

Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods

R. R. Coifman*†, S. Lafon*, A. B. Lee*, M. Maggioni*, B. Nadler*, F. Warner*, and S. W. Zucker‡

*Department of Mathematics, Program in Applied Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06510; and ‡Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06510

Contributed by R. R. Coifman, February 2, 2005

In the companion article, a framework for structural multiscale geometric organization of subsets of \mathbb{R}^n and of graphs was introduced. Here, diffusion semigroups are used to generate multiscale analyses in order to organize and represent complex structures. We emphasize the multiscale nature of these problems and build scaling functions of Markov matrices (describing local transitions) that lead to macroscopic descriptions at different scales. The process of iterating or diffusing the Markov matrix is seen as a generalization of some aspects of the Newtonian paradigm, in which local infinitesimal transitions of a system lead to global macroscopic descriptions by integration. This article deals with the construction of fast-order N algorithms for data representation and for homogenization of heterogeneous structures.

In the companion article (1), it is shown that the eigenfunctions of a diffusion operator, A , can be used to perform global analysis of the set and of functions on a set. Here, we present a construction of a multiresolution analysis of functions on the set related to the diffusion operator A . This allows one to perform a local analysis at different diffusion scales.

This is motivated by the fact that in many situations one is interested not in the data themselves but in functions on the data, and in general these functions exhibit different behaviors at different scales. This is the case in many problems in learning, in analysis on graphs, in dynamical systems, etc. The analysis through the eigenfunctions of Laplacian considered in the companion article (1) are global and are affected by global characteristics of the space. It can be thought of as global Fourier analysis. The multiscale analysis proposed here is in the spirit of wavelet analysis.

We refer the reader to (2–4) for further details and applications of this construction, as well as a discussion of the many relationships between this work and the work of many other researchers in several branches of mathematics and applied mathematics. Here, we would like to at least mention the relationship with fast multiple methods (5, 6), algebraic multigrid (7), and lifting (8, 9).

Multiscale Analysis of Diffusion

Construction of the Multiresolution Analysis. Suppose we are given a self-adjoint diffusion operator A as in ref. 1 acting on L^2 of a metric measure space (X, d, μ) . We interpret A as a dilation operator and use it to define a multiresolution analysis. It is natural to discretize the semigroup $\{A^t\}_{t \geq 0}$ of the powers of A at a logarithmic scale, for example at the times

$$t_j = 1 + 2 + 2^2 + \dots + 2^j = 2^{j+1} - 1. \quad [1]$$

For a fixed $\varepsilon \in (0, 1)$, we define the approximation spaces by

$$V_j = \overline{\langle \{\phi_i: \lambda_i^j \geq \varepsilon\} \rangle}, \quad [2]$$

where the ϕ_i s are the eigenvectors of A , ordered by decreasing eigenvalue. We will denote by P_j the orthogonal projection onto V_j . The set of subspaces $\{V_j\}_{j \in \mathbb{Z}}$ is a multiresolution analysis in the sense that it satisfies the following properties:

- $\lim_{j \rightarrow -\infty} V_j = L^2(X, \mu)$, $\lim_{j \rightarrow +\infty} V_j = \overline{\langle \{\phi_i: \lambda_i = 1\} \rangle}$.
- $V_{j+1} \subseteq V_j$ for every $j \in \mathbb{Z}$.
- $\{\phi_i: \lambda_i^j \geq \varepsilon\}$ is an orthonormal basis for V_j .

We can also define the detail subspaces W_j as the orthogonal complement of V_j in V_{j+1} , so that we have the familiar relation between approximation and detail subspaces as in the classical wavelet multiresolution constructions:

$$V_{j+1} = V_j \oplus^\perp W_j.$$

This is very much in the spirit of a Littlewood–Paley decomposition induced by the diffusion semigroup (10). However, in each subspace V_j and W_j we have the orthonormal basis of eigenfunctions, but we would like to replace them with localized orthonormal bases of scaling functions as in wavelet theory. Generalized Heisenberg principles (see *Extension of Empirical Functions of the Data Set*) put a lower bound on how much localization can be achieved at each scale j , depending on the spectrum of the operator A and on the space on which it acts. We would like to have basis elements as much localized as allowed by the Heisenberg principle at each scale, and spanning (approximately) V_j . We do all this while avoiding computation of the eigenfunctions.

We start by fixing a precision $\varepsilon > 0$ and assume that A is represented on the basis $\Phi_0 = \{\delta_k\}_{k \in X}$. We consider the columns of A , which can be interpreted as the set of functions $\Phi_1 = \{A\delta_k\}_{k \in X}$ on X . We use a local multiscale Gram–Schmidt procedure, described below, to carefully but efficiently orthonormalize these columns into a basis $\Phi_1 = \{\varphi_{1,k}\}_{k \in X_1}$ (X_1 is defined as this index set) for the range of A up to precision ε . This is a linear transformation we represent by a matrix G_0 . This yields a subspace that is ε -close to V_1 . Essentially, Φ_1 is a basis for a subspace that is ε -close to the range of A , the basis elements that are well localized and orthogonal. Obviously, $|X_1| \leq |X|$, but the inequality may already be strict since part of the range of A may be below the precision ε . Whether this is the case or not, we have then a map M_0 from X to X_1 , which is the composition of A with the orthonormalization by G_0 . We can also represent A in the basis Φ_1 : We denote this matrix by A_1 and compute A_1^2 . See the diagram in Fig. 1.

We now proceed by looking at the columns of A_1^2 , which are $\Phi_2 = \{A_1^2 \delta_k\}_{k \in X_1} = \{A^2 \varphi_{1,k}\}_{k \in X_1}$ up to precision ε , by unraveling the bases on which the various elements are represented. Again we can apply a local Gram–Schmidt procedure to orthonormalize this set: This yields a matrix G_1 and an orthonormal basis $\Phi_2 = \{\varphi_{2,k}\}_{k \in X_2}$ for the range of A_1^2 up to precision ε , and hence for the range of A^3 up to precision 2ε . Moreover, depending on the decay of the spectrum of A , $|X_2| \ll |X_1|$. The matrix M_1 , which is the composition of G_1 with A_1^2 , is then of size $|X_2| \times |X_1|$, and $A_2^2 = M_1 M_1^T$ is a representation of A^4 acting on Φ_2 .

After j steps in this fashion, we will have a representation of $A^{1+2+2^2+\dots+2^j} = A^{2^{j+1}-1}$ onto a basis $\Phi_j = \{\varphi_{j,k}\}_{k \in X_j}$ that spans

†To whom correspondence should be addressed. E-mail: coifman-ronald@yale.edu.

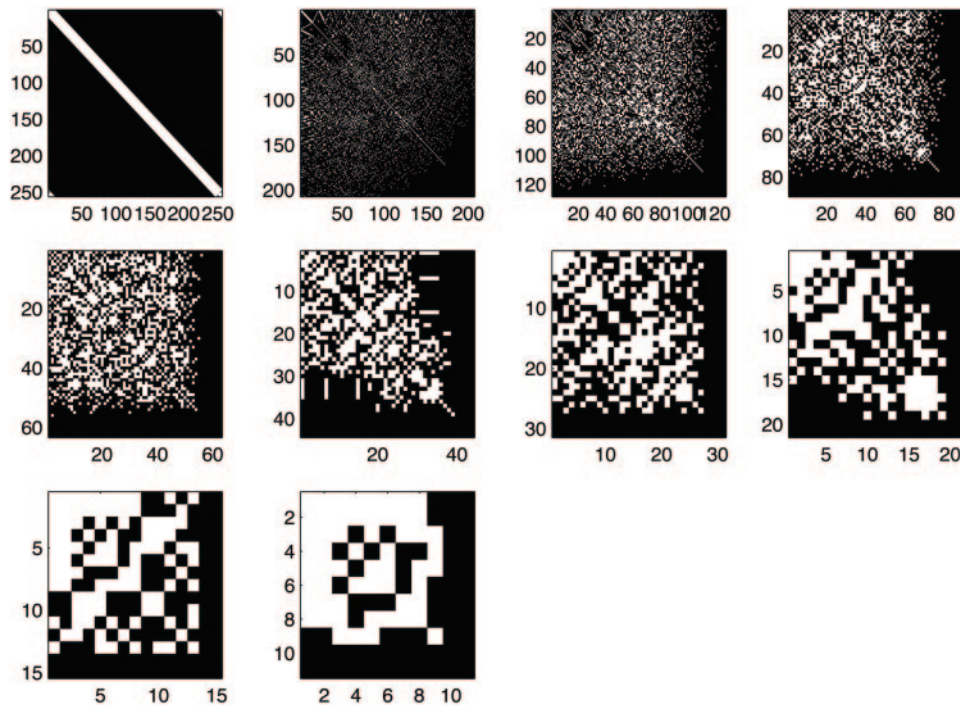


Fig. 3. Diffusion multiresolution analysis on the circle. We plot the compressed matrices representing powers of the diffusion operator, in white are the entries above working precision (here set to 10^{-8}). Notice the shrinking of the size of the matrices which are being compressed at the different scales.

if we let $S_K = \sum_{k=1}^{2^K} A^k$, we see that

$$S_{K+1} = S_K + A^{2^K} S_K = \prod_{k=0}^K (I + A^{2^k})f,$$

and each term of the product can be applied fast to f .

The construction of the multiscale bases can be done in time $O(n \log^2 n)$, where $n = |X|$, if the spectrum of A has fast enough decay. The decomposition of a function f onto the scaling functions and wavelets we construct can be done in the same time, and so does the computation of $(I - A)^{-1} f$.

The Orthogonalization Process. We sketch here how the orthogonalization works (for details, refer to refs. 2 and 3). Suppose we start from a δ -local basis $\Phi = \{\varphi_x\}_{x \in T}$ (in our case, φ_x is going to be a bump $A^l \delta_x$). We greedily build a first layer of basis functions $\Phi_0 = \{\tilde{\varphi}_{0,x_k}\}_{x_k \in \mathcal{K}_0}$, $\mathcal{K}_0 \subseteq T$ as follows. We let φ_{0,x_0} be a basis function with greatest L^2 -norm. Then we let φ_{0,x_1} be a basis function with biggest L^2 -norm among the basis functions with support disjoint from the support of φ_{0,x_0} but not farther than δ from it. By induction, after $\varphi_{0,x_0}, \dots, \varphi_{0,x_l}$ have been chosen, we let $\varphi_{0,x_{l+1}}$ be a scaling function with largest L^2 -norm among those having a support that does not intersect any of the supports of the basis functions already constructed but is not farther than δ from the closest such support. We stop when no such choice can be made. One can think of \mathcal{K}_0 roughly as a 2δ lattice.

At this point, Φ_0 in general spans a subspace much smaller than the one spanned by Φ . We construct a second layer $\Phi_1 = \{\tilde{\varphi}_{1,x_k}\}_{x_k \in \mathcal{K}_1}$, $\mathcal{K}_1 \subseteq T \setminus \mathcal{K}_0$ as follows. Orthogonalize each $\{\varphi_x\}_{x \in T \setminus \mathcal{K}_0}$ to the functions $\{\varphi_{0,x_k}\}_{x_k \in \mathcal{K}_0}$. Observe that since the support of φ_x is small, this orthogonalization is local, in the sense that each φ_x needs to be orthogonalized only to the few φ_{0,x_k} s that have an intersecting support. In this way, we get a set $\tilde{\Phi}_1$, orthogonal to Φ_0 but not orthogonal itself. We orthonormalize it exactly as we did to get Φ_0 from Φ . We proceed by building as many layers as necessary to span the whole space $\langle \Phi \rangle$ (up to the specified precision ε).

Wavelets. We would like to construct bases $\{\psi_{j,k}\}_k$ for the spaces W_j , $j \geq 1$, such that $V_j \oplus^\perp W_j = V_{j+1}$. To achieve this, after having built $\{\varphi_{j,k}\}_{k \in \mathcal{K}_j}$ and $\{\varphi_{j+1,k}\}_{k \in \mathcal{K}_{j+1}}$, we can apply our modified Gram-Schmidt procedure with geometric pivoting to the set of functions

$$\{(P_j - P_{j+1})\varphi_{j,k}\}_{k \in \mathcal{K}_j},$$

which will yield an orthonormal basis of wavelets for the orthogonal complement of V_{j+1} in V_j . Observe that each wavelet is a result of an orthogonalization process that is local, so the

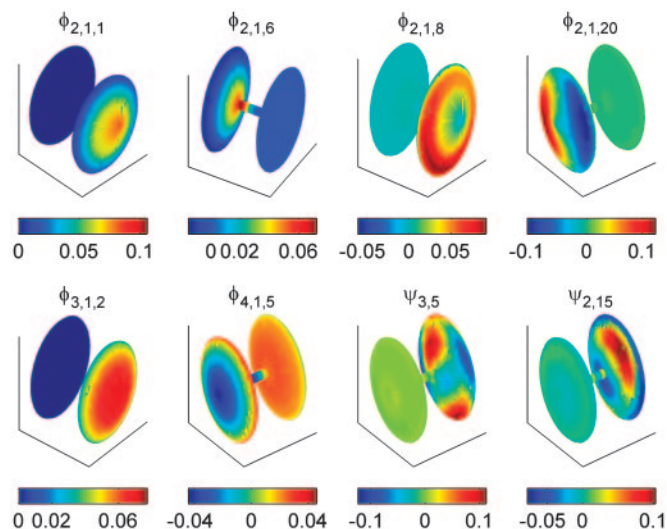


Fig. 4. Some diffusion scaling functions and wavelets at different scales on a dumbbell-shaped manifold sampled at 1,400 points.

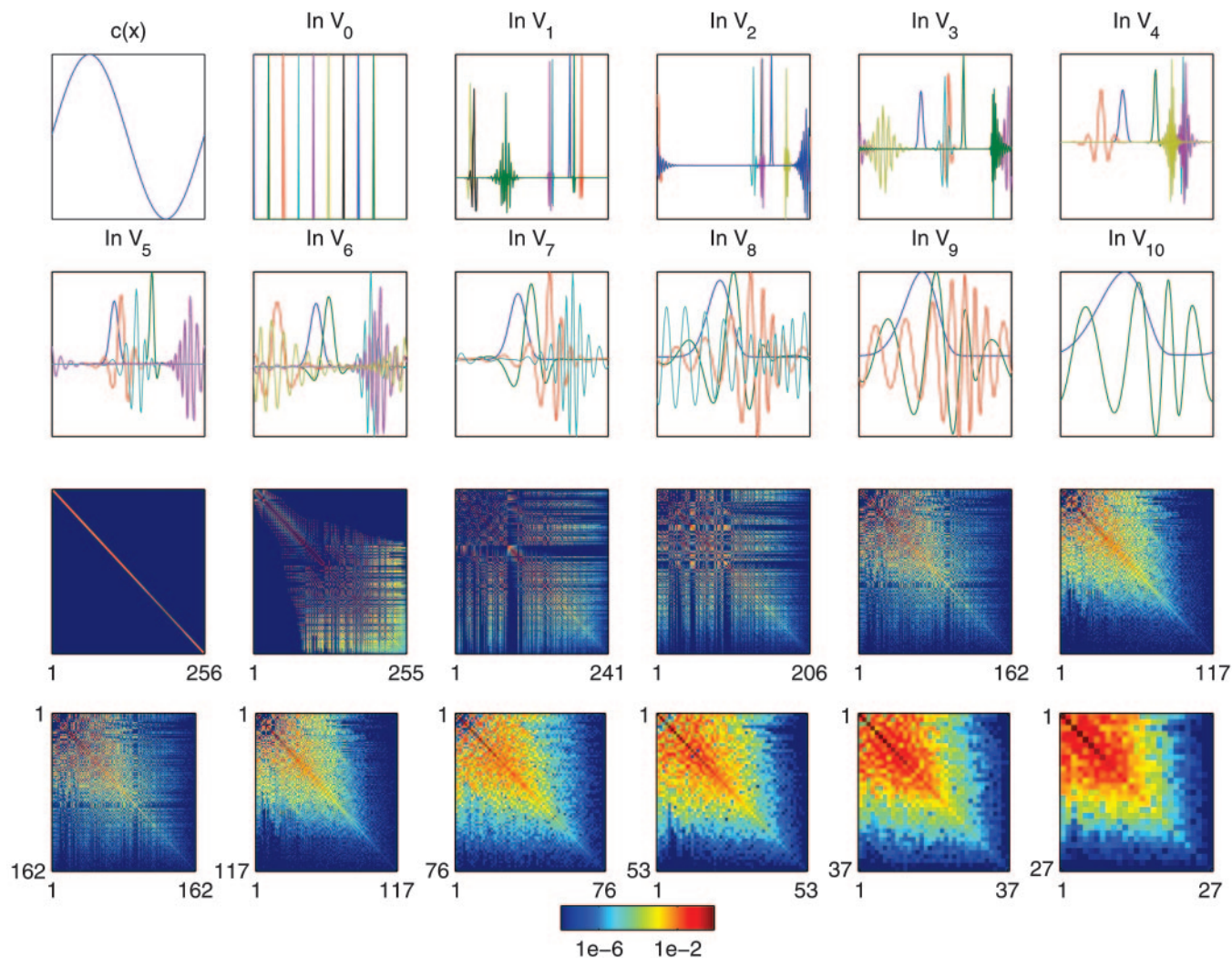


Fig. 5. Multiresolution diffusion on a circular medium with non-constant diffusion coefficient. (*Upper*) Several scaling functions and wavelets in different approximation subspaces V_j . Notice that scaling functions at the same diffusion scale exhibit different spatial localization, which depends on the local diffusion coefficient. (*Lower*) Matrix compression of the dyadic powers of T on the scaling function bases of the V_j . Notice the size of the matrices shrinking with scale.

computation is again fast. To achieve numerical stability, we orthogonalize at each step the remaining $\varphi_{j+1,k}$ s to both the wavelets built so far and $\varphi_{j,k}$. Wavelet subspaces can be recursively split further to obtain diffusion wavelet packets (4), which allow the application of the classical fast algorithms (13) for denoising (14), compression (15), and discrimination (16).

Examples and Applications

Example 3.1: Multiresolution diffusion on the homogeneous circle. To compare with classical constructions of wavelets, we consider the unit circle, sampled at 256 points, and the classical isotropic heat diffusion on it. The initial orthonormal basis Φ_0 is given by the set of δ -functions at each point, and we build the diffusion wavelets at all scales, which clearly relate to splines and multi-wavelets. The spectrum of the diffusion operator does not decay very fast (see Figs. 2 and 3).

Example 3.2: Dumbbell. We consider a dumbbell-shaped manifold, sampled at 1,400 points, and the diffusion associated to the (discretized) Laplace–Beltrami operator as discussed in ref. 1. See Fig. 4 for the plots of some scaling functions and wavelets: They exhibit the expected locality and multiscale features, dependent on the intrinsic geometry of the manifold.

Example 3.3: Multiresolution diffusion on a nonhomogeneous circle. We can apply the construction of diffusion wavelets to nonisotropic diffusions arising from partial differential equations, to tackle problems of homogenization in a natural way. The literature on homogenization is vast (see, e.g., refs. 17–21 and references therein).

Our definition of scales, driven by the differential operator, in general results in highly nonuniform and nonhomogeneous spatial and spectral scales, and in corresponding coarse equations of the system, which have high precision.

For example, we can consider the nonhomogeneous heat equation on the circle

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(c(x) \frac{\partial}{\partial x} u \right), \quad [4]$$

where $c(x)$ is a positive function close 0 at certain points and almost 1 at others. We want to represent the intermediate and large scale/time behavior of the solution by compressing powers of the operator representing the discretization of the spatial differential operator $\partial/\partial x(c(x)(\partial/\partial x))$. The *spatial* differential operator on the right-hand side of Eq. 4 is a matrix T that, when properly normalized, can be interpreted as a nontranslation invariant random walk.

$$k_B(x, y) = \int_{\|\xi\| < B} e^{2i\pi(\xi, x)} e^{-2i\pi(\xi, y)} d\xi = \frac{J_{n/2}(2\pi B\|x - y\|)}{\|x - y\|^{n/2}},$$

where x and y belong to X , and J_ν is the Bessel function of the first type and of order ν . From the first equality sign, we see that the geometric harmonics arise from a principal component analysis of the set of all restrictions of B -bandlimited complex exponentials to X .

It can be verified that, in addition to being orthogonal on the set X , these B -bandlimited geometric harmonics are also orthogonal over the whole space \mathbb{R}^n . Moreover, ψ_j minimizes the Rayleigh quotient

$$\frac{\int_{\mathbb{R}^n} |\bar{f}(x)|^2 dx}{\int_X |f(x)|^2 dx}$$

under the constraint that f be orthogonal to $\{\psi_0, \psi_1, \dots, \psi_{j-1}\}$. In other words, ψ_0 is the B -bandlimited extension of ψ_j that has minimal energy on \mathbb{R}^n . As a consequence, \bar{f} is the B -bandlimited extension of f that has minimal energy off the set X . This type of extension is optimal in the sense that it is the average of all B -bandlimited extensions of f . It also suggests that this extension satisfies Occam's razor in that it is the "simplest" among all bandlimited extensions: Any other extension is equal to \bar{f} plus an orthogonal bandlimited function that vanishes on X .

Multiscale Extension. For a given function f on X , we have constructed a minimal energy B -bandlimited extension \bar{f} . In the case when X is a smooth compact submanifold of \mathbb{R}^n , we can now relate the spectral theory on the set X to that on \mathbb{R}^n .

On the one hand, any band limited function of band $B > 0$ restricted to X can be expanded to exponential accuracy in terms of the eigenfunctions of the Laplace–Beltrami operator Δ with eigenvalues ν_j^2 not exceeding CB^2 for some small constant $C >$

0. On the other hand, it can be shown that every eigenfunction of the Laplace–Beltrami operator satisfying this condition extends as a bandlimited function with band $C'B$. Both of these statements can be proved by observing that eigenfunctions on the manifold are well approximated by restrictions of bandlimited functions.

We conclude that any empirical function f on X that can be approximated as a linear combination of eigenfunctions of Δ , and these eigenfunctions can be extended to different distances: If the eigenvalue is ν^2 , then the corresponding eigenfunction can be extended as a ν -bandlimited function off the set X to a distance $C\nu^{-1}$. This observation constitutes a formulation of the Heisenberg principle involving the Fourier analysis on and off the set X , and which states that any empirical function can be extended as a sum of "atoms" whose numerical supports in the ambient space is related to their frequency content on the set.

The generalized Heisenberg principle is illustrated on Fig. 6, where we show the extension of the functions $f_j(\theta) = \cos(2\pi j\theta)$ for $j = 1, 2, 3$ and 4, from the unit circle to the plane. For each function, we used Gaussian kernels, and the scale was adjusted as the maximum scale that would preserve a given accuracy.

Conclusion

We have introduced a multiscale structure for the efficient computation of large powers of a diffusion operator, and its Green function, based on a generalization of wavelets to the general setting of discretized manifolds and graphs. This has application to the numerical solution of partial differential equations and to the analysis of functions on large data sets and learning. We have shown that a global (with eigenfunctions of the Laplacian) or local (with diffusion wavelets) analysis on a manifold embedded in Euclidean space can be extended outside the manifold in a multiscale fashion by using band-limited functions.

We thank James C. Bremer, Jr., Naoki Saito, Raanan Schul, and Arthur D. Szlam for their useful comments and suggestions during the preparation of the manuscript. This work was partially funded by the Defense Advanced Research Planning Agency and the Air Force Office of Scientific Research.

- Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F. & Zucker, S. W. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 7426–7431.
- Coifman, R. R. & Maggioni, M. (2004) *Tech. Rep. YALE/DCS/TR-1289* (Dept. Comput. Sci., Yale Univ., New Haven, CT).
- Coifman, R. R. & Maggioni, M. (2004) *Tech. Rep. YALE/DCS/TR-1303*, Yale Univ. and *Appl. Comput. Harmonic Anal.*, in press.
- Bremer, J. C., Jr., Coifman, R. R., Maggioni, M. & Szlam, A. D. (2004) *Tech. Rep. YALE/DCS/TR-1304*, Yale Univ. and *Appl. Comput. Harmonic Anal.*, in press.
- Beylkin, G., Coifman, R. R. & Rokhlin, V. (1991) *Comm. Pure Appl. Math* **44**, 141–183.
- Greengard, L. & Rokhlin, V. (1987) *J. Comput. Phys.* **73**, 325–348.
- Brandt, A. (1986) *Appl. Math. Comp.* **19**, 23–56.
- Sweldens, W. (1996) *Appl. Comput. Harmonic Anal.* **3**, 186–200.
- Sweldens, W. (1997) *SIAM J. Math. Anal.* **29**, 511–546.
- Stein, E. (1970) *Topics in Harmonic Analysis Related to the Littlewood–Paley Theory* (Princeton Univ. Press, Princeton), Vol. 63.
- Meyer, Y. (1990) *Ondelettes et Opérateurs* (Hermann, Paris).
- Daubechies, I. (1992) *Ten Lectures on Wavelets* (Soc. for Industrial and Appl. Math., Philadelphia).
- Coifman, R. R., Meyer, Y., Quake, S. & Wickerhauser, M. V. (1993) in *Progress in Wavelet Analysis and Applications (Toulouse, 1992)* (Frontières, Gif-Sur-Yvette, France), pp. 77–93.
- Donoho, D. L. & Johnstone, I. M. (1994) *Ideal Denoising in an Orthonormal Basis Chosen from a Library of Bases* (Stanford Univ., Stanford, CA).
- Coifman, R. R. & Wickerhauser, M. V. (1992) *IEEE Trans. Info. Theory* **38**, 713–718.
- Coifman, R. R. & Saito, N. (1994) *C. R. Acad. Sci.* **319**, 191–196.
- Babuska, I. (1976) *Numerical Solutions of Partial Differential Equations* (Academic, New York), Vol. 3, pp. 89–116.
- Gilbert, A. C. (1998) *Appl. Comp. Harmonic Anal.* **5**, 1–35.
- Beylkin, G. & Coult, N. (1998) *Appl. Comp. Harmonic Anal.* **5**, 129–155.
- Hackbusch, W. (1985) *Multigrid Methods and Applications* (Springer, New York).
- Knapek, S. (1998) *SIAM J. Sci. Comput.* **20**, 515–533.
- Coifman, R. R. & Lafon, S. (2004) *Appl. Comp. Harmonic Anal.*, in press.
- Fowlkes, C., Belongie, S., Chung, F. & Malik, J. (2004) *IEEE Pattern Anal. Machine Intell.* **26**, 214–225.
- Williams, C. & Seeger, M. (2001) in *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, eds. Leen, T. K., Dietterich, T. G. & Tresp, V. (MIT Press, Cambridge, MA), pp. 682–688.
- Slepian, D. & Pollack, H. O. (1961) *Bell Syst. Tech. J.* **40**, 43–64.
- Slepian, D. (1964) *Bell Syst. Tech. J.* **43**, 3009–3058.