Invited Paper

SMOOTHNESS SPACES AND WAVELET DECOMPOSITION

RONALD A. DEVORE
Department of Mathematics
University of South Carolina
Columbia, South Carolina 29208
devore@math.scarolina.edu

BRADLEY J. LUCIER
Department of Mathematics
Purdue University
West Lafayette, Indiana 47907
lucier@math.purdue.edu

ABSTRACT

Multiresolution and wavelets are being promoted as a possible aid in various numerical applications including signal and image processing. One of the theoretical advantages of wavelet bases is that they provide a simple characterization of smoothness classes in terms of the coefficients in wavelet decompositions. In this presentation, we shall give examples of how this characterization can be used to advantage in various numerical applications.

1. WAVELET DECOMPOSITIONS

We begin by explaining what we shall mean by a wavelet. Various generalizations are possible, but the following is sufficient for our purposes. Given a function ψ defined on $\mathbb{R}^d$, we use the two elementary operations of translation and dilation to form a family of functions $\psi_{j,k} := 2^{kd/2}\psi(2^k\cdot - j)$, $j \in \mathbb{Z}^d$, $k \in \mathbb{Z}$. (The symbol ":=" means "defined by" and the $\cdot$ denotes where the argument of a function is placed; e.g., $\psi_{j,k}(x) = 2^{kd/2}\psi(2^k x - j)$.) We are interested in when such a family generated by one or several functions ψ is a basis for $L_2(\mathbb{R}^d)$.

We say that a finite collection Ψ of functions is a wavelet set if $\{\psi_{j,k} : j \in \mathbb{Z}^d,\ k \in \mathbb{Z}, \psi \in \Psi\}$ forms a stable (i.e., Riesz) basis for $L_2(\mathbb{R}^d)$. This means that each $f \in L_2(\mathbb{R}^d)$ admits the decomposition (in the sense of $L_2(\mathbb{R}^d)$ convergence)

$$f = \sum_{k\in\mathbb{Z}} \sum_{\psi\in\Psi} \sum_{j\in\mathbb{Z}^d} c_{j,k,\psi}(f)\psi_{j,k} \tag{1.1}$$

with the $c_{j,k,\psi}$ coefficient functionals. Stability of this basis means that

$$\sum_{j,k,\psi} |c_{j,k,\psi}(f)|^2 \approx \|f\|^2_{L_2(\mathbb{R}^d)} := \int_{\mathbb{R}^d} f^2(x)\,dx. \tag{1.2}$$

The notation $A \approx B$ means that the quotient A/B is bounded above and below by positive constants independent of the variables involved; in the present case independent of f.

We mention some important special cases.

1.1. Orthogonal wavelet sets.

In this case, the functions $\psi_{j,k}$, $j \in \mathbb{Z}^d$, $k \in \mathbb{Z}$, $\psi \in \Psi$, are assumed to form a complete orthonormal system for $L_2(\mathbb{R}^d)$. In one space dimension, it is usually the case that $\Psi = \{\psi\}$ is a singleton and ψ is then called an *orthogonal wavelet*. The simplest example is the Haar function $\psi := H$ where $H := \chi_{(0,1/2]} - \chi_{(1/2,1]}$, which takes the value 1 on $(0,1/2]$ and -1 on $(1/2,1]$. Other important examples of univariate orthogonal wavelets with higher smoothness were constructed by Daubechies using multiresolution, which we shall now discuss.

1.2. Multiresolution.

Mallat and Meyer (see for example [Ma]) introduced a general method, called "multiresolution," for constructing wavelet sets. It begins with a function $\phi \in L_2(\mathbb{R}^d)$ that has $L_2(\mathbb{R}^d)$ stable shifts (here and later a shift means an

0-8194-1031-4/92/$4.00

integer translate). This means that the collection of functions $\{\phi(\cdot - j) : j \in \mathbb{Z}^d\}$ forms a stable basis for the shift-invariant space $\mathcal{S} := \mathcal{S}(\phi)$ it generates. It is also assumed that ϕ should satisfy:

$$\phi(x) = \sum_{j\in\mathbb{Z}^d} a(j)\phi(2x - j). \tag{1.3}$$

with $a \in \ell_2(\mathbb{Z}^d)$. While (1.3) is a severe restriction on the function ϕ, it is nevertheless satisfied by many familiar functions, most notably the B-splines and box splines. The identity (1.3) is called a refinement equation in Computer Aided Geometric Design and forms the starting point for the construction of subdivision algorithms in that subject (see, e.g., the monograph of Cavaretta, Dahmen, and Micchelli [CDM]). Two other conditions must apply, however, and they are always satisfied if, for example, ϕ has compact support and nonvanishing integral (see de Boor, DeVore, and Ron [BDR]), which we shall always assume.

Given a space X of functions defined on $\mathbb{R}^d$, we shall denote by X^k the dilated space

$$X^k := \{ f(2^k\cdot) : f \in X \}, \quad k = 0, \pm1, \ldots. \tag{1.4}$$

The importance of the refinement equation (1.3) is that it implies $\mathcal{S} \subset \mathcal{S}^1$ or more generally by dilation

$$\mathcal{S}^k \subset \mathcal{S}^{k+1}, \quad k = 0, \pm1, \ldots. \tag{1.5}$$

This allows us to create the wavelet space $W := \mathcal{S}^1 \ominus \mathcal{S}$, which is the orthogonal complement of $\mathcal{S}$ in $\mathcal{S}^1$. Being a subspace of $\mathcal{S}^1$, it is clear that W is closed under half-integer translates. A more subtle argument shows that it is is actually closed under shifts. Therefore, its dilates

$$W^k := \mathcal{S}^{k+1} \ominus \mathcal{S}^k, \quad k = 0, \pm1, \ldots \tag{1.6}$$

are closed under $2^{-k}\mathbb{Z}^d$ translates.

We see from (1.6) that W^k is the orthogonal complement of $\mathcal{S}^k$ in the space $\mathcal{S}^{k+1}$. The conditions of multiresolution then imply the orthogonal decomposition

$$L_2(\mathbb{R}^d) = \bigoplus_{k\in\mathbb{Z}} W^k. \tag{1.7}$$

In particular, any function in W^k is orthogonal to any function in $W^{k'}$, $k \neq k'$. We find a wavelet set Ψ from this multiresolution by finding a suitable set of generators for the shift invariant space W. That is, we seek a set Ψ such that $W = \mathcal{S}(\Psi)$, where for any finite set Φ of $L_2(\mathbb{R}^d)$ functions, the shift invariant space $\mathcal{S}(\Phi)$ is defined as the $L_2(\mathbb{R}^d)$-closure of the finite linear combinations of the functions $g \in \Phi$ and their shifts. For example, if the collection of functions $\psi(\cdot - j)$, $j \in \mathbb{Z}^d$, $\psi \in \Psi$, form an orthonormal basis for W, then Ψ is easily seen to be an orthogonal wavelet set and, in particular,

$$\|f\|^2_{L_2(\mathbb{R}^d)} = \sum_{j,k,\psi} |c_{j,k,\psi}(f)|^2.$$

In the univariate case ($d = 1$), there is a simple recipe given by Mallat for finding a single function ψ that is an orthogonal generator of W. Namely, if the function ϕ has orthogonal shifts, then

$$\psi := \sum_{j\in\mathbb{Z}} (-1)^j \bar{a}(1 - j)\phi(2\cdot - j), \tag{1.8}$$

with $\bar{a}$ the conjugate of the refinement mask (1.3), is such an orthogonal generator and hence an orthogonal wavelet.

For example, if we begin with the function $\phi := \chi_{[0,1]}$, which trivially satisfies the conditions of multiresolution, then $W = \mathcal{S}(H)$ with H the Haar function. More generally, Daubechies [D] has found *compactly supported* univariate functions ϕ of arbitrary finite orders of smoothness that have orthogonal shifts and that satisfy the refinement equation (1.3) with a *finite* refinement mask a. The resulting orthogonal wavelets ψ have compact support and the same smoothness as ϕ.

It is also possible to construct univariate orthogonal wavelets from multiresolution by starting with a function ϕ that does not have orthogonal shifts. For this, we find from ϕ a new function ϕ_0 that has orthogonal shifts and $\mathcal{S}(\phi_0) = \mathcal{S}(\phi)$. By applying the multiresolution construction of wavelets to ϕ_0, we obtain an orthogonal wavelet. The disadvantage of this approach however is that in general neither ϕ_0 nor the resulting orthogonal wavelet will have compact support even if ϕ does. Thus for example, starting with ϕ as a B-spline, the resulting orthogonal wavelet ψ will not have compact support. This malady is one of the motivating factors for studying other generators ψ for W that are known as prewavelets.

1.3. Prewavelets.

A set of generators Ψ for W is a *prewavelet* set if the collection of functions $\psi(\cdot - j)$, $j \in \mathbb{Z}^d$, $\psi \in \Psi$, is $L_2(\mathbb{R}^d)$-stable (but not necessarily orthonormal). Then, the functions $\{\psi_{j,k} := 2^{k/2}\psi(2^k \cdot - j) : j \in \mathbb{Z}^d,\ k \in \mathbb{Z},\ \psi \in \Psi\}$ form a stable basis for $L_2(\mathbb{R}^d)$. While they do not form an orthonormal system, they still possess orthogonality between levels:

$$\int_{\mathbb{R}^d} \psi_{j,k}\psi_{j',k'}\,dx = 0, \quad k \neq k'.$$

There is a formula (see [CW1] or [M]) similar to (1.8) that associates to any univariate function ϕ for multiresolution a univariate prewavelet ψ. If ϕ has compact support and a finite refinement mask then ψ will also have compact support and the same smoothness as ϕ. For example, if ϕ is a cardinal B-spline of order r, then as was shown by Chui and Wang [CW], the prewavelet associated to ϕ is a cardinal spline function of order r with support in $[0, 2r]$.

1.4. Multivariate orthogonal wavelets and prewavelets.

The construction of multivariate wavelet sets is more difficult. In $\mathbb{R}^d$, a wavelet set Ψ usually consists of $2^d - 1$ functions that generate W, that is $W = \mathcal{S}(\Psi)$. The situation is complicated by the fact that there are many such sets of generators and it does not seem easy to extract a canonical set with the desired properties such as compact support or orthogonality (see [JM] or [BDR]).

On the other hand, there is one case for which it is easy to construct satisfactory wavelet sets. This occurs when the function ϕ of multiresolution is the tensor product of a univariate function η that satisfies the conditions of multiresolution in one variable. Then $\phi(x) = \eta(x_1)\cdots\eta(x_d)$, $x = (x_1, \ldots, x_d)$. If μ is a univariate prewavelet (or orthogonal wavelet) associated to η, we can obtain a wavelet set as follows. We let V denote the vertices of the cube $[0,1]^d$ and let $\eta_0 := \eta$ and $\eta_1 := \mu$. Then, the functions

$$\psi_v(x) := \prod_{i=1}^{d} \eta_{v_i}(x_i), \quad v \in V,\ v \neq 0,$$

form a wavelet set in $\mathbb{R}^d$.

In this way, one easily obtains from univariate wavelets such as the Haar function, the Daubechies wavelets, or the Chui-Wang spline prewavelets, a wavelet set in $\mathbb{R}^d$. For example, from the Haar function, we obtain (for the case $d = 2$) the two dimensional Haar wavelet set. It consists of three functions supported on the unit square $[0,1]^2$. The one corresponding to $v = (1,0)$ is 1 on the left half of the square and -1 on the right; the second corresponding to $v = (0,1)$ is 1 on the top half of the square and -1 on the bottom half; the third corresponding to $v = (1,1)$ takes the values ± 1 in a checkerboard fashion.

1.5. Decompositions in $L_p(\mathbb{R}^d)$, $p \neq 2$.

Wavelet decompositions would not be of much use if it were not for the fact that they hold in a much more general setting than the $L_2(\mathbb{R}^d)$ description given above. For example, the wavelet decomposition (1.1) will be valid in the sense of $L_p(\mathbb{R}^d)$ convergence $p > 1$, for functions $f \in L_p(\mathbb{R}^d)$, whenever the functions in the wavelet set Ψ have sufficient decay properties. (A function f is in $L_p(\mathbb{R}^d)$ whenever

$$\|f\|^p_{L_p(\mathbb{R}^d)} := \int_{\mathbb{R}}^d |f(x)|^p \, dx < \infty.)$$

Moreover, wavelet theory can be modified to deal also with the case $p \leq 1$ (see for example [DP], [DJP], [DJL]). While, we are not able to go into these variants in detail here, some use of them is made in the following discussion of wavelets and smoothness and in later numerical applications.

2. SMOOTHNESS AND WAVELET DECOMPOSITIONS

There are several ways of mathematically measuring the smoothness of a function f, such as derivatives, moduli of smoothness, and potentials. These lead to spaces of functions that gather together all functions with a common smoothness. One of the advantages of wavelets and wavelet decompositions is that it is often possible to describe these spaces solely in terms of the coefficients in wavelets decompositions. One of the main points of this paper is to indicate how such characterizations can be used to advantage in various numerical applications.

We shall restrict our discussion to the class of Besov spaces. This class is rich enough to allow the measurement of smoothness of all orders and in all L_p spaces, $0 < p \leq \infty$. The Besov spaces are usually defined in terms of moduli of smoothness. But to spare the reader this technicality, we shall proceed immediately to their equivalent definition in terms of wavelet coefficients.

We let $B^\alpha_p := B^\alpha_p(L_p(\mathbb{R}^d))$ denote the Besov space of functions with smoothness of order $\alpha > 0$ in $L_p(\mathbb{R}^d)$, $0 < p \leq \infty$. Actually, the full definition of Besov spaces involves a second fine tuning parameter q, which we are taking equal to p. This Besov space should be viewed heuristically as the set of all functions whose derivatives of order α are in $L_p(\mathbb{R}^d)$. However, as noted, the precise definition of these spaces depends on moduli of smoothness of functions and applies to all $\alpha > 0$ and all $0 < p \leq \infty$ (see [DP]).

We state now a generic theorem that characterizes the Besov spaces B^α_p in terms of wavelet coefficients. For the statement of this theorem, it is useful to use the L_p-normalized wavelets $\psi_{j,k,p} := 2^{kd/p}\psi(2^k \cdot - j)$ (we have previously used $\psi_{j,k} = \psi_{j,k,2}$).

THEOREM 2.1. *Let Ψ be a wavelet set whose functions have compact support and smoothness of order r. Then a function $f \in L_p(\mathbb{R}^d)$, $1 \leq p \leq \infty$, $0 < \alpha < r$, is in the Besov space B^α_p if and only if it has the wavelet decomposition*

$$f = \sum_{k \in \mathbb{Z}} \sum_{\psi \in \Psi} \sum_{j \in \mathbb{Z}^d} c_{j,k,\psi,p}(f) \psi_{j,k,p}$$

and

$$|f|^p_{B^\alpha_p(L_p(\mathbb{R}^d))} \approx \sum_{k \in \mathbb{Z}} 2^{k\alpha p} \sum_{\psi \in \Psi} \sum_{j \in \mathbb{Z}^d} |\gamma_{j,k,\psi,p}(f)|^p. \qquad (2.1)$$

We shall take (2.1) as the defining quasi-semi-norm for B^α_p.

A proof of this theorem can be found in Meyer [Me]. An outline for proving this theorem can be found in DeVore and Lucier [DL]. Usually, it is assumed that polynomials of order r are contained in the space $S(\phi)$. However, this automatically holds if ϕ has compact support, nonvanishing integral, and satisfies a refinement equation with a finite mask (see Theorem 8.3 of Cavaretta, Dahmen, and Micchelli [CDM]) It is also important to note that the above theorem holds also for $p < 1$ (see [K]) provided the smoothness index α is large enough that the functions in B^α_p are in some L_s space with $s > 1$ (such embeddings follow from the Sobolev embedding theorem for Besov spaces). Moreover, for other sets of functions Ψ, constructed by a variant of multiresolution, this theorem will hold for $p < 1$ and all $\alpha < r$ (see [DJP] or [DP]).

The main point of this paper is to show how Theorem 2.1 can be exploited in numerical algorithms for image and signal processing. We shall consider two canonical problems: the compression of surfaces/images and the removal of noise from images.

3. COMPRESSION

Wavelets can be used for surface/image compression as follows. We find a mathematical representation of the surface/image as a wavelet decomposition. Then, we retain a relatively small number of terms from this decomposition as the compressed surface/image. We begin by discussing how one obtains such representations of surfaces/images.

3.1. Representation of surfaces and images.

To describe how to represent surfaces and images, we return for a moment to the multiresolution construction of wavelets. Let ϕ be a generating function for multiresolution that satisfies the assumptions of multiresolution and has compact support. We have associated to ϕ, the shift-invariant space $\mathcal{S} := \mathcal{S}(\phi)$ and its dilates $\mathcal{S}^k$. One of the conditions of multiresolution implies that the spaces $\mathcal{S}^k$ fill up $L_2(\mathbb{R}^d)$ as $k \to \infty$. The same can be proved for $L_p(\mathbb{R}^d)$, $1 \leq p \leq \infty$, under suitable decay conditions on ϕ.

Let P be the orthogonal projector that maps $L_2(\mathbb{R}^d)$ onto $\mathcal{S}$ with norm one. By dilation we obtain the norm-one orthogonal projectors P_k that map $L_2(\mathbb{R}^d)$ onto $\mathcal{S}^k$, $k = 0, \pm 1, \ldots$. It follows that $P_k f \to f$, $k \to \infty$, in the sense of $L_2(\mathbb{R}^d)$, for each $f \in L_2(\mathbb{R}^d)$. We can therefore write

$$f = P_0 f + \sum_{k=0}^{\infty}(P_{k+1}f - P_k f) = P_0 f + \sum_{k=0}^{\infty}\sum_{\psi\in\Psi}\sum_{j\in\mathbb{Z}^d} c_{j,k,\psi}(f)\psi_{j,k}. \tag{3.1}$$

This representation also holds in $L_p(\mathbb{R}^d)$, $1 \leq p \leq \infty$.

It follows that if N is large enough then

$$P_N(f) = P_0(f) + \sum_{k=0}^{N}\sum_{\psi\in\Psi}\sum_{j\in\mathbb{Z}^d} c_{j,k,\psi}(f)\psi_{j,k} \tag{3.2}$$

approximates f to arbitrarily prescribed tolerance.

The decompositions (3.1) and (3.2) actually hold in a more general setting. For example, if we take $\mathcal{S} = \mathcal{S}(\phi)$ with ϕ a box spline or a B-spline and let P be any bounded projector onto this space, then we obtain the decompositions (3.2) and (3.3) with ϕ in place of ψ. In this case, we do not have a wavelet set because the functions $\phi_{j,k}$ are not even linearly independent. However, the applications discussed later in this paper remain valid in this setting. This is for example the approach used for surface compression in DeVore, Jawerth, and Lucier [DJL1].

Suppose now that we are given a surface described as the graph of a real valued function of *compact support* Ω defined on $\mathbb{R}^d$ (a similar analysis holds for parametrically defined surfaces as well). Since we are assuming that the wavelets in Ψ all have compact support, (3.2) will involve only a finite number of terms that are not identically zero on Ω. If N is large enough this sum will represent our surface to any desired accuracy. We pick such an N and call (3.2) the representation of the surface. There are a lot of nontrivial questions that can be asked concerning the numerical implementation of this representation. For example, how large do we need to choose N, how can we calculate the coefficients $c_{j,k}$ numerically, etc? We shall not answer these here but refer the reader to [DJL1] where these questions are discussed in some detail. Let us only mention that there are fast wavelet transforms to calculate the wavelet coefficients $c_{j,k,\psi}(f)$ from the coefficients in the representation of $P_N f$ as a linear combination of the functions $\phi(2^N\cdot - j)$ (see e.g. [DL]).

We can use similar ideas to come to a wavelet representation of images. In fact the situation here is a bit simpler. A digitized grey scale image can be considered as an array of pixel (picture element) values. Each pixel value p_j,

$j = (j_1, j_2)$, is an integer between 0 (white) and 255 (black) depicting the level of grey scale intensity. Typical arrays are of size $2^m \times 2^m$, with $m = 8, 9, 10$. We view the digitized image as arising from an intensity function F defined on the unit square $\Omega := [0,1]^2$. The pixel value p_j can be considered as the average of F over the square with lower left corner $2^{-m}(j_1, j_2)$ and side length 2^{-m}.

Let again ϕ be a function of compact support that admits multiresolution. We represent the image as a finite sum

$$f = \sum a_j \phi(2^m \cdot - j) \tag{3.3}$$

where the sum in (3.3) is taken over all indices j for which $\phi(2^m \cdot - j)$ does not vanish identically on Ω. The coefficients a_j are to be determined in a reasonable way from the pixel values. One possibility is to let $a_j := p_j$ for all j such that $(j_1 2^{-N}, j_2 2^{-N})$ is a point in Ω and use some extension of these values to assign coefficients corresponding to points outside of Ω. Once we have a representation (3.3) for the image, we can use the fast wavelet transform to change to the wavelet basis and obtain a representation like (3.2).

3.2. Measurement of compression error.

Before deciding how we are going to compress the wavelet decomposition (3.2), we must first decide how we are going to measure the compression error. For surfaces that arise in computer aided design, the uniform metric ($= L_\infty$ norm) is often most suitable. In image compression, the L_2 norm is often chosen by default. However, we argue in [DJL] that the L_1 norm better matches the human visual system (at least in the middle frequency range). In what follows, we shall assume that the error is to be measured in one of the L_p norms with $1 \le p \le \infty$.

3.3. A compression algorithm.

Let f denote the sum (3.2) that is the wavelet decomposition of our representation of the surface or image under consideration. There are two equivalent ways to view the compression problem for wavelet decompositions. We can fix a prescribed error tolerance and ask to retain the fewest number of terms in the wavelet decomposition to meet this tolerance or we can prescribe the number of terms we wish to retain and ask to minimize the error over all sums with this number of terms. Numerically, one usually works with the first formulation but it is notationally simpler to deal with the second formulation. For this, we introduce the nonlinear space Σ_n consisting of all functions $S = \sum_{(j,k,\psi)\in\Lambda} a_{j,k,\psi}\psi_{j,k}$ where the cardinality of Λ is n. We note that Σ_n is not a linear space since the sets Λ vary (only there cardinality n is held fixed). The optimal compression is given by the solution to the minimization problem

$$\sigma_n(f)_p := \inf_{S\in\Sigma_n} \|f - S\|_{L_p} \tag{3.4}$$

This is a nonlinear approximation problem that was studied by DeVore, Jawerth, and Popov [DJP] and the algorithms discussed here are based on their results.

If we choose to measure the compression error with the L_2-norm, and if we are using an orthogonal wavelet set, then the solution to the minimization problem (3.4) is obvious. We should choose the n largest terms in the wavelet decomposition. Indeed this minimizes the error, which is the square root of the sum of the remaining coefficients squared. We should note the following bonus. We were willing to take any coefficients in our approximant S from Σ_n but we find that the best approximant has its coefficient as the wavelet coefficients of f.

It is perhaps surprising and certainly very useful that the simple strategy described above for the case $p = 2$ has an analogue for $p \neq 2$ and for more general wavelets (not necessarily orthogonal) that is near optimal in a sense to be made more precise below. To describe this algorithm, let $\Lambda^* := \Lambda^*(f,p)$ be the set of n values (j,k,ψ) appearing in (3.2) for which

$$\|c_{j,k,\psi}\psi_{j,k}\|_{L_p(\mathbb{R}^d)} \quad \text{is largest} \tag{3.5}$$

(breaking a tie in an arbitrary manner) and let $S^* := S^*(f) := \sum_{(j,k,\psi)\in\Lambda^*} c_{j,k,\psi}(f)\psi_{j,k}$. Then the error for approximation by S^* is given by

$$\tilde{\sigma}_n(f)_p := \|f - S^*\|_{L_p}.$$

decide how good this algorithm is we would like to compare $\bar{\sigma}_n(f)_p$ and $\sigma_n(f)_p$. For an individual f, it is impossible make comparisons. Therefore, we shall instead look at classes of surfaces/images that can be compressed well and : how these two algorithms compare for such a class.

1. Classification of surfaces/images.

It is natural to classify surfaces/images by how well they can be compressed. Let $\alpha > 0$ and let $\mathcal{A}(\alpha, p)$ denote ɜ class of those $f \in L_p(\mathbb{R}^d)$ for which

$$\sigma_n(f)_p = O(n^{-\alpha/d}). \tag{3.6}$$

ıalogously, we have the class $\bar{\mathcal{A}}(\alpha, p)$ obtained by replacing σ_n by $\bar{\sigma}_n$. Clearly, the largest the value of α, the better : can compress f (at least asymptotically as $n \to \infty$). We would like to be able to decide which f are in $\mathcal{A}(\alpha, p)$. hile we can come close to doing this, we have to go to a slightly altered class to obtain an exact characterization.

For each $\alpha > 0$, $1 \le p \le \infty$ and $0 < \tau \le \infty$, we let $\mathcal{A}(\alpha, p, \tau)$ be the class of all function in $L_p(\mathbb{R}^d)$ that satisfy

$$\sum_{n=1}^{\infty} [n^{\alpha/d} \sigma_n(f)_p]^{\tau} \frac{1}{n} < \infty \tag{3.7}$$

ith the usual change when $\tau = \infty$; we use a similar definition to define $\bar{\mathcal{A}}(\alpha, p, \tau)$. The sets $\mathcal{A}(\alpha, p, \tau)$ are a ıneralization of (3.6) . Indeed, $\mathcal{A}(\alpha, p, \infty) = \mathcal{A}(\alpha, p)$ and for other τ, the sets get slightly smaller as τ decreases. ɔr each α and p, there is one value of τ for which we can characterize the class $\mathcal{A}(\alpha, p, \tau)$. Namely, for $1 \le p \le \infty$,

$$\mathcal{A}(\alpha, p, \tau) = B^{\alpha}_{\tau}, \quad \text{provided} \quad \tau := (\alpha/d + 1/p)^{-1}. \tag{3.8}$$

To prove (3.8) (see [DJP] for the case Ψ is a singleton), one utilizes the characterization of Besov spaces by wavelet ɔefficients given in Theorem 2.1. We cannot go into the proof in detail but only make the following remarks about ıe direct portion of the proof (i.e. estimating $\sigma_n(f)_p$ given that $f \in B^{\alpha}_{\tau}$). From Theorem 2.1, we have that a ınction f is in B^{α}_{τ} if and only if

$$\sum_{k \in \mathbb{Z}} \sum_{\psi \in \Psi} \sum_{j \in \mathbb{Z}^d} |c_{j,k,\psi,p}(f)|^{\tau} < \infty. \tag{3.9}$$

:learly, it would be sufficient to provide an estimate for $\sigma_n(f)_p$ in the case that the sum in (3.9) is one. Assuming his is the case, we choose as our approximant to f, the portion of the wavelet sum that corresponds to the n largest ɜrms in (3.8). The error function is then $\sum e_{j,k,\psi} \psi_{j,k,p}$ and the coefficients satisfy $|e_{j,k,\psi}| \le n^{-1/\tau}$. From this, it an be deduced that

$$\|\sum e_{j,k,\psi} \psi_{j,k}\|_p \le C n^{-\alpha/d}, \tag{3.10}$$

/hich can then be used to prove one half of the equivalence (3.8).

The equivalence (3.8) tells us precisely which images can be compressed with a compression error $\sigma_n(f)_p$ in $L_p(\mathbb{R}^d)$ hat satisfies (3.7). Namely, they correspond to $f \in B^{\alpha}_{\tau}$ with $\tau = (\alpha/d + 1/p)^{-1}$. It is important to note that the ·alue of τ is generally less than one (it is always less than one if $p = 1$). Also, functions in B^{α}_{τ} can be discontinuous. [he range of α that allows discontinuous functions increases as τ decreases. This remark is particularly important in mage compression since images by necessity correspond to discontinuous functions f. In some sense, (3.8) explains ɤhy images can be compressed. See [DJL] for a more complete discussion of the smoothness of images.

1.5. Optimal compression.

While the error function $\sigma_n(f)_p$ gives the best error we can expect in wavelet compression using n terms, there is ıo numerical recipe for finding the best approximant from Σ_n. However, the simple compression algorithm given in §3 ınd based on the selection criteria (3.5) is in a certain sense comparable with best approximation. Namely, the proof ɔf (3.8) actually showed that this algorithm achieves the error estimate for (3.8). It follows that the characterization

(3.8) also holds for the class $\tilde{\mathcal{A}}(\alpha,p,\tau)$. Thus, we must have $\mathcal{A}(\alpha,p,\tau) = \tilde{\mathcal{A}}(\alpha,p,\tau)$. In other words, the optimal algorithm and the algorithm associated to the selection (3.4) perform the same on the class B_τ^α of surfaces/images. In this sense, the algorithm of §3.3 is near optimal provided $p < \infty$. (We use the terminology "near optimal" to mean within multiplicative constants of being optimal in the sense under consideration.) In the case $p = \infty$, there is a modification of this algorithm that is near optimal (see [DPY]).

There is an even larger sense in which the algorithm of §3.3 is optimal. If we consider the class of all nonlinear algorithms for compression that are numerically stable, then a slight modification of the algorithm of §3.3 has been shown in [DKLT] to be near optimal for the classes B_s^α for any s satisfying $s \geq (\alpha/d - 1/p)^{-1}$. This latter restriction on s is necessary to have this class embedded in $L_p(\mathbb{R}^d)$. Thus, it is not only among wavelet based algorithms that we have optimality but among general nonlinear algorithms.

3.6. Numerical implementation.

We wish to make a few remarks concerning the significance of the results of §3.3 and §3.4 for numerical algorithms for compression. First of all, we have given no analysis of the size of numerical constants that arise in estimates of compression error for the algorithm of §3.3. Since image compression involves only a few dyadic levels in the wavelet decomposition, such numerical constants can be of critical importance in ascertaining which wavelets perform best in practice or more generally in analyzing how wavelet compression compares with other methods of compression such as the Discrete Cosine Transform. Our results only show that asymptotically wavelets perform within a multiplicative constant as well as any numerically stable algorithm.

We have already noted that the actual numerical implementation of the compression algorithm of §3.3 does not proceed to find the n largest coefficients (according to the criteria (3.4)). Indeed this would require expensive sorting. Instead, one chooses an error tolerance ϵ and then selects all coefficients $c_{j,k,\psi,p}(f)$ that satisfy

$$\|c_{j,k,\psi,p}(f)\|_p \geq \epsilon \tag{3.11}$$

Such a selection criteria is called thresholding by engineers. Note that it only depends on p and the dyadic level k. Thus, the algorithm of §3.3 tells how to threshold to attain near optimal compression (in the sense described above) when the error is measured in an $L_p(\mathbb{R}^d)$ norm.

While the analysis of the performance of the compression algorithm of §3.3 involves the Besov spaces, the algorithm itself is devoid of any reference to these spaces. The algorithm of §3.3 says that we should choose in our compressed wavelet approximation certain terms in the wavelet decomposition that depend on the size of the wavelet coefficients.

There is a variant to thresholding suggested by the proof of the equivalence (3.8). Namely, it is not necessary to retain the entire wavelet coefficient $c_{j,k,\psi,p}(f)$ in order to achieve the near optimal error analysis. It is only necessary that $c_{j,k,\psi,p}(f)$ be replaced by a coefficient $\bar{c}_{j,k,\psi,p}$ that satisfies

$$\|c_{j,k,\psi,p}(f) - \bar{c}_{j,k,\psi,p}\|_p < \epsilon$$

This can be viewed as follows. We retain a certain number of bits of the coefficient $c_{j,k,\psi,p}$ depending on p and the dyadic level k. As k increases, we retain fewer bits. Such strategies are called quantization by engineers. The algorithm of §3.3 tells how to perform quantization if the error is to be measured in the $L_p(\mathbb{R}^d)$ norm.

4. REMOVING NOISE FROM IMAGES

As a second application of the characterization of smoothness spaces by wavelet coefficients, we shall briefly discuss strategies for removing Gaussian noise from images. We assume that the original pixel values p_j are corrupted by Gaussian noise so that what we observe are pixel values $\tilde{p}_j = p_j + \epsilon_j$, where the ϵ_j are independent, identically distributed, normal, random variables with mean 0 and variance σ_0^2.

The fast wavelet transform that determines the wavelet coefficients $c_{j,k,\psi}$ from the pixel values $2^{-m}p_j$ is an orthonormal transformation (see, e.g., [DL]) and therefore the noisy image $\tilde{f}$ has the representation

$$\tilde{f} = \sum_{k\in\mathbb{Z}} \sum_{\psi\in\Psi} \sum_{j\in\mathbb{Z}^d} \tilde{c}_{j,k,\psi}\psi_{j,k} \tag{4.1}$$

(throughout this section, we shall use the $L_2(\mathbb{R}^d)$ normalized wavelets $\psi_{j,k}$) where $\tilde{c}_{j,k,\psi} = c_{j,k,\psi}(f) + \delta_{j,k,\psi}$ with the $\delta_{j,k,\psi}$ independent, identically distributed, normal, random variables with variance $2^{-2m}\sigma_0^2$.

Heuristically, to remove the noise from $\tilde{f}$, we would like to retain large coefficients in the wavelet expansion (4.1) since they contain the most image information and discard small coefficients since they are most likely noise. But just how the decision process should be formulated is the question. We are motivated to have the reconstructed image be smoother than the observed image since noise typically degrades smoothness. This leads us to the following general extremal problem.

We let F be the original image intensity function defined on $[0,1]^2$ (recall that we obtain the representation f from F by averaging F to form the pixel values). We consider two spaces. The first space X will measure error (by using $\|\cdot\|_X$) and the second space Y will measure smoothness (by using $\|\cdot\|_Y$). For a real number $\lambda > 0$, we seek $g = g(\lambda, X, Y) \in Y$ to solve the minimization

$$\inf_{g\in Y}[\|\tilde{f} - g\|_X + \lambda\|g\|_Y]. \tag{4.2}$$

The expression (4.2) occurs frequently in mathematical analysis and is known as a K-functional. Here, we are at liberty to choose the three quantities X, Y, and λ. After we have chosen X and Y, we look for the best λ. Our criteria in chosing these three quantities is to minimize the expected error

$$E(\|F - g\|_X).$$

As usual, we cant expect to solve this problem for a single function F. Rather, we assume that our observations $\tilde{f}$ come from functions F in a fixed function class (usually a unit ball in a smoothness space). Extremal problems of this type are well studied in statistics (see for example the monograph of Wahba [W]).

All of this is just a nice exercise unless we can numerically solve this extremal problem. However, because of the characterization of smoothness spaces in terms of wavelet coefficients, this is sometimes possible. We mention two cases discussed in [DL] in which $X = L_2(\mathbb{R}^d)$ and Y is one of the Besov spaces.

4.1. Linear estimators.

We take $Y = B_2^\beta$, for some $\beta > 0$. This is the classical Sobolev space if β is an integer. We recall that an equivalent norm for Y is given in terms of wavelet coefficients by (2.1). Thus, the minimization problem (4.2) is equivalent to the problem of finding $g = \sum_{j,k,\lambda} d_{j,k,\psi}\psi_{j,k}$ that minimizes

$$\sum_{j,k,\psi} |\tilde{c}_{j,k,\psi} - d_{j,k,\psi}|^2 + \lambda^2 \sum_{j,k,\psi} 2^{2k\beta}|d_{j,k,\psi}|^2 \tag{4.3}$$

This is a simple problem in calculus to solve (4.3) for the $d_{j,k,\psi}$. A near optimal solution is to choose $d_{j,k,\psi} = \tilde{c}_{j,k,\psi}$ provided $k \le K$ and $d_{j,k,\psi} = 0$ otherwise, with K the smallest integer for which $\lambda 2^{\beta K} \ge 1$.

How do we choose λ? For this, as noted before, we need to assume more about the underlying image intensity function F. If we assume that the F under consideration all come from some smoothness space B_2^α and

$$\|F\|_{B_2^\alpha} \le 1, \tag{4.4}$$

then an easy calculation shows that the expected error

$$E(\|F - g\|_{L_2(\mathbb{R}^d)}) \le 2^{2K-2m}\sigma_0^2 + 2^{-2\alpha K} \tag{4.5}$$

with K given as above and depending on λ. The best choice of λ is to balance the two terms in the sum (4.5), which gives the error estimate

$$E(\|F-g\|_{L_2(\mathbb{R}^d)}) \le 2(2^{-2m}\sigma_0^2)^{\alpha/(\alpha+1)}.$$

We say that this algorithm is a linear estimator since its solution is $g = P_K\tilde{f}$ where P_K is the linear projector from $L_2(\mathbb{R}^d)$ onto the space S^K. Similar problems involving minimization by splines rather than wavelet sums are considered in [W]. We reiterate that the wavelet case is much simplified by the ability to calculate norms in terms of the wavelet coefficients.

4.2. Nonlinear estimators.

Motivated by the image compression algorithms discussed in §3, we consider next the case $Y = B^\beta_\tau$ with $\tau := (\beta/d + 1/2)^{-1}$. We recall that these spaces arise in the characterization of approximation from the the nonlinear space Σ_n in the norm of $L_2(\mathbb{R}^d)$. Again, we can use the sequence norm (2.1) as an equivalent norm on Y. This reduces the minimization problem (4.2) to finding a sequence $d_{j,k,\psi}$ that minimizes

$$\sum_{j,k,\psi} |\tilde{c}_{j,k,\psi} - d_{j,k,\psi}|^2 + \lambda \sum_{j,k,\psi} |d_{j,k,\psi}|^\tau \tag{4.6}$$

A simple exercise in calculus again shows that a near optimal solution to (4.6) is to choose $d_{j,k,\psi} = \tilde{c}_{j,k,\psi}$ if $|\tilde{c}_{j,k,\psi}| > \epsilon$ and $d_{j,k,\psi} = 0$ otherwise, with $\epsilon := \lambda^{1/(2-\tau)}$. Thus, for each λ, a thresholding strategy provides a near optimal solution to our extremal problem.

Again, we can ask for the best selection of λ. For this, we assume that the images F have norm ≤ 1 in the smoothness space $B^\alpha_{\tau(\alpha)}$ with $\tau(\alpha) := (\alpha/d - 1/2)^{-1}$. The analysis here is not as simple as for the linear estimators. However, near optimal choices of λ been found by DeVore and Lucier [DL] for this class. The ensuing algorithm, with this choice of λ, yields the error estimate

$$E(\|F-g\|_{L_2(\mathbb{R}^d)}) \le C(2^{-2m}\sigma_0^2)^{\alpha/(\alpha+1)}(\log 2^{2m})^{\alpha/(\alpha+1)}.$$

This is much the same error estimate as for the linear estimator but it holds for a much larger class of functions.

The wavelet-based nonlinear noise removal algorithm given here was given first by Donoho and Johnstone [DJ] and later rediscovered by the present authors based on the K-functional approach. Donoho and Johnstone call a slight variant of this algorithm wavelet shrinkage and show that it is near optimal in several statistical senses. Their analysis was not applied to the Besov spaces $B^\alpha_{\tau(\alpha)}$ but rather to smaller Besov spaces B^α_s, $s \ge 1$, which are *compactly* embedded into $L_2(\mathbb{R}^d)$.

REFERENCES

[BDR] C. de Boor, R. DeVore and A. Ron, *Wavelets and prewavelets*, Constructive Approximation 8 (1993).

[BDR1] C. de Boor, R. DeVore and A. Ron, *Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$*, preprint.

[CDM] A. Cavaretta, W. Dahmen and C. Micchelli, *Stationary Subdivision*, Memoirs of the Amer. Math. Society 453 (1992).

[CW] C. K. Chui and J. Z. Wang, *On compactly supported spline wavelets and a duality principle*, Trans. Amer. Math. Soc. 330 (1992), 903–915.

[CW1] C. K. Chui and J. Z. Wang, *A general framework for compactly supported splines and wavelets*, Jour. of Approximation Theory (to appear).

[D] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. on Pure and Appl. Math. XLI (1988), 909–996.

[DJ] D. Donoho and I. Johnstone, *Minimax estimation by wavelet shrinkage*, preprint.

[DJL] R. DeVore, B. Jawerth, and B. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. on Information Theory 38 (1992), 719–746.

[DJL1] R. DeVore, B. Jawerth, and B. Lucier, *Surface compression*, Computer Aided Geometric Design 9 (1992), 219–239.

[DJP] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. Jour. Math. (to appear).

[DKLT] R. DeVore, G. Kyriazis, D. Leviatan, and V.M. Tikhomirov, *Compression and nonlinear n-widths*, preprint.

[DL] R. DeVore and B. Lucier, *Wavelets*, Acta Numerica 1 (1992), 1–56.

[DL1] R. DeVore and B. Lucier, *Fast wavelet techniques for near-optimal image processing*, Proceedings of MIL-COM 92 (to appear).

[DPY] R. DeVore, P. Petrushev and X.M. Yu, *Nonlinear wavelet approximation in the space* $C(\mathbb{R}^d)$, Progress in Approximation Theory: An International Perspective (Proceedings of the US-Soviet Conference on Approximation Theory, March 19–22, 1990) (A. A. Gončar and E. B. Saff, eds.), Springer Series on Computational Mathematics, vol. 19, Springer Verlag, New York, 1992, pp. 261–284.

[DP] R. DeVore and V. Popov, *Interpolation of Besov spaces*, Trans. AMS 305 (1988), 397–414.

[JM] R-Q. Jia and C. Micchelli, *Using the refinement equation for the construction of prewavelets: power of two*, Curves and Surfaces (A. LeMéhauté and L. Schumaker, eds.), Academic Press, New York, 1991, pp. 209–246.

[K] G. Kyriazis, *Wavelet decompositions and spaces of functions*, Ph.D. Thesis University of South Carolina, August, 1992.

[Ma] S. G. Mallat, *Multiresolution approximations and wavelet orthonormal bases of* $L^2(\mathbb{R})$, Trans. Amer. Math. Soc 315 (1989), 69–87.

[Me] Y. Meyer, *Ondelettes et Opérateurs I: Ondelettes*, Hermann Éditeurs, 1990.

[M] C. Micchelli, *Using the refinement equation for the construction of pre-wavelets*, Numerical Algorithms 1 (1991), 75–116.

[W] G. Wahba, *Spline Models for Observation Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59, SIAM, Philadelphia, 1990.