mouth University, 1992.
25. S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets", *IEEE Trans. Info. Theory.* **38**, 2 (1992) 617-643.
26. Y. Meyer, *Ondelettes et opérateurs I: Ondelettes*, Hermann, Paris, 1990.
27. P. Moulin, "Wavelets as a regularization technique for spectral density estimation", in *Time-Frequency and Time-Scale Analysis*, IEEE, New York, 1992, 73-76.
28. E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, "Shiftable multiscale transforms", *IEEE Trans. Info. Theory* **38**, 2, 587-607.
29. C. Stein, "Estimation of the mean of a multivariate normal distribution", *Ann. Statist.* **9** (1981) 1135-1151.
30. G. Wahba and S. Wold, "A completely automatic French curve", *Commun. Statist.* **4** (1975) 1-17.
31. G. Wahba, "Automatic smoothing of the log periodogram", *J. Amer. Statist. Assn.* **75** (1980) 122-132.

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CA
E-mail: donoho@playfair.stanford.edu

References

1. L. Andersson, N. Hall, B. Jawerth and G. Peters, "Wavelets on closed subsets of the real line", in *Recent Advances in Wavelet Analysis*, Larry L. Schumaker and Glenn Webb (eds.), Academic Press, Boston, 1993.
2. A. Cohen, I. Daubechies, B. Jawerth and P. Vial, "Multiresolution analysis, wavelets, and fast algorithms on an interval", *Comptes Rendus Acad. Sci. Paris* (A) **316** (1992) 417-421.
3. R. R. Coifman and Y. Meyer, "Remarques sur l'analyse de Fourier à fenêtre", *Comptes Rendus Acad. Sci. Paris* (A) **312** (1991) 259-261.
4. R. R. Coifman, Y. Meyer and M. V. Wickerhauser, "Wavelet analysis and signal processing", pp. 153–178 in *Wavelets and Their Applications*, M. B. Ruskai et al. (eds.), Jones and Bartlett (Boston) 1992.
5. R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best-basis selection", *IEEE Trans. Info. Theory* **38** (1992) 713-718.
6. R. A. DeVore and B. J. Lucier, "Fast wavelet techniques for near-optimal image processing", *Proc. IEEE Mil. Commun. Conf.*, Oct. 1992. IEEE Communications Society, NY, 1992.
7. D. L. Donoho, *De-Noising via Soft-Thresholding*, Tech. Rept., Statistics, Stanford, 1992.
8. D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Tech. Rept., Statistics, Stanford, 1992.
9. D. L. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Tech. Rept., Statistics, Stanford, 1992.
10. D. L. Donoho and I. M. Johnstone, *Minimax risk over ℓ_p-balls*, Tech. Rept., Statistics, Univ. Calif., Berkeley, 1990.
11. D. Donoho, "Smooth wavelet decompositions with blocky coefficient kernels", to appear in *Advances in Wavelet Analysis*, L. L. Schumaker and G. Webb (eds.), Academic Press, Boston, 1993.
12. D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Tech. Rept., Statistics, Stanford, 1992.
13. D. L. Donoho and I. M. Johnstone, *New minimax theorems, thresholding, and adaptation*, Tech. Rept., Statistics, Stanford, 1992.
14. D. L. Donoho and I. M. Johnstone, *Minimax estimation by wavelet shrinkage*, Tech. Rept., Statistics, Stanford, 1992.
15. D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness by wavelet shrinkage*, Tech. Rept., Statistics, Stanford, 1992.
16. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, *Wavelet Shrinkage: Asymptopia?*, Tech. Rept., Statistics, Stanford, 1993.
17. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, *Density Estimation via Wavelet Shrinkage*, Tech. Rept., Statistics, Stanford, 1993.
18. B. Efron and C. Morris, "Data analysis using Stein's estimator and its generalizations", *J. Amer. Statist. Assn.* **70** (1975) 311–319.
19. J. Froment and S. Mallat, "Second-generation compact image coding with wavelets", in *Wavelets: a Tutorial in Theory and Applications*, C. Chui (ed.), Academic, Boston, 1992, 655-678.
20. Hong-ye Gao, *Choice of Thresholds for wavelet estimation of the log-spectrum*, Tech. Rept., Statistics, Stanford, 1993.
21. Hong-ye Gao, *Spectral Density Estimation via Wavelet Shrinkage*, Tech. Rept., Statistics, Stanford, 1993.
22. I. M. Johnstone, G. Kerkyacharian and D. Picard, "Estimation d'une densité de probabilité par méthode d'ondelettes", *Comptes Rendus Acad. Sciences Paris* (A) **315** (1992) 211-216.
23. G. Kerkyacharian and D. Picard, "Density estimation in Besov Spaces", *Statistics and Probability Letters* **13** (1992) 15-24.
24. Jian Lu, Yansun Xu, J. B. Weaver and D. M. Healy, Jr., *Noise reduction by constrained reconstructions in the wavelet-transform domain.* Department of Mathematics, Dart-

Andrew Bruce and Carl Taswell for many discussions about wavelet software. The NMR datasets were provided by Chris Raphael (Figure 1) and Jeff Hoch (Figure 11), the ESCA dataset by Jean-Paul Bibérian, the image dataset by Ingrid Daubechies, the seismic dataset by Paul Donoho. Many thanks to Tina Sharp for intense last-minute editorial work.

FIGURE 27.

FIGURE 28.

FIGURE 26.

Consider now a simple recursive nonlinear multiresolution scheme based on decimating by factors of 3. The fine-to-coarse mapping is obtained by grouping the signal in triplets of successive points, and replacing each group of three by a single number – the median of the group of 3. (This is a sort of nonlinear Haar analysis, since dyadic Haar wavelets correspond to grouping data in pairs and keeping only the mean of each pair.) This triadic nonlinear coarsening operator gives rise in an obvious way to a telescoping nonlinear multiresolution decomposition. Figure 28 shows the result of setting to zero the fine scale coefficients of this nonlinear triadic transform applied to the noisy data in Figure 26(b). This clearly performs much better than the linear recovery in Figure 27.

Theoretical work to date on nonlinear multiresolution analysis has been done by Ron DeVore (S. Carolina) and Bradley Lucier (Purdue). Interesting applied work with mammograms has been done by Rich Richardson (Univ. of Texas at San Antonio). Doug Martin and Andrew Bruce (Univ. Washington, Seattle), along with the author, have developed a variety of algorithms based on ideas from robust statistics.

Acknowledgements.

It is a pleasure to thank Iain Johnstone, Gérard Kerkyacharian, and Dominique Picard, with whom much of the theory described here has been developed, and to thank Yves Meyer, Ronald Coifman, and Ingrid Daubechies for encouragement at key moments. Specific inspirations provided by the work of Ronald DeVore and Björn Jawerth are also gratefully acknowledged, as well as stimulating conversations with Albert Cohen and Bradley Lucier. Thanks to

Of course, in Figure 25 we are showing what happens when the wavelet transform is segmented exactly at the point of discontinuity. How are we to obtain, in analyzing noisy data, information about the proper segmentation point? Viewing the collection of segmented wavelet transforms with different values of t as a collection of bases $\mathcal{B}^{(t)}$, this is really a problem of selecting a best basis. Therefore we propose a *best-basis segmentation*

$$\mathrm{SURE}(y, \hat{t}) = \min_t \mathrm{SURE}(y, \mathcal{B}^{(t)}).$$

For the dataset in question, the visual performance of the resulting estimate is indistinguishable from that in Figure 25.

FIGURE 25.

8.3. Beyond wavelets III: Nonlinear multiresolution analysis. In our discussion so far, "noise" has always meant Gaussian noise, or something close to Gaussian so that Central Limit Theorem considerations apply. In some sense, linear, orthogonal wavelet analysis is naturally tied to an assumption of Gaussianity. If we instead have very non-Gaussian noise, then *nonlinear* multiresolution analysis becomes practically de rigueur. To see why, we consider Cauchy noise (the symmetric stable law of index 1), independent and identically distributed, with density $f(t) = \pi^{-1}(1+t^2)^{-1}$. In Figure 26 we superpose such noise on a sine-curve. The occasional large noise spikes completely dominate the plot scaling, and nothing of interest remains visible. If we try to "smooth away" the noise by setting to zero the noisy wavelet coefficients at fine scales and inverting the wavelet transform, we still don't see anything useful (Figure 27).

For simplicity we discuss the 1-d case. A *segmented multiresolution analysis* V_j^t of $[0,1]$ is a multiresolution analysis in which $[0,t]$ and $[t,1]$ are somehow kept essentially separate, so that functions in V_j^t need not be continuous at the point t. (Compare the notions of splitting and merging in Andersson, Hall, Jawerth, and Peters [**1**], which could be used to implement segmented multi-resolution decomposition. The specific segmented multi-resolution we use is based on the average-interpolating multi-resolutions in [**11**]).

Obviously, in analyzing objects with discontinuities at the points t, a multiresolution analysis which permits discontinuities at the point t permits better approximation and, ultimately, better compression. Figure 24 makes this point. We have a piecewise linear function, together with a standard (average-interpolating) multi-resolution refinement, and a segmented refinement. The standard multi-resolution smooths out the edge; the segmented refinement preserves the edge.

FIGURE 24.

Corresponding to the segmented refinement is a segmented wavelet transform, with coefficients $\alpha_{j,k}^t$, which depend on the segmentation point t. When we analyze a function π^t which is piecewise polynomial, a discontinuity being allowed only at t, then $P_j^t \pi^t = \pi^t$, which implies that the wavelet coefficients $\alpha_{j,k}^t$ are identically zero for such an object (when the parameter t of the transform and the site t of the discontinuity are really the same.) Figure 25 shows the behavior of de-noising applied to a segmented wavelet transform, and a comparison with de-noising of a non-segmented wavelet transform. The improvement in the neighborhood of the discontinuity is dramatic: a significant reduction in the Gibbs phenomena there.

Figure 22.

Figure 23.

Figure 20.

Figure 21.

ground, we recommend [**12**] where the concept of ideal risk is discussed more carefully, and precedents for nearly attaining it are given.

In [**15**], Donoho and Johnstone have introduced a method of selecting among a family of estimators for use in the wavelet basis – the selection is based on Stein's Unbiased Estimation of Risk (SURE). SURE had previously been used in selecting among families of linear estimates; [**15**] showed that it could be used in selecting among soft-thresholding estimates, achieving results nearly as good as could be obtained with an oracle.

In the basis selection problem, we fix the threshold $t_n = \sqrt{2\log_e(n\log_2(n))}$, and we define

$$\mathrm{SURE}(y,\mathcal{B}) = \epsilon^2 \cdot (n-2)\sum_i (y_i/\epsilon) 1_{|y_i|\le t_n} + \sum \min((y_i/\epsilon)^2, 1).$$

This has the property that when the model $y_i = \theta_i + \epsilon z_i$ holds in the basis $\mathcal{B}$, with z_i i.i.d. $N(0,1)$, then

$$E\,\mathrm{SURE}(y,\mathcal{B}) = E\|(\eta_{t_n}(y_i)) - \theta_i\|_2^2.$$

This is true for all θ; it is the unbiasedness property which explains the initials "URE" in SURE. We therefore have tried selecting a basis by minimizing the SURE of the soft-thresholding estimator in that basis, i.e. finding a basis by the principle

$$\mathrm{SURE}(y,\hat{\mathcal{B}}) = \min_{\mathcal{B}} \mathrm{SURE}(y,\mathcal{B}).$$

It is easy to find a best basis by adapting the Best-Basis algorithm of Coifman and Wickerhauser to the SURE cost function.

To show how this works in examples, Figure 20 presents four signals – two highly oscillatory, where it is expected that wavelets will not work very well, and *Bumps* and *Doppler* from earlier examples, where we have already seen that wavelets work well. Figure 21 shows the same signals with noise. Figure 22 shows the same signals reconstructed using η_{t_n} in the Wavelet Packet basis selected by SURE. Note that the same software has been used, with the same settings on all four signals. The range of reconstructions is made possible by adaptive choice of basis.

For comparison, Figure 23 shows the results of reconstruction of the object *Mishmash* by Splines with automatically chosen tension parameter, by Wavelet Shrinkage, by Fourier Shrinkage, and by SURE selection of wavelet packet basis. Visually, at least, the adaptive basis method outperforms the more traditional techniques (if "wavelet shrinkage" can now be called traditional).

8.2. Beyond wavelets II: Segmented multiresolution analysis. One can view the "Best Basis" methods of Coifman, Meyer, and Wickerhauser as establishing an interesting paradigm with applications far beyond the wavelet-packets – cosine packets setting. For example, we describe work in progress on minimum entropy segmentation.

important mathematical operators (e.g. certain convolutional operators) are almost diagonal in a wavelet basis. The implication for us is that when we need to solve an inverse problem involving such an operator, wavelets are almost as good as eigenfunctions at representing the operator under study, but in our applications they typically are far better than eigenfunctions at representing the object to be recovered, hence the WVD approach beats the traditional SVD approach.

8. Beyond wavelets.

So far, we have emphasized the use of wavelet bases, and the development of methods which are simple, yet in some sense provably optimal for use in those bases. These developments prove that wavelets solve theoretical problems which had attracted considerable activity over many years, and which resisted solution by non-wavelet techniques.

Now we turn to applications which are more complicated and for which the theory is not yet complete. The applications involve modifying or extending wavelets in various ways. Here we present computational examples indicating some of the motivation and some of the possibilities.

8.1. Beyond wavelets I: Adaptive choice of basis. As indicated in the introduction, Coifman, Meyer and Wickerhauser have introduced a family of orthogonal bases, of which wavelet and Fourier bases are special cases, for which there exist fast transforms and which offer promise for analysis of signals which exhibit moderate-duration oscillatory phenomena.

These ideas, translated into a statistical setting, pose a number of interesting issues. As indicated in section 3, if $\mathcal{B}$ denotes an orthogonal basis, the best mean-squared error among linear estimates diagonal in the basis $\mathcal{B}$ is essentially

$$R(f, \mathcal{B}) = \sum_i \min(\theta_i^2, \epsilon^2),$$

where ϵ^2 is the noise level. This goal is attainable only if we know a priori which coordinates are large and which are small, and hence only with the aid of a *coordinate-oracle*. We never have such an oracle at our disposal; nevertheless this is the goal we shall consider. We call it the "Ideal Risk", since it is the Risk attainable by an ideal oracle-assisted algorithm.

Now, when a whole collection of bases is available, it becomes of interest to consider what could be obtained with the aid of a *basis-oracle*, which presents us with the basis $\mathcal{B}^*$ obeying

$$R(f, \mathcal{B}^*) = \min_{\mathcal{B}} R(f, \mathcal{B}).$$

To what extent can we develop a method which approaches the ideal risk, i.e. the risk attainable with a coordinate-and-basis oracle?

Preliminary experiments with the use of Stein's Unbiased Estimate of Risk indicate that we can come surprisingly close to this ideal performance. As backminimax rate only over special subsets of the full range of Besov and Triebel classes [**14**].

Traditional methods, except for the "amount of smoothing" issue, are linear, and cannot compete effectively with the wavelet shrinkage method in cases of high spatial variability – either in practice (e.g. Figures 6, 7, 8) or in theory. In estimating functions of bounded variation, linear methods cannot attain the optimal rate, nor can methods with ideal choice of "amount to smooth"; the wavelet shrinkage method of section 2 attains a mean-squared error of size $(log(n)/n)^{2/3}$ based on n observations, while linear and adaptive linear methods attain only an error of size $n^{-1/2}$.

For inverse problems, WVD has parallel optimality properties. An example of its quantitative advantages is the ability to recover objects in the 2-dimensional Bump Algebra from Radon data, with an error of order $n^{-4/7}$ from n samples, while the SVD and traditional linear methods only achieve the rate $n^{-2/5}$; see [**7**]. Presumably this means that filtered backprojection and similar linear methods now employed in medical scanners can be outperformed by wavelet shrinkage, when the object to be recovered is spatially variable – possessing edges and highly localized features.

7.2. What's so special about wavelets II: Mathematical properties. An interesting aspect of the above theorems in mathematical statistics is how they rely on fundamental facts about wavelet bases derived by mathematicians for other purposes. For the reader's convenience, we briefly point out that various mathematical results on the special properties of wavelet bases imply corresponding statistical results just stated.

Unconditional basis property. A primary preoccupation of Meyer-Lemarie and Frazier and Jawerth has been in showing that wavelets offer unconditional bases of L^2 and also of many smoothness spaces as well. As we have seen, this property of being simultaneously an unconditional basis of many spaces means that shrinkage of wavelet coefficients is a smoothing operation in many different norms simultaneously. [**9**] and [**16**] have shown how this property leads explicitly to the near-minimaxity results quoted above.

Spatial adaptation property. A primary preoccupation of Ron DeVore has been to show that wavelets are good at representing objects in certain special Besov spaces B_τ. The statistical implication: wavelet shrinkage has therefore at least the same ability to estimate spatially adaptive phenomena as various adaptive partitioning and variable-bandwidth kernel estimation schemes in common use in statistics, and conjectured to have good behavior, but for which rigorous theory is harder to get than for wavelets [**12**].

Almost-diagonality property. A considerable body of research by French waveleticians Yves Meyer, Stephane Jaffard, Philippe Tchamitchian, and U.S. waveleticians Michael Frazier and Björn Jawerth has been to show that many a point, the estimator $\hat{f}(t_0)$ attains within a constant factor the minimax behavior among all measurable procedures. Formally this goes as follows. Let $\mathcal{F}(C, \alpha)$ denote the collection of all functions Hölder(-Zygmund) continuous with exponent α and Hölder seminorm bounded by C. For $0 < \alpha < 1$ this means $|f(x) - f(y)| \leq C|x - y|^{\alpha}$, with obvious extensions to $\alpha \geq 1$. Then

$$\sup_{\mathcal{F}(C,\alpha)} E(\hat{f}^*(t_0) - f(t_0))^2 \leq Const \cdot \log(n)^r \cdot \infty_{\hat{f}} \sup_{\mathcal{F}(C,\alpha)} E(\hat{f}(t_0) - f(t_0))^2,$$

where $r = 2\alpha/(2\alpha + 1)$, valid for $0 < C < \infty$, and for $0 < \alpha < \alpha_0$, where α_0 is set by the regularity of the underlying wavelets. Hence a single estimator is within a logarithmic factor of minimax over every Hölder ball. Recent results in statistical decision theory due to Lepskii and to Brown and Low show that this logarithmic factor cannot be removed. No estimator can do essentially better than this uniformly over such a broad range of balls.

[10] shows that the same estimator attains, within logarithmic factors, the optimal rate of convergence in a global ℓ^2 norm simultaneously over all Besov and Triebel balls in a certain range; this range is limited by the wavelet employed. If now $\mathcal{F}(C)$ denotes a Besov ball $B^{\sigma}_{p,q}(C)$ with smoothness degree σ obeying $1/p < \sigma < R$, where R is the regularity of the wavelet employed, then

$$\sup_{B^{\sigma}_{p,q}(C)} E \sum_i (\hat{f}^*(t_i) - f(t_i))^2 \leq Const \cdot \log(n)^r \cdot \infty_{\hat{f}} \sup_{B^{\sigma}_{p,q}(C)} E \sum_i (\hat{f}(t_i) - f(t_i))^2,$$

where $r = 2\sigma/(2\sigma + 1)$.

[13] shows that there is a way to remove these logarithmic factors by clever choice of thresholds; **[15]** shows how to use Stein's Unbiased estimate of Risk **[29]** in order to do so in a practical way.

[16] shows that the same conclusions hold in a wide variety of norms in the Besov and Triebel scales; it is not necessary to use L^2-type losses.

In **[23]**, **[22]**, and **[17]**, Johnstone, Kerkyacharian, and Picard have discovered a variety of nice properties of wavelet shrinkage in the density model, $X_1, \ldots, X_n$ i.i.d. f, though not with the estimator described above, and by completely different methods of proof.

These properties are unprecedented in several ways. For many years, statisticians in the USA, Europe, and Russia have developed techniques for smoothing noisy data for the purpose of signal extraction. Typically, they were working with convolutional smoothers, stiffness-penalized splines, or Fourier-domain damping, and so the questions of how much to smooth, penalize, or damp were paramount **[30]**. Wavelet shrinkage completely avoids these issues, is much simpler, and has very broad near-optimality properties never dreamed of before, and not attainable by older methods. The method achieves, within a logarithmic factor, the minimax risk over each functional class in a wide variety of smoothness classes and with respect to a wide variety of losses **[16]**. Older methods achieve the with threshold

$$t_j = \sqrt{2log(2^j)}\widehat{\mathrm{SDEV}}([y, u_{j,k}])$$

which is an abstract generalization of the earlier examples.

To understand when this all works, compare (6.7)–(6.8) with the usual SVD relations

$$Kf = \sum [Kf, f_\nu]\lambda_\nu f_\nu$$

and

$$f = \sum [Kf, f_\nu]\lambda_\nu^{-1} e_\nu;$$

here the e_ν are eigenfunctions of the operator $K^\star K$ and $f_\nu = Ke_\nu/\|Ke_\nu\|$.

In some sense the approach works when wavelets are "almost eigenfunctions" of $K^\star K$. That is, when the WVD may be defined, we have

$$K\psi_{j,k} = \kappa_j v_{j,k}; \qquad K^\star u_{j,k} = \kappa_j \psi_{j,k};$$

so K is mapping wavelets into vaguelettes and $K^\star$ is mapping vaguelettes into wavelets. Only special operators K will exhibit such character (in the same way that "only" Calderón-Zygmund operators map "atoms" into "molecules"). When one has such an operator, wavelets offer an almost-SVD, where we give up exact invariance under $K^\star K$ in order to get much better representation of the objects f of interest. Examples where the WVD may be defined include Radon transform, Fractional Integration, and various convolution operators.

7. What's so special about wavelets?

Many groups have independently developed methods for noise suppression which are also based on wavelet thresholding in some sense. I think here of Mallat and collaborators (Courant), Coifman and collaborators (Yale), and Healy and collaborators (Dartmouth). These other groups have found that wavelet thresholding methods work well in problems ranging from photographic image restoration to medical imaging. R.A. DeVore (South Carolina) and B.J. Lucier (Purdue) have also come to thresholding, motivated by approximation-theoretic arguments. P. Moulin of Bell Labs has introduced wavelet thresholding techniques for radar imaging.

This agreement of diverse empirical, engineering, and mathematical work is very encouraging, and suggests that wavelet shrinkage will soon have a large impact on how scientists treat noisy data. There is also theoretical work in mathematical statistics, which we describe in a moment, which "proves" that wavelet shrinkage offers special properties.

7.1. What's so special about wavelets I: Advantages in statistical theory. Wavelet shrinkage possesses a disarming simplicity. In fact it achieves many theoretical goals simultaneously. For example, in the context of section 2, [**14**] shows that in estimating a function of unknown Hölder smoothness at

FIGURE 19.

The WVD starts from the representers $\gamma_{j,k}$ solving the quadrature relations (6.4) and identifies constants κ_j so that the functions

$$u_{j,k} = \gamma_{j,k} \cdot \kappa_j$$

make a set of functions with norms bounded above and below. Then the functions $v_{j,k} = K\psi_{j,k}/\kappa_j$ are biorthogonal to $u_{j,k}$ in the data space:

$$[u_{j,k}, v_{j',k'}] = \delta_{j,k;j',k'}.$$

Next one checks that the two sets $(u_{j,k})$ and $(v_{j,k})$ are almost-orthogonal, in the sense that

$$\left\| \sum a_{j,k} u_{j,k} \right\|_{L^2(dt)} \asymp \|(a_{j,k})\|_{\ell^2} \asymp \left\| \sum a_{j,k} v_{j,k} \right\|_{L^2(dt)} .$$

It results that the formal relations

$$Kf = \sum [Kf, u_{j,k}] \kappa_j v_{j,k} \tag{6.7}$$

and

$$f = \sum [Kf, u_{j,k}] \kappa_j^{-1} \psi_{j,k} \tag{6.8}$$

have a content which can be made rigorous. When this is so, inversion from noisy data may be defined by soft thresholding

$$\hat{f} = \sum \eta_{t_j}([y, u_{j,k}]) \kappa_j^{-1} \psi_{j,k}$$

in other words, the linear functional $c_{j,k}$ applied to noiseless data gives the corresponding wavelet coefficient of f. Then applying these to noisy data

$$y_{j,k} = c_{j,k}(d), \tag{6.5}$$

gives noisy measurements of the wavelet coefficients

$$y_{j,k} = \langle \psi_{j,k}, f \rangle + \tilde{z}_{j,k} \tag{6.6}$$

where $\tilde{z}_{j,k}$ is an induced noise process. (6.5)–(6.6) make much better sense than (6.2)–(6.3), and one can follow the three-step de-noising procedure of section 2, using the MAD idea to obtain resolution-level dependent thresholds. This gives a practical method for dealing with rather general inverse problems.

When we apply this formalism to the Radon transform, the results are interesting. A two-dimensional tensor product wavelet basis has indices j, $k = (k_x, k_y)$, and also a directional preference $\epsilon \in \{1, 2, 3\}$. The functionals that solve the quadrature problem

$$c^{(\epsilon)}_{j,k}(Kf) = \langle \psi^{(\epsilon)}_{j,k}, f \rangle$$

have Riesz representers. To describe these, recall the set-up of the tomography problem. We have data

$$d(u, \theta) = (P_\theta f)(u) + z(u, \theta)$$

where $\theta \in [0, 2\pi]$ has to do with the projection angle, and $u \in \mathbf{R}$ with the foot of the projection ray. The representers $\gamma_{(j,k,\epsilon)}$ of the $c^{(\epsilon)}_{j,k}$ have the form

$$\gamma_{(j,k,\epsilon)}(u, \theta) = 2^j \cdot \gamma_{(0,0,\epsilon)}(2^j u - \cos(\theta)k_x - \sin(\theta)k_y).$$

The $\gamma_{(j,k,\epsilon)}$ are all "twisted" dilations of three fixed "mother representers" $\gamma_{(0,0,\epsilon)}$. As j increases, they concentrate around certain sine-curves in the (u, θ) plane. These sine-curves $2^{-j}(\cos(\theta)k_x - \sin(\theta)k_y)$ name certain positions $2^{-j}(k_x, k_y)$ in the original image space.

Figure 19 shows the three mother representers, and an example of a twisted dilation. The diagonal in the direction East-NorthEast is θ, the one in direction North-NorthWest is u. The directional sensitivity of the original wavelets is responsible for the fact that the representers effectively vanish for certain ranges of θ.

[**7**] develops a general formalism for addressing inverse problems using wavelets which generates the above examples as special cases. The idea is to develop a decomposition of the forward operator K in terms of wavelets and *vaguelettes* which, at a formal level, resembles the Singular Value Decomposition (SVD), but which uses a wavelet basis instead of an eigenfunction basis. The idea is that an eigenfunction basis, like the Fourier basis, will have trouble representing objects with spatial variability, and therefore a *Wavelet-Vaguelette decomposition* (WVD) will be a better way to represent many problems than the SVD.

FIGURE 18.

With this resolution-dependent thresholding, the noise is heavily damped, while the main structure in object *Bumps* persists.

6. Continuous inverse problems.

General inverse problems can be conceptualized as observations

$$d(t) = (Kf)(t) + z(t), \qquad t \in \mathcal{T} \tag{6.1}$$

where the index set might even be continuous. Mimicking section 5, we would ideally like to have an operator K^{-1} such that

$$y(\xi) = (K^{-1}d)(\xi) \tag{6.2}$$

satisfies

$$y(\xi) = f(\xi) + \tilde{z}(\xi) \tag{6.3}$$

where $\tilde{z} = K^{-1}z$ is a non-white noise. Unfortunately, in all the really interesting cases K^{-1} does not exist as a bounded operator on spaces to which the noise belongs.

We seek instead to mimick (6.2)–(6.3) in the wavelet domain. We want functionals $c_{j,k}$ with the property that

$$c_{j,k}(Kf) = \langle \psi_{j,k}, f \rangle; \tag{6.4}$$

FIGURE 17.

a formal convolution inverse k^{-1}; we may attempt to invert this relation, forming

$$y_i = (k^{-1} \star d)_i.$$

This is equivalent to observing

$$y_i = x_i + \sigma \cdot (k^{-1} \star z)_i;$$

i.e. observations in a non-white noise.

We propose to reconstruct (x_i) by a three-step process similar to section 2, only again with a threshold that is *level-dependent*. We choose this threshold by the rule

$$t_{j,n} = \sqrt{2\log(n)} \cdot \mathrm{MAD}((w_{j,k})_k)/.6745,$$

where $\mathrm{MAD}((v_i)_i) = \mathrm{Median}((|v_i|)_i)$.

We apply this idea to the system where k is a finite length recursive filter and k^{-1} a finite-length moving average $(1, -1.8, .81)$. This gives the reconstruction that was depicted earlier in Figure 2. The situation in the wavelet domain is depicted in Figure 18. Note that the threshold is again much larger at high-resolution levels than at low ones.

The motivation for this thresholding scheme is similar to that in §5.1. The noise in the wavelet transform is, at each resolution level, a Gaussian noise which is again approximately stationary. We now estimate the variance of the noise by assuming that "most" of the empirical wavelet coefficients at each resolution level are noise, and hence that the median absolute deviation reflects the size of the typical noise. The MAD/.6745 is an estimate of the noise standard deviation.

FIGURE 15.

FIGURE 16.

where Kf is a transformation of f. Examples include: Fourier transformation (magnetic resonance imaging), Laplace transformation (fluorescence spectroscopy), Radon Transformation (medical imaging) and many convolutional transformations (gravity anomalies, infrared spectroscopy, extragalactic astronomy).

Luckily, wavelet methods extend to handle various inverse problems as well. In some sense, such problems become problems of recovering wavelet coefficients in the presence of *non-white* noise. I will briefly discuss two simple examples.

5.1. Numerical differencing. Suppose we wish to reconstruct the discrete signal $(x_i)_{i=0}^{n-1}$, but we have only noisy data about the cumulative of x:

$$d_i = \left(\sum_{t=0}^{i} x_t\right) + \sigma z_i, \ldots i = 1, \ldots, n,$$

where z_i is a standard white Gaussian noise. We may attempt to invert this relation, forming the differences

$$y_i = d_i - d_{i-1},$$

with $y_0 = d_0$, of course. This is equivalent to observing

$$y_i = x_i + \sigma \cdot (z_i - z_{i-1}).$$

i.e. observations in a non-white noise.

We propose to reconstruct (x_i) by a three-step process similar to section 2, only with a threshold that is *level-dependent*. We choose this threshold by the rule

$$t_{j,n} = \sqrt{2\log(n)} \cdot (2\sigma)/\sqrt{n} \cdot 2^{(J-j)/2}, \qquad j = j_0, \ldots, J;$$

this gives the reconstruction depicted in Figure 15. The situation in the wavelet domain is depicted in Figure 16. Note that the threshold is much larger at high-resolution levels than at low ones.

This scheme for thresholding may be motivated as follows. The noise in the wavelet transform is, at each resolution level, a Gaussian noise which is approximately stationary. The variance of the noise at level j grows roughly like 2^j (this is visually apparent). With this resolution-dependent thresholding, the noise is heavily damped, while the main structure in object "Bumps" persists. If we try traditional approaches instead, we get the results in Figure 17. Ideal Fourier damping is unable to suppress the noise.

5.2. Discrete-time deconvolution. Suppose we wish to reconstruct the discrete signal $(x_i)_{i=0}^{n-1}$, but we have only noisy data about a blurred-out x:

$$d_i = (k \star x)_i + \sigma z_i, \qquad i = 1, \ldots, n,$$

where $k \star x$ denotes a discrete convolution $\sum_u k_u x_{t-u}$ and z_i is a standard white Gaussian noise. (We cut corners by ignoring edge-effects.) Assume that we have called the "Log-o-Gram". We then treat the y_k as if they were Gaussian white noise data, with mean $\log(f(\xi_k)) \cdot \frac{\sqrt{6}}{\pi}$ and variance 1; here $\xi_k = 2\pi k/n$. The results of doing this for an AR(6) process which has roots near the unit circle are indicated in Figure 14. This general approach to time series spectra has been investigated by Hong-Ye Gao in his Berkeley Ph.D. thesis. Independently, P. Moulin [**27**] has suggested an approach based on this idea, and proposed a number of variations on choice of threshold, and has generalized the approach to the study of problems in radar imaging.

FIGURE 14.

In this rapid tour, we are cutting a few corners. A careful analysis of the theory underlying the Gaussian white noise model shows that for treating the density and spectral density case, we ought to use resolution-dependent thresholds which depend on the large-deviations properties of Poisson and Exponential noise. Otherwise we will tend to see tiny noise-induced 'blips' in an otherwise smooth curve (compare Figure 14). Hong-Ye Gao is writing his Ph.D. thesis at Berkeley in part on a finer analysis of this question in the time series setting; Eric Kolaczyk is writing his Ph.D. thesis at Stanford in part on an analysis of this question in the density setting.

5. Discrete inverse problems.

Many interesting problems having to do with noisy data involve *indirect* measurements. Here we obtain measurements

$$y_i = (Kf)(t_i) + \epsilon z_i$$

Similarly, suppose we have a random sample $X_1, \ldots, X_m$, iid f, where f is an unknown density on $[0,1]$. Partition $[0,1]$ into $n = 2^{J+1}$ intervals , where $n \approx m/4$, and let N_i be the count of observations falling into the i-th interval. Then set

$$y_i = 2 \cdot \sqrt{N_i + 3/8}, \quad i = 1, ..., n,$$

and behave as if the y_i were Gaussian with mean $2 \cdot \sqrt{f(i/n)}$ and variance 1. This is connected with John Tukey's "Rootogram".

Data Courtesy of Jean-Paul Bibérian (Marseille)

FIGURE 13.

In another direction, suppose we have time series data $(x_t)_{t=0}^{n-1}$, $n = 2^{J+1}$, and we wish to estimate the spectral density function $f(\xi)$ of the (supposed) underlying second-order stationary process. We calculate the periodogram

$$I_k = n^{-2} \left| \sum_t x_t e^{i2\pi(t-1)(k-1)/n} \right|^2, \qquad k = 0, \ldots, n-1,$$

and apply the Wahba (1980) variance-stabilizing transformation to the log-periodogram:

$$y_k = (\log(I_k) + \gamma) \cdot \frac{\sqrt{6}}{\pi}, \qquad k = 1, ..., n/2-1, n/2+1, \ldots, n-1$$

where $\gamma \sim .57721...$ is the Euler-Mascheroni constant, and a modification is required for the exceptional Fourier frequencies $k = 0, n/2$. This object might be they propose $C\sqrt{\log(n)}\sigma$. Mallat's multiscale-edge denoising involves a related thresholding principle, compare [**25**].)

In low lighting, the photon counting model $N_{i_1,i_2} \sim Poisson(f(i_1/m, i_2/m))$ is appropriate. To such data we would apply the Anscombe (1948) variance-stabilizing transformation

$$y_{i_1,i_2} = 2 \cdot \sqrt{N_{i_1,i_2} + 3/8}, \quad i_1, i_2 = 0, ..., m-1$$

and act as if the data arose from the Gaussian white noise model, with $\sigma = 1$.

Figure 12.

The results of doing this, in 1-dimension, on an ESCA spectrum, are shown in Figure 13.

FIGURE 11.

"noise" throughout the signal – a very bad pointwise behavior. This illustrates that simple thresholding in the Fourier domain will not only give worse mean-squared errors, it will give unacceptable visual artifacts as well.

We conclude that the shrinkage of coefficients by soft-thresholding is in some sense visually adapted to use with the wavelet transform. Compare also [**10**].

4. Extensions: images, photon counts, densities, spectra.

The de-noising method of section 2 applies surprisingly widely. For example, if we had two-dimensional image data $y_{i_1,i_2} = f(i_1/m, i_2/m) + \epsilon z_{(i_1,i_2)}$ $i_1, i_2 = 0, ..., m-1$ with $z_{(i_1,i_2)}$ white Gaussian noise, we would just use a 2-d pyramid filtering, and proceed as before, using the same three-step formalism with $n = m^2$. Figure 12 presents a 2-d image de-noising example.

(The application of thresholding to 2-d wavelet-like transforms has been discussed by Simoncelli, et al.[**28**] and by DeVore and Lucier [**6**]. DeVore and Lucier even find, by a different route, thresholds of the same general form we suggest;

Fourier basis. Traditional methods of smoothing are effectively little else than

FIGURE 10.

(non-ideal) diagonal projectors in the Fourier basis. At the same time, soft-thresholding closely mimicks an ideal diagonal projector in the wavelet basis [**12**]. The compression advantages of the wavelet basis are responsible for the mean-squared error advantages of wavelet shrinkage.

3.2. Why it works II: Unconditional basis. A very special feature of wavelet bases is that they serve as unconditional bases, not just of L^2, but of a wide range of smoothness spaces, including Sobolev and Hölder classes.

As a consequence, "shrinking" the coefficients of an object towards zero, as with soft-thresholding, acts as a "smoothing operation" in any of a wide range of smoothness measures.

The same can not be said of the Fourier basis. Kahane, Katznelson, and De Leeuw (see reference in Y. Meyer's book, volume 1, page 1) have shown that for functions on the circle, given any sequence of Fourier coefficients in ℓ^2 – perhaps the coefficients of an object that has square-integrable singularities on a countable dense subset of the circle – there is a continuous function that has each one of its Fourier coefficients *larger* than the given coefficients. In other words, the *smaller* coefficients correspond to the more bizarre object.

This can all be illustrated by example on the computer. In Figure 11 we display two signals – one a signal gathered by a seismic exploration crew, another an NMR spectrum. We also display reconstructions using only the 100 largest coefficients in the wavelet domain and in the Fourier domain, respectively. Note that reconstructions from thresholding in the Fourier domain display a kind of

FIGURE 9.

amplitude than ϵ, and "kills" all coefficients where θ_i is smaller in amplitude than ϵ. (This ideal is unattainable, since it requires knowledge of θ, which we don't know). The ideal mean squared error is

$$R(\hat{\theta}^{IDEAL}, \theta) = \sum_i \min(\theta_i^2, \epsilon^2).$$

Define the "compression number" c_n as follows. With $|\theta|_{(k)} = k$-th largest amplitude in vector (θ_i), set $c_n \equiv \sum_{k>n} |\theta|_{(k)}^2$. This is a measure of how well the vector θ can approximated by a vector with n nonzero entries. Setting $N(\epsilon) = \#\{i : |\theta_i| \geq \epsilon\}$,

$$\begin{aligned} \sum_i \min(\theta_i^2, \epsilon^2) &= \epsilon^2 \cdot \#\{i : |\theta_i| \geq \epsilon\} \\ + \quad \sum_i \theta_i^2 1_{\{i:|\theta_i| \leq \epsilon\}} &= \epsilon^2 \cdot N(\epsilon) + c_{N(\epsilon)}, \end{aligned}$$

so this ideal risk is explicitly a measure of the extent to which the energy is compressed into a few big coefficients. (For more on this connection, see [**9**].)

Figure 10 shows the extent to which the different orthogonal bases compress the objects. The logarithm of the compression numbers is shown, plotted against n. The medium heavy line shows compression numbers in the Fourier Basis; the very heavy line (consisting of very closely spaced '+' signs – see 10(a)) marks the Haar compression numbers; and the thin line marks the compression using nearly-symmetric Daubechies wavelets having 8 vanishing moments. The wavelet basis generally wins, though with object *Blocks*, the Haar basis wins. Hence, ideal diagonal projectors work better in the wavelet basis than in the linear smoothers, one based on fitting splines under tension with adaptively chosen tension parameter, and one based on truncating the empirical Fourier series with adaptively chosen truncation. (Adaptation using Stein's Unbiased Estimates of Risk [**29**]). The adaptive spline under tension suppresses noise, but at the expense of significantly broadening, and in fact erasing, certain features. The adaptive Fourier Series estimate leaves features sharp, but does not really suppress the noise.

FIGURE 8.

3. Why it works.

3.1. Why it works: Data compression. A depiction of the wavelet shrinkage method in operation is given in Figure 9. Here we use Haar-basis shrinkage on a noisy version of object *Blocks*. The figure shows the original, noisy data (a), the noisy Haar coefficients (c), the thresholded coefficients (d), and the reconstruction (b). The method works because the Haar transform of the noiseless object *Blocks* compresses the ℓ^2 energy of the signal into a very small number of (consequently) very large coefficients. On the other hand, Gaussian white noise in any one orthogonal basis is again a white noise in any other (and with the same amplitude). Thus, in the Haar basis, the few nonzero signal coefficients really stick up above the noise. Therefore, the thresholding has the effect that it kills the noise while not killing the signal.

For a more formal argument, suppose we have data $d_i = \theta_i + \epsilon z_i$, $i = 1, ..., n$, where z_i is a standard white noise, and we wish to recover (θ_i). The ideal diagonal projector is the one which "keeps" all coefficients where θ_i is larger in

FIGURE 6.

FIGURE 7.

Figure 4.

Figure 5.

a response to such outliers, based on a nonlinear multiresolution analysis centered around L^1 or median fits. When the data are contaminated by extremely long-tailed error distributions, such as the Cauchy distribution, such nonlinear multiresolutions provide plausible reconstructions in cases where standard linear multiresolutions behave horribly.

Overview. In this paper we aim only to show that wavelets and associated ideas can make serious contributions to problem areas where there is already a considerable amount of interest, and to show that wavelets and associated ideas open up totally new questions in other areas. There are many other applications of wavelets in data analysis and signal processing, but we limit ourselves here to those areas where the author is directly involved. We attempt to reference a variety of work on wavelets in reconstruction and recovery, so that the reader may also find out about what others are doing. The wide variety of activities in the areas we discuss makes us hopeful that wavelets will soon have a large impact on the way in which scientists and engineers treat noisy and indirect observations.

2. De-noising by soft-thresholding.

Suppose we are interested in a function $f(t)$ on the unit interval $t \in [0,1]$ and we have $n = 2^{J+1}$ data $y_i = f(t_i) + \sigma z_i$, $i = 1, \ldots, n$; here the t_i are equispaced and the z_i a white noise. Donoho and Johnstone [**12**] propose a three step method for recovery of $f(t)$.

(1) Perform the pre-conditioned, interval-adapted, pyramid wavelet filtering of Cohen, Daubechies, Jawerth, and Vial [**2**] to the data $\beta_{J+1,k} = y_k/\sqrt{n}$, yielding noisy wavelet coefficients $w_{j,k}$, $j = j_0, \ldots, J$, $k = 0, \ldots, 2^j - 1$.

(2) Apply the soft-threshold nonlinearity $\eta_t(w) = sgn\,(w)(|w| - t)_+$ to the noisy empirical wavelet coefficients, with threshold $t = \sqrt{2\log(n)}\sigma/\sqrt{n}$, yielding estimates $\hat{\alpha}_{j,k}$.

(3) Set all wavelet coefficients $\hat{\alpha}_{j,k} = 0$ for $j > J$, invert the wavelet transform, producing the estimate $\hat{f}(t)$, $t \in [0,1]$.

This method shrinks the empirical wavelet coefficients towards zero. Statisticians consider this an example of multivariate shrinkage estimates, e.g. Efron and Morris [**18**], Stein [**29**].

To see how this works, we take four functions, *Blocks*, *Bumps*, *HeaviSine*, and *Doppler*, illustrated in Figure 4. Here $n = 2048 = 2^{11}$. Noisy versions are depicted in Figure 5. Reconstructions by the method are depicted in Figure 6. The reconstructions have two properties.

(1) The noise has been almost entirely suppressed.

(2) Features sharp in the original remain sharp in reconstruction.

This behavior is very different from traditional linear methods of smoothing, which achieve noise suppression only by broadening features significantly. For comparison, Figures 7 and 8 show the results of two state-of-the-art adaptive that minimizes a certain measure of the "entropy" of the sequence, leading to a transform which in some rhetorical sense renders the data maximally simple.

Transposing these ideas into a statistical setting leads to the question of selecting a best basis for de-noising a given dataset. In section 8.1 below, we describe a method based on Stein's Unbiased Risk Estimate. Figure 3 shows the results in recovering a signal which is a superposition of moderate-duration oscillatory phenomena, from data containing both signal and white noise. Reconstruction by denoising in the adaptively-selected wavelet packet basis is much better than in the wavelet basis.

FIGURE 3.

Minimum entropy segmentation. Wavelet methods are often used in the analysis of objects containing edges – for example in 2-d image processing. Wavelets behave well but not ideally in the presence of such 2-d edges. In section 8.2 we describe a response to this, based on defining edge-preserving multiresolution operators and corresponding edge-adapted wavelet bases. The issue of adapting to edges in the 2-d transform is then, in principle, simply one of selecting that edge-adapted basis which optimally compresses the object at hand. Such a selection may be obtained by a minimum entropy criterion (noiseless data), or else by minimizing the SURE (noisy data). As a side benefit, denoising in the selected basis does not erode the edges present in the images of the object, as denoising in the wavelet basis is sometimes said to do.

Nonlinear multi-resolutions. Wavelet methods are sometimes used in the analysis of data contaminated by severe outliers. In section 8.3 we describe

Sections 5 and 6 below discuss this approach to inverse problems in more detail.

FIGURE 2.

Wavelet optimality. Wavelet methods for de-noising and de-blurring have recently attracted considerable interest; the author is aware of efforts, in fields ranging from medical imaging to synthetic aperture radar, where some variant of thresholding of wavelet coefficients is being tried. A complement to such promising empirical work is recent theoretical work which shows that from a variety of points of view these simple nonlinear wavelet methods outperform the traditional linear methods such as splines, Fourier series, and kernel-based smoothers. Some of this theoretical work, and the relation with mathematical results on wavelet bases, is described in section 7 below.

Wavelet packet de-noising. Wavelet bases are not well-suited to representing objects containing sinusoidal oscillations of moderate duration. Coifman and Meyer [**3**] have introduced *local cosine* bases and Coifman, Meyer, and Wickerhauser [**4**] *wavelet packet* bases. In constructing these bases the technical ideas underlying the wavelet transformation are deployed in different ways, leading to a large number of orthogonal basis, each one a serious alternative to classical Fourier analysis and classical wavelet basis. These transforms are better suited than wavelets for certain specific problems; an example might be in the analysis of acoustic phenomena consisting of moderate-duration damped sinusoids. There are many such transforms, though, and it is important to select those which are well-adapted to the signal at hand. Coifman and Wickerhauser [**5**] have introduced a method of selecting among all these transforms for the one

Figure 1 gives an illustration of this method in action. An NMR signal is transformed, thresholded, and inverse transformed. The result has a noise-free visual appearance; this has been achieved without broadening features. Sections 2-4 below deal with wavelet shrinkage, its heuristic basis, and its diverse applications.

Data Provided by Chris Raphael (Stanford)

FIGURE 1.

Wavelet solution of linear inverse problems. Often scientific data are indirectly observed, as well as noisy; either the object is blurred (e.g. by a convolution operator or "point-spread function"), or else it is observed in an entirely different domain (think of the Radon transform, which gives data on line integrals of the object rather than the object itself). Such problems are typically ill-posed, in that naive attempts to undo the blurring or indirection give, in the presence of even small amounts of noise, completely useless reconstructions.

Recently the author [**7**] has proposed a Wavelet-Vaguelette Decomposition of inverse problems. Using this, one transforms the noisy, blurred data, using vaguelettes, into the wavelet domain, then thresholds the wavelet coefficients, and then applies an inverse wavelet transform. Figure 2 shows the use of this method in operation on a deconvolution problem. The improvement over naive deconvolution is evident.

Proceedings of Symposia in Applied Mathematics
Volume 00, 1993

Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data

DAVID L. DONOHO

ABSTRACT. We describe wavelet methods for recovery of objects from noisy and incomplete data. The common themes: (a) the new methods utilize nonlinear operations in the wavelet domain; (b) they accomplish tasks which are not possible by traditional linear/Fourier approaches to such problems. We attempt to indicate the heuristic principles, theoretical foundations, and possible application areas for these methods. Areas covered: (1) Wavelet De-Noising. (2) Wavelet Approaches to Linear Inverse Problems. (4) Wavelet Packet De-Noising. (5) Segmented Multi-Resolutions. (6) Nonlinear Multi-resolutions.

1. Introduction.

With the rapid development of computerized scientific instruments comes a wide variety of interesting problems for data analysis and signal processing. In fields ranging from Extragalactic Astronomy to Molecular Spectroscopy to Medical Imaging to Computer Vision, one must recover a signal, curve, image, spectrum, or density from incomplete, indirect, and noisy data.

What can wavelets contribute to this already intensely developed and rapidly advancing field? As it turns out quite a lot – both in theory and practice. In this paper we will give a brief discussion of several contributions.

Wavelet shrinkage. Wavelet shrinkage refers to reconstructions obtained by wavelet transformation, followed by shrinking the empirical wavelet coefficients towards zero, followed by inverse transformation.

1991 *Mathematics Subject Classification.* Primary 62A99, 62M10, 62M20, 62P99; Secondary 46B10.

Partially supported by NSF DMS 92-09130 (Stanford), and by ONR N00014-92-0066 (Statistical Sciences, Inc.).