September 7, 1992

COMPUTATION WITH ADAPTED TIME-FREQUENCY ATOMS

Mladen Victor Wickerhauser*
Department of Mathematics, Washington University
Campus Box 1146, St. Louis, Missouri 63130

Introduction

Operators can be represented by their coefficients with respect to an orthonormal basis of functions well localized in both time and frequency. Such building blocks will be called *time-frequency atoms*. We will insist that they be approximately as well localized in both time and frequency as the Heisenberg uncertainty principle allows. Bases of functions of this type abound and one may be chosen to minimize the number of nonnegligible coefficients for a given operator. We have reason to believe that the equations of motion for important physical systems have a simpler form when posed as evolutions of time-frequency atoms. In this note we will use particular libraries of such atoms to represent operators efficiently and thereby obtain a "fast" functional calculus. We hope to obtain lower complexity matrix application and matrix multiplication algorithms from the new representation. Of course the method is not perfectly general: the speedup depends very much upon the problem. However, others have shown that for broad classes of operators we can expect an order of magnitude asymptotic complexity reduction for matrix application.

The main idea is to lift an $N \times N$ matrix, which maps $\mathbf{R}^N \to \mathbf{R}^N$, into the space of maps on $\mathbf{R}^{N \log N} \to \mathbf{R}^{N \log N}$. Any of a large number of $N \times N$-coefficient subsets of this lifting can be used to represent the operator, so we may pick the subset in which the matrix is most sparse. The choice is made with the "best-basis" algorithm and is itself a fast algorithm. The best-basis implementations described here can use either wavelet packets or adapted local trigonometric libraries. They provide a generalization of the "nonstandard form" introduced by Beylkin, Coifman, and Rokhlin.

* Research supported in part by AFOSR Grant F49620-92-J-0106

Typeset by $\mathcal{AMS}$-TEX

Definition. A *time-frequency atom* (over $\mathbf{R}$) is a function $f = f(x)$ belonging to $L^2(\mathbf{R})$ for which $\|Df\|^2\|Xf\|^2 < \infty$, where D is the derivative operator and X is multiplication by x. For each time-frequency atom f we may compute a location in time and in frequency

$$x_f = \|f\|^{-2} \int_{\mathbf{R}} x|f(x)|^2\,dx, \qquad \xi_f = \|\hat{f}\|^{-2} \int_{\mathbf{R}} \xi|\hat{f}(\xi)|^2\,d\xi; \tag{1}$$

as well as uncertainties in time localization and frequency localization

$$\Delta x_f = \sqrt{\|f\|^{-2}\|Xf\|^2 - x_f^2}, \qquad \Delta\xi_f = \sqrt{\|f\|^{-2}\|Df\|^2 - \xi_f^2}. \tag{2}$$

The Heisenberg uncertainty principle asserts that $\Delta\xi_f \Delta x_f \geq 1/2$; we now require that a basis of time-frequency atoms also satisfy a uniform upper bound for this *Heisenberg functional*, namely that $\Delta\xi_b \Delta x_b \leq C$ for all functions b in the basis, with C being not too large. A sufficiently smooth wavelet basis satisfies this requirement; the Heisenberg functional is a constant for all elements of the basis. In particular, the Heisenberg functional is preserved by translation and dilation.

Wavelet packets. We define wavelet packets over l^2 as in a previous paper[2]. For a pair $H = \{h_i\}, G = \{g_i\}$ of quadrature mirror filters (QMFs) satisfying standard orthogonality and decay conditions[5], there is a unique solution to the functional equation

$$\phi(t) = \sqrt{2} \sum_{j \in \mathbf{Z}} h_j \phi(2t - j). \tag{3}$$

Put $w = w_{0,0,0} = \phi$, and define recursively

$$\begin{aligned} w_{2n,0,0}(t) &= \sqrt{2} \sum_{j \in \mathbf{Z}} h_j w_{n,0,0}(2t - j), \\ w_{2n+1,0,0}(t) &= \sqrt{2} \sum_{j \in \mathbf{Z}} g_j w_{n,0,0}(2t - j). \end{aligned} \tag{4}$$

Then set $w_{mnk}(t) = 2^{m/2} w_{0n0}(2^m t - k)$. The collection of functions so defined are called *wavelet packets*. The quadrature mirror filters H, G may be chosen so that this set is a basis in many common function spaces. Using the Haar filters $H = \{1/\sqrt{2}, 1/\sqrt{2}\}, G = \{1/\sqrt{2}, -1/\sqrt{2}\}$ produces a basis of $L^p(\mathbf{R})$ for $1 < p < \infty$. Longer filters can generate smoother solutions ϕ to the fixed point equation[5]; since wavelet packets are just finite linear

combinations of translates and dilates of ϕ, they can also provide bases for Sobolev spaces, Hardy spaces, etc.

The Heisenberg functional of the wavelet packets $\{w_{mn_0k} : m, k \in \mathbf{Z}\}$ is constant in m, k for each fixed n_0. Unfortunately, it grows with n_0 because of the frequency spreading of highly oscillatory wavelet packets[3]. This growth is slow, however, and can be ignored in practice since for sampled signals the frequency will never exceed the Nyquist cutoff.

Wavelet packets have a nominal position k and frequency n, which we can use in lieu of Eqs.(1,2), i.e., pretending that $x_f = 2^{-m}k$, $\xi_f = 2^m n'$ (where n is the Gray-code value[7] of n'), $\Delta x_f = 2^{-m}$, and $\Delta\xi_f = 2^m$ for $f = w_{mnk}$. These will of course be in error, but only by a small fixed amount and this is compensated by the gain of not having to compute the actual quantities.

Localized trigonometric functions. For definiteness let

$$b(x) = \begin{cases} \sin\frac{\pi}{4}(1 + \sin\pi x), & \text{if } -\frac{1}{2} < x < \frac{3}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

This function is symmetric about the value $x = \frac{1}{2}$. It is smooth on $(-\frac{1}{2}, \frac{3}{2})$ with vanishing derivatives at the boundary points, so that it has a continuous derivative on $\mathbf{R}$. Let $c_n(t) = \sqrt{2}\cos\pi(n + \frac{1}{2})t$, and consider the set of functions $u_{mnk}(t) = 2^{m/2}b(2^m t - k)c_n(2^m t - k)$ for $n \geq 0$, $m, k \in \mathbf{Z}$, $t \in \mathbf{R}$. These *localized cosine functions* all have the same (small) Heisenberg functional. As with wavelet packets, we can pretend that $x_f = 2^{-m}k$, $\xi_f = 2^m n$, $\Delta x_f = 2^{-m}$, and $\Delta\xi_f = 2^m$ for $f = u_{mnk}$.

Several remarks must be made. First, the collection $\{u_{nm_0k} : n, k \in \mathbf{Z}\}$ is an orthonormal basis for $L^2(\mathbf{R})$. With one of two modifications—either using undilated b and matching windows[7] or using multiple folding[6]—we can obtain orthogonality across scales m as well. Second, it is not necessary to use cosine functions: sine works as a direct replacement, and with the modification described in [W4] any periodic functions can be used.

Adapted basis subsets. Define a *basis subset* σ of the indices (m, n, k) to be any subcollection with the property that $\{w_{mnk} : (m, n, k) \in \sigma\}$ is a Hilbert basis for $L^2(\mathbf{R})$. Many basis subsets can be characterized by disjoint dyadic covers of an interval[8]. Abusing notation, we shall also refer to the collection of wavelet packets $\{w_{mnk} : (m, n, k) \in \sigma\}$ as a basis subset.

The collection of tensor products $\{w_X \otimes w_Y\}$ of time-frequency atoms is dense in the space of Hilbert-Schmidt operators. Call $\sigma \subset \mathbf{Z}^6$ a (two-dimensional) basis subset if the collection $\{w_{n_X m_X k_X} \otimes w_{n_Y m_Y k_Y} : (n_X, m_X, k_X, n_Y, m_Y, k_Y) \in \sigma\}$ forms a Hilbert basis for $L^2(\mathbf{R}^2)$. Many two-dimensional basis subsets can be characterized by subtrees of a quadtree[9]. Roughly speaking, they correspond to disjoint covers of the unit square by rectangles whose dimensions are powers of $\frac{1}{2}$, with corners at dyadic rational points. Three such decompositions are displayed in Figure 1.

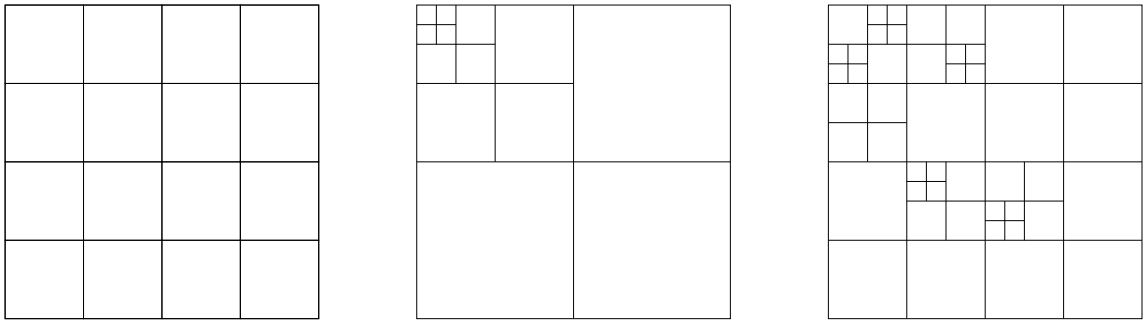


Figure 1. Representations of 3 bases of time-frequency atoms.

We can choose basis subsets to minimize some cost functional, such as the entropy of the coefficient sequence for a particular operator with respect to the basis. There is a fast algorithm for this purpose[4], one that has complexity $O(N)$ for basis selection in $\mathbf{R}^N$. This gives a representation of the operator in a *best-adapted basis of time-frequency atoms*.

Computation

Orthogonal bases of time-frequency atoms are quite useful for computation. We shall briefly discuss the technical details of transforming an operator into such a basis.

Ordering time-frequency atoms. For definiteness we restrict our attention to wavelet packets w_{mnk}. These can be totally ordered. We say that $w < w'$ if $(m, n, k) < (m', n', k')$. The triplets are compared lexicographically, counting the scale parameter m as most significant. Tensor products of wavelet packets inherit this total order. Write $w_X = w_{n_X m_X k_X}$, etc.

Then we will say that $w_X \otimes w_Y < w'_X \otimes w'_Y$ if $w_X < w'_X$ or else if $w_X = w'_X$ but $w_Y < w'_Y$. This is equivalent to $(m_X, n_X, k_X, m_Y, n_Y, k_Y) < (m'_X, n'_X, k'_X, m'_Y, n'_Y, k'_Y)$ comparing lexicographically from left to right. We can also define the *adjoint order* $<^*$ by exchanging X and Y indices, namely $w_X \otimes w_Y <^* w'_X \otimes w'_Y$ if and only if $w_Y \otimes w_X < w'_Y \otimes w'_X$. This is also a total order.

Injections and projections. Let $\mathcal{W}^1$ denote the space of bounded sequences indexed by the three wave packet indices m, n, k. With the ordering above, we obtain a natural isomorphism between l^∞ and $\mathcal{W}^1$. There is also a natural injection $J^1 : L^2(\mathbf{R}) \hookrightarrow \mathcal{W}^1$ given by $c_{mnk} = \langle v, w_{mnk} \rangle$, for $v \in L^2(\mathbf{R})$ and wavelet packets w_{mnk}. If σ is a basis subset, then the composition J^1_σ of J^1 with projection onto the subsequences indexed by σ is also injective. J^1_σ is an isomorphism of $L^2(\mathbf{R})$ onto $l^2(\sigma)$, which is defined to be the square summable sequences of $\mathcal{W}^1$ whose indices belong to σ. In the rank-N case where there are $N \log N$ wavelet packets, these maps are $J^1 : \mathbf{R}^N \to \mathbf{R}^{N \log N}$ and $J^1_\sigma : \mathbf{R}^N \to \mathbf{R}^N$.

We also have a map $R^1 : \mathcal{W}^1 \to L^2(\mathbf{R})$ defined by $R^1 c(t) = \sum_{(m,n,k) \in \mathbf{Z}^3} c_{mnk} w_{mnk}(t)$. This map is defined and bounded on the closed subspace of $\mathcal{W}^1$ isomorphic to l^2 under the natural isomorphism mentioned above. In particular, R^1 is defined and bounded on the range of J^1_σ for every basis subset σ. The related restriction $R^1_\sigma : \mathcal{W}^1 \to L^2(\mathbf{R})$ defined by $R^1_\sigma c(t) = \sum_{(m,n,k) \in \sigma} c_{mnk} w_{mnk}(t)$ is a left inverse for J^1 and J^1_σ. In addition, $J^1 R^1_\sigma$ is a projection of $\mathcal{W}^1$. Likewise, if $\sum_i \alpha_i = 1$ and $R^1_{\sigma_i}$ is one of the above maps for each i, then $J^1 \sum_i \alpha_i R^1_{\sigma_i}$ is also a projection of $\mathcal{W}^1$. It is an orthogonal projection on any finite subset of $\mathcal{W}^1$. In the rank-N case we have $R^1 : \mathbf{R}^{N \log N} \to \mathbf{R}^N$ and $R^1_\sigma : \mathbf{R}^N \to \mathbf{R}^N$.

Similarly, writing $\mathcal{W}^2$ for $\mathcal{W}^1 \times \mathcal{W}^1$, the ordering of tensor products gives a natural isomorphism between l^∞ and $\mathcal{W}^2$. The space $L^2(\mathbf{R}^2)$, i.e., the Hilbert-Schmidt operators, inject into this sequence space $\mathcal{W}^2$ in the obvious way, namely $M \mapsto \langle M, w_{n_X m_X k_X} \otimes w_{n_Y m_Y k_Y} \rangle$. Call this injection J^2. If σ is a basis subset of $\mathcal{W}^2$, then the composition J^2_σ of J^2 with projection onto subsequences indexed by σ is also injective. J^2_σ is an isomorphism of $L^2(\mathbf{R}^2)$ onto $l^2(\sigma)$, the square summable sequences of $\mathcal{W}^2$ whose indices belong to σ. In the $N \times N$ matrix case we have $J^2 : \mathbf{R}^{N^2} \to \mathbf{R}^{N^2 \log N}$ and $J^2_\sigma : \mathbf{R}^{N^2} \to \mathbf{R}^{N^2}$.

The map $R^2 : \mathcal{W}^2 \to L^2(\mathbf{R}^2)$ given by $R^2 c(x, y) = \sum c_{XY} w_X(x) w_Y(y)$, is bounded on

that subset of $\mathcal{W}^2$ naturally isomorphic to l^2. In particular, it is bounded on the range of J^2_σ for every basis subset σ. We may also define the restrictions R^2_σ of R^2 to subsequences indexed by σ, defined by $R^2_\sigma c(x,y) = \sum_{(w_X,w_Y)\in\sigma} c_{XY} w_X(x) w_Y(y)$. There is one for each basis subset σ of $\mathcal{W}^2$. Then R^2_σ is a left inverse of J^2 and J^2_σ, and $J^2 R^2_\sigma$ is a projection of $\mathcal{W}^2$. As before, if $\sum_i \alpha_i = 1$ and σ_i is a basis subset for each i, then $J^2 \sum_i R^2_{\sigma_i}$ is also a projection of $\mathcal{W}^2$. It is an orthogonal projection on any finite subset of $\mathcal{W}^2$. In the $N \times N$ matrix case we have $R^2 : \mathbf{R}^{N^2 \log N} \to \mathbf{R}^{N^2}$ and $R^2_\sigma : \mathbf{R}^{N^2} \to \mathbf{R}^{N^2}$.

Applying matrices to vectors. For definiteness, let X and Y be two named copies of $\mathbf{R}^N$. Let $v \in X$ be a vector, whose coordinates with respect to wavelet packets form the sequence $J^1 v = \{\langle v, w_X\rangle\}$ for some wavelet packet basis $\{w_X\}$ of X.

Let $M : X \to Y$ be an $N \times N$ matrix with respect to the standard basis. Its coefficients with respect to the complete set of tensor products of wavelet packets form the sequence $J^2 M = \{\langle M, w_X \otimes w_Y\rangle : w_X \in X, w_Y \in Y\}$. We obtain the identity

$$\langle Mv, w_Y\rangle = \sum_{w_X \in X} \langle M, w_X \otimes w_Y\rangle \langle v, w_X\rangle \tag{6}$$

This identity generalizes to a linear action of $\mathcal{W}^2$ on $\mathcal{W}^1$ defined by

$$c(v)_{mnk} = \sum_{(n'm'k')} c_{mnkm'n'k'} v_{m'n'k'}. \tag{7}$$

Now, images of matrices form a proper submanifold of $\mathbf{R}^{N^2 \log N}$. Likewise, images of vectors form a submanifold $\mathbf{R}^{N \log N}$. We can lift the action of M on v to these larger spaces via the commutative diagram

$$\begin{array}{ccc} \mathbf{R}^{N \log N} & \xrightarrow{J^2_\sigma M} & \mathbf{R}^{N \log N} \\ {\scriptstyle J^1}\uparrow & & \downarrow{\scriptstyle R^1} \\ \mathbf{R}^N & \xrightarrow[M]{} & \mathbf{R}^N \end{array} \tag{8}$$

The significance of this lift is that by a suitable choice of σ we can reduce the complexity of the map $J^2_\sigma M$, and therefore the complexity of the operator application. Figure 2 shows a schematic representation of $J^2 M$ and its action on the space $\mathcal{W}^1 = \mathbf{R}^{N \log N}$. Here M is a 1024×1024 sine transform matrix, and the black dots represent nonnegligible coefficients—those greater than 0.01—in the best isoscale wavelet packet basis. These comprise about 10% of the total number of coefficients in the original matrix.

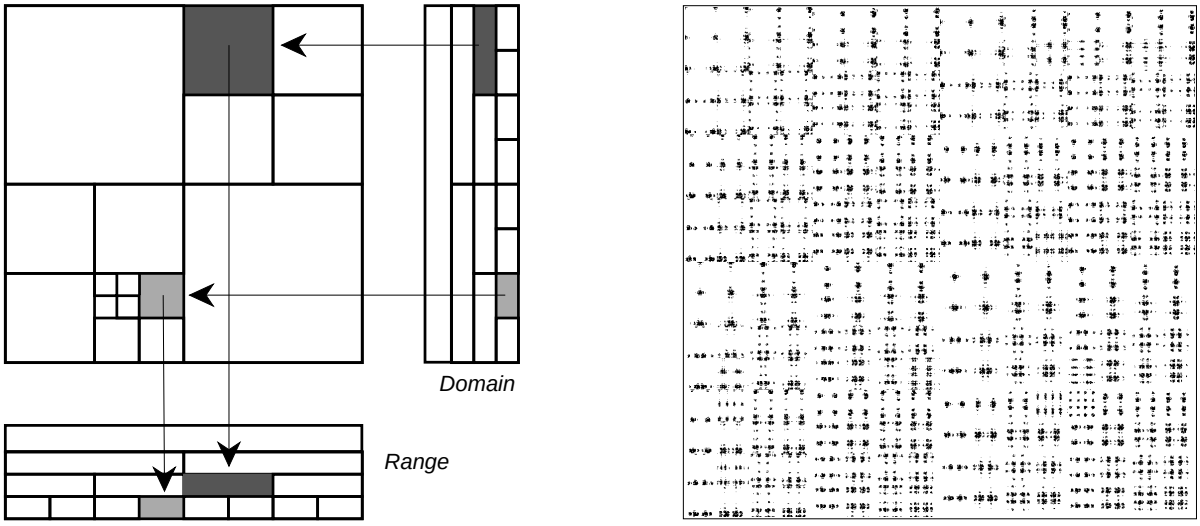


Figure 2. Applying a matrix in its best basis to a vector.

Remark. Alternatively, each row of a matrix can be expanded into its best basis and then only the nonnegligible elements of each row applied to the vector. This is the same as using the basis subset $\{w_X \otimes w_Y\}$ where $w_Y \in \{w_{(\log N)0k} : 0 \le k < N\}$. The left part of Figure 3 shows this application schematically; the right part shows (as black dots) how many of the largest "best-rows" coefficients are needed to capture 99% of the row l^2 norm for a 1024×1024 sine transform matrix. The dots comprise about 7% of the matrix.

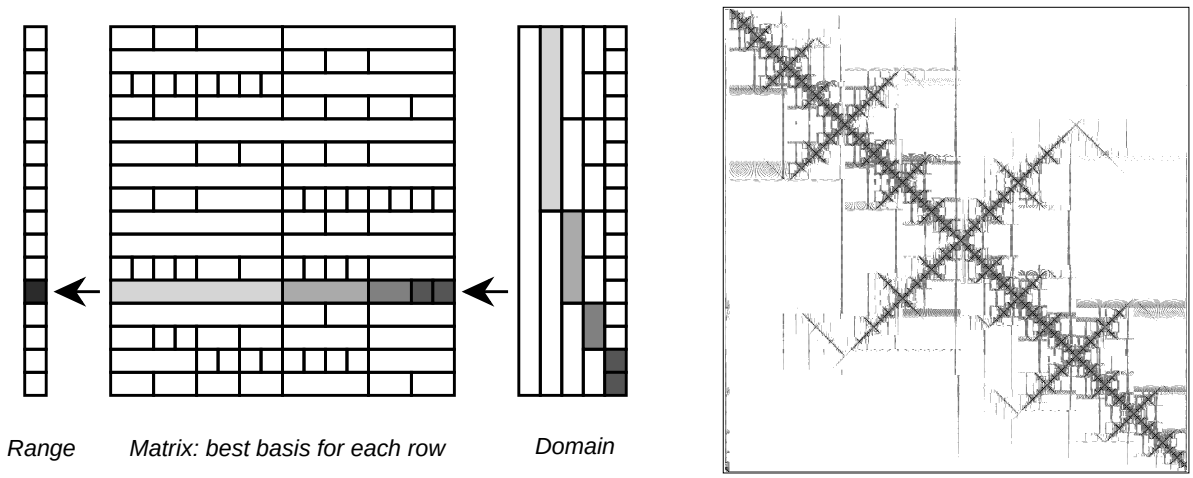


Figure 3. Applying a matrix in its best-rows form to a vector.

Composing operators. Let X, Y, Z be three named copies of $\mathbf{R}^N$. Suppose that $M : X \to Y$ and $P : Y \to Z$ are $N \times N$ matrices. We have the identity

$$\langle PM, w_X \otimes w_Z \rangle = \sum_{w_Y \in Y} \langle P, w_Y \otimes w_Z \rangle \langle M, w_X \otimes w_Y \rangle. \tag{9}$$

This generalizes to an action of $\mathcal{W}^2$ on $\mathcal{W}^2$, which in this case are both $\mathbf{R}^{N^2 \log N}$. The action is defined by the formula

$$c(d)_{mnkm'n'k'} = \sum_{m''n''k''} d_{mnkm''n''k''} c_{m''n''k''m'n'k'}, \tag{10}$$

where c and d are sequences in $\mathcal{W}^2$. Using J^2, we can lift multiplication by P to an action on these larger spaces via the commutative diagram

$$\begin{array}{ccc} \mathcal{W}^2 & \xrightarrow{J^2_\sigma P} & \mathcal{W}^2 \\ {\scriptstyle J^2}\uparrow & & \downarrow{\scriptstyle R^2} \\ \mathbf{R}^{N^2} & \xrightarrow[P]{} & \mathbf{R}^{N^2} \end{array} \tag{11}$$

Again, by a suitable choice of σ the complexity of the operation may be reduced to below that of ordinary operator composition.

Operation counts

Applying a matrix to a vector. Suppose that M is a non-sparse $N \times N$ matrix. Ordinary multiplication of a vector by M takes at least $O(N^2)$ operations. On the other hand, the injection J^2 will require $O(N^2 \log N)$ operations, and each of J^1 and R^1 require $O(N \log N)$ operations. For a fixed basis subset σ of $\mathcal{W}^2$, the application of $J^2_\sigma M$ to $J^1 v$ requires at most $\#|J^2_\sigma M|$ operations, where $\#|U|$ denotes the number of nonzero coefficients in U. We may choose our wavelet library so that $\#|J^2_\sigma M| = O(N^2)$. Thus the multiplication method described above costs an initial investment of $O(N^2 \log N]^2)$, plus at most an additional $O(N^2)$ per right-hand side. Thus the method has asymptotic complexity $O(N^2)$ per vector in its exact form, as expected for what is essentially multiplication by a conjugated matrix.

We can obtain lower complexity if we take into account the finite accuracy of our calculation. Given a fixed matrix of coefficients C, write C_δ for the same matrix with all coefficients set to 0 whose absolute values are less than δ. By the continuity of the Hilbert-Schmidt norm,

for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|C - C_\delta\|_{HS} < \epsilon$. Given M and ϵ as well as a library of wavelet packets, we can choose a basis subset $\sigma \subset \mathcal{W}^2$ so as to minimize $\#|(J^2_\sigma M)_\delta|$. The choice algorithm has complexity $O(N^2 \log N)$, as shown in [W3]. For matrices derived from Calderón–Zygmund operators, it has been shown[1] that there is a library of wavelet packets—the smooth wavelets—such that for every fixed $\delta > 0$ we have

$$\#|(J^2_\sigma M)_\delta| = O(N \log N), \tag{12}$$

with the constant depending, of course, on δ. For the moment, call the class of operators with property (12) the *sparsifiable* Hilbert-Schmidt operators $\mathcal{S}$. By the estimate above, finite-precision multiplication by sparsifiable rank-N operators has asymptotic complexity $O(N \log N)$. Recently Coifman, Grossman, and Meyer have proved that other classes of operators—including some oscillatory-kernel integral operators—are sparsifiable.

Composing operators. Suppose that M and P are rank-N operators. Standard multiplication of P and M has complexity $O(N^3)$. The complexity of injecting P and M into $\mathcal{W}^2$ is $O(N^2 \log N)$. The action of $J^2_\sigma P$ on $J^2 M$ has complexity $O(\sum_{mnk} \#|J^2_\sigma P_{YZ} : (m_Y, n_Y, k_Y) = (m, n, k)| \#|J^2 M_{XY} : (m_Y, n_Y, k_Y) = (m, n, k)|)$. The second factor is a constant $N \log N$, while the first when summed over all mnk is exactly $\#|J^2_\sigma P|$. Thus the complexity of the nonstandard multiplication algorithm, including the conjugation into the basis set σ, is $O(\#|J^2_\sigma P|\, N \log N)$. Since the first factor is N^2 in general, the complexity of the exact algorithm is $O(N^3 \log N)$ for generic matrices, reflecting the extra cost of conjugating into the basis set σ.

For the approximate algorithm, the complexity is $O(\#|(J^2_\sigma N)_\delta|\, N \log N)$. For the sparsifiable matrices, this can be reduced by a suitable choice of σ to a complexity of $O(N^2 \log N)$ for the complete algorithm. Since choosing σ and evaluating J^2_σ each have this complexity, it is not possible to do any better by this method.

References

[1] G. Beylkin, R. R. Coifman, and V. Rokhlin, *Fast wavelet transforms and numerical algorithms I*, Communications on Pure and Applied Math **XLIV** (1991), 141–183.

[2] R. R.Coifman, Y. Meyer, and M. V. Wickerhauser, *Wavelet analysis and signal processing*, Wavelets and Their Applications, ed. Ruskai et al., ISBN 0-86720-225-4, Jones and Bartlett, Boston, 1992, pp. 153–178..

[3] R. R.Coifman, Y. Meyer, and M. V. Wickerhauser, *Size properties of wavelet packets*, Wavelets and their Applications, ed. Ruskai, et al., ISBN 0-86720-225-4, Jones and Bartlett, Boston, 1992, pp. 453–470.

[4] M. V. Wickerhauser and R. R. Coifman, *Entropy based methods for best basis selection*, IEEE Transactions on Information Theory **32** (March, 1992), 712–718.

[5] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics **XLI** (1988), 909–996.

[6] X. Fang and E. Seré, *Multiple folding local sine transform*, preprint, Washington University, 1992.

[7] M. V. Wickerhauser, *INRIA lectures on wavelet packet algorithms*, lecture notes, INRIA, Roquencourt, France, 1991.

[8] M. V. Wickerhauser, *Acoustic signal compression with wavelet packets*, Wavelets–A Tutorial in Theory and Applications, C. K. Chui (ed.) ISBN 0-12-174590-2, Academic Press, Boston, 1992, pp. 679–700.

[9] M. V. Wickerhauser, *High-resolution still picture compression*, to appear, Digital Signal Processing, a review journal.