fourier transform is essentially supported on a disc of radius n.

2. Proof and discussion of Theorems 3 and 5. Let L be the differential operator: $\sum_{i=0}^n c_i\, \partial^n/\partial x^i\, \partial y^{n-i}$. By hypothesis, the polynomial

$$\sum_{i=0}^{n} c_i x^i y^{n-i} = y^n \sum_{i=0}^{n} c_i \left(\frac{x}{y}\right)^i$$

splits into distinct linear factors. Thus, L can be written as $\prod [b_i\, \partial/\partial x - a_i\, \partial/\partial y]$, with the lines $a_i x + b_i y$ distinct. It must be shown that f can be represented as $\sum_{i=1}^n g_i(a_i x + b_i y)$. The proof is by induction on n. For $n = 1$, suppose without real loss that $a_1 \neq 0$. Then $f(x, y) = g(a_1 x + b_1 y)$ with $g(z) = f(z/a_1, 0)$. One way to show this is to fix (x, y) and define $h(t) = f(x + (b_1/a_1)y - (b_1/a_1)yt, ty)$. Then $h(0) = f(x + (b_1/a_q)y, 0) = g(a_1 x + b_1 y)$; $h(1) = f(x, y)$ and $h'(t) \equiv 0$, for $0 \leqq t \leqq 1$. The fundamental theorem of calculus gives $h(1) = \int_0^1 h' + h(0)$. Suppose the result is true for operators of degree $\leqq n - 1$. To prove it for degree n, write

$$\prod_{i=1}^{n} \left(b_i \frac{\partial}{\partial x} - a_i \frac{\partial}{\partial y}\right) f = \left\{\prod_{i=1}^{n-1} \left(b_i \frac{\partial}{\partial x} - a_i \frac{\partial}{\partial y}\right)\right\} \left(b_n \frac{\partial}{\partial x} - a_n \frac{\partial}{\partial y}\right) f \equiv 0.$$

By the induction hypotheses, there are functions g_i, $1 \leqq i \leqq n - 1$ satisfying

$$\left(b_n \frac{\partial}{\partial x} - a_n \frac{\partial}{\partial y}\right) f = \sum_{i=1}^{n-1} g_i(a_i x + b_i y). \tag{2.1}$$

A solution f^* of (2.1) of the form

$$f^*(x, y) = \sum_{i=1}^{n-1} h_i(a_i x + b_i y)$$

is found by choosing $h_i(t) = (b_n a_i - a_n b_i)^{-1} \int_0^t g_i(s)\, ds$. This is well-defined because the lines are distinct. Now $\{b_n\, \partial/\partial x - a_n\, \partial/\partial y\}(f - f^*) \equiv 0$ can be solved explicitly with $(f - f^*)(x, y) = h_n(a_n x + b_n y)$ by the argument for $n = 1$. It follows that $f = f^* + h_n$ can be written in the required form. □

Remarks on explicit computations. If f is of the form (1.2) then Theorem 3 gives the existence of numbers $c_0, \cdots, c_n$ such that $\sum c_j(\partial^n/\partial x^i\, \partial y^{n-i})(f) \equiv 0$. The c_i can be found by fixing $n + 1$ distinct pairs (x_i, y_i), calculating $\partial^n/\partial x^i\, \partial y^{n-i}|_{(x_i, y_i)}$ and solving the resulting system of equations for c_i. It is feasible to check if the polynomial $c_0 + \cdots + c_n z^n$ has distinct real roots using techniques in Henrici (1977, Chap. 6). Each stage of the procedure is feasible by a finite algorithm. If the procedure fails at any stage, then equality is impossible. Given feasible $c_0, \cdots, c_n$, it may be possible to find the roots of the associated polynomial. This determines directions (a_i, b_i).

In simple examples there is often enough freedom of choice to make determination of (a_i, b_i) possible. Consider $f(x, y) = xy$ for $n = 2$,

$$\prod_{i=1}^{2}\left(b_i\frac{\partial f}{\partial x} - a_i\frac{\partial f}{\partial y}\right) = b_1b_2\frac{\partial^2 f}{\partial x^2} - (b_1a_2 + b_2a_1)\frac{\partial^2 f}{\partial x\,\partial y} + a_1a_2\frac{\partial^2 f}{\partial y^2}.$$

Since $\partial^2 f/\partial x^2 = \partial^2 f/\partial y^2 = 0$, $\partial^2 f/\partial x\,\partial y = 1$; any distinct choice of a_i and b_i with $b_1a_2 = -b_2a_1$ works. Taking $a_1 = b_1 = 1$, $a_2 = -b_2 = 1$, we are led to solve

$$f(x, y) = g_1(x + y) + g_2(x - y).$$

Applying $\partial/\partial x - \partial/\partial y$ to both sides leads to $y - x = 2g_2'(x - y)$; setting $y = 0$; $g_2'(x) = -x/2$, $g_2 = -x^2/4 + c_2$. Similarly, $g_1(x) = x^2/4 + c_1$ and the result is $xy = \frac{1}{4}(x+y)^2 + c_1 - \frac{1}{4}(x-y)^2 + c_2$ where $c_1 + c_2 = 0$ is forced. In general, if $f = \sum_{i=1}^{n} g_i(a_ix + b_iy)$; $\prod_{j\neq i}(b_i\,\partial/\partial x - a_i\,\partial/\partial y)f = c_ig_i^{(n-1)}(a_ix + b_iy)$, for an explicit c_i. This determines g_i up to an essentially free choice of an $n-1$ degree polynomial.

In the case of a polynomial f, some additional tricks become available. For a multinomial x^ay^b let $a + b = n$; only sums of the form $\sum_{i=1}^{n}\alpha_i(x + \beta_jy)^n$ need be considered. Expanding out and equating coefficients gives

$$\sum \alpha_i = 0, \qquad \sum \alpha_i\beta_i = 0 \cdots \sum \alpha_i\beta_j^a = \frac{1}{\binom{n}{j}} \cdots \sum \alpha_i\beta_j^n = 0.$$

This gives $n+1$ equations in $2n$ unknowns. These are linear in the α's for given β's and may be solved explicitly because the matrix is a Vandermonde with a well-known inverse. See Gautschi (1963).

The proof of Theorem 5 was outlined by H. Royden. The proof follows from three lemmas. Throughout a^r are distinct nonzero directions in $\mathbb{R}^3$.

LEMMA 1. *If* $\sum_{r=1}^{l} g_r(a^r \cdot x) \equiv 0$ *then* $g_r^{(l-1)} \equiv 0$ *and* g_r *is a polynomial of degree at most* $l-2$.

Proof. Fix r. For each $j \neq r$ there is $\rho^j \in \Pi^j$ but $\rho^j \cdot a^r \neq 0$. Apply the operator $\prod_{j\neq r}(\rho^j \cdot \nabla)$ to the sum to conclude $\prod_{j\neq r}(\rho^j \cdot a^j)g^{(l-1)}(\rho^j \cdot x) \equiv 0$. The coefficient is nonzero, so the conclusion follows. □

In the next two lemmas, the notation f_j means $\partial f/\partial x_j$.

LEMMA 2. *Let* P *and* Q *be polynomials of degree* $\leqq k$ *in* (x_1, x_2, x_3). *Suppose that in some open* $\mathcal{O} \subset \mathbb{R}^3$

$$P_3 = Q_2.$$

Then there is polynomial H *of degree at most* $k+1$ *such that*

$$P = H_2 \quad \text{and} \quad Q = H_3.$$

Proof. Argue in a cube $\{a \leqq x_1 \leqq \alpha,\ b \leqq x_2 \leqq \beta,\ c \leqq x_3 \leqq \gamma\}$ contained in $\mathcal{O}$. Let Γ be a path connecting (x_1, a, b) to (x_1, x_2, x_3) which lies entirely in the plane of constant x_1. The line integral

$$\int_\Gamma P(x_1, y, z)\,dy + Q(x, y, z)\,dz = H(x_1, x_2, x_3)$$

is independent of Γ in view of the hypothesis and Green's theorem. Furthermore,

$$dH = P\,dx_2 + Q\,dx_3.$$

Here d is exterior differentiation in the plane $x_1 = \text{constant}$. Therefore

$$\frac{\partial H}{\partial x_2} = P \quad \text{and} \quad \frac{\partial H}{\partial x_3} = Q.$$

In particular, let Γ be the path

$$\begin{aligned} &t \to (x_1, a + t(x_2 - a), b), && 0 \leqq t \leqq 1, \\ &t \to (x_1, x_2, b + (t-1)(x_3 - b)), && 1 \leqq t \leqq 2. \end{aligned}$$

Then,

$$\begin{aligned} H(x_1, x_2, x_3) = (x_2 - a) \int_0^1 P(x_1, a + t(x_2 - a), b)\, dt \\ + (x_3 - b) \int_1^2 P(x_1, x_2, (t-1)(x_3 - b))\, dt \end{aligned}$$

which is clearly a polynomial of degree $\leqq k+1$. $\square$

LEMMA 3. *Let $f \in C^{n+2}(\mathbb{R}^3)$ have the following properties: for n distinct directions a^r, with a^r distinct from* $(1, 0, 0)$,

$$f_2(x) = \sum_{r=1}^{n} g_r(a^r \cdot x) + P(x), \tag{2.2a}$$

$$f_3(x) = \sum_{r=1}^{n} h_r(a^r \cdot x) + Q(x). \tag{2.2b}$$

With P and Q polynomials of degree at most $n-1$; then there are univariate functions G_r, $1 \leqq r \leqq n+1$, and a polynomial H of degree at most n such that

$$f(x) = \sum_{r=1}^{n} G_r(a^r \cdot x) + G_{n+1}(x_1) + H(x).$$

Proof. Because $f_{23} = f_{32}$, conditions (2.2a) and (2.2b) translate into

$$0 = \sum_{r=1}^{n} (a_3^r g_r^1 - a_2^r h_r^1)(a^r \cdot x) + P_3(x) - Q_2(x).$$

By hypothesis, for $i \neq r$ there are vectors $\rho^i \in \Pi^i$ with $\rho^i \cdot a^r \neq 0$. Let $A = \prod_{i \neq r} (\rho^i \cdot \nabla)$. Applying A to $P_3(x) - Q_2(x)$ gives zero because this polynomial is of degree at most $n-2$. Thus,

$$0 \equiv c\{a_3^r g_r^{(n)} - a_2^r h_r^{(n)}\},$$

where

$$c = \prod_{i \neq r} (\rho^i \cdot a^r) \neq 0.$$

It follows that

$$a_3^r g_r(a^r \cdot x) - a_2^r h_r(a^r \cdot x) = P_r(a^r \cdot x) \tag{2.3}$$

for P a polynomial of degree at most $n-1$. Because the a^r are distinct from $(1, 0, 0)$, either $a_2^r \neq 0$ or $a_3^r \neq 0$. Define the n functions G_r by

$$G_r' = g_r / a_2^r \quad \text{if } a_2^r \neq 0, \tag{2.4a}$$

$$G_r' = h_r / a_3^r \quad \text{if } a_2^r = 0. \tag{2.4b}$$

Consider

$$\phi(x)=f(x)-\sum_{r=1}^{n} G_r(a^r\cdot x).$$

From the hypothesis, (2.3) and (2.4),

$$\phi_2(x)=f_2(x)-\sum a_2^r G_r'(a^r\cdot x)=\sum_{a_2^r=0} g_r(a^r\cdot x)+P(x)=P^*(x)$$

with P^* a polynomial of degree at most $n-1$. Further,

$$\phi_3(x)=f_3(x)-\sum a_3^r G'(a^r\cdot x)=\sum\{h_r(a^r\cdot x)-a_3^r G'(a^r\cdot x)\}+Q.$$

Each term in the sum is a polynomial. If $a_2^r\neq 0$, the rth term equals

$$h_r(a^r\cdot x)-a_3^r g_r(a^r\cdot x)/a_2^r$$

a polynomial from (2.3). If $a_2^r=0$, the rth term is zero. It follows that

$$\phi_3=Q^*(x)$$

with Q^* a polynomial of degree at most $n-1$. To finish off, observe that $\phi_{23}=\phi_{32}$ gives $P_3^*=Q_2^*$. From Lemma 2, there is a polynomial H of degree at most n, such that $H_2=P^*$ and $H_3=Q^*$. The function $\psi=\phi-H$ has $\psi_2=\psi_3=0$. Thus, ψ is only a function of x_1, as required. □

Proof of Theorem 5. Clearly, if f can be represented as a sum of n univariate functions plus a polynomial the differential operator kills f. The proof of the converse is by induction on n. For $n=1$, we know that if $f_2=f_3=0$ then f is a function of x_1 only. Rotating to bring the plane Π^1 into $\{\rho:\rho_2=\rho_3=0\}$ proves the general case. Suppose that the result is true for $n-1$. Let $a^1, a^2, \cdots, a^{n-1}, a^n$ be n distinct nonzero directions. By rotating, we may assume that $a^n=(1\ 0\ 0)$. Then, for any $\rho^i\in\Pi^i$, $1\leqq i\leqq n-1$,

$$\Pi(\rho^i\cdot\nabla)f_2\equiv\Pi(\rho^i\cdot\nabla)f_3\equiv 0.$$

The induction hypothesis yields that f satisfies conditions (2.2) of Lemma 3. The theorem follows. □

PROPOSITION 1. *Let p and k be positive integers. Let* $r=\binom{m+p-1}{m}$. *There are r distinct directions* $a^1, a^2, \cdots, a^r$ *in* $\mathbb{R}^p$ *such that any homogeneous polynomial f of degree m can be written as*

$$f(x)=\sum_{j=1}^{r}\alpha_j(a^j\cdot x)^m \textit{ for some real numbers } \alpha_r.$$

Proof. The space of homogeneous polynomials of degree m is an r-dimensional vector space over the real numbers. Let $m_i(x)$, $1\leqq i\leqq r$ be an enumeration of the monomials. For each monomial, let D_i be the associated differential operator (e.g., if $m_i(x)=x_1^2x_2x_3$, $D_i=\partial^4/\partial x_1^2\,\partial x_2\,\partial x_3$). Observe that $D_i(a^r\cdot x)^m=m!\,m_i(a^r)$. For dimension reasons, to prove the proposition it suffices to show that directions a^i can be chosen so that the polynomials $(a^j\cdot x)^m$, $j=1,\cdots,r$, are linearly independent. Suppose

$$\sum c_j(a^j\cdot x)^m=0.$$

Applying D_i we get

$$\sum_{j=1}^{r} m_i(a^j)c_j=0 \quad \text{for all } i=1,\cdots,r.$$

For this system to have a nontrivial solution $(c_1, \cdots, c_r)$, we must have

$$(*) \qquad \det(m_i(a^j)) = 0.$$

We write $a^j = (a_1^j, \cdots, a_p^j)$. Since $\det(m_i(a^j))$ is a nontrivial algebraic expression with rational coefficients, if we choose the pr real numbers a^i in such a way that they are algebraically independent over Q, then

$$\det(m_i(a^j)) \neq 0,$$

contradicting $(*)$. □

3. Some generalizations. The theory presented so far does not apply to identities between nondifferentiable functions. Most of the results remain valid if differentiation is interpreted in the sense of distributions. Consider the identity

$$\max(x, y) = \tfrac{1}{2}(|x+y| + |x-y|).$$

For fixed y the function $\max(x, y)$ is constant for $x \leqq y$ and equal to x for larger x. Thus

$$\frac{\partial}{\partial x}\max(x, y) = \begin{cases} 0, & x < y, \\ 1, & x > y. \end{cases}$$

This function is also called the Heavyside function shifted to y. Its derivative is well known to be the delta function concentrated on the line $x = y$. This acts on $\phi \in C_0^\infty(\mathbb{R}^2)$ by $\delta(\phi) = \int \phi(t, t)\, dt$. Similarly, $\partial^2/\partial y^2 \max(x, y) = \delta$, so $\max(x, y)$ is a solution of the wave equation

$$\frac{\partial^2}{\partial x^2}U - \frac{\partial^2}{\partial y^2}U = 0.$$

The only solutions of this equation are of the form $U = f_1(x+y) + f_2(x-y)$ where f_1 and f_2 are distributions (see Schwartz (1966, p. 9) for some history). We further show, here and more generally, that if the solution U is a sufficiently well-behaved function, then the f_i are functions.

Any undefined terms in the following discussion can be found in Barros-Neto (1973) or Schwartz (1966). Let $\mathcal{D}(\mathbb{R}^2)$ be the space of test functions—compactly supported C^∞ functions. The dual space $\mathcal{D}'(\mathbb{R}^2)$ is the space of distributions on $\mathbb{R}^2$. For $\gamma = (a, b)$, the translate of $T \in \mathcal{D}'$ by γ is written T_γ. This acts on $\phi \in \mathcal{D}$ by $T_\gamma\{\phi(x)\} = T\{\phi(x-\gamma)\}$. The distribution $T \in D'(\mathbb{R}^2)$ *depends only on* $ax + by$ if for all real t, $T_{(bt, -at)} = T$. The following theorem collects together several results in Schwartz (1966, § II.5). It is the case $m = 1$ of the theorem at which we are aiming.

THEOREM 6. *Let $T \in \mathcal{D}'(\mathbb{R}^2)$. For (a, b) nonzero, the following conditions are equivalent*:

(a) *T depends only on $ax + by$.*

(b) $(b\,\partial/\partial x - a\,\partial/\partial y)T = 0$.

(c) *There is a distribution $g \in \mathcal{D}'(\mathbb{R}^2)$ such that for all $\phi \in \mathcal{D}(\mathbb{R}^2)$,*

$$T(\phi) = g\left\{\int \phi(au + bv, bu - av)\, dv\right\}$$

where g operates on the function of u inside the brackets.

Remark. If T and g are functions with $T(x, y) = g(ax + by)$ and $a^2 + b^2 = 1$, then part (c) becomes

$$T(\phi) = \iint T(x, y)\phi(x, y)\, dx\, dy = \iint g(ax + by)\phi(x, y)\, dx\, dy$$

$$= \iint g(u)\phi(au + bv, bu - av)\, du\, dv = g\left\{\int \phi(au + bv, bu - av)\, dv\right\}.$$

Notation. If T satisfies any of the three conditions of Theorem 6 we write $T = q_\gamma^*(g)$ where $\gamma = (a, b)$ and q_γ is the linear map from $\mathbb{R}^2 \to \mathbb{R}^1$ given by $q_\gamma(x, y) = ax + by$. We can now state the distribution version of Theorem 3.

THEOREM 7. *Let $T \in \mathcal{D}'(\mathbb{R}^2)$. Suppose that for some real numbers $c_0, c_1, \cdots, c_m$, the operator $\sum_{i=0}^m c_i\, \partial^m/\partial x^i\, \partial y^{m-i}$ applied to T is zero. If the polynomial $\sum c_i z^i$ has distinct real zeros then there are distinct nonzero $\gamma_1, \gamma_2, \cdots, \gamma_m$; $\gamma_i = (a_i, b_i)$ such that, writing q_i for q_{γ_i},*

$$T = \sum_{i=1}^m q_i^*(g_i) \quad \textit{with } g_i \in \mathcal{D}'(\mathbb{R}). \tag{3.1}$$

Conversely, if (3.1) *holds, then* $\prod (b_i\, \partial/\partial x - a_i\, \partial/\partial y)$ *applied to* f *is zero.*

Proof. One direction is clear; the argument for the other direction is by induction on m. The case $m = 1$ follows from Theorem 6. Thus, assume the result for $m - 1$. Without loss of generality, assume $a_i^2 + b_i^2 = 1$ for $1 \leqq i \leqq m$. Then

$$\prod_{i=2}^m \left(b_i \frac{\partial}{\partial x} - a_i \frac{\partial}{\partial y}\right)\left(b_1 \frac{\partial}{\partial x} - a_1 \frac{\partial}{\partial y}\right) f = 0.$$

This implies

$$\left(b_1 \frac{\partial}{\partial x} - a_1 \frac{\partial}{\partial y}\right) f = \sum_{i=2}^m q_i^*(g_i). \tag{3.2}$$

We will show that (3.2) has a solution f^* of form

$$f^* = \sum_{i=2}^m q_i^*(h_i).$$

Supposing this, $(b_1\, \partial/\partial x - a_1\, \partial/\partial y)(f - f^*) = 0$, so by the result for $m = 1, f - f^* = q_1^*(g_1)$ for $g_1 \in \mathcal{D}'(\mathbb{R})$ giving the theorem. To complete the proof, let h_i be a distribution solution to

$$\frac{dh_i}{dt} = \frac{1}{b_1 a_i - a_1 b_i} g_i.$$

A solution exists by Schwartz (1966, Thm. IX, p. 130). We claim

$$\sum_{i=2}^m q_i^*(h_i) = f^*$$

is a solution to (3.2). To show this, we need the following relation:

$$aq^*(h') = \frac{\partial}{\partial x} q^*(h).$$

To prove this, consider $\phi \in C_0^\infty(\mathbb{R}^2)$. Now

$$aq^*(h')(\phi) = ah'\left(\int \phi(au+bv, bu-av)\,dv\right)$$
$$= -h\left(\int [a\phi_1(au+bv, bu-av) + b\phi_2(au+bv, bu-av)]\,dv\right).$$

Since ϕ is compactly supported

$$\int \frac{d}{dv}(\phi(au+bv, bu-av))\,dv = 0.$$

Thus,

$$\int [b\phi_1(au+bv, bu-av) - a\phi_2(au+bv, bu-av)]\,dv = 0.$$

Using this gives

$$aq^*(h')(\phi) = -h\left(\int [a^2\phi_1(au+bv, bu-av) + b^2\phi_1(au+bv, bu-av)]\,dv\right),$$
$$= -h\left(\int \phi_1(au+bv, bu-av)\,dv\right)$$
$$= \frac{\partial}{\partial x} q^*(h)(\phi).$$

Thus, $(b_1\,\partial/\partial x - a_1\,\partial/\partial y)q_i^*(h_i) = q_i^*(g_i)$. The claim regarding f^* follows. □

The next theorem shows that if the equation

$$f(x, y) = \sum_{i=1}^{n} q_i^*(g_i)$$

holds in the sense that the two sides are equal as distributions, and if $f(x, y)$ is a sufficiently regular function, then each of the distributions g_i can be realized as a function on $\mathbb{R}$. Theorems of this sort may be described as results on propagation of singularities of partial differential equations.

The notion of "sufficiently regular" which we adopt involves the Sobolev spaces H^s; the definitions involve Fourier transforms, and so the space $\mathcal{S}$ of C^∞ functions that, together with all derivatives, tend to zero at infinity faster than any polynomial. The dual of $\mathcal{S}$, denoted $\mathcal{S}'$ is the space of tempered distributions. The Fourier transform of $\phi \in \mathcal{S}(\mathbb{R}^2)$ is

$$\hat{\phi}(\lambda) = \iint e^{-i\lambda\cdot x}\phi(x)\,dx,$$

where dx is $1/2\pi$ times Lebesgue measure. The Fourier inversion theorem becomes

$$\phi(x) = \iint e^{i\lambda\cdot x}\hat{\phi}(\lambda)\,d\lambda.$$

The Fourier transform of a tempered distribution $\theta \in \mathcal{S}'$ is defined by

$$\hat{\theta}(\phi) = \theta(\hat{\phi}).$$

For real s, $-\infty < s < \infty$, the Sobolev space $H^s(\mathbb{R}^2)$ is the set of tempered distributions $\theta \in S'(\mathbb{R}^2)$ such that $(1+|\rho|^2+|\eta|^2)^{s/2}\hat{f}(\rho,\eta) \in L^2(\mathbb{R}^2)$. There are various embedding theorems that say when a distribution is a function. For example, Taylor (1980, Chap. 1, § 3) gives:

(a) If $s > n/2$, then each $\theta \in H^s(\mathbb{R}^n)$ is a bounded continuous function that vanishes at infinity.

(b) If $s > n/2+k$, then $H^s(\mathbb{R}^n) \subset C^k(\mathbb{R}^n)$.

(c) For $0<\alpha<1$, define C^α as the set of bounded functions u such that $|u(x+y)-u(y)| < C|y|^\alpha$ for $|y| \leqq 1$. If $s = n/2+\alpha$, $0<\alpha<1$, then $H^2(\mathbb{R}^n) \subset C^\alpha(\mathbb{R}^n)$.

(d) If $0 \leqq s < n/2$, $H^s(\mathbb{R}^n) \subset L^q(\mathbb{R}^n)$, $q = 2n/(n-2s)$.

We have chosen the route of interpolating between integer values of s by means of the Fourier transform. There are other routes. See Adams (1975) for discussion.

For U open in $\mathbb{R}^2$, $H^s_{\text{loc}}(U)$ is the set of distributions $\theta \in \mathscr{D}'(U)$ such that for each compactly supported $\phi \in C_0^\infty(U)$, $\phi \cdot \theta \in H^s(\mathbb{R}^2)$. For example, $\max(x,y) \in H^1_{\text{loc}}(\mathbb{R}^2)$. With this notation, we can state the main result.

THEOREM 8. *Let $\gamma_i = (a_i, b_i)$, $1 \leqq i \leqq m$ be distinct nonzero directions in $\mathbb{R}^2$. Let q_i denote projection in the direction γ_i, so $q_i(x,y) = a_i x + b_i y$. Let U be open in $\mathbb{R}^2$. Let U_i be open sets in $\mathbb{R}$ with $q_i^{-1}(U_i) \supset U$ for all i. Suppose $f \in H^s_{\text{loc}}(U)$ can be written*

$$f = \sum_{i=1}^m q_i^*(g_i),$$

where $g_i \in \mathscr{D}'(U_i)$. Then $g_i \in H^s_{\text{loc}}(U_i)$.

The proof of Theorem 8 will be given following two preliminary lemmas. Let (a,b) be a unit vector in $\mathbb{R}^2$. Let $q(x,y) = ax+by$ denote the projection. For $g \in \mathscr{S}'(\mathbb{R})$, the distribution q^*g acts on $\psi \in \mathscr{S}(\mathbb{R}^2)$ as $g(\int \psi(au+bv, bu-av)\,dv)$. The distribution p_*g acts on functions $\psi \in \mathscr{S}(\mathbb{R}^2)$ by $g(\psi(at,bt))$. We have the next lemma.

LEMMA 4. $(\widehat{q^*g}) = p_*(\hat{g})$.

Proof. $\widehat{q^*g}(\psi) = q^*g(\hat{\psi}) = g\{\int \hat{\psi}(au+bv, bu-av)\,dv\}$. Now the integral equals

$$\begin{aligned}\int e^{-i(au+bv)x-i(bu-av)y}\phi(x,y)\,dx\,dy\,dv &= \int e^{-i(ax+by)u-i(bx-ay)v}\phi(x,y)\,dx\,dy\,dv\\ &= \int e^{-isu-itv}\phi(as+bt, bs-at)\,ds\,dt\,dv\\ &= \int e^{-isu}\left\{\int e^{-itv}\phi(as+bt, bs-at)\,dt\,dv\right\}ds.\end{aligned}$$

The inner integral equals $\phi(as, bt)$; indeed for any function $g \in C_0^\infty(\mathbb{R})$, $\int e^{itv}g(t)\,dt\,dv = \hat{\delta}(\hat{g}) = \delta(g) = g(0)$. Making this substitution, proves the result. □

The next lemma is the case $m = 1$ of Theorem 8.

LEMMA 5. *Let $\pi: \mathbb{R}^2 \to \mathbb{R}$ be the projection $\pi(x,y) = x$, $U_1 \subset \mathbb{R}$ and $U \supset \pi^{-1}(U_1)$. Let $T \in \mathscr{D}'(U)$ and assume there is $g \in \mathscr{D}'(U_1)$ such that*

$$T = \pi^*(g).$$

If $T \in H^s_{\text{loc}}(U)$, then $g \in H^s_{\text{loc}}(U_1)$.

Proof. Without loss of generality we may assume $U = \pi^{-1}(U_1)$ since $T = \pi^*(g)$. Let $V_1 \subset \bar{V}_1 \subset U_1$ and V_1 open, $\bar{V}_1$ compact, and $\chi \in C_0^\infty(U_1)$ with $\chi \equiv 1$ on $\bar{V}_1$. Then it suffices to show $\chi g \in H^s(\mathbb{R})$. Let $f = \chi g$ and $\theta = \pi^*(\chi g)$. Then the hypothesis on T implies

$$\theta \in H^s_{\text{loc}}(\mathbb{R}^2).$$

Let $\phi(x, y) = \phi_1(y)$ with $\phi_1 \in C_0^\infty(\mathbb{R})$ such that

$$\phi_1(y) = \begin{cases} 1, & |y| \leqq r, \\ 0, & |y| \geqq R, \end{cases}$$

for some constants $0 < r < R$. Then $\phi\chi \in C_0^\infty(U)$, so

$$\phi\theta \in H^s(\mathbb{R}^2).$$

This means $\widehat{\phi\theta}$ is a function and

$$\iint |(\widehat{\phi\theta})(\rho, \eta)|^2(1 + |\rho|^2 + |\eta|^2)^{s/2}\, d\rho\, d\eta < \infty. \tag{3.3}$$

We will argue that (3.3) implies $f \in H^s(\mathbb{R})$. Lemma 4 implies that $\hat{\theta} = p_*(\hat{f})$. Moreover, $\hat{f}$ is an analytic function of one variable, being the Fourier transform of a distribution of compact support (see Barros-Neto (1973, § 4.5)). Thus

$$\hat{\theta}(\rho, \eta) = \hat{f}(\rho)\delta_0(\eta). \tag{3.4}$$

Also,

$$\hat{\phi}(\rho, \eta) = \delta_0(\rho)\hat{\phi}_1(\eta). \tag{3.5}$$

Using $(\widehat{\phi\theta}) = \hat{\phi} * \hat{\theta}$ with (3.4) and (3.5),

$$(\widehat{\phi\theta})(\rho, \eta) = \hat{\phi}_1(\eta) \cdot \hat{f}(\rho).$$

Thus, (3.3) becomes

$$\iint |\hat{\phi}_1(\eta)|^2 |\hat{f}(\rho)|^2 (1 + |\rho|^2 + |\eta|^2)^{s/2} < \infty. \tag{3.6}$$

Now elementary arguments show that for any real s there are positive constants ρ_1, ρ_2 such that

$$\rho_1 |\rho|^s < \int |\hat{\phi}_1(\eta)|^2 (1 + |\rho|^2 + |\eta|^2)^{s/2}\, d\eta < \rho_2 |\rho|^s \quad \text{for } |\rho| \geqq 1.$$

Using this and (3.6) gives

$$\int_{-\infty}^{-1} |\hat{f}(\rho)|^2 |\rho|^s\, d\rho + \int_1^\infty |\hat{f}(\rho)|^2 |\rho|^s\, d\rho < \infty.$$

Hence, the desired result:

$$\int |\hat{f}(\rho)|^2 (1 + |\rho|^2)^{s/2}\, d\rho < \infty. \qquad \square$$

Proof of Theorem 8. We may assume γ_i are unit vectors. The case $m = 1$ follows from Lemma 5 via linear transformation. For the general case, suppose

$$f = \sum_{i=1}^m q_i^*(g_i).$$

Let $D_i = \prod_{j \neq i} (b_j\, \partial/\partial x - a_j\, \partial/\partial y)$. Then, for a nonzero constant C_i,

$$D_i f = D_i q_i^*(g_i) = C_i q_i^*(g_i^{(m-1)}).$$

Since $f \in H^s_{\text{loc}}(U)$, $D_i f \in H^{s-m+1}_{\text{loc}}(U)$. By Lemma 5,

$$g_i^{(m-1)} \in H^{s-m+1}_{\text{loc}}(U_1).$$

This implies that $g_i \in H^s_{\text{loc}}(U_1)$, see for example Treves (1966, Thm. 7.6).

Acknowledgment. We thank David Donoho, Jerry Friedman, Bob Hulquist, Winni Li and Bruce Reznick for helpful discussions. Two extremely helpful referees found a gap in the original version of Theorem 5. We are grateful to Halsey Royden for allowing us to use his elegant proof of the corrected version.

Note added in proof. Halsey Royden has communicated the following conditions necessary and sufficient for a function f of m variables to be representable as a sum of N univariate functions: Let $\alpha = (\alpha_1, \cdots, \alpha_m)$ denote a multi-index of weight $\sum \alpha_j$. Let f_α denote the appropriate partial derivative and $f_{\alpha j} = (\partial/\partial x_j) f_\alpha$.

THEOREM (H. Royden). *Let $N = \binom{m+l-1}{l}$. Then a smooth function $f(x_1, \cdots, x_m)$ can be written*

$$f = \sum_{\nu=1}^{N} g_\nu\left(\sum a_k^\nu x_k\right)$$

where the N vectors $a^\nu = (a_k^\nu)$ do not lie on a hypersurface of degree l in projective $m-1$ space, if and only if there are functions h_ν, an invertible $N \times N$ constant matrix $C = [C_\alpha^\nu]$ where α runs over the N multi-indices of weight l) and N $m \times m$ constant invertible matrices $B_\nu = [B^j_{k\nu}]$ such that

$$\sum_{\alpha, j} c_\nu^\alpha f_{\alpha j} B^j_{k\nu} = \delta^1_k h_\nu(x_1, \cdots, x_m).$$

Here δ^1_k is Kronecker's delta function. The functions $g_\nu^{(l)}$ are then uniquely defined (so the representation is unique up to polynomials of degree $l-1$) and the directions a_ν are unique for those ν's with $g_\nu^{(l)} \not\equiv 0$.

REFERENCES

R. A. ADAMS (1975), *Sobolev Spaces*, Academic Press, New York.

J. M. ASH (1976), *Studies in Harmonic Analysis*, American Mathematical Society, Providence, RI.

J. BARROS-NETO (1973), *An Introduction to the Theory of Distributions*, Marcel Dekker, New York.

W. G. DOTSON (1968), *Decomposability of positive functions on $\mathbb{R}^n$*, Amer. Math. Monthly, pp. 350–357.

J. FRIEDMAN AND J. TUKEY (1974), *A projection pursuit algorithm for exploratory data analysis*, IEEE Transactions Computers, C-23, pp. 881–889.

J. FRIEDMAN AND W. STUETZLE (1981a), *Projection pursuit regression*, J. Amer. Statist. Assoc., 76, pp. 817–823.

——— (1981b), *Projection pursuit methods for data analysis*, Technical Report, Stanford Linear Accelerator, Stanford, CA.

J. FRIEDMAN, E. GROSSE AND W. STUETZLE (1983), *Multi-dimensional additive spline approximation*, this Journal, 4 (1983), pp. 291–301.

W. A. GAUTSCHI (1963), *On inverses of Vandermonde and confluent Vandermonde matrices*, Numer. Math., 5, pp. 425–430.

P. HENRICI (1977), *Applied and Computational Complex Analysis*, Vol. I, John Wiley, New York.

P. HUBER (1981a), *Density estimation and projection pursuit methods*, Tech. Rep. PJH-7, Dept. Statistics, Harvard Univ., Cambridge, MA.

—— (1981b), *Projection pursuit*, Tech. Rep. Dept. Statistics, Harvard, Univ., Cambridge, MA.

B. F. LOGAN (1975), *The uncertainty principle in reconstructing functions from projections*, Duke Math. J., 42, pp. 661–706.

B. F. LOGAN AND L. A. SHEPP (1975), *Optimal reconstruction of a function from its projections*, Duke Math. J., 42, pp. 645–660.

G. G. LORENTZ (1966), *Approximation of Functions*, Holt, Rinehart, and Winston, New York.
——— (1976), *The 13th problem of Hilbert*, in Mathematical Developments Arising from Hilbert Problems, Proc. Symp. Pure Math. XXVIII, American Mathematical Society, Providence, RI.
L. SCHWARTZ (1966), *Théorie des distributions*, 2nd edition, Hermann, Paris.
M. E. TAYLOR (1981), *Pseudo-differential Operators*, Princeton Univ. Press, Princeton, NJ.
F. TREVES (1966), *Linear Partial Differential Equations with Constant Coefficients*, Gordon-Breach, New York.
D. G. VITUSHKIN (1977), *On representation of functions by means of superpositions and related topics*, L'Enseignement Math., July–Dec. 1977, pp. 256–319.
K. ZYGMUND (1959), *Trigonometric Series*, 2nd ed., Cambridge Univ., Cambridge.