WAVELETS IN NUMERICAL ANALYSIS

G. BEYLKIN
University of Colorado at Boulder
Program in Applied Mathematics
Campus Box 526 Boulder, Colorado

R. COIFMAN[1]
Yale University, Department of Mathematics, P.O.Box 2155 Yale Station, New Haven, Connecticut

V. ROKHLIN[2]
Yale University, Department of Computer Science, P.O.Box 2158 Yale Station, New Haven, Connecticut

I INTRODUCTION

The use of wavelet based algorithms in numerical analysis is superficially similar to other transform methods, in which, instead of representing a vector or an operator in the usual way it is expanded in a wavelet basis, or it's matrix representation is computed in this basis. It turns out, however, that because of the localization of wavelet bases in both space and wave number domains, wavelet expansions organize transformations efficiently in terms of proximity on a given scale (wave number) and interactions between different neighbouring scales. Such organization of transformations (both linear and non-linear) has been a powerful tool in Harmonic Analysis and usually referred to as Littlewood-Paley, and Calderón-Zygmund theories (see e.g. [1]).

Initially, the relations between computation and Calderón-Zygmund theory was described in [2], [3], [4], where the Fast Multipole algorithm for

[1]Research partially supported by ONR grant N00014-88-K0020

[2]Research partially supported by ONR grants N00014-89-J1527, N00014-86-K0310 and IBM grant P00038436

computing potential interactions has made explicit many of the ingredients of Calderón-Zygmund theory. In that paper a fast algorithm of order N to compute all sums

$$p_j = \sum_{i=1}^{N} \frac{g_i g_j}{|x_i - x_j|} \quad \text{where} \quad x_i \in \mathbf{R}^3 \quad i, j = 1, \ldots, N$$

is constructed. Naively it would seem to be impossible to do this calculation in less than N^2 computations, since this is the number of interactions. It was observed that the far field effect of a cloud of charges located in a box can be described, to any accuracy, by the effect of a single multipole at the center of the box, requiring only a few numbers (Taylor coefficients of the field at the center of external boxes removed from the source). All boxes were organized in a dyadic hierarchy enabling an efficient $O(N)$ algorithm. This algorithm is $O(N)$ independently of the configuration of the charges, therefore providing a substantial improvement over FFT, even though the problem is to evaluate a convolution.

Wavelet based algorithms provide a systematic elegant generalization of the fast multipole method, in which the geometric and cancellation structure of the basis functions (which may be thought of as multipoles) provide for automatic adaptability and economy in computation [5]. Another novel aspect of transform analysis appearing naturally in connection with wavelets is the so-called non-standard form, in which a transformation is analyzed as a combination of successive contributions from different scales. We start with an initial smooth or blurred input and output vector which is then upgraded successively in input and output to higher and higher resolution, in much the same way as the pyramid scheme in image processing. This non-standard form corresponds algebraically to an imbedding of the original vector of length N into a $2N - 2$ dimensional space where all scales are uncoupled and in which the original transformation becomes sparse, followed by a projection into N dimensional space, in which scale interactions are introduced.

Even in standard form, i.e. the usual matrix realization of an operator in the wavelet basis, we gain a remarkable insight about operator compressions. In fact, already for the case of the Haar basis we see that the numerical manipulations needed to convert a given matrix to the Haar basis, involve a succesion of difference operations between neighbouring columns thus taking advantage of smoothness to reduce numerical complexity and, thereby, providing a general method for selecting an orthonormal basis for numerical compression of operators [5], [6], [7].

The class of operators which can be efficiently treated with wavelet bases

includes Calderón-Zygmund and pseudo-differential operators; these operators are well behaved under translations and dilations (they satisfy translation and scale invariant size estimates). The numerical implementations described in this paper are the beginning of a program for the conversion of pseudo-differential calculus into a numerical tool. The main idea here is the conversion of smoothing operators (error terms in pseudo-differential calculus) into sparse matrices with a small number of significant entries. Various numerical examples and applications are described in [5], [6], [8].

II PRELIMINARY REMARKS

Computing in the Haar basis, $h_{j,k}(x) = 2^{-j/2}h(2^{-j}x - k)$ $j, k \in \mathbf{Z}$, where

$$h(x) = \begin{cases} 1 & \text{for } 0 < x < 1/2 \\ -1 & \text{for } 1/2 \leq x < 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.1}$$

offers a glimpse of the algorithms that we will review.

First, we note that the decomposition of a function into the Haar basis is an order N procedure. Given $N = 2^n$ "samples" of a function, which can for simplicity be thought of as values of scaled averages

$$s_k^0 = 2^{n/2} \int_{2^{-n}k}^{2^{-n}(k+1)} f(x)dx, \tag{2.2}$$

of f on intervals of length 2^{-n}, we obtain the Haar coefficients

$$d_k^{j+1} = \frac{1}{\sqrt{2}}(s_{2k-1}^j - s_{2k}^j) \tag{2.3}$$

and averages

$$s_k^{j+1} = \frac{1}{\sqrt{2}}(s_{2k-1}^j + s_{2k}^j) \tag{2.4}$$

for $j = 0, \ldots, n-1$ and $k = 0, \ldots, 2^{n-j-1} - 1$. It is easy to see that evaluating the whole set of coefficients d_k^j, s_k^j in (2.3), (2.4) requires $2(N-1)$ additions and $2N$ multiplications.

Second, we note that in two dimensions, there are two natural ways to construct Haar systems. The first is simply the tensor product $h_{j,j',k,k'}(x,y) = h_{j,k}(x)h_{j',k'}(y)$, so that each basis function $h_{j,j',k,k'}(x,y)$ is supported on a rectangle. This basis leads to what we call the standard representation of an operator.

The second basis is defined by the set of three kinds of basis functions supported on squares: $h_{j,k}(x)h_{j,k'}(y)$, $h_{j,k}(x)\chi_{j,k'}(y)$, and $\chi_{j,k}(x)h_{j,k'}(y)$,

where $\chi(x)$ is the characteristic function of the interval $(0,1)$ and $\chi_{j,k}(x) = 2^{-j/2}\chi(2^{-j}x - k)$. This basis leads to what we call the non-standard representation of an operator (the terminology will become clear later).

Third, we note that if we consider an integral operator

$$T(f)(x) = \int K(x,y)f(y)dy, \tag{2.5}$$

and expand its kernel in a two-dimensional Haar basis we find (for a wide class of operators) that the decay of entries as a function of the distance from the diagonal is faster in these representations than that in the original kernel. This decay depends on the number of vanishing moments of the functions of the basis. The Haar functions have only one vanishing moment, $\int h(x)dx = 0$, and for this reason the gain in the decay is insufficient to make computing in the Haar basis practical.

To have a faster decay, it is necessary to use a basis in which the elements have several vanishing moments. This is accomplished by wavelets. In particular, the orthonormal bases of compactly supported wavelets constructed by I. Daubechies [9] following the work of Y. Meyer [10] and S. Mallat [11] prove to be very useful. We outline here the properties of compactly supported wavelets and refer for the details to [9] and [12].

The orthonormal basis of compactly supported wavelets of $\mathbf{L}^2(\mathbf{R})$ is formed by the dilation and translation of a single function $\psi(x)$

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \tag{2.6}$$

where $j,k \in \mathbf{Z}$. The function $\psi(x)$ has a companion, the scaling function $\varphi(x)$, and these functions satisfy the following relations:

$$\varphi(x) = \sqrt{2}\sum_{k=0}^{L-1} h_k\varphi(2x-k), \tag{2.7}$$

$$\psi(x) = \sqrt{2}\sum_{k=0}^{L-1} g_k\varphi(2x-k), \tag{2.8}$$

where

$$g_k = (-1)^k h_{L-k-1}, \qquad k = 0,\ldots,L-1 \tag{2.9}$$

and

$$\int_{-\infty}^{+\infty} \varphi(x)dx = 1. \tag{2.10}$$

In addition, the function ψ has M vanishing moments

$$\int_{-\infty}^{+\infty} \psi(x)x^m dx = 0, \qquad m = 0, \ldots, M-1. \tag{2.11}$$

The number of coefficients L in (2.7) and (2.8) is related to the number of vanishing moments M. For the wavelets in [9] $L = 2M$. If additional conditions are imposed (see [5] for an example), then the relation might be different, but L is always even.

The decomposition of a function into the wavelet basis is an order N procedure. Given the coefficients s_k^0, $k = 0, 1, \ldots, N-1$ as "samples" of the function f, the coefficients s_k^j and d_k^j on scales $j \geq 1$ are computed at a cost proportional to N via

$$s_k^j = \sum_{n=0}^{n=L-1} h_n s_{n+2k}^{j-1}, \tag{2.12}$$

and

$$d_k^j = \sum_{n=0}^{n=L-1} g_n s_{n+2k+1}^{j-1}, \tag{2.13}$$

where s_k^j and d_k^j are viewed as periodic sequences with the period 2^{n-j}.

We note that the Haar system is a degenerate case of Daubechies's wavelets. There is, however, a different way to construct orthonormal bases which generalize the Haar system and yields basis functions with several vanishing moments. We will discuss this construction in Section III in greater detail, since it approaches the problem of vanishing moments directly and does not require any prior knowledge of the wavelet theory.

To discuss the standard and non-standard representations, and compression of operators in Sections IV-VI, we use Daubechies's bases. Effectively, these representations yield two schemes for the numerical evaluation of integral operators. The first uses the standard representation and leads to numerical schemes which are, generally, of order $N \log(N)$, even for such simple operators as multiplication by a function. Another class of algorithms is obtained using the non-standard representation, which leads to numerical schemes of order N. Also, the non-standard representation leads to a proof of the celebrated "$T(1)$ theorem" of David and Journé (see [13]) (necessary and sufficient conditions for a Calderon-Zygmund operator to be bounded in $\mathbf{L}^2(\mathbf{R})$), and to uniform estimates for the error of the numerical algorithms.

The non-standard forms of many basic operators, such as fractional derivatives, Hilbert and Riesz transforms, etc., may be computed explic-

itly [8]. In Section VII we give an example of constructing the non-standard form for differential operators.

In Section VIII we show how to multiply two standard forms, and in Section IX describe a fast iterative algorithm for constructing the generalized inverse. This algorithm as well as several examples of Section X contain the beginning of the program for conversion of the pseudo-differential calculus into a numerical tool.

III BASES WITH VANISHING MOMENTS

Let us describe very simple bases for $\mathbf{L}^2([0,1])$ which are composed of functions with several vanishing moments. Our construction generalizes the Haar basis and does not require any prior knowledge of the wavelet theory [6], [14].

Using the notion of multiresolution analysis [15], [16], we define $\mathbf{V}_j^M$ to be a space of piecewise polynomial functions,

$$\mathbf{V}_j^M = \{f: \text{ the restriction of } f \text{ to the interval } (2^j n, 2^j(n+1)) \text{ is a polynomial of degree less than } M, \text{ for } n = 0, \dots, 2^{-j}-1, \text{ and } f \text{ vanishes elsewhere}\}, \tag{3.1}$$

where M is a positive integer and $j = 0, -1, -2, \dots$. The space $\mathbf{V}_j^M$ has dimension $2^{-j}M$,

$$\mathbf{V}_0^M \subset \mathbf{V}_{-1}^M \subset \cdots \subset \mathbf{V}_j^M \subset \cdots,$$

and

$$\mathbf{L}^2([0,1]) = \overline{\bigcup_{j\le 0} \mathbf{V}_j}.$$

We define the $2^{-j}M$-dimensional space $\mathbf{W}_j^M$ to be the orthogonal complement of $\mathbf{V}_j^M$ in $\mathbf{V}_{j-1}^M$,

$$\mathbf{V}_{j-1}^M = \mathbf{V}_j^M \bigoplus \mathbf{W}_j^M,$$

and obtain

$$\mathbf{L}^2([0,1]) = \mathbf{V}_0^M \bigoplus_{j\le 0} \mathbf{W}_j^M. \tag{3.2}$$

If functions $h_1, \dots, h_M : [0,1] \to \mathbf{R}$ form an orthogonal basis for $\mathbf{W}_0^M$, then the orthogonality of $\mathbf{W}_0^M$ to $\mathbf{V}_0^M$ implies that the first M moments of $h_1, \dots, h_M$ vanish,

$$\int_0^1 h_i(x)\, x^m\, dx = 0, \qquad m = 0, 1, \dots, M-1.$$

The $2M$-dimensional space $\mathbf{W}_{-1}^M$ is spanned by the $2M$ orthogonal functions $h_1(2x), \ldots, h_M(2x), h_1(2x-1), \ldots, h_M(2x-1)$, of which M are supported on the interval $[0, \frac{1}{2}]$ and M on $[\frac{1}{2}, 1]$. In general, the space $\mathbf{W}_j^M$ is spanned by $2^{-j}M$ functions obtained from $h_1, \ldots, h_M$ by translation and dilation. There is some freedom in choosing the functions $h_1, \ldots, h_M$ within the constraint that they be orthogonal; by requiring normality and additional vanishing moments, we specify them uniquely (up to sign).

First let us construct M functions $f_1, \ldots, f_M : \mathbf{R} \to \mathbf{R}$ supported on the interval $[-1, 1]$ and such that

1. The restriction of f_i to the interval $(0, 1)$ is a polynomial of degree $M - 1$.

2. The function f_i is extended to the interval $(-1, 0)$ as an even or odd function according to the parity of $i + M - 1$.

3. The functions $\{f_i\}_{i=1}^{i=M}$ are orthonormal,

$$\int_{-1}^{1} f_i(x)\, f_l(x)\, dx = \delta_{il}, \quad i, l = 1, \ldots, M.$$

4. The function f_i has vanishing moments,

$$\int_{-1}^{1} f_i(x)\, x^m\, dx = 0, \qquad m = 0, 1, \ldots, i + M - 2.$$

Properties 1 and 2 imply that there are M^2 polynomial coefficients that determine the functions $f_1, \ldots, f_M$, while properties 3 and 4 provide M^2 constraints. It turns out that the equations uncouple to give M nonsingular linear systems that may be solved to obtain the coefficients, yielding the functions uniquely (up to sign).

We now determine $f_1, \ldots, f_M$ constructively by starting with $2M$ functions which span the space of polynomials of degree less than M on the interval $(0, 1)$ and on $(-1, 0)$, then orthogonalize M of them, first to the functions $1, x, \ldots, x^{M-1}$, then to the functions $x^M, x^{M+1}, \ldots, x^{2M-1}$, and finally among themselves. We define $f_1^1, f_2^1, \ldots, f_M^1$ by the formula

$$f_m^1(x) = \begin{cases} x^{m-1}, & x \in (0, 1), \\ -x^{m-1}, & x \in (-1, 0), \\ 0, & \text{otherwise}, \end{cases}$$

and note that the $2M$ functions $1, x, \ldots, x^{M-1}, f_1^1, f_2^1, \ldots, f_M^1$ are linearly independent.

1. By the Gram-Schmidt process we orthogonalize f_m^1 with respect to $1, x, \ldots, x^{M-1}$, to obtain f_m^2, for $m = 1, \ldots, M$. This orthogonality is preserved by the remaining orthogonalizations, which only produce linear combinations of the f_m^2.

2. The following sequence of steps yields $M-1$ functions orthogonal to x^M, of which $M-2$ functions are orthogonal to x^{M+1}, and so forth, down to one function which is orthogonal to x^{2M-2}. First, if at least one of f_m^2 is not orthogonal to x^M, we reorder the functions so that it appears first, $\langle f_1^2, x^M \rangle \neq 0$. We then define $f_m^3 = f_m^2 - a_m \cdot f_0^2$ where a_m is chosen so $\langle f_m^3, x^M \rangle = 0$ for $m = 2, \ldots, M$, achieving the desired orthogonality to x^M. Similarly, we orthogonalize to $x^{M+1}, \ldots, x^{2M-2}$, each in turn, to obtain $f_1^2, f_2^3, f_3^4, \ldots, f_M^{M+1}$ such that $\langle f_m^{m+1}, x^i \rangle = 0$ for $i \leq m + M - 2$.

3. Finally, we do Gram-Schmidt orthogonalization on $f_M^{M+1}, f_{M-1}^M, \ldots, f_1^2$, in that order, and normalize to obtain $f_M, f_{M-1}, \ldots, f_1$.

It is easy to see that functions $\{f_m\}_{m=1}^{m=M}$ satisfy properties 1-4 of the previous paragraph. Defining $h_1, \ldots, h_M$ as

$$h_m(x) = 2^{1/2} f_m(2x-1), \qquad m = 1, \ldots, M,$$

we obtain

$$\mathbf{W}_j^M = \text{linear span } \{h_{m,j}^n : \quad h_{m,j}^n(x) = 2^{-j/2} h_m(2^{-j}x - n), \quad m = 1, \ldots, M;\ n = 0, \ldots, 2^{-j} - 1\}. \tag{3.3}$$

Let $\{u_m\}_{m=1}^{m=M}$ denote an orthonormal basis for $\mathbf{V}_0^M$. Combining it with (3.3) (in view of (3.2)) we obtain an orthonormal basis of $\mathbf{L}^2([0,1])$. We refer to this basis as the *multi-wavelet* basis of order M.

It is easy to see that the orthonormal set

$$\{h_{m,j}^n : \ h_{m,j}^n(x) = 2^{-j/2} h_m(2^{-j}x - n),\ m = 1, \ldots, M;\ n \in Z\}.$$

is an orthonormal basis of $\mathbf{L}^2(\mathbf{R})$.

Algorithms, various numerical examples and applications utilizing bases of this Section are described in [6] and [14].

We now outline the construction of bases for $\mathbf{L}^2[0,1]^d$ and $\mathbf{L}^2(\mathbf{R}^d)$, for any dimension d. We describe this extension by giving the basis for $\mathbf{L}^2([0,1]^2)$, which is illustrative of the construction for any finite-dimensional space. Let us define the space $\mathbf{V}_j^{M,2}$ as

$$\mathbf{V}_j^{M,2} = \mathbf{V}_j^M \times \mathbf{V}_j^M, \qquad j = 0, -1, -2, \ldots,$$

where $\mathbf{V}_j^M$ is given in (3.1), and the space $\mathbf{W}_j^{M,2}$ as the orthogonal complement of $\mathbf{V}_j^{M,2}$ in $\mathbf{V}_{j-1}^{M,2}$,

$$\mathbf{V}_{j-1}^{M,2} = \mathbf{V}_j^{M,2} \bigoplus \mathbf{W}_j^{M,2}.$$

The space $\mathbf{W}_0^{M,2}$ is spanned by the orthonormal basis

$$\{u_i(x)h_l(y),\ h_i(x)u_l(y),\ h_i(x)h_l(y):\ \ i,l=1,\ldots,M\}.$$

Among these $3M^2$ basis elements each element $v(x,y)$ has vanishing moments,

$$\int_0^1 \int_0^1 v(x,y)\, x^i\, y^l\, dx\, dy = 0, \qquad i,l = 0,1,\ldots,M-1.$$

The space $\mathbf{W}_j^{M,2}$ is spanned by dilations and translations of the $v(x,y)$ and the basis of $\mathbf{L}^2([0,1]^2)$ consists of these functions and the low-order polynomials $\{u_i(x)u_l(y):\ \ i,l=1,\ldots,M\}$.

IV THE NON-STANDARD FORM

The two-dimensional multi-wavelet basis described in Section III (with the functions of the basis supported on squares) requires $3M^2$ different combinations of one-dimensional basis functions, where M is the number of vanishing moments. On the other hand, the two-dimensional bases obtained using compactly supported wavelets [9] require only three such combinations. Thus, we will use Daubechies' bases to review the non-standard form [5].

The wavelet basis induces a multiresolution analysis on $\mathbf{L}^2(\mathbf{R})$ [15], [16], i.e., the decomposition of the Hilbert space $\mathbf{L}^2(\mathbf{R})$ into a chain of closed subspaces

$$\ldots \subset \mathbf{V}_2 \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots \tag{4.1}$$

such that

$$\bigcap_{j\in\mathbf{Z}} \mathbf{V}_j = \{0\}, \quad \overline{\bigcup_{j\in\mathbf{Z}} \mathbf{V}_j} = \mathbf{L}^2(\mathbf{R}). \tag{4.2}$$

By defining $\mathbf{W}_j$ as an orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$,

$$\mathbf{V}_{j-1} = \mathbf{V}_j \bigoplus \mathbf{W}_j, \tag{4.3}$$

the space $\mathbf{L}^2(\mathbf{R})$ is represented as a direct sum

$$\mathbf{L}^2(\mathbf{R}) = \bigoplus_{j\in\mathbf{Z}} \mathbf{W}_j. \tag{4.4}$$

On each fixed scale j, the wavelets $\{\psi_{j,k}(x)\}_{k\in\mathbf{Z}}$ form an orthonormal basis of $\mathbf{W}_j$ and the functions $\{\varphi_{j,k}(x) = 2^{-j/2}\varphi(2^{-j}x - k)\}_{k\in\mathbf{Z}}$ form an orthonormal basis of $\mathbf{V}_j$.

If there is the coarsest scale n, then the chain of the subspaces (4.1) is replaced by

$$\mathbf{V}_n \subset \ldots \subset \mathbf{V}_2 \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots, \quad \mathbf{L}^2(\mathbf{R}) = \mathbf{V}_n \bigoplus_{j\le n} \mathbf{W}_j. \tag{4.5}$$

If there are finitely many scales, then without loss of generality we set the scale $j = 0$ to be the finest scale. Instead of (4.5) we then have

$$\mathbf{V}_n \subset \ldots \subset \mathbf{V}_2 \subset \mathbf{V}_1 \subset \mathbf{V}_0, \quad \mathbf{V}_0 \subset \mathbf{L}^2(\mathbf{R}). \tag{4.6}$$

In numerical realizations the subspace $\mathbf{V}_0$ is finite dimensional.

Let T be an operator

$$T : \mathbf{L}^2(\mathbf{R}) \to \mathbf{L}^2(\mathbf{R}), \tag{4.7}$$

with the kernel $K(x, y)$. We define projection operators on the subspace $\mathbf{V}_j$, $j \in \mathbf{Z}$,

$$P_j : \mathbf{L}^2(\mathbf{R}) \to \mathbf{V}_j, \tag{4.8}$$

as follows

$$(P_j f)(x) = \sum_k \langle f, \varphi_{j,k}\rangle \varphi_{j,k}(x). \tag{4.9}$$

Expanding T in a "telescopic" series, we obtain

$$T = \sum_{j\in\mathbf{Z}} (Q_j T Q_j + Q_j T P_j + P_j T Q_j), \tag{4.10}$$

where

$$Q_j = P_{j-1} - P_j \tag{4.11}$$

is the projection operator on the subspace $\mathbf{W}_j$. If there is the coarsest scale n, then instead of (4.10) we have

$$T = \sum_{j=-\infty}^{n} (Q_j T Q_j + Q_j T P_j + P_j T Q_j) + P_n T P_n, \tag{4.12}$$

and if the scale $j = 0$ is the finest scale, then

$$T_0 = \sum_{j=1}^{n} (Q_j T Q_j + Q_j T P_j + P_j T Q_j) + P_n T P_n, \tag{4.13}$$

where $T \sim T_0 = P_0 T P_0$ is a discretization of the operator T on the finest scale.

The non-standard form is a representation (see [5]) of the operator T as a chain of triplets

$$T = \{A_j, B_j, \Gamma_j\}_{j \in \mathbf{Z}} \tag{4.14}$$

acting on the subspaces $\mathbf{V}_j$ and $\mathbf{W}_j$,

$$A_j : \mathbf{W}_j \to \mathbf{W}_j, \tag{4.15}$$

$$B_j : \mathbf{V}_j \to \mathbf{W}_j, \tag{4.16}$$

$$\Gamma_j : \mathbf{W}_j \to \mathbf{V}_j. \tag{4.17}$$

The operators $\{A_j, B_j, \Gamma_j\}_{j \in \mathbf{Z}}$ are defined as $A_j = Q_j T Q_j$, $B_j = Q_j T P_j$ and $\Gamma_j = P_j T Q_j$.

The operators $\{A_j, B_j, \Gamma_j\}_{j \in \mathbf{Z}}$ admit a recursive definition (see [5]) via the relation

$$T_j = \begin{pmatrix} A_{j+1} & B_{j+1} \\ \Gamma_{j+1} & T_{j+1} \end{pmatrix}, \tag{4.18}$$

where operators $T_j = P_j T P_j$,

$$T_j : \mathbf{V}_j \to \mathbf{V}_j. \tag{4.19}$$

If there is a coarsest scale n, then

$$T = \{\{A_j, B_j, \Gamma_j\}_{j \in \mathbf{Z}: j \le n}, T_n\}, \tag{4.20}$$

where $T_n = P_n T P_n$. If the number of scales is finite, then $j = 1, 2, \ldots, n$ in (4.20) and the operators are organized as blocks of the matrix (see Figures 1 and 2).

We will now make the following observations:

1). The map (4.15) implies that the operator A_j describes the interaction on the scale j only, since the subspace $\mathbf{W}_j$ is an element of the direct sum in (4.4).

2). The operators B_j, Γ_j in (4.16) and (4.17) describe the interaction between scale j and all coarser scales. Indeed, the subspace $\mathbf{V}_j$ contains all the subspaces $\mathbf{V}_{j'}$ with $j' > j$ (see (4.1)).

3). The operator T_j is an "averaged" version of the operator T_{j-1}.

The operators A_j, B_j and Γ_j are represented by the matrices α^j, β^j and γ^j,

$$\alpha^j_{k,k'} = \int\int K(x,y)\, \psi_{j,k}(x)\, \psi_{j,k'}(y)\, dxdy, \tag{4.21}$$

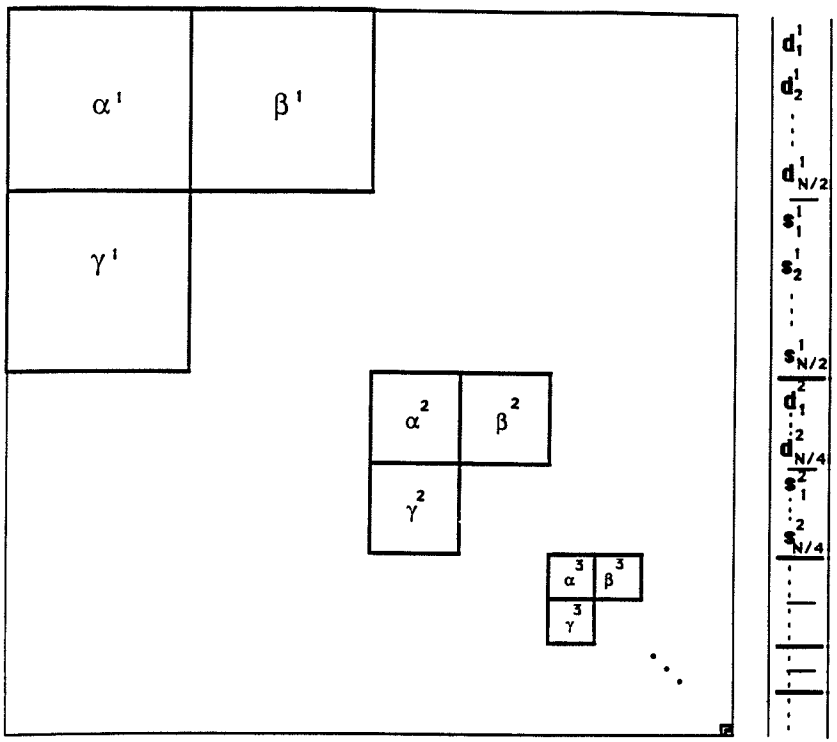

FIGURE 1. Structure of the non-standard form of a matrix. Submatrices α, β and γ on different scales are the only nonzero submatrices. In fact, most of the entries of these submatrices can be set to zero given the desired accuracy.

$$\beta^j_{k,k'} = \int\int K(x,y)\,\psi_{j,k}(x)\,\varphi_{j,k'}(y)\;dxdy, \tag{4.22}$$

and

$$\gamma^j_{k,k'} = \int\int K(x,y)\,\varphi_{j,k}(x)\,\psi_{j,k'}(y)\;dxdy. \tag{4.23}$$

The operator T_j is represented by the matrix s^j,

$$s^j_{k,k'} = \int\int K(x,y)\,\varphi_{j,k}(x)\,\varphi_{j,k'}(y)\;dxdy. \tag{4.24}$$

Given a set of coefficients $s^0_{k,k'}$ with $k,k' = 0,1,\ldots,N-1$, repeated application of the formulae (2.12), (2.13) produces

$$\alpha^j_{i,l} = \sum_{k,m=0}^{L-1} g_k g_m s^{j-1}_{k+2i,m+2l}, \tag{4.25}$$

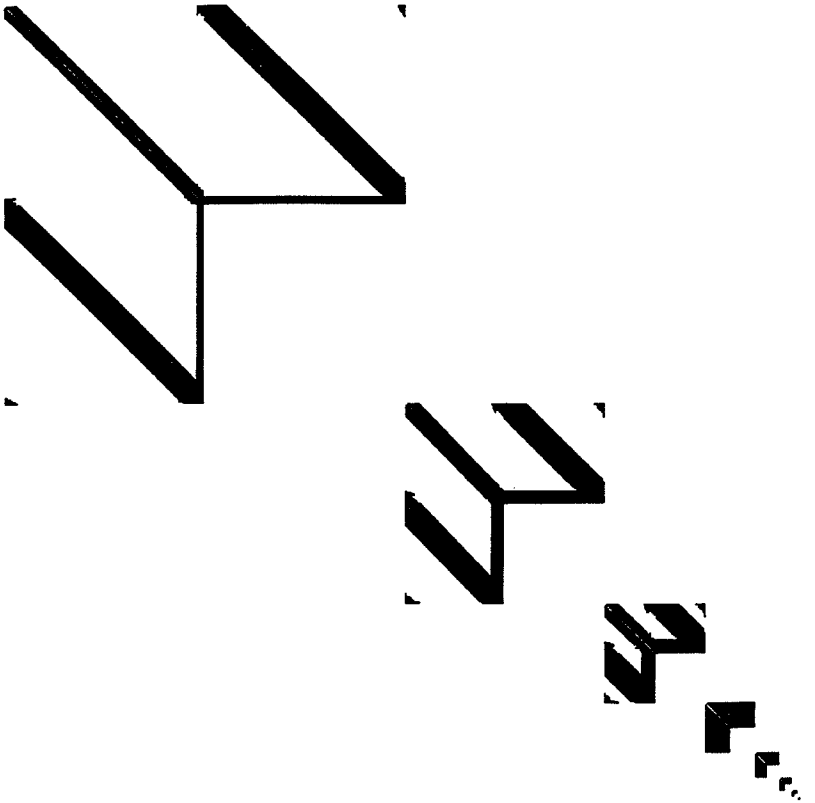

FIGURE 2. An example of a matrix in the non-standard form, $A_{ij} = 1/(i-j), i \neq j$. Entries above the threshold of 10^{-7} are shown black. Note that the width of the bands does not grow with the size of the matrix.

$$\beta^j_{i,l} = \sum_{k,m=0}^{L-1} g_k h_m s^{j-1}_{k+2i,m+2l}, \tag{4.26}$$

$$\gamma^j_{i,l} = \sum_{k,m=0}^{L-1} h_k g_m s^{j-1}_{k+2i,m+2l}, \tag{4.27}$$

$$s^j_{i,l} = \sum_{k,m=0}^{L-1} h_k h_m s^{j-1}_{k+2i,m+2l}, \tag{4.28}$$

with $i,l = 0,1,\ldots,2^{n-j}-1$, $j = 1,2,\ldots,n$. Clearly, formulae (4.25) – (4.28) provide an order N^2 scheme for the evaluation of the elements of all matrices $\alpha^j, \beta^j, \gamma^j$ with $j = 1,2\ldots,n$.

To compute the coefficients $s^0_{k,k'}$, we refer to [5], where wavelet-based quadratures for the evaluation of these coefficients are developed. Also, we refer to [5] for a fast algorithm (order N) for constructing the non-standard form for operators with known singularities, and to [8] for the direct evaluation of non-standard forms of several basic operators (see Section VII).

V THE STANDARD FORM

The standard form is obtained by representing

$$\mathbf{V}_j = \bigoplus_{j'>j} \mathbf{W}_{j'}, \tag{5.1}$$

and considering for each scale j the operators $\{B^{j'}_j, \Gamma^{j'}_j\}_{j'>j}$,

$$B^{j'}_j : \mathbf{W}_{j'} \to \mathbf{W}_j, \tag{5.2}$$

$$\Gamma^{j'}_j : \mathbf{W}_j \to \mathbf{W}_{j'}. \tag{5.3}$$

If there is the coarsest scale n, then instead of (5.1) we have

$$\mathbf{V}_j = \mathbf{V}_n \bigoplus_{j'=j+1}^{j'=n} \mathbf{W}_{j'}. \tag{5.4}$$

In this case, the operators $\{B^{j'}_j, \Gamma^{j'}_j\}$ for $j' = j+1,\ldots,n$ are as in (5.2) and (5.3) and, in addition, for each scale j there are operators $\{B^{n+1}_j\}$ and $\{\Gamma^{n+1}_j\}$,

$$B^{n+1}_j : \mathbf{V}_n \to \mathbf{W}_j, \tag{5.5}$$

$$\Gamma^{n+1}_j : \mathbf{W}_j \to \mathbf{V}_n. \tag{5.6}$$

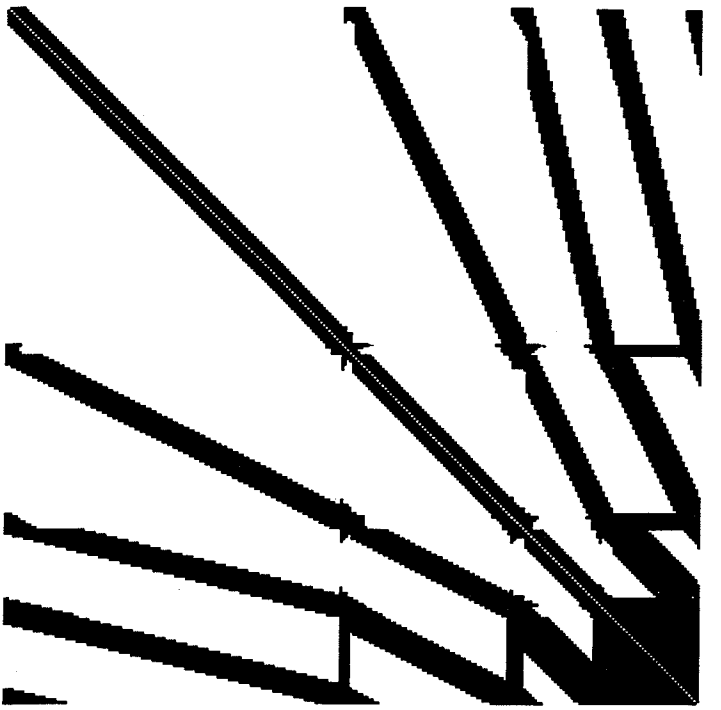

FIGURE 3. An example of a matrix in the standard form, $A_{ij} = 1/(i-j), i \neq j$. Different "finger" bands represent "interactions" between scales.

(In this notation, $\Gamma_n^{n+1} = \Gamma_n$ and $B_n^{n+1} = B_n$). If there are finitely many scales and $\mathbf{V}_0$ is finite dimensional, then the standard form is a representation of $T_0 = P_0 T P_0$ as

$$T_0 = \{A_j, \{B_j^{j'}\}_{j'=j+1}^{j'=n}, \{\Gamma_j^{j'}\}_{j'=j+1}^{j'=n}, B_j^{n+1}, \Gamma_j^{n+1}, T_n\}_{j=1,\ldots,n}. \tag{5.7}$$

The operators (5.7) are organized as blocks of the matrix (see Figure 3).

If the operator T is a Calderón-Zygmund or a pseudo-differential operator then, for a fixed accuracy, all the operators in (5.7) (except T_n) are banded. As a result, the standard form has several "finger" bands which correspond to the interaction between different scales. For a large class of operators (pseudo-differential, for example), the interaction between different scales characterized by the size of the coefficients of "finger" bands, decays as the distance $j' - j$ between the scales increases. Therefore, if the scales j and j' are well separated, then for a given accuracy, the operators $B_j^{j'}, \Gamma_j^{j'}$ can be neglected.

There are two ways of computing the standard form of a matrix. First consists in applying the one-dimensional transform (see (2.12) and (2.13)) to each column (row) of the matrix and, then, to each row (column) of the result. Alternatively, one can compute the non-standard form and then apply the one-dimensional transform to each row of all operators B^j and each column of all operators Γ_j. We refer to [5] for details.

VI COMPRESSION OF OPERATORS

If the operator T is a Calderon-Zygmund or a pseudo-differential operator, then by using the wavelet basis with M vanishing moments, we force operators $\{A_j, B_j, \Gamma_j\}_{j\in\mathbf{Z}}$ to decay roughly as $1/d^{M+1}$, where d is a distance from the diagonal. For example, let the kernel satisfy the conditions

$$|K(x,y)| \leq \frac{1}{|x-y|}, \tag{6.1}$$

$$|\partial_x^M K(x,y)| + |\partial_y^M K(x,y)| \leq \frac{C_0}{|x-y|^{1+M}} \tag{6.2}$$

for some $M \geq 1$. Then by choosing the wavelet basis with M vanishing moments, the coefficients $\alpha_{i,l}^j, \beta_{i,l}^j, \gamma_{i,l}^j$ of the non-standard form (see (4.21) - (4.23)) satisfy the estimate

$$|\alpha_{i,l}^j| + |\beta_{i,l}^j| + |\gamma_{i,l}^j| \leq \frac{C_M}{1+|i-l|^{M+1}}, \tag{6.3}$$

for all

$$|\, i-l \,| \geq 2M. \tag{6.4}$$

If, in addition to (6.1), (6.2),

$$\Big|\int_{I\times I} K(x,y)\, dxdy\Big| \leq C|I| \tag{6.5}$$

for all dyadic intervals I (this is the "weak cancellation condition", see [10]), then (6.3) holds for all i, l.

If T is a pseudo-differential operator with symbol $\sigma(x,\xi)$ defined by the formula

$$T(f)(x) = \sigma(x,D)f = \int e^{ix\xi}\, \sigma(x,\xi)\hat{f}(\xi)\, d\xi = \int K(x,y)f(y)\, dy, \tag{6.6}$$

where K is the distributional kernel of T, then assuming that the symbols σ of T and σ^* of T^* satisfy the standard conditions

$$|\, \partial_\xi^\alpha\, \partial_x^\beta\, \sigma(x,\xi)\,| \leq C_{\alpha,\beta}(1+|\,\xi\,|)^{\lambda-\alpha+\beta} \tag{6.7}$$

$$| \partial_\xi^\alpha \partial_x^\beta \sigma^*(x,\xi) | \leq C_{\alpha,\beta}(1+ | \xi |)^{\lambda-\alpha+\beta}, \tag{6.8}$$

we have the inequality

$$|\alpha_{i,l}^j| + |\beta_{i,l}^j| + |\gamma_{i,l}^j| \leq \frac{2^{\lambda j} C_M}{(1+|i-l|)^{M+1}}, \tag{6.9}$$

for all integer i, l.

Suppose now that we approximate the operator T_0 by the operator T_0^B obtained from T_0 by setting to zero all coefficients of matrices α^j, β^j and γ^j outside of bands of width $B \geq 2M$ around their diagonals. We obtain

$$\|T_0^B - T_0\| \leq \frac{C}{B^M} \log_2 N, \tag{6.10}$$

where C is a constant determined by the kernel K. In most numerical applications, the accuracy ε of calculations is fixed, and the parameters of the algorithm (in our case, the band width B and order M) have to be chosen in such a manner that the desired precision of calculations is achieved. If M is fixed, then we choose B so that

$$B \geq \left(\frac{C}{\varepsilon} \log_2 N\right)^{1/M}. \tag{6.11}$$

In other words, T_0 has been approximated to precision ε with its truncated version, which can be applied to arbitrary vectors for a cost proportional to $N\left((C/\varepsilon)\log_2 N\right)^{1/M}$, which for all practical purposes does not differ from N.

A more detailed investigation [5] permits the estimate (6.10) to be replaced with the estimate

$$\|T_0^B - T_0\| \leq \frac{C}{B^M}, \tag{6.12}$$

making the application of the operator T_0 to an arbitrary vector with arbitrary fixed accuracy into a procedure of order N. Obtaining this uniform estimate leads to a proof of

Theorem (G. David, J.L. Journé) Suppose that the operator

$$T(f) = \int K(x,y)\ f(y)\ dy \tag{6.13}$$

satisfies the conditions (6.1), (6.2), (6.5). Then a necessary and sufficient condition for T to be bounded on L^2 is that

$$\beta(x) = T(1)(x), \tag{6.14}$$

$$\gamma(y) = T^*(1)(y) \tag{6.15}$$

belong to dyadic $B.M.O.$, i.e. satisfy condition

$$\sup_J \frac{1}{|J|} \int_J |\beta(x) - m_J(\beta)|^2 dx \le C, \tag{6.16}$$

where J is a dyadic interval and

$$m_J(\beta) = \frac{1}{|J|} \int_J \beta(x) dx. \tag{6.17}$$

Again we refer to [5] for details.

VII THE OPERATOR d/dx IN WAVELET BASES

As an example, we construct the non-standard form of the operator d/dx [8]. The matrix elements α_{il}^j, β_{il}^j, and γ_{il}^j of A_j, B_j, and Γ_j, where $i, l, j \in \mathbf{Z}$ for the operator d/dx are easily computed as

$$\alpha_{il}^j = 2^{-j} \int_{-\infty}^{\infty} \psi(2^{-j}x - i)\, \psi'(2^{-j}x - l)\, 2^{-j} dx = 2^{-j} \alpha_{i-l}, \tag{7.1}$$

$$\beta_{il}^j = 2^{-j} \int_{-\infty}^{\infty} \psi(2^{-j}x - i)\, \varphi'(2^{-j}x - l)\, 2^{-j} dx = 2^{-j} \beta_{i-l}, \tag{7.2}$$

and

$$\gamma_{il}^j = 2^{-j} \int_{-\infty}^{\infty} \varphi(2^{-j}x - i)\, \psi'(2^{-j}x - l)\, 2^{-j} dx = 2^{-j} \gamma_{i-l}, \tag{7.3}$$

where

$$\alpha_l = \int_{-\infty}^{+\infty} \psi(x - l)\, \frac{d}{dx} \psi(x)\, dx, \tag{7.4}$$

$$\beta_l = \int_{-\infty}^{+\infty} \psi(x - l)\, \frac{d}{dx} \varphi(x)\, dx, \tag{7.5}$$

and

$$\gamma_l = \int_{-\infty}^{+\infty} \varphi(x - l)\, \frac{d}{dx} \psi(x)\, dx. \tag{7.6}$$

Moreover, using (2.7) and (2.8) we have

$$\alpha_i = 2 \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} g_k\, g_{k'}\, r_{2i+k-k'}, \tag{7.7}$$

$$\beta_i = 2 \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} g_k \, h_{k'} \, r_{2i+k-k'}, \tag{7.8}$$

and

$$\gamma_i = 2 \sum_{k=0}^{L-1} \sum_{k'=0}^{L-1} h_k \, g_{k'} \, r_{2i+k-k'}, \tag{7.9}$$

where

$$r_l = \int_{-\infty}^{+\infty} \varphi(x-l) \, \frac{d}{dx} \varphi(x) \, dx, \quad l \in \mathbf{Z}. \tag{7.10}$$

Therefore, the representation of d/dx is completely determined by r_l in (7.10) or in other words, by the representation of d/dx on the subspace $\mathbf{V}_0$.

Rewriting (7.10) in terms of $\hat{\varphi}(\xi)$, where

$$\hat{\varphi}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(x) \, \mathrm{e}^{\mathrm{i}x\xi} \, dx, \tag{7.11}$$

we obtain

$$r_l = \int_{-\infty}^{+\infty} |\hat{\varphi}(\xi)|^2 (\mathrm{i}\xi) \mathrm{e}^{-\mathrm{i}l\xi} \, d\xi. \tag{7.12}$$

The following proposition [8] reduces the computation of the coefficients r_l to solving a system of linear algebraic equations.

1. *If the integrals in (7.10) or (7.12) exist, then the coefficients* r_l, $l \in \mathbf{Z}$ *in (7.10) satisfy the following system of linear algebraic equations*

$$r_l = 2 \left[r_{2l} + \tfrac{1}{2} \sum_{k=1}^{L/2} a_{2k-1} (r_{2l-2k+1} + r_{2l+2k-1}) \right], \tag{7.13}$$

and

$$\sum_l l \, r_l = -1, \tag{7.14}$$

where the coefficients a_{2k-1},

$$a_{2k-1} = 2 \sum_{i=0}^{L-2k} h_i \, h_{i+2k-1}, \quad k = 1, \ldots, L/2. \tag{7.15}$$

2. *If* $M \geq 2$, *then equations (7.13) and (7.14) have a unique solution with a finite number of non-zero* r_l, *namely,* $r_l \neq 0$ *for* $-L+2 \leq l \leq L-2$

and

$$r_l = -r_{-l}, \tag{7.16}$$

Solving equations (7.13), (7.14), we present the results for Daubechies' wavelets with $M = 2, 3, 4, 5$. For further examples we refer to [8].

1. $M = 2$

$$a_1 = \frac{9}{8}, \quad a_3 = -\frac{1}{8},$$

and

$$r_1 = -\frac{2}{3}, \quad r_2 = \frac{1}{12},$$

We note, that the coefficients $(-1/12, 2/3, 0, -2/3, 1/12)$ of this example can be found in many books on numerical analysis as a choice of coefficients for numerical differentiation.

2. $M = 3$

$$a_1 = \frac{75}{64}, \quad a_3 = -\frac{25}{128}, \quad a_5 = \frac{3}{128},$$

and

$$r_1 = -\frac{272}{365}, \quad r_2 = \frac{53}{365}, \quad r_3 = -\frac{16}{1095}, \quad r_4 = -\frac{1}{2920}.$$

3. $M = 4$

$$a_1 = \frac{1225}{1024}, \quad a_3 = -\frac{245}{1024}, \quad a_5 = \frac{49}{1024}, \quad a_7 = -\frac{5}{1024},$$

and

$$r_1 = -\frac{39296}{49553}, \quad r_2 = \frac{76113}{396424}, \quad r_3 = -\frac{1664}{49553},$$

$$r_4 = \frac{2645}{1189272}, \quad r_5 = \frac{128}{743295}, \quad r_6 = -\frac{1}{1189272}.$$

4. $M = 5$

$$a_1 = \frac{19845}{16384}, \quad a_3 = -\frac{2205}{8192}, \quad a_5 = \frac{567}{8192}, \quad a_7 = -\frac{405}{32768}, \quad a_9 = \frac{35}{32768},$$

and

$$r_1 = -\frac{957310976}{1159104017}, \quad r_2 = \frac{265226398}{1159104017}, \quad r_3 = -\frac{735232}{13780629},$$

$$r_4 = \frac{17297069}{2318208034}, \quad r_5 = -\frac{1386496}{5795520085}, \quad r_6 = -\frac{563818}{10431936153},$$

$$r_7 = -\frac{2048}{8113728119}, \quad r_8 = -\frac{5}{18545664272}.$$

Remark 1. If $M = 1$, then equations (7.13) and (7.14) have a unique solution but the integrals in (7.10) or (7.12) may not be absolutely convergent. For the Haar basis ($h_1 = h_2 = 2^{-1/2}$) $a_1 = 1$ and $r_1 = -1/2$ and we obtain the simplest finite difference operator $(1/2, 0, -1/2)$. In this case the function φ is not continuous and

$$\hat{\varphi}(\xi) = \frac{1}{\sqrt{2\pi}} \frac{\sin \frac{1}{2}\xi}{\frac{1}{2}\xi} e^{i\frac{1}{2}\xi}.$$

Remark 2. For the coefficients $r_l^{(n)}$ of d^n/dx^n, $n > 1$, the system of linear algebraic equations is similar to that for the coefficients of d/dx. This system (and (7.13)) may be written in terms of

$$\hat{r}(\xi) = \sum_l r_l^{(n)} e^{il\xi}, \tag{7.17}$$

as

$$\hat{r}(\xi) = 2^n \left(|m_0(\xi/2)|^2 \hat{r}(\xi/2) + |m_0(\xi/2 + \pi)|^2 \hat{r}(\xi/2 + \pi) \right), \tag{7.18}$$

where m_0 is the 2π-periodic function

$$m_0(\xi) = 2^{-1/2} \sum_{k=0}^{k=L-1} h_k e^{ik\xi}, \tag{7.19}$$

and h_k are the wavelet coefficients. Considering the operator M_0 on 2π-periodic functions

$$(M_0 f)(\xi) = |m_0(\xi/2)|^2 f(\xi/2) + |m_0(\xi/2 + \pi)|^2 f(\xi/2 + \pi), \tag{7.20}$$

we rewrite (7.18) as

$$M_0 \hat{r} = 2^{-n} \hat{r}, \tag{7.21}$$

so that $\hat{r}$ is an eigenvector of the operator M_0 corresponding to the eigenvalue 2^{-n}. Thus, finding the representation of the derivatives in the wavelet basis is equivalent to finding trigonometric polynomial solutions of (7.21) and vice versa [8].

Remark 3. While theoretically it is well understood that the operators with homogeneous symbols have an explicit diagonal preconditioner

and $\hat{T}$. For example,

$$\Gamma_1^2 = \tilde{\Gamma}_1^2 \hat{A}_1 + \tilde{A}_2 \hat{\Gamma}_1^2 + \sum_{j'=3}^{j'=n+1} \tilde{B}_2^{j'} \hat{\Gamma}_1^{j'}. \tag{8.4}$$

If the operators $\tilde{T}$ and $\hat{T}$ are Calderón-Zygmund or pseudo-differential operators, then all the blocks of (8.1) and (8.2) (except for $\tilde{T}_n$ and $\hat{T}_n$) are banded and it is clear that Γ_1^2 is banded. This example is generic for all operators in (8.3) except for B_j^{n+1}, Γ_j^{n+1}, $(j = 1, \ldots, n)$ and T_n. The latter are dense due to the terms involving $\tilde{T}_n$ and $\hat{T}_n$. It is easy now to estimate the number of operations necessary to compute T. It takes no more than $O(N \log^2 N)$ operations to obtain T, where $N = 2^n$.

If, in addition, when the scales j and j' are well separated, the operators $B_j^{j'}, \Gamma_j^{j'}$ can be neglected for a given accuracy (as in the case of pseudo-differential operators), then the number of operations reduces asymptotically to $O(N)$.

We note, that we may set to zero all the entries of T below the threshold of accuracy and, thus, prevent the widening of the bands in the product. On denoting $\tilde{T}_\epsilon$ and $\hat{T}_\epsilon$ the approximations to $\tilde{T}$ and $\hat{T}$ obtained by setting all entries that are less than ϵ to zero, and assuming (without a loss of generality) $||\tilde{T}|| = ||\hat{T}|| = 1$, we obtain using the result of [5]

$$||\tilde{T} - \tilde{T}_\epsilon|| \leq \epsilon, \quad ||\hat{T} - \hat{T}_\epsilon|| \leq \epsilon, \tag{8.5}$$

and, therefore,

$$||\tilde{T}\hat{T} - (\tilde{T}_\epsilon \hat{T}_\epsilon)_\epsilon|| \leq \epsilon + \epsilon(1+\epsilon) + \epsilon(1+\epsilon)^2. \tag{8.6}$$

The right hand side of (8.6) is dominated by 3ϵ. For example, if we compute T^4 then we might lose one significant digit.

IX FAST ITERATIVE CONSTRUCTION OF THE GENERALIZED INVERSE

The fast multiplication algorithm of Section VIII gives a second life to a number of iterative algorithms. As an example, we consider an iterative construction of the generalized inverse. In order to construct the generalized inverse $A^\dagger$ of the matrix A we use the following result [17]:

. Let σ_1 be the largest singular value of the $m \times n$ matrix A. Consider the sequence of matrices X_k

$$X_{k+1} = 2X_k - X_k A X_k \tag{9.1}$$

with

$$X_0 = \alpha A^*, \tag{9.2}$$

where A^* is the adjoint matrix and α is chosen so that the largest eigenvalue of $\alpha A^* A$ is less than two. Then the sequence X_k converges to the generalized inverse $A^\dagger$.

When this result is combined with the fast multiplication algorithm of Section VIII, we obtain an algorithm for constructing the generalized inverse in at most $O(N \log^2 N \log R)$ operations, where R is the condition number of the matrix. (By the condition number we understand the ratio of the largest singular value to the smallest singular value above the threshold of accuracy).

The details of this algorithm (in the context of computing in wavelet bases) will be described in [18]. We note that throughout the iteration (9.1), it is necessary to maintain the "finger" band structure of the standard form of matrices X_k. Hence, the standard form of both the operator and its generalized inverse must admit such structure. We note that the pseudo-differential operators satisfy this condition.

Size $N \times N$	SVD	FWT Generalized Inverse	L_2-Error
128×128	20.27 sec.	25.89 sec.	$3.1 \cdot 10^{-4}$
256×256	144.43 sec.	77.98 sec.	$3.42 \cdot 10^{-4}$
512×512	1,155 sec. (est.)	242.84 sec.	$6.0 \cdot 10^{-4}$
1024×1024	9,244 sec. (est.)	657.09 sec.	$7.7 \cdot 10^{-4}$
...	...	...	...
$2^{15} \times 2^{15}$	9.6 years (est.)	1 day (est.)	

TABLE 3.

We now present an example. The table above contains timings and accuracy comparison of the construction of the generalized inverse via the singular value decomposition (SVD), which is $O(N^3)$ procedure, and via the iteration (9.1)–(9.2) in the wavelet basis using Fast Wavelet Transform (FWT). The computations were performed on Sun Sparc workstation and we used a routine from LINPACK for computing the singular value decomposition. For tests we used the following full rank matrix

$$A_{ij} = \begin{cases} \frac{1}{i-j} & i \neq j \\ \\ 1 & i = j \end{cases},$$

where $i, j = 1, \ldots, N$. The accuracy theshold was set to 10^{-4}, i.e., entries of X_k below 10^{-4} were systematically removed after each iteration.

X SOME PRELIMINARY RESULTS AND DIRECTIONS OF RESEARCH

In this section we describe several iterative algorithms indicating that numerical functional calculus with operators can be implemented efficiently (at least for pseudo-differential operators). Numerical results and relative performance of these algorithms will be reported separately.

Remark on iterative computation of the projection operator on the null space.

We present here a fast iterative algorithm for computing P_{null} for a wide class of operators compressible in the wavelet bases.

Let us consider the following iteration

$$X_{k+1} = 2X_k - X_k^2 \tag{10.1}$$

with

$$X_0 = \alpha A^* A, \tag{10.2}$$

where A^* is the adjoint matrix and α is chosen so that the largest eigenvalue of $\alpha A^* A$ is less than two.

Then $I - X_k$ converges to P_{null}. This can be shown either directly or by combining an invariant representation for $P_{null} = I - A^*(AA^*)^\dagger A$ with the iteration (9.1)–(9.2) to compute the generalized inverse $(AA^*)^\dagger$. The fast multiplication algorithm makes the iteration (10.1)–(10.2) fast for a wide class of operators (with the same complexity as the algorithm for the generalized inverse). The important difference is, however, that (10.1)-(10.2) does not require compressibility of the inverse operator but only of the powers of the operator.

Remark on iterative computation of a square root of an operator.

Let us describe an iteration to construct both $A^{1/2}$ and $A^{-1/2}$, where A is, for simplicity, a self-adjoint and non-negative definite operator. We consider the following iteration

$$Y_{l+1} = 2Y_l - Y_l X_l Y_l, \tag{10.3}$$

$$X_{l+1} = \frac{1}{2}(X_l + Y_l A), \tag{10.4}$$

with

$$\begin{aligned} Y_0 &= \frac{\alpha}{2}(A+I), \\ X_0 &= \frac{\alpha}{2}(A+I), \end{aligned} \tag{10.5}$$

where α is chosen so that the largest eigenvalue of $\frac{\alpha}{2}(A+I)$ is less than $\sqrt{2}$.

The sequence X_l converges to $A^{1/2}$ and Y_l to $A^{-1/2}$. By writing $A = V^*DV$, where D is a diagonal and V is a unitary, it is easy to verify that both X_l and Y_l can be written as $X_l = V^*P_lV$ and $Y_l = V^*Q_lV$, where P_l and Q_l are diagonal and

$$\begin{aligned} Q_{l+1} &= 2Q_l - Q_lP_lQ_l, \\ P_{l+1} &= \frac{1}{2}(P_l + Q_lD), \end{aligned} \tag{10.6}$$

with

$$\begin{aligned} Q_0 &= \frac{\alpha}{2}(D+I), \\ P_0 &= \frac{\alpha}{2}(D+I). \end{aligned} \tag{10.7}$$

Thus, the convergence need to be checked only for the scalar case, which we leave to the reader. If the operator A is a pseudo-differential operator, then the iteration (10.3)–(10.4) leads to a fast algorithm due to the same considerations as in the case of the generalized inverse in Section IX.

Remark on fast algorithms for exponential, sine and cosine of a matrix

The exponential of a matrix (or an operator), as well as sine and cosine functions are among the first to be considered in any calculus of operators. In this section we present a fast algorithm for computing the exponential, cosine and sine functions of a matrix. Again, as in the case of the generalized inverse, we use previously known algorithms (see e.g. [19]) which obtain completely different complexity estimates when we use them in conjunction with the wavelet representations. We do not study these algorithms in detail since our main goal in this paper is limited to pointing out the advantages of computing in the wavelet bases.

The algorithm for the exponential is based on the identity

$$\exp(A) = \left[\exp(2^{-L}A)\right]^{2^L}. \tag{10.8}$$

First, $\exp(2^{-L}A)$ is computed by, for example, using the Taylor series. The number L is chosen so that the largest singular value of $2^{-L}A$ is less than

one. At the second stage of the algorithm the matrix $2^{-L}A$ is squared L times to obtain the result.

Similarly, sine and cosine of a matrix can be computed using the elementary double-angle formulas. On denoting

$$Y_l = \cos(2^{l-L}A) \tag{10.9}$$

$$X_l = \sin(2^{l-L}A), \tag{10.10}$$

we have for $l = 0, \ldots, L-1$

$$Y_{l+1} = 2Y_l^2 - I \tag{10.11}$$

$$X_{l+1} = 2Y_l X_l, , \tag{10.12}$$

where I is the identity. Again, we choose L so that the largest singular value of $2^{-L}A$ is less than one, compute the sine and cosine of $2^{-L}A$ using the Taylor series, and then use (10.11) and (10.12).

Ordinarily, such algorithms require at least $O(N^3)$ operations, since a number of multiplications of dense matrices has to be performed [19]. Fast multiplication algorithm of Section VIII reduces complexity to not more than $O(N \log^2 N)$ operations.

To acheive such perfomance it is necessary to maintain the "finger" band structure of the standard form throughout the iteration. Whether it is possible to do depends on the particular operator and, usually, can be verified analytically.

Unlike the algorithm for the generalized inverse, the algorithms of this remark are not self-correcting. Thus, it is necessary to maintain sufficient accuracy initially so as to obtain the desired accuracy after all the multiplications have been performed.

REFERENCES

[1] Y. Meyer, Ondelettes et Operateurs, Hermann, Paris, 1990.

[2] V. Rokhlin, *Rapid Solution of Integral Equations of Classical Potential Theory*, Journal of Computational Physics, vol. 60, 2, 1985.

[3] L. Greengard and V. Rokhlin, *A Fast Algorithm for Particle Simulations*, Journal of Computational Physics, 73(1), **325**, 1987.

[4] J. Carrier, L. Greengard and V. Rokhlin *A Fast Adaptive Multipole Algorithm for Particle Simulations*, Yale University Technical Report, YALEU/DCS/RR-496 (1986), SIAM Journal of Scientific and Statistical Computing, 9 (4), 1988.

[5] G. Beylkin, R. R. Coifman and V. Rokhlin, *Fast wavelet transforms and numerical algorithms I.* Yale University Technical Report YALEU/DCS/RR-696, August 1989, Comm. on Pure and Applied Math., vol. **XLIV**, 141-183, 1991.

[6] B. Alpert, G. Beylkin, R. R. Coifman and V. Rokhlin, *Wavelets for the fast solution of second-kind integral equations*, Technical report, Department of Computer Science, Yale University, New Haven, CT, 1990.

[7] R. R. Coifman, Y. Meyer and V. Wickerhauser, *Wavelet Analysis and Signal Processing*, this volume.

[8] G. Beylkin, *On the representation of operators in bases of compactly supported wavelets*, preprint, to appear in SIAM J. on Numerical Analysis.

[9] I. Daubechies, *Orthonormal Bases of Compactly Supported Wavelets*, Comm. Pure and Applied Math., **XL1**, 1988.

[10] Y. Meyer, *Wavelets and Operators*, Analysis at Urbana, vol.1, edited by E. Berkson, N.T. Peck and J. Uhl, London Math. Society, Lecture Notes Series 137, 1989.

[11] S. Mallat, *Review of Multifrequency Channel Decomposition of Images and Wavelet Models*, Technical Report 412, Robotics Report 178, NYU (1988).

[12] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Series in Applied Mathematics, SIAM, (1991 in press).

[13] R. Coifman and Yves Meyer, *Non-linear Harmonic Analysis, Operator Theory and P.D.E.*, Annals of Math Studies, Princeton, 1986, ed. E. Stein.

[14] B. Alpert, *Sparse Representation of Smooth Linear Operators*, PhD thesis, Yale University, 1990.

[15] Y. Meyer, Ondelettes et functions splines. Technical Report, Séminaire EDP, Ecole Polytechnique, Paris, France, 1986.

[16] S. Mallat, *Multiresolution approximation and wavelets*, Technical report, GRASP Lab, Dept. of Computer and Information Science, University of Pennsylvania.

[17] G. Schulz, *Iterative Berechnung der reziproken Matrix*, Z. Angew. Math. Mech. 13, 57-59, 1933.

[18] G. Beylkin, R. R. Coifman and V. Rokhlin, *Fast wavelet transforms and numerical algorithms II.*, in progress.

[19] R. C. Ward, *Numerical computation of the matrix exponential with accuracy estimates*, SIAM. J. Numer. Anal. vol.14, 4, 600-610, 1977.