Applied and Computational Harmonic Analysis **6,** 197–218 (1999)
Article ID acha.1998.0248, available online at http://www.idealibrary.com on IDEAL®

Harmonic Analysis of Neural Networks

Emmanuel J. Candès

Stanford University, California 94305

Communicated by Charles K. Chui

Received February 3, 1997; revised January 28, 1998

It is known that superpositions of ridge functions (single hidden-layer feedforward neural networks) may give good approximations to certain kinds of multivariate functions. It remains unclear, however, how to effectively obtain such approximations. In this paper, we use ideas from harmonic analysis to attack this question. We introduce a special admissibility condition for neural activation functions. The new condition is not satisfied by the sigmoid activation in current use by the neural networks community; instead, our condition requires that the neural activation function be oscillatory. Using an admissible neuron we construct linear transforms which represent quite general functions f as a superposition of ridge functions. We develop

- a continuous transform which satisfies a Parseval-like relation;
- a discrete transform which satisfies frame bounds.

Both transforms represent f in a stable and effective way. The discrete transform is more challenging to construct and involves an interesting new discretization of time–frequency–direction space in order to obtain frame bounds for functions in $L^2(A)$ where A is a compact set of $\mathbf{R}^n$. Ideas underlying these representations are related to Littlewood–Paley theory, wavelet analysis, and group representation theory.

1. INTRODUCTION

Let $f(x)$: $\mathbf{R}^n \to \mathbf{R}$ be a function of n variables. In this paper, we are interested in constructing convenient approximations to f using systems called *neural networks.* A single hidden-layer feedforward neural network is the name given to a function of n-variables constructed by the rule

$$f_m(x) = \sum_{i=1}^{m} \alpha_i \rho(\langle k_i, x\rangle - b_i),$$

where the m terms in the sum are called neurons; the α_i and b_i, scalars; and the k_i, n-vectors. Each neuron maps a multivariate input $x \in \mathbf{R}^n$ into a real-valued output by

1063-5203/99 $30.00

composing a simple linear projection $x \to \langle k_i, x\rangle - b_i$ with a scalar nonlinearity ρ, called the activation function. Traditionally, ρ has been given a sigmoid shape, $\rho(t) = e^t/(1 + e^t)$, modeled after the activation mechanism of biological neurons. The vectors k_i specify the "connection strengths" of the n inputs to the ith neuron; the b_i specify activation thresholds. The use of this model for approximating functions in applied sciences, engineering, and finance is large and growing; for examples, see journals such as *IEEE Trans. Neural Networks.*

From a mathematical point of view, such approximations amount to taking finite linear combinations of atoms from the dictionary $\mathcal{D}_{Ridge} = \{\rho(\langle k, x\rangle - b);\ k \in \mathbf{R}^n,\ b \in \mathbf{R}\}$ of elementary *ridge functions.* As is known [6, 18], any function of n variables can be approximated arbitrarily well by such combinations. As far as constructing these combinations, a frequently discussed approach is the greedy algorithm that, starting from $f_0(x) = 0$, operates in a stepwise fashion running through steps $i = 1, \ldots m$; at the ith stage it augments the approximation f_{i-1} by adding a term from the dictionary $\mathcal{D}_{Ridge}$ which results in the largest decrease in approximation error, i.e., minimizes $\| f - (f_{i-1} + \alpha \cdot \rho(\langle k, x\rangle - b))\|_{L^2}$ over all choices of (k, α, b). It is known that when $f \in L^2(D)$ with D a compact set, the greedy algorithm converges [15]; it is also known that for a relaxed variant of the greedy algorithm, the convergence rate can be controlled under certain assumptions [1, 16]. There are, unfortunately, two problems with the conceptual basis of such results.

First, they lack the constructive character which one ordinarily associates with the word "algorithm." In any assumed implementation of minimizing $\| f - (f_{i-1} + \alpha \cdot \rho(\langle k, x\rangle - b))\|_{L^2}$, one would need to search for a minimum within a discrete collection of k and b. What are the properties of procedures restricted to such collections? Or, more directly, how finely discretized must the collection be so that a search over that collection gives results similar to a minimization over the continuum? In some sense, the word "algorithm" used to mean abstract minimization procedures in the absence of an understanding of this issue is a misnomer.

Second, even if one is willing to forgive the lack of constructivity in such results, one must still face the lack of stability of the resulting decomposition. An approximant $f_N(x) = \sum_{i=1}^{N} \alpha_i \rho(\langle k_i, x\rangle - b_i)$ has coefficients which in no way are continuous functionals of f and do not necessarily reflect the size and organizations of f [20].

Our goal in this paper is to apply the concepts and methods of modern harmonic analysis to the problem of constructing neural networks. Using techniques developed in group representations theory and wavelet analysis, we develop two concrete and stable representations of functions f as superpositions of ridge functions.

1.1. A Continuous Representation

First, we develop the concept of *admissible neural activation function* ψ: $\mathbf{R} \to \mathbf{R}$. Unlike traditional sigmoidal neural activation functions which are positive and monotone increasing, such an admissible activation function is oscillating, taking both positive and negative values. In fact, our condition requires for ψ a number of vanishing moments which is proportional to the dimension n, so that an admissible ψ has zero integral, zero "average slope," zero "average curvature," etc., in high dimensions.

We show that if one is willing to abandon the traditional sigmoidal neural activation function ρ, which typically has no vanishing moments and is not in L^2, and replace it with an admissible neural activation function ψ, then any reasonable function f may be represented exactly as a *continuous* superposition from the dictionary $\mathcal{D}_{Ridgelet} = \{\psi_\gamma: \gamma \in \Gamma\}$ of *ridgelets* $\psi_\gamma(x) = a^{-1/2}\psi\left(\frac{\langle u, x\rangle - b}{a}\right)$, where the ridgelet parameter $\gamma = (a, u, b)$ runs through the set $\Gamma \equiv \{(a, u, b);\ a, b \in \mathbf{R},\ a > 0,\ u \in \mathbf{S}^{n-1}\}$ with $\mathbf{S}^{n-1}$ denoting the unit sphere of $\mathbf{R}^n$. In short, we establish a continuous reproducing formula

$$f = c_\psi \int \langle f, \psi_\gamma\rangle \psi_\gamma \mu(d\gamma), \tag{1}$$

for $f \in L^1 \cap L^2(\mathbf{R}^n)$, where c_ψ is a constant which depends only on ψ and $\mu(d\gamma) \propto da/a^{n+1} du db$ is a kind of uniform measure on Γ; for details, see below. We also establish a Parseval relation

$$\|f\|^2 = c_\psi \int |\langle f, \psi_\gamma\rangle|^2 \mu(d\gamma). \tag{2}$$

Integral representations like (1) have been independently discovered in Murata [22]. These two formulas mean that we have a well-defined *continuous Ridgelet transform* $\mathcal{R}(f)(\gamma) = \langle f, \psi_\gamma\rangle$ taking functions on $\mathbf{R}^n$ isometrically into functions of the ridgelet parameter $\gamma = (a, u, b)$.

1.2. Discrete Representation

We next develop somewhat stronger admissibility conditions on ψ (which we call *frameability* conditions) and replace this continuous transform with a discrete transform. Let D be a fixed compact set in $\mathbf{R}^n$. We construct a special countable set $\Gamma_d \subset \Gamma$ such that every $f \in L^2(D)$ has a representation

$$f = \sum_{\gamma \in \Gamma_d} \alpha_\gamma \psi_\gamma, \tag{3}$$

with equality in the $L^2(D)$ sense. This representation is stable in the sense that the coefficients change continuously under perturbations of f which are small in $L^2(D)$ norm. Underlying the construction of such a discrete transform is of course a quasi-Parseval relation, which in this case takes the form

$$A\|f\|^2_{L^2(D)} \le \sum_{\gamma \in \Gamma_d} |\langle f, \psi_\gamma\rangle_{L^2(D)}|^2 \le B\|f\|^2_{L^2(D)}. \tag{4}$$

Equation (3) follows by use of the standard machinery of frames [7, 10]. Frame machinery also shows that the coefficients α_γ are realizable as bounded linear functionals $\alpha_\gamma(f)$ having Riesz representers $\tilde{\psi}_\gamma(x) \in L^2(D)$. These representers are not ridge functions themselves; but by the convergence of Neumann series underlying the frame operator, we are entitled to think of them as *molecules* made up of linear combinations of ridge atoms, where the linear concentrate on atoms with parameters γ' "near" γ.

1.3. Applications

As a result of this work, we are, roughly speaking, in a position to efficiently construct finite approximations by ridgelets which give good approximations to a given function $f \in L^2(D)$. Although we do not attempt to go so far in this paper, one can see where these tools are heading: from the exact series representation (3), one aims to extract a finite linear combination which is a good approximation to the infinite series; once such a representation is available, one has a stable, mathematically tractable method of constructing approximate representations of functions f based on systems of neuron-like elements. We hope to report on this program in a later paper.

1.4. Innovations

Underlying our methods is the inspiration of modern harmonic analysis—ideas like the Calderón reproducing formula and the theory of frames. We shall briefly describe what is new here—that which is not merely an "automatic" consequence of existing ideas.

First, there is, of course, a general machinery for obtaining continuous reproducing formulas like (1), via the theory of square-integrable group representations [8, 11]. Such a theory has been applied to develop wavelet-like representations over groups other than the usual $ax + b$ group on $\mathbf{R}^n$; see [3]. However, the particular geometry of ridge functions does not allow the identification of the action of Γ on ψ with a linear group representation (notice that the argument of ψ is real, while the argument of ψ_γ is a vector in $\mathbf{R}^n$). As a consequence, the possibility of a straightforward application of well-known results is ruled out. As an example of the difference, our condition for admissibility of a neural activation function for the continuous ridgelet transform is much stronger—requiring about $n/2$ vanishing moments in dimension n—than the usual condition for admissibility of the mother wavelet for the continuous wavelet transform, which requires only one vanishing moment in any dimension.

Second, in constructing frames of ridgelets, we have been guided by the theory of wavelets, which holds that one can turn continuous transforms into discrete expansions by adopting a strategy of discretizing frequency space into dyadic coronae [7, 8]; this goes back to Littlewood–Paley [13]. Our approach indeed uses such a strategy for dealing with the location and scale variables in the Γ_d dictionary. However, in dealing with ridgelets there is also an issue of discretizing the directional variable u that seems to be a new element: u must be discretized more finely as the scale becomes finer. The existence of frame bounds under our discretization shows that we have achieved, in some sense, the "right" discretization, and we believe this to be new and of independent interest.

In a discussion section we describe limitations, possible improvements, and possible directions for further work.

2. THE RIDGELET TRANSFORM

In this section we present results regarding the existence and the properties of the continuous representation (1). The measure $\mu(d\gamma)$ on neuron parameter space Γ is defined by $\mu(d\gamma) = (da/a^{n+1})\sigma_n du db$, where σ_n is the surface area of the unit sphere $\mathbf{S}^{n-1}$ in dimension n and du the uniform probability measure on $\mathbf{S}^{n-1}$. As usual, $\hat{f}(\xi) = \int e^{-i\langle x,\xi\rangle} f(x)dx$ denotes the Fourier transform of f and $\mathcal{F}(f)$ as well. To simplify notation we will consider only the case of multivariate $x \in \mathbf{R}^n$ with $n \geq 2$. Finally, we will always assume that $\psi: \mathbf{R} \to \mathbf{R}$ belongs to the Schwartz space $\mathcal{S}(\mathbf{R})$. Most of what follows holds under weaker conditions on ψ but we avoid study of various technicalities in this paper.

DEFINITION 1. Let $\psi: \mathbf{R} \to \mathbf{R}$ satisfy the condition

$$K_\psi = \int \frac{|\hat{\psi}(\xi)|^2}{|\xi|^n} d\xi < \infty. \tag{5}$$

Then ψ is called an *admissible neural activation function.*

THEOREM 1 (Reconstruction). *Suppose that f and $\hat{f} \in L^1(\mathbf{R}^n)$. If ψ is admissible, then*

$$f = c_\psi \int \langle f, \psi_\gamma\rangle \psi_\gamma \mu(d\gamma), \tag{6}$$

where $c_\psi = (2\pi)^{-(n-1)} K_\psi^{-1}$.

Remark 1. In fact, for $\psi \in \mathcal{S}(\mathbf{R})$, the admissibility condition (5) is essentially equivalent to the requirement of vanishing moments:

$$\int t^k \psi(t)dt = 0, \qquad k \in \left\{0, 1, \ldots, \left[\frac{n+1}{2}\right] - 1\right\}.$$

This clearly shows the similarity of (5) to the one-dimensional wavelet admissibility condition [7, p. 24]; however, unlike wavelet theory, the number of necessary vanishing moments grows linearly in the dimension n.

Remark 2. If $\rho(t)$ is the sigmoid function $e^t/(1 + e^t)$, then ρ is *not admissible.* Actually no formula like (6) can hold if one uses neurons of the type commonly employed in the theory of neural networks. However, $\rho^{(m)}(t)$ is an admissible activation function for $m \geq \left[\frac{n}{2}\right] + 1$. Hence, sufficiently high derivatives of the functions used in neural networks theory do lead to good reconstruction formulas.

We will call the ridge function ψ_γ generated by an admissible ψ a *ridgelet.*

Proof of Theorem 1. The proof uses the Radon transform P_u defined by $P_u f(t) = \int f(tu + U^\perp s)ds$ with $s = (s_1, \ldots, s_{n-1}) \in \mathbf{R}^{n-1}$ and $U^\perp$ an $n \times (n-1)$ matrix containing as columns an orthonormal basis for $u^\perp$.

With a slight abuse of notation, let $\psi_a(x) = a^{-1/2}\psi\left(\frac{x}{a}\right)$ and $\tilde{\psi}(x) = \psi(-x)$. Put $w_{a,u}(b) = \tilde{\psi}_a * P_u f(b)$ and let $\mathrm{I} = \int \langle f, \psi_\gamma\rangle \psi_\gamma(x)\mu(d\gamma) = \int \psi_a(\langle u, x\rangle - b) w_{a,u}(b)(da/a^{n+1})\sigma_n du db$. Recall $\widehat{P_u f} = \hat{f}(\xi u)$ and, hence, if $\hat{f} \in L^1(\mathbf{R}^n)$, $\widehat{P_u f} \in L^1(\mathbf{R})$. Then, $\mathrm{I} =$

$\int \psi_a * (\tilde{\psi}_a * P_u f)(\langle u, x\rangle)(da/a^{n+1})\sigma_n du$. Noting that $\psi_a * (\tilde{\psi}_a * P_u f) \in L^1(\mathbf{R})$ and that its one-dimensional Fourier transform is given by $a|\hat{\psi}(a\xi)|^2 \hat{f}(\xi u)$, we have

$$\mathrm{I} = \frac{1}{2\pi} \int \exp\{i\xi\langle u, x\rangle\} \hat{f}(\xi u) a \big|\hat{\psi}(a\xi)\big|^2 \frac{da}{a^{n+1}} \sigma_n du d\xi.$$

If ψ is real valued, $\overline{\hat{\psi}(-\xi)} = \hat{\psi}(\xi)$; hence,

$$\mathrm{I} = \frac{1}{\pi} \int \exp\{i\xi\langle u, x\rangle\} \hat{f}(\xi u) a \big|\hat{\psi}(a\xi)\big|^2 1_{\{\xi>0\}} \frac{da}{a^{n+1}} \sigma_n du d\xi.$$

Then, by Fubini,

$$\begin{aligned}
\mathrm{I} &= \frac{1}{\pi} \int \exp\{i\xi\langle u, x\rangle\} \hat{f}(\xi u) \left\{ \int \big|\hat{\psi}(a\xi)\big|^2 \frac{da}{a^n} \right\} 1_{\{\xi>0\}} d\xi \sigma_n du \\
&= \frac{1}{2\pi} \int \exp\{i\xi\langle u, x\rangle\} \hat{f}(\xi u) K_\psi |\xi|^{n-1} 1_{\{\xi>0\}} d\xi \sigma_n du \\
&= \frac{1}{2\pi} K_\psi \int_{\mathbf{R}^n} \exp\{i\langle x, k\rangle\} \hat{f}(k) dk \\
&= \frac{1}{2\pi} K_\psi (2\pi)^n f(x). \quad \blacksquare
\end{aligned}$$

THEOREM 2 (Parseval Relation). *Assume $f \in L^1 \cap L^2(\mathbf{R}^n)$ and ψ admissible. Then*

$$\|f\|_2^2 = c_\psi \cdot \int |\langle f, \psi_\gamma\rangle|^2 \mu(d\gamma).$$

Proof. With $w_{a,u}(b)$ defined as in the proof of Theorem 1, we then have

$$\int |\langle f, \psi_\gamma\rangle|^2 \mu(d\gamma) = \int |w_{a,u}(b)|^2 \frac{da}{a^{n+1}} \sigma_n du db = \mathrm{I},$$

say. Using Fubini's theorem for positive functions,

$$\int |w_{a,u}(b)|^2 \frac{da}{a^{n+1}} \sigma_n du db = \int \|w_{a,u}\|_2^2 \frac{da}{a^{n+1}} \sigma_n du. \tag{7}$$

$w_{a,u}$ is integrable, being the convolution between two integrable functions, and belongs to $L^2(\mathbf{R})$ since $\|w_{a,u}\|_2 \leq \|f\|_1 \|\psi_a\|_2$; its Fourier transform is then well defined and $\hat{w}_{a,u}(\xi)$

$= \bar{\hat{\psi}}_{\hat{a}}(\xi)\hat{f}(\xi u)$. By the usual Plancherel theorem, $\int |w_{a,u}(b)|^2 db = \frac{1}{2\pi} \int |\hat{w}_{a,u}(\xi)|^2 d\xi$ and, hence,

$$\mathrm{I} = \frac{1}{2\pi} \int |\hat{f}(\xi u)|^2 |\hat{\psi}_a(\xi)|^2 \frac{da}{a^{n+1}} \sigma_n du d\xi = \frac{2}{2\pi} \int_{\{\xi>0\}} |\hat{f}(\xi u)|^2 |\hat{\psi}(a\xi)|^2 \frac{da}{a^n} \sigma_n du d\xi.$$

Since $\int |\hat{\psi}(a\xi)|^2 \frac{da}{a^n} = K_\psi |\xi|^{n-1}/2$ (admissibility), we have

$$\mathrm{I} = \frac{K_\psi}{2\pi} \int |\hat{f}(\xi u)|^2 \xi^{n-1} d\xi du = K_\psi (2\pi)^{n-1} \| f \|_2^2. \quad \blacksquare$$

The assumptions on f in the above two theorems are somewhat restrictive, and the basic formulas can be extended to an even wider class of objects. It is classical to define the Fourier transform first for $f \in L^1(\mathbf{R}^n)$ and only later to extend it to all of L^2 using the fact that $L^1 \cap L^2$ is dense in L^2. By a similar density argument, one obtains

PROPOSITION 1. *There is a linear transform $\mathcal{R}$: $L^2(\mathbf{R}^n) \to L^2(\Gamma, \mu(d\gamma))$ which is an L^2-isometry and whose restriction to $L^1 \cap L^2$ satisfies*

$$\mathcal{R}(f)(\gamma) = \langle f, \psi_\gamma \rangle.$$

For this extension, a generalization of the Parseval relationship (2) holds.

PROPOSITION 2 (Extended Parseval). *For all $f, g \in L^2(\mathbf{R}^n)$,*

$$\langle f, g \rangle = c_\psi \int \mathcal{R}(f)(\gamma) \mathcal{R}(g)(\gamma) \mu(d\gamma). \tag{8}$$

We will give the proof in the Appendix. Notice that one need only prove the property for a dense subspace of $L^2(\mathbf{R}^n)$, i.e., $L^1 \cap L^2(\mathbf{R}^n)$.

Relation (8) allows identification of the integral $c_\psi \int \langle f, \psi_\gamma \rangle \psi_\gamma \mu(d\gamma)$ with f by duality. In fact, taking the inner product of $c_\psi \int \langle f, \psi_\gamma \rangle \psi_\gamma \mu(d\gamma)$ with any $g \in L^2(\mathbf{R}^n)$ and exchanging the order of inner product and integration over γ, one obtains

$$\left\langle c_\psi \left[\int \langle f, \psi_\gamma \rangle \psi_\gamma \mu(d\gamma) \right], g \right\rangle = c_\psi \int \langle f, \psi_\gamma \rangle \langle g, \psi_\gamma \rangle \mu(d\gamma) = \langle f, g \rangle,$$

which, by the Riesz theorem, leads to $f \equiv c_\psi \int \langle f, \psi_\gamma \rangle \psi_\gamma \mu(d\gamma)$ in the prescribed weak sense.

The theory of wavelets and Fourier analysis contain results of a similar flavor: for example, the Fourier inversion theorem in $L^2(\mathbf{R}^n)$ can be proven by duality. However, there exists a more concrete proof of the Fourier inversion theorem. Recall, in fact, that if $f \in L^1 \cap L^2(\mathbf{R}^n)$ and if we consider the truncated Fourier expansion $\hat{f}_K(\xi) = \hat{f}(\xi) 1_{\{|\xi| \le K\}}$, then $\hat{f}_K \in L^1(\mathbf{R}^n)$ and $\|\bar{\mathcal{F}}(\hat{f}_K) - (2\pi)^n f\|_{L^2} \to 0$ as $K \to \infty$. This argument provides an interpretation of the Fourier inversion formula that reassures us of its practical relevance.

We now give a similar result for the convergence of truncated ridgelet expansions. For each $\varepsilon > 0$, define $\Gamma_\varepsilon := \{\gamma = (a, u, b) : \varepsilon \le a \le \varepsilon^{-1}, u \in \mathbf{S}^{n-1}, b \in \mathbf{R}\} \subset \Gamma$.

PROPOSITION 3. *Let $f \in L^1(\mathbf{R}^n)$ and $\{\alpha_\gamma\} = \{\langle f, \psi_\gamma\rangle\}_{(\gamma\in\Gamma)}$; then for every $\varepsilon > 0$,*

$$\alpha_\gamma 1_{\Gamma_\varepsilon}(\gamma) \in L^1(\Gamma, \mu(d\gamma)).$$

Proof. Notice that $\alpha_\gamma = (\tilde{\psi}_a * P_u f)(b)$; then

$$\int_{\Gamma_\varepsilon} |\alpha_\gamma| \mu(d\gamma) = \int |w_{a,u}(b)| \frac{da}{a^{n+1}} \sigma_n du db \le \sigma_n \|f\|_1 \int_\varepsilon^{\varepsilon^{-1}} \|\psi\|_1 \frac{da}{a^{n+1/2}} < \infty,$$

where we have used $\|w_{a,u}\|_1 \le \|\tilde{\psi}_\alpha\|_1 \|f\|_1 = a^{1/2}\|\psi\|_1\|f\|_1$. ∎

The above proposition shows that for any $f \in L^1(\mathbf{R}^n)$, the expression

$$f_\varepsilon \equiv c_\psi \int_{\Gamma_\varepsilon} \langle f, \psi_\gamma\rangle \psi_\gamma \mu(d\gamma)$$

is meaningful, since $\{\psi_\gamma\}_{\gamma\in\Gamma}$ is uniformly L^∞ bounded over Γ_ε. The next theorem, whose proof is given in the Appendix, makes more precise the meaning of the reproducing formula.

THEOREM 3. *Suppose $f \in L^1 \cap L^2(\mathbf{R}^n)$ and ψ admissible.*

(1) $f_\varepsilon \in L^2(\mathbf{R}^n)$, *and*
(2) $\|f - f_\varepsilon\|_2 \to 0$ *as* $\varepsilon \to 0$.

3. THE DISCRETE TRANSFORM: FRAMES OF RIDGELETS

The previous section described a class of neurons, the ridgelets $\{\psi_\gamma\}_{\gamma\in\Gamma}$, such that

(i) any function f can be reconstructed from the continuous collection of its coefficients $\langle f, \psi_\gamma\rangle$, and
(ii) any function can be decomposed in a continuous superposition of neurons ψ_γ.

The purpose of this section is to achieve similar properties using only a discrete set of neurons $\Gamma_d \subset \Gamma$.

3.1. Generalities about Frames

The theory of frames [7, 27] deals precisely with questions of this kind. In fact, if $\mathcal{H}$ is a Hilbert space and $\{\varphi_n\}_{n\in N}$ a frame, an element $f \in \mathcal{H}$ is completely characterized by its coefficients $\{\langle f, \varphi_n\rangle\}_{n\in N}$ and can be reconstructed from them via a simple and numerically stable algorithm. In addition, the theory provides an algorithm to express f as a linear combination of the frame elements φ_n.

DEFINITION 2. Let $\mathcal{H}$ be a Hilbert space and let $\{\varphi_n\}_{n\in N}$ be a sequence of elements of $\mathcal{H}$. Then $\{\varphi_n\}_{n\in N}$ is a frame if there exist $0 < A, B < \infty$ such that for any $f \in \mathcal{H}$,

$$A\|f\|_{\mathcal{H}}^2 \leq \sum_{n\in N} |\langle f, \varphi_n\rangle_{\mathcal{H}}|^2 \leq B\|f\|_{\mathcal{H}}^2, \tag{9}$$

in which case A and B are called *frame bounds.*

Let $\mathcal{H}$ be a Hilbert space and $\{\varphi_n\}_{n\in N}$ a frame with bounds A and B. Notice that $A\|f\|_{\mathcal{H}}^2 \leq \sum |\langle f, \varphi_n\rangle|^2$ implies that $\{\varphi_n\}_{n\in N}$ is a complete set in $\mathcal{H}$. A frame $\{\varphi_n\}_{n\in N}$ is said to be tight if we can take $A = B$ in Definition 2. Furthermore, if $\{\varphi_n\}_{n\in N}$ is a basis for $\mathcal{H}$, it is called a Riesz basis. Simple examples of frames include orthonormal basis, Riesz basis, concatenation of several Riesz bases, etc.

The following results are stated without proofs and can be found in Daubechies [7, p. 56] and Young [27, p. 184]. Define the coefficient operator $F\colon \mathcal{H} \to l^2(N)$ by $F(f) = (\langle f, \varphi_n\rangle)_{n\in N}$. Suppose that F is a bounded operator ($\|Ff\| \leq B\|f\|_{\mathcal{H}}$). Let F^* be the adjoint of F and let $G = F^*F$ be the *frame operator;* then $A\,\mathrm{Id} \leq G \leq B\,\mathrm{Id}$ in the sense of orders of positive definite operators. Hence, G is invertible and its inverse G^{-1} satisfies $B^{-1}\mathrm{Id} \leq G^{-1} \leq A^{-1}\mathrm{Id}$. Define $\tilde{\varphi}_n = G^{-1}\varphi_n$; then $\{\tilde{\varphi}_n\}_{n\in N}$ is also a frame (with frame bounds B^{-1} and A^{-1}) and the following holds:

$$f = \sum_{n\in N} \langle f, \tilde{\varphi}_n\rangle_{\mathcal{H}}\varphi_n = \sum_{n\in N} \langle f, \varphi_n\rangle_{\mathcal{H}}\tilde{\varphi}_n. \tag{10}$$

Moreover, if $f = \sum_{n\in N} a_n\varphi_n$ is an another decomposition of f, then $\sum_{n\in N} |\langle f, \tilde{\varphi}_n\rangle|^2 \leq \sum_{n\in N} |a_n|^2$. To rephrase Daubechies, the frame coefficients are the most economical in an L^2 sense. Finally, $G = \frac{A+B}{2}(I - R)$, where $\|R\| < 1$, and so G^{-1} can be computed as $G^{-1} = \frac{2}{A+B}\sum_{k=0}^{\infty} R^k$.

3.2. *Discretization of* Γ

The special geometry of ridgelets imposes differences between the organization of ridgelet coefficients and the organization of traditional wavelet coefficients.

With a slight change of notation, we recall that $\psi_\gamma = a^{1/2}\psi(a(\langle u, x\rangle - b))$. We are looking for a countable set Γ_d and some conditions on ψ such that the quasi-Parseval relation (4) holds. Let $\mathcal{R}(f)(\gamma) = \langle f, \psi_\gamma\rangle$; then $\mathcal{R}(f)(\gamma) = \langle P_u f, \psi_{a,b}\rangle$ with $\psi_{a,b}(t) = a^{1/2}\psi(a(t - b))$. Thus, the information provided by a ridgelet coefficient $\mathcal{R}(f)(\gamma)$ is the one-dimensional wavelet coefficient of $P_u f$, the Radon transform of f. Applying Plancherel, $\mathcal{R}(f)(\gamma)$ may be expressed as

$$\mathcal{R}(f)(\gamma) = \frac{1}{2\pi}\langle \widehat{P_u f}, \hat{\psi}_{a,b}\rangle = \frac{a^{-1/2}}{2\pi}\int \hat{f}(\xi u)\hat{\psi}(\xi/a)\exp\{ib\xi\}d\xi, \tag{11}$$

which corresponds to a one-dimensional integral in the frequency domain (see Fig. 1).

In fact, this is the line integral of $\hat{f}\hat{\psi}_{a,0}$, modulated by $\exp\{ib\xi\}$, along the line $\{tu : t \in \mathbf{R}\}$. If, as in the Littlewood–Paley theory [13], $a = 2^j$ and $\mathrm{supp}(\psi) \subset [\frac{1}{2}, 2]$, it emphasizes a certain dyadic segment $\{t : 2^j \le t \le 2^{j+1}\}$. In contrast, in the multidimensional wavelets case, where the wavelet $\psi_{a,b} = a^{-n/2}\psi\left(\frac{x-b}{a}\right)$ with $a > 0$ and $b \in \mathbf{R}^n$, the analogous inner product $\langle f, \psi_{a,b}\rangle$ corresponds to the average of $\hat{f}\hat{\psi}_a$ over the whole frequency domain, emphasizing the dyadic corona $\{\xi : 2^j \le |\xi| \le 2^{j+1}\}$.

Now, the underlying object $\hat{f}$ must certainly satisfy specific smoothness conditions in order for its integrals on dyadic segments to make sense. Equivalently, in the original domain f must decay sufficiently rapidly at ∞. In this paper, we take for our decay condition that f be compactly supported so that $\hat{f}$ is band limited. From now on, we will only consider functions supported on the unit cube $Q = \{x \in \mathbf{R}^n, \|x\|_\infty \le 1\}$ with $\|x\|_\infty = \max_i |x_i|$. Thus $\mathcal{H} = L^2(Q)$.

Guided by the Littlewood–Paley theory, we choose to discretize the scale parameter a as $\{a_0^j\}_{j \ge j_0}$ ($a_0 > 1$, j_0 being the coarsest scale) and the location parameter b as $\{kb_0a_0^{-j}\}_{k, j \ge j_0}$. Our discretization of the sphere will also depend on the scale: the finer the scale, the finer the sampling over $\mathbf{S}^{n-1}$. At scale a_0^j, our discretization of the sphere, denoted Σ_j, is an ε_j-net of $\mathbf{S}^{n-1}$ with $\varepsilon_j = \epsilon_0 a_0^{-(j-j_0)}$ for some $\epsilon_0 > 0$. We assume that for any $j \ge j_0$, the sets Σ_j satisfy the following *equidistribution property:* two constants k_n, $K_n > 0$ must exist s.t. for any $u \in \mathbf{S}^{n-1}$ and r such that $\epsilon_j \le r \le 2$,

$$k_n\left(\frac{r}{\varepsilon_j}\right)^{n-1} \le |\{B_u(r) \cap \Sigma_j\}| \le K_n\left(\frac{r}{\varepsilon_j}\right)^{n-1}. \tag{12}$$

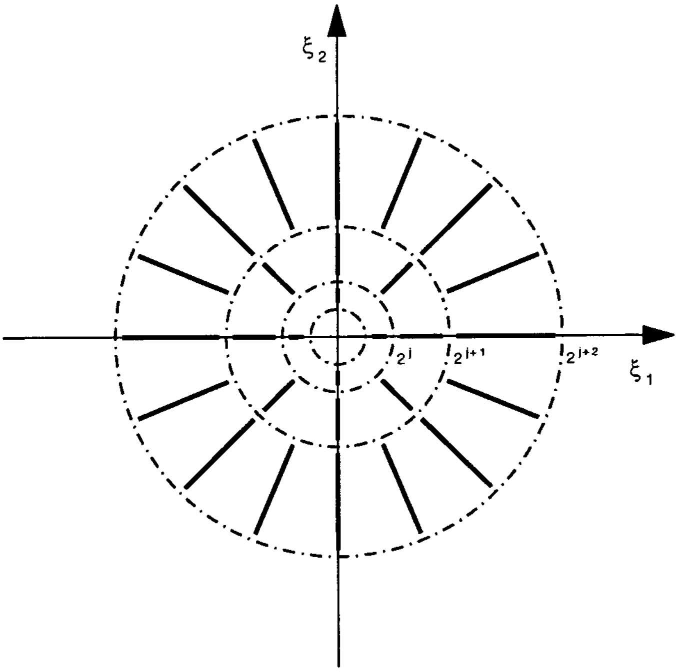

FIG. 1. Diagram schematically illustrating the ridgelet discretization of the Frequency plane (two-dimensional case). The circles represent the scales 2^j (we have chosen $a_0 = 2$) and the different segments essentially correspond to the support of different coefficient functionals. There are more segments at finer scales.

On the other hand, if $r \le \epsilon_j$, then from $B_u(r) \subset B_u(\epsilon_j)$ and the above display, $|\{B_u(r) \cap \Sigma_j\}| \le K_n$. Furthermore, the number of points N_j satisfies $k_n\left(\frac{2}{\varepsilon_j}\right)^{n-1} \le N_j \le K_n\left(\frac{2}{\varepsilon_j}\right)^{n-1}$. Essentially, our condition guarantees that Σ_j is a collection of N_j almost equispaced points on the sphere $\mathbf{S}^{n-1}$, N_j being of order $a_0^{(j-j_0)(n-1)}$. The discrete collection of ridgelets is then given by

$$\psi_\gamma(x) = a_0^{j/2}\psi(a_0^j\langle u, x\rangle - kb_0), \qquad \gamma \in \Gamma_d = \{(a_0^j, u, kb_0a_0^j), j \ge j_0, u \in \Sigma_j, k \in \mathbf{Z}\}. \tag{13}$$

In our construction, the coarsest scale is determined by the dimension of the space $\mathbf{R}^n$. Defining d as $\sup\left\{\frac{\pi}{2k}, k \in N \text{ and } \frac{\pi}{2k} < \frac{\log 2}{2n}\right\}$, we choose j_0 s.t. $a_0^{j_0+1} \le d < a_0^{j_0+2}$. Finally, we will set $\epsilon_0 = \frac{1}{2}$ so that $\epsilon_j = a_0^{-(j-j_0)}/2$.

3.3. *Main Result*

We now introduce a condition that allows us to construct frames.

DEFINITION 3. The function ψ is called *frameable* if $\psi \in C^1(\mathbf{R})$ and

- $\inf_{1\le|\xi|\le a_0} \sum_{j\ge 0} |\hat{\psi}(a_0^{-j}\xi)|^2 |a_0^{-j}\xi|^{-(n-1)} > 0$;
- $|\hat{\psi}(\xi)| \le C|\xi|^\alpha(1 + |\xi|)^{-\gamma}$, where $\alpha > \frac{n-1}{2}$, $\gamma > 2 + \alpha$.

This type of condition bears a resemblance to conditions in the theory of wavelet frames (compare, for example, [7, p. 55]). In addition, this condition looks like a discrete version of the admissible neural activation condition described in the previous section.

There are many frameable ψ. For example, sufficiently high derivatives (larger than $n/2 + 1$) of the sigmoid are frameable.

THEOREM 4 (Existence of Frames). *Let ψ be frameable. Then there exists $b_0^* > 0$ so that for any $b_0 < b_0^*$, we can find two constants $A, B > 0$ (depending on ψ, a_0, b_0, and n) so that, for any $f \in L^2(Q)$ (where Q denotes the unit cube of $\mathbf{R}^n$),*

$$A\|f\|_2^2 \le \sum_{\gamma\in\Gamma_d} |\langle f, \psi_\gamma\rangle|^2 \le B\|f\|_2^2. \tag{14}$$

The theorem is proved in several steps. We first show:

LEMMA 1.

$$\left|\sum_{\gamma\in\Gamma_d} |\langle f, \psi_\gamma\rangle|^2 - \frac{1}{2\pi b_0}\int_{\mathbf{R}} \sum_{j\ge j_0, u\in\Sigma_j} |\hat{f}(\xi u)|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi\right|$$
$$\le \frac{1}{2\pi}\sqrt{\int_{\mathbf{R}} \sum_{j\ge j_0, u\in\Sigma_j} |\hat{f}(\xi u)|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi} \sqrt{\int_{\mathbf{R}} \sum_{j\ge j_0, u\in\Sigma_j} |\hat{f}(\xi u)|^2 |a_0^{-j}\xi|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi}. \tag{15}$$

The argument is a simple application of the analytic principle of the large sieve [21]. Note that it presents an alternative to Daubechies' proof of one-dimensional dyadic affine frames [7]. We first recall an elementary lemma that we state without proof.

LEMMA 2. *Let f be a real-valued function in $C^1[0, \delta]$ for some $\delta > 0$: then,*

$$\left| f(\delta/2) - \frac{1}{\delta}\int_0^\delta f(x)dx \right| \leq \frac{1}{2}\int_0^\delta |f'(x)|dx.$$

Again, let $\psi_j(x)$ be $a_0^{j/2}\psi(a_0^j x)$. The ridgelet coefficient is then $\langle f, \psi_\gamma\rangle = (P_u f * \psi_j)(kb_0 a_0^{-j})$. For simplicity we denote $F_j = |P_u f * \psi_j|^2$. Applying the lemma gives

$$\left| F_j(kb_0a_0^{-j}) - \frac{a_0^j}{b_0}\int_{(k-1/2)b_0a_0^{-j}}^{(k+1/2)b_0a_0^{-j}} F_j(b)db \right| \leq \frac{1}{2}\int_{(k-1/2)b_0a_0^{-j}}^{(k+1/2)b_0a_0^{-j}} |F_j'(b)|db.$$

Now, we sum over k:

$$\left| \sum_k |(P_u f * \psi_j)(kb_0a_0^{-j})|^2 - \frac{a_0^j}{b_0}\int_{\mathbf{R}} |(P_u f * \psi_j)(b)|^2\, db \right|$$
$$\leq \int_{\mathbf{R}} |(P_u f * \psi_j)(b)|\, |(P_u f * (\psi_j)')(b)|db \leq \|P_u f * \psi_j\|_2 \|(P_u f * (\psi_j)')\|_2.$$

Applying Plancherel, we have

$$\left| \sum_k |(P_u f * \psi_j)(kb_0a_0^{-j})|^2 - \frac{1}{2\pi b_0}\int_{\mathbf{R}} |\hat{f}(\xi u)|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi \right|$$
$$\leq \frac{1}{2\pi}\sqrt{\int_{\mathbf{R}} |\hat{f}(\xi u)|^2|\hat{\psi}(a_0^{-j}\xi)|^2 d\xi}\ \sqrt{\int_{\mathbf{R}} |\hat{f}(\xi u)|^2 |a_0^{-j}\xi|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi}.$$

Hence, if we sum the above expression over $u \in \Sigma_j$ and j and apply the Cauchy–Schwartz inequality to the right-hand side, we get the desired result. ■

We then show that there exist $A', B' > 0$ s.t. for any $f \in L^2(Q)$, we have

$$A'\|\hat{f}\|_2^2 \leq \sum_{j\geq j_0, u\in\Sigma_j}\int_{-\infty}^{\infty} |\hat{f}(\xi u)|^2|\hat{\psi}(a_0^{-j}\xi)|^2 d\xi \leq B'\|\hat{f}\|_2^2; \tag{16}$$

$$\sum_{j\geq j_0, u\in\Sigma_j}\int_{-\infty}^{\infty} |\hat{f}(\xi u)|^2 |a_0^{-j}\xi|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi \leq B'\|\hat{f}\|_2^2. \tag{17}$$

Thus, if b_0 is chosen small enough, Theorem 4 holds.

3.4. *Irregular Sampling Theorems*

Relationship (16) is, in fact, a special case of a more abstract result which holds for general multivariate entire functions of exponential type. An excellent presentation of entire functions may be found in Boas [4]. In the present section, $B_1^2(\mathbf{R}^n)$ denotes the set of square-integrable functions whose Fourier transform is supported in $[-1, 1]^n$ and $Q_a(d) = \{x, \|x - a\|_\infty \leq d\}$, the cube of center a and volume $(2d)^n$. Finally, let $\{z_m\}_{m\in\mathbf{Z}^n}$ be the grid on $\mathbf{R}^n$ defined by $z_m = 2dm$.

THEOREM 5. *Suppose $F \in B_1^2(\mathbf{R}^n)$ and $d < \dfrac{\log 2}{n}$ with $\dfrac{\pi}{2d}$ an integer; then $\forall a \in \mathbf{R}^n$,*

$$\sum_{m\in\mathbf{Z}^n} \min_{Q_{a+z_m}(d)} |F(x)|^2 \geq c_d^2 \sum_{m\in\mathbf{Z}^n} \max_{Q_{a+z_m}(d)} |F(x)|^2, \tag{18}$$

where c_d can be chosen equal to $2e^{-nd} - 1$.

In fact, a more general version of this result holds for any exponent $p > 0$. (In this case, the constants d and c_d will depend on p.) The requirement that $\pi/2d$ must be an integer simplifies the proof but this assumption may be dropped.

Proof of Theorem 5. First, note that by making use of $F_a(x) = F(x - a)$, we only need to prove the result for $a = 0$. The proof is then based on the lemma stated below which is an extension to the multivariate case of a theorem of Paley and Wiener on nonharmonic Fourier series [27, p. 38]. Then with $|F(\lambda_m^-)| = \min_{Q_{z_m}(d)} |F(x)|$ (resp. $|F(\lambda_m^+)| = \max_{Q_{z_m}(d)} |F(x)|$), we have (using Lemma 3)

$$\sum_{m\in\mathbf{Z}^n} |F(\lambda_m^-)|^2 \geq (1/2d)^n(1-\rho_d)^2\|F\|_2^2 \geq \left(\frac{1-\rho_d}{1+\rho_d}\right)^2 \sum_{m\in\mathbf{Z}^n} |F(\lambda_m^+)|^2,$$

and $(1 - \rho_d)/(1 + \rho_d) = 2e^{-nd} - 1$.

LEMMA 3. *Let $F \in B_1^2(\mathbf{R}^n)$ and $\{\lambda_m\}_{m\in\mathbf{Z}^n}$ be a sequence of $\mathbf{R}^n$ such that $\sup_{m\in\mathbf{Z}^n} \|\lambda_m - m\pi\|_\infty < \dfrac{\log 2}{n}$; then*

$$(1-\rho_d)^2\pi^{-n}\|F\|_2^2 \leq \sum_{m\in\mathbf{Z}^n} |F(\lambda_m)|^2 \leq (1+\rho_d)^2\pi^{-n}\|F\|_2^2, \tag{19}$$

for $\rho_d = e^{nd} - 1 < 1$.

Proof of Lemma 3. The Polya–Plancherel theorem (see [25, p. 116]) gives that

$$\sum_{m\in\mathbf{Z}^n} |F(m\pi)|^2 = \pi^{-n}\|F\|_2^2.$$

Let k denote the usual multi-index $(k_1, \ldots, k_n)$ and let $|k| = k_1 + \cdots + k_n$, $k! = k_1! \cdots k_n!$ and $x^k = x_1^{k_1} \cdots x_n^{k_n}$. For any k, $\partial^k F$ is an entire function of type π. Moreover, Bernstein's inequality gives $\|\partial^k F\|_2 \leq \|F\|_2$; see [4, p. 211] for a proof. Since F is an entire function of exponential type, F is equal to its absolutely convergent Taylor expansion. Letting s be a constant to be specified below, we have

$$F(\lambda_m) - F(m\pi) = \sum_{|k|\geq 1} \frac{\partial^k F(m\pi)}{k!} (\lambda_m - m)^k$$
$$= \sum_{|k|\geq 1} \frac{\partial^k F(m\pi)}{k!} (\lambda_m - m)^k \frac{s^{|k|}}{s^{|k|}}.$$

Applying Cauchy–Schwarz and summing over m, we get

$$\sum_{m\in\mathbf{Z}^n} |F(\lambda_m) - F(m\pi)|^2 \leq \sum_{m\in\mathbf{Z}^n} \sum_{|k|\geq 1} \frac{|\partial^k F(m\pi)|^2}{k! s^{2|k|}} \sum_{|k|\geq 1} \frac{\|\lambda_m - m\|_\infty^{2|k|} s^{2|k|}}{k!}$$
$$\leq \sum_{|k|\geq 1} \frac{\pi^{-n}\|F\|_2^2}{k! s^{2|k|}} \sum_{|k|\geq 1} \frac{d^{2|k|} s^{2|k|}}{k!}$$
$$= \pi^{-n}\|F\|_2^2 (e^{n(1/s^2)} - 1)(e^{nd^2 s^2} - 1).$$

We choose $s^2 = \dfrac{1}{d}$. If $\rho_d = e^{nd} - 1 < 1$, then

$$\sum_{m\in\mathbf{Z}^n} |F(\lambda_m) - F(m\pi)|^2 \leq \rho_d^2 \pi^{-n} \|F\|_2^2$$

and, by the triangle inequality, the expected result follows.

Let μ be a measure on $\mathbf{R}^n$; μ will be called *d-uniform* if there exist α, $\beta > 0$ such that $\alpha \leq \mu(Q_{z_m}(d))/(2d)^n \leq \beta$. The following result is completely equivalent to the previous theorem.

COROLLARY 1. *Fix $d < \dfrac{\log 2}{n}$ with $\dfrac{\pi}{2d}$ an integer. Let $F \in B_1^2(\mathbf{R}^n)$ and μ be a d-uniform measure with bounds α, β. Then*

$$\alpha c_d \|F\|_2^2 \leq \int |F|^2 d\mu \leq \frac{\beta}{c_d} \|F\|_2^2. \tag{20}$$

3.5. Proof of the Main Result

We notice that the frameability condition implies that

(i)

$$\sup_{1\leq|\xi|\leq a_0} \sum_{j\in\mathbf{Z}} \frac{|\hat{\psi}(a_0^j \xi)|^2}{|a_0^j \xi|^{n-1}} < \infty$$

and

(ii)

$$\sup_{1\le|\xi|\le a_0} \sum_{j\ge 0} |\hat{\psi}(a_0^j\xi)|^2 < \infty,$$

and, respectively, (i′) and (ii′), where $\hat{\psi}(\xi)$ is replaced by $\xi\hat{\psi}(\xi)$.

For any measurable set A, let μ_ψ be the measure defined as

$$\mu_\psi(A) = \sum_{j\ge j_0, u\in\Sigma_j} \int |\hat{\psi}(a_0^{-j}\xi)|^2 1_A(\xi u)\, d\xi.$$

Similarly, we can define μ'_ψ by changing $\hat{\psi}(\xi)$ into $\xi\hat{\psi}(\xi)$. Then,

$$\sum_{j\ge j_0, u\in\Sigma_j} \int |\hat{f}(\xi u)|^2 |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi = \int |\hat{f}|^2 d\mu_\psi$$

and likewise for μ'_ψ.

Proposition 4. *If ψ is frameable, μ_ψ and μ'_ψ are d-uniform and therefore there exist $A', B' > 0$ s.t.* (16)–(17) *hold.*

We only give proof for the measure μ_ψ. The proof for μ'_ψ is exactly the same. Let ρu be the standard polar form of x. In this section, we will denote by $\Delta_x(r, \delta)$ the sets defined by $\Delta_x(r, \delta) = \{y = \rho' u', 0 \le \rho' - \rho \le r, \|u' - u\| \le \delta\}$. These sets are truncated cones. The proof uses the technical Lemma 4.

Lemma 4. *For ψ frameable,*

$$0 < \inf_{\|x\|\ge d} \mu_\psi\left(\Delta_x\left(d, \frac{d}{2\|x\|}\right)\right) \le \sup_{\|x\|\ge d} \mu_\psi\left(\Delta_x\left(d, \frac{d}{2\|x\|}\right)\right) < \infty$$

and respectively for μ'_ψ.

Proof. To simplify the notation, we will use ρ for $\|x\|$ and u for $x/\|x\|$. Let j_x be defined by $a_0^{-(j_x-j_0)} \le d/\rho < a_0 a_0^{-(j_x-j_0)}$. Hence, if $j \ge j_x$, $\forall \epsilon \in \{-1, 1\}$, the *equidistribution property* (12) implies that

$$k_n\left(\frac{a_0^{(j-j_0)}d}{\rho}\right)^{n-1} \le |\{B_{\epsilon u}(d/2\rho) \cap \Sigma_j\}| \le K_n\left(\frac{a_0^{(j-j_0)}d}{\rho}\right)^{n-1}.$$

We have

$$\begin{aligned}\mu_\psi(\Delta_x(d, d/2\rho)) &= \sum_{j\geq j_0\leq, u\in\Sigma_j} \int |\hat{\psi}(a_0^{-j}\xi)|^2 1_{\Delta_x(d,d/2\rho)}(\xi u)\,d\xi \\ &\geq \sum_{j\geq j_x} k_n\left(\frac{a_0^{(j-j_0)}d}{\rho}\right)^{n-1} \int_{\rho\leq|\xi|\leq\rho+d} |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi \\ &\geq k_n(a_0^{-j_0}d)^{n-1} \int_{\rho\leq|\xi|\leq\rho+d} \left(\frac{|\xi|}{\rho}\right)^{n-1} \sum_{j'\geq 0} \frac{|\hat{\psi}(a_0^{-j'}a_0^{-j_x}\xi)|^2}{|a_0^{-j'}a_0^{-j_x}\xi|^{n-1}}\,d\xi.\end{aligned}$$

Now, since by assumption, $d \leq \rho$, we have $\forall|\xi| \in [\rho, \rho + d]$, $da_0^{-(j_0+1)} \leq |a_0^{-j_x}\xi| \leq 2da_0^{-j_0}$. We recall that $da_0^{-(j_0+1)} \geq 1$. Therefore,

$$\begin{aligned}\mu_\psi(\Delta_x(d, d/2\rho)) &\geq k_n(a_0^{-j_0}d)^{n-1}2d \inf_{da_0^{-(j_0+1)}\leq|\xi|\leq 2da_0^{-j_0}} \sum_{j'\geq 0} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}} \\ &\geq k_n(a_0^{-j_0}d)^{n-1}2d \inf_{1\leq|\xi|\leq a_0} \sum_{j'\geq 0} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\sum_{j\geq j_x, u\in\Sigma_j} \int |\hat{\psi}(a_0^{-j}\xi)|^2 1_{\Delta_x(d,d/2\rho)}(\xi u)\,d\xi &\leq K_n(a_0^{-j_0}d)^{n-1}2^{n-1}2d \sup_{da_0^{-(j_0+1)}\leq|\xi|\leq 2da_0^{-j_0}} \sum_{j'\geq 0} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}} \\ &\leq K_n(a_0^{-j_0}d)^{n-1}2^{n-1}2d \sup_{1\leq|\xi|\leq a_0} \sum_{j'\in\mathbf{Z}} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}}.\end{aligned}$$

We finally consider the case of the j's s.t. $j_0 \leq j < j_x$. We recall that in this case, we have $|\{B_{\varepsilon u}(d/2\rho) \cap \Sigma_j\}| \leq K_n$, and thus

$$\begin{aligned}\sum_{j_0\leq j<j_x, u\in\Sigma_j} \int |\hat{\psi}(a_0^{-j}\xi)|^2 1_{\Delta_x(d,d/2\rho)}(\xi u)\,d\xi &\leq K_n \int_{\rho\leq|\xi|\leq\rho+d} \sum_{j_0\leq j<j_x} |\hat{\psi}(a_0^{j_x-j}a_0^{-j_x}\xi)|^2 \\ &\leq K_n 2d \sup_{da_0^{-(j_0+1)}\leq|\xi|\leq 2da_0^{-j_0}} \sum_{j'>0} |\hat{\psi}(a_0^{j'}\xi)|^2 \\ &\leq K_n 2d \sup_{1\leq|\xi|\leq a_0} \sum_{j'>0} |\hat{\psi}(a_0^{j'}\xi)|^2.\end{aligned}$$

The lemma follows. ■

Proof of Proposition 4. Now, we recall that $\{z_m\}_{m\in\mathbf{Z}^n}$ is the grid on $\mathbf{R}^n$ defined by $z_m = 2dm$ and we show that $\sup_m \mu_\psi(Q_{z_m}(d)) < \infty$ and that $\inf_m \mu_\psi(Q_{z_m}(d)) > 0$.

Again, we shall use the polar coordinates, i.e., $z_m = \rho_m u_m$. For $m \neq 0$, let z'_m be $\rho'_m u_m$ with $\rho'_m = \rho_m - d/2$. Then, we have that $\Delta_{z'_m}(d, d/2\rho'_m) = \{\rho' u' \text{ s.t. } |\rho' - \rho_m| \leq d/2, \|u' - u_m\| \leq d/2\rho'_m\} \subset B_{z_m}(d) \subset Q_{z_m}(d)$. To see the first inclusion, we can check that $\|\rho' u' - \rho_m u_m\|^2 = (\rho' - \rho_m)^2 + \rho' \rho_m \|u' - u_m\|^2$. Then we use the fact that $\rho'/\rho'_m \leq \frac{5}{3}$ and $\rho_m/\rho'_m \leq \frac{4}{3}$ to prove the inclusion.

For $m \neq 0$, let $\{x_j^{(m)}\}_{1 \leq j \leq J_m}$ with $\|x_j^{(m)}\| \geq d$ s.t. $Q_{z_m}(d) \subset \cup_{1 \leq j \leq J_m} \Delta_{x_j}^{(m)}(d, d/2\|x_j^{(m)}\|)$ and $T_{n,m}$ be the minimum number of j's such that the above inclusion is satisfied. By rescaling, we see that the numbers $T_{n,m}$ are independent of d. Moreover, it is easy to check that if δ is chosen small enough, then any set $\Delta_x(d, d/2\|x\|)$ (where again $\|x\| \geq d$) contains a ball of radius δ. (Although we do not prove it here, δ may be chosen equal to $d/2$.) Therefore, the numbers $T_{n,m}$ are bounded above and we let $T_n = \sup_{m \neq 0} T_{n,m}$. It follows that for all $m \neq 0$ $(m \in \mathbf{Z}^n)$, we have

$$0 < \inf_{\|x\| \geq d} \mu_\psi\left(\Delta_x\left(d, \frac{d}{2\|x\|}\right)\right) \leq \mu_\psi(\Delta_{z'_m}(d, d/2\rho'_m)) \leq \mu_\psi(Q_{z_m}(d))$$

$$\leq T_n \sup_{\|x\| \geq d} \mu_\psi\left(\Delta_x\left(d, \frac{d}{2\|x\|}\right)\right) < \infty.$$

Finally, we need to prove the result for the cube $Q_0(d)$. In order to do so, we need to establish two last estimates:

$$\mu_\psi(B_0(d)) = \sum_{j \geq j_0} |\Sigma_j| \int_{\{|\xi| \leq d\}} |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi$$

$$\geq k_n a_0^{(j-j_0)(n-1)} \int_{\{|\xi| \leq d\}} \sum_{j \geq j_0} |\hat{\psi}(a_0^{-j}\xi)|^2 d\xi$$

$$= k_n \int_{\{|\xi| \leq d\}} |a_0^{-j_0}\xi|^{n-1} \sum_{j' \geq 0} \frac{|\hat{\psi}(a_0^{-j'} a_0^{-j_0}\xi)|^2}{|a_0^{-j'} a_0^{-j_0}\xi|^{n-1}} d\xi$$

$$\geq k_n \int_{\{d/a_0 \leq |\xi| \leq d\}} |a_0^{-j_0}\xi|^{n-1} \sum_{j' \geq 0} \frac{|\hat{\psi}(a_0^{-j'} a_0^{-j_0}\xi)|^2}{|a_0^{-j'} a_0^{-j_0}\xi|^{n-1}} d\xi$$

$$\geq k_n 2d(1 - 1/a_0)(da_0^{-(j_0+1)})^{n-1} \inf_{da_0^{-(j_0+1)} \leq |\xi| \leq da_0^{-j_0}} \sum_{j' \geq 0} \frac{|\hat{\psi}(a_0^{-j'} a_0^{-j_0}\xi)|^2}{|a_0^{-j'} a_0^{-j_0}\xi|^{n-1}}.$$

Repeating the argument of Lemma 4 finally gives

$$\mu_\psi(B_0(d)) \geq k_n 2d(1 - 1/a_0)(da_0^{-(j_0+1)})^{n-1} \inf_{1 \leq |\xi| \leq a_0} \sum_{j' \geq 0} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}}.$$

After similar calculations, we can prove that

$$\mu_\psi(B_0(d)) \le K_n 2d(da_0^{-j_0})^{n-1} \sup_{da_0^{-(j_0+1)} \le |\xi| \le da_0^{-j_0}} \sum_{j' \ge 0} \frac{|\hat{\psi}(a_0^{-j'}\xi)|^2}{|a_0^{-j'}\xi|^{n-1}}.$$

Again let $\{x_j\}_{1 \le j \le J}$ with $\|x_j\| \ge d$ s.t. $Q_0(d) \subset \cup_{1 \le j \le J} \Delta_{x_j}(d, d/2\|x_j\|) \cup B_0(d)$ and T_n^0 be the minimum number of j's needed. We then have

$$0 < \mu_\psi(B_0(d)) \le \mu_\psi(Q_0(d)) \le \mu_\psi(B_0(d)) + T_n^0 \sup_{\|x\| \ge d} \mu_\psi\left(\Delta_x\left(d, \frac{d}{2\|x\|}\right)\right) < \infty.$$

This completes the proof of Proposition 4. ■

Although we do not prove it here, we may replace the frameability condition by one slightly weaker. For any traditional one-dimensional wavelet φ which satisfies the sufficient conditions listed in Daubechies [7, pp. 68–69], define ψ via $\hat{\psi}(\xi) \equiv \operatorname{sgn}(\xi)|\xi|^{(n-1)/2}(1 + \xi^2)^{-(n-1)/4}\hat{\varphi}(\xi)$; then Theorem 4 holds for such a ψ.

4. DISCUSSION

4.1. Quantitative Improvements

Our goal in this paper has been merely to provide a qualitative result concerning the existence of frames of ridgelets. However, quantitative refinements will undoubtedly be important for practical applications.

The coefficients a_γ in a frame expansion may be computed via a Neumann series expansion for the frame operator; see Daubechies [7]. For computational purposes, the closer the ratio of the upper and lower frame bounds to 1, the fewer terms will be needed in the Neumann series to compute a dual element within an accuracy of ϵ. Thus for computational purposes, it may be desirable to have good control of the frames bound ratio. Of course, the proof presented in Section 3 provides only crude estimates for the upper bound of the frame bound ratio. The interest of this method is that it uses general ideas, stated in Section 3.4, which may be applied in a variety of different settings. The author is confident that further detailed studies will allow proof of versions of Theorem 4 with tighter bounds. Such refinements are beyond the scope of the present study.

The redundancy of the frame that one can construct by this strategy depends heavily on the quality of the underlying “quasi-uniform” sampling of the sphere at each scale j. The construction of quasi-uniform discrete point sets on spheres has received considerable attention in the literature; see Sloane and Conway [5] and additional references given in the bibliography. Quantitative improvements of our results would follow from applying some of the known results obtained in that field.

Another area for investigation has to do with rapid calculation of groups of coefficients. Note that if the sets Σ_j for $j \ge j_0$ present some symmetries, it may not be necessary to compute $\tilde{\psi}_\gamma$ for all $\gamma \in \Gamma_d$; many dual elements would simply be translations, rotations, and rescalings of each other. This type of relationship would be important to pursue for practical applications.

4.2. *Finite Approximations*

The frame dictionary $\mathcal{D}_{\Gamma_d} = \{\psi_\gamma,\ \gamma \in \Gamma_d\}$ may be used for constructing approximations of certain kinds of multivariate functions. It would be interesting to know the "approximation space" associated to this frame, that is, the collection of multivariate functions f obeying

$$\| f - f_N\|_2 \leq CN^{-r}, \tag{21}$$

where f_N is an appropriately chosen superposition of dictionary elements

$$f_N = \sum_{i=1}^{N} \lambda_{i,N}\psi_{\gamma_{i,N}}. \tag{22}$$

Based on obvious analogies with the orthogonal basis case, one naturally expects that functions f of this type can be characterized by their frame coefficients, saying (21) is possible if, and only if, the frame coefficients $\{\alpha_\gamma\}_{\gamma\in\Gamma_d}$ belong to the Lorentz weak l^p space $l_{p,\infty}$, with $r = (1/p - \frac{1}{2})_+$. Work to establish those conditions under which the above would hold is in progress.

It would also be interesting to establish results which state that (21) is equivalent to a weak l^p condition on the frame coefficients even when the approximant (22) is not restricted to using only $\gamma \in \Gamma_d$. If one could establish that any continuous choices $\gamma_{i,N} \in \Gamma$ would still only lead to f with weak-l^p conditions on frame coefficients, then one would know that the frame system is really an effective way of obtaining high-quality nonlinear approximations.

APPENDIX

Proof of Proposition 2. Let $f, g \in L^1 \cap L^2$; then we can write

$$\int \mathcal{R}(f)(\gamma)\mathcal{R}(g)(\gamma)\mu(d\gamma) = \int \langle \tilde\psi_a * f,\ \tilde\psi_a * g\rangle \frac{da}{a^{n+1}}\,\sigma_n du = \mathrm{I}.$$

Applying Plancherel,

$$\begin{aligned}\mathrm{I} &= \frac{1}{2\pi}\int \langle \widehat{\tilde\psi_a * f},\ \widehat{\tilde\psi_a * g}\rangle \frac{da}{a^{n+1}}\,\sigma_n du\\ &= \frac{1}{2\pi}\int \hat f(\xi u)\,\hat g(\xi u)\,a|\hat\psi(a\xi)|^2\,\frac{da}{a^{n+1}}\,\sigma_n du d\xi,\end{aligned}$$

and, by Fubini, we obtain the desired result. ∎

Proof of Theorem 3. *Step 1.* Letting $\phi_\lambda(x) = \left(\frac{1}{2\pi\lambda}\right)^{n/2}\exp\left\{-\frac{\|x\|^2}{2\lambda}\right\}$ and defining f_ε^λ as

$$f_\varepsilon^\lambda = c_\psi \int_{\Gamma_\varepsilon} \langle f*\phi_\lambda, \psi_\gamma\rangle \psi_\gamma \mu(d\gamma),$$

we start proving that $f_\varepsilon^\lambda \in L^2(\mathbf{R}^n)$. Notice that $P_u(f*\phi_\lambda) = P_u f*P_u\phi_\lambda$ and $P_u\phi_\lambda(t) = 1/(2\pi\lambda)^{1/2}\exp\{-t^2/2\lambda\}$. Now

$$\mathcal{F}(P_u f*P_u\phi_\lambda)(\xi) = (\widehat{P_u f}\cdot\widehat{P_u\phi_\lambda})(\xi) = \hat{f}(\xi u)\exp\left\{-\frac{\lambda}{2}\xi^2\right\}.$$

Repeating the argument in the proof of Theorem 1, we obtain

$$f_\varepsilon^\lambda = \frac{c}{\pi}\int_{\{\xi>0\},\mathbf{S}^{n-1}} \left\{\int_{\varepsilon\le a\le\varepsilon^{-1}} \frac{da}{a^n}|\hat{\psi}(a\xi)|^2\right\}\exp\left\{i\xi\langle u, x\rangle - \frac{\lambda}{2}\xi^2\right\}\hat{f}(\xi u)\sigma_n d\xi du.$$

Note that for $\xi \neq 0$, we have $\int_\varepsilon^{\varepsilon^{-1}} |\hat{\psi}(a\xi)|^2 \frac{da}{a^n} = |\xi|^{n-1}\int_{\varepsilon|\xi|}^{\varepsilon^{-1}|\xi|} |\hat{\psi}(t)|^2\frac{dt}{t^n}$ (which we will abbreviate as $K_\psi/2|\xi|^{n-1}c_\varepsilon(|\xi|)$) and $c_\varepsilon(|\xi|)\uparrow 1$ as $\varepsilon\to 0$. After the change of variable $k = |\xi|u$, we obtain

$$f_\varepsilon^\lambda = \frac{c_\psi}{2\pi}K_\psi\int \exp\left\{i\langle k, x\rangle - \frac{\lambda\|k\|^2}{2}\right\}c_\varepsilon(\|k\|)\hat{f}(k)dk,$$

which allows the interpretation of f_ε^λ as the "conjugate" Fourier transform of an L^2 element and therefore the conclusion $f_\varepsilon^\lambda \in L^2(\mathbf{R}^n)$.

Step 2. We aim to prove that $f_\varepsilon^\lambda \to f_\varepsilon$ pointwise and in $L^2(\mathbf{R}^n)$. The dominated convergence theorem leads to

$$c_\varepsilon(\|k\|)\hat{f}(k)\exp\left\{-\frac{\lambda}{2}\|k\|^2\right\}\to c_\varepsilon(\|k\|)\hat{f}(k) \qquad \text{in } L^2(\mathbf{R}^n) \text{ as } \lambda\to 0.$$

Then by the Fourier transform isometry, we have $f_\epsilon^\lambda \to (2\pi)^{-n}\bar{F}(c_\epsilon\hat{f})$ in $L^2(\mathbf{R}^n)$. It remains to be proved that this limit, which we will abbreviate with g_ε, is indeed f_ϵ:

$$\begin{aligned}
\left| f_{\varepsilon}^{\lambda}(x) - f_{\varepsilon}(x) \right| &= c_{\psi} \int_{\Gamma_{\varepsilon}} (\langle f * \phi_{\lambda}, \psi_{\gamma} \rangle - \langle f, \psi_{\gamma} \rangle) \psi_{\gamma} \mu(d\gamma) \\
&\leq c_{\psi} \sup_{\gamma \in \Gamma_{\varepsilon}} \left| \psi_{\gamma}(x) \right| \int_{\varepsilon}^{\varepsilon^{-1}} \int_{\mathbf{S}^{n-1}} \| \tilde{\psi}_a * (P_u f * P_u \phi_{\lambda} - P_u f) \|_1 \frac{da}{a^{n+1}} \sigma_n du \\
&\leq c_{\psi} \varepsilon^{-1/2} \| \psi \|_{\infty} \int_{\varepsilon}^{\varepsilon^{-1}} \int_{\mathbf{S}^{n-1}} \| \tilde{\psi}_a \|_1 \| P_u f * P_u \phi_{\lambda} - P_u f \|_1 \frac{da}{a^{n+1}} \sigma_n du \\
&= c_{\psi} \varepsilon^{-1/2} \| \psi \|_{\infty} \int_{\varepsilon}^{\varepsilon^{-1}} \frac{da}{a^{n+1/2}} \| \psi \|_1 \int_{\mathbf{S}^{n-1}} \| P_u f * P_u \phi_{\lambda} - P_u f \|_1 \sigma_n du.
\end{aligned}$$

Then for a fixed u, $\|P_u f * P_u \phi_{\lambda} - P_u f\|_1 \to 0$ as $\lambda \to 0$ and

$$\begin{aligned}
\|P_u f * P_u \phi_{\lambda} - P_u f\|_1 &\leq \|P_u f\|_1 + \|P_u f * P_u \phi_{\lambda}\|_1 \\
&\leq 2\|P_u f\|_1 \leq 2\|f\|_1.
\end{aligned}$$

Thus by the dominated convergence theorem, $\int_{\mathbf{S}^{n-1}} \|P_u f * P_u \phi_{\lambda} - P_u f\|_1 \sigma_n du \to 0$.

From $\left| f_{\varepsilon}^{\lambda}(x) - f_{\varepsilon}(x) \right| \leq \delta(\varepsilon) \|\psi\|_{\infty} \psi\|_1 \int_{\mathbf{S}^{n-1}} \|P_u f * P_u \phi_{\lambda} - P_u f\|_1 \sigma_n du$, we obtain $\|f_{\varepsilon}^{\lambda} - f_{\varepsilon}\|_{\infty} \to 0$ as $\lambda \to 0$. Note that the convergence is in $C(\mathbf{R}^n)$ as the functions are continuous.

Finally, we get $f_{\varepsilon} = g_{\varepsilon}$ and, therefore, f_{ε} is in $L^2(\mathbf{R}^n)$ by completeness.

To show that $\|f_{\varepsilon} - f\|_2 \to 0$ as $\varepsilon \to 0$, it is necessary and sufficient to show that $\|\hat{f}_{\varepsilon} - \hat{f}\|_2 \to 0$,

$$\|\hat{f}_{\varepsilon} - \hat{f}\|_2^2 = \int |\hat{f}(k)|^2 (1 - c_{\varepsilon}(\|k\|))^2 dk.$$

Recalling that $0 \leq c_{\varepsilon} \leq 1$ and that $c_{\varepsilon} \uparrow 1$ as $\varepsilon \to 0$, the convergence follows. ■

ACKNOWLEDGMENTS

I thank David Donoho for serving as my adviser and suggesting this topic. It is a pleasure to acknowledge conversations with Ytzhak Katznelson and Iain Johnstone. Thanks to the referees, whose helpful suggestions have very much improved the clarity of the argument. This work was partially supported by NSF DMS-95-05151, AFOSR MURI-F49620-96-1-0028, and a fellowship from the D.R.E.T. (French Authority). These results were briefly described at the Montreal meeting on Spline Functions and the Theory of Wavelets, March 1996.

REFERENCES

1. A. R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Trans. Inform. Theory* **39** (1993), 930–945.
2. A. Benveniste and Q. Zhang, Wavelet networks, *IEEE Trans. Neural Networks* **3** (1992), 889–898.

3. D. Bernier and K. F. Taylor, Wavelets from square-integrable representations, *SIAM J. Math. Anal.* **27** (1996), 594–608.
4. R. P. Boas, Jr., "Entire Functions," Academic Press, New York, 1952.
5. J. H. Conway and N. J. A. Sloane, "Sphere Packings, Lattices and Groups," Springer-Verlag, New York, 1988.
6. G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems* **2** (1989), 303–314.
7. I. Daubechies, "Ten Lectures on Wavelets," SIAM, Philadelphia, 1992.
8. I. Daubechies, A. Grossmann, and Y. Meyer, Painless nonorthogonal expansions, *J. Math. Phys.* **27** (1986), 1271–1283.
9. D. L. Donoho, Unconditional bases are optimal bases for data compression and for statistical estimation, *Appl. Comput. Harmon. Anal.* **1** (1993), 100–115.
10. R. J. Duffin and A. C. Schaeffer, A class of nonharmonic Fourier Series, *Trans. Amer. Math. Soc.* **72** (1952), 341–366.
11. M. Duflo and C. C. Moore, On the regular representation of a nonunimodular locally compact group, *J. Funct. Anal.* **21** (1976), 209–243.
12. H. G. Feichtinger and K. Gröchenig, A unified approach to atomic decompositions via integrable group representations, *in* "Function Spaces and Applications (Lund, 1986)," Lecture Notes in Mathematics, Vol. 1302, Springer, Berlin/New York, 1988.
13. M. Frazier, B. Jawerth, and G. Weiss, Littlewood theory and the study of function spaces, *in* "NSF-CBMS Regional Conf. Ser. in Mathematics," Vol. 79, Amer. Math. Soc., Providence, RI, 1991.
14. M. Holschneider, Inverse Radon transforms through inverse wavelet transforms, *Inverse Problems* **7** (1991), 853–861.
15. L. K. Jones, On a conjecture of Huber concerning the convergence of projection pursuit regression. *Ann. Statist.* **15** (1987), 880–882.
16. L. K. Jones, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, *Ann. Statist.* **20** (1992), 608–613.
17. Y. Katznelson, "An Introduction to Harmonic Analysis," Wiley, New York, 1968.
18. M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks* **6** (1993), 861–867.
19. S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* **41** (1993), 3397–3415.
20. Y. Meyer, "Wavelets and Operators," Cambridge Univ. Press, Cambridge, UK, 1992.
21. H. L. Montgomery, The analytic principle of the large sieve, *Bull. Amer. Math. Soc.* **84** (1978), 547–567.
22. N. Murata, An integral representation of functions using three-layered networks and their approximation bounds, *Neural Networks* **9** (1996), 947–956.
23. Y. C. Pati and P. S. Krishnaprasad, Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations, *IEEE Trans. Neural Networks* **4** (1993), 73–85.
24. F. Peyrin, M. Zaim, and R. Goutte, Construction of wavelet decompositions for tomographic images, *J. Math. Imaging Vision* **3** (1993), 105–122.
25. M. Plancherel and G. Pólya, Fonctions entières et intégrales de Fourier multiples, *Comment. Math. Helv.* **10** (1938), 110–163.
26. G. Wagner, On a new method for constructing good point sets on spheres, *Discrete Comput. Geom.* **9** (1993), 111–129.
27. R. M. Young, "An Introduction to Nonharmonic Fourier Series," Academic Press, New York, 1980.