Wavelet Shrinkage and W.V.D.: A 10-minute tour

David L. Donoho

Stanford University

Contents

Based on presentation at the International Conference on Wavelets and Applications, Toulouse, France, June, 1992. Supported by NSF DMS 92-09130. With appreciation to S. Roques for patience and Y. Meyer for encouragement. It is a pleasure to thank Iain Johnstone with whom many of these theoretical results have been derived, and Carl Taswell with whom Johnstone and I have developed the software used here.

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.

Recently, it has been shown by the author and his collaborators Iain Johnstone (Stanford)/, Gérard Kerkyacharian (Amiens), and Dominique Picard (Paris VII) that shrinking noisy wavelet coefficients via thresholding offers very attractive alternatives to existing methods of recovering signals from noisy data. Our new methods have theoretical properties of adaptive minimaxity that far surpass anything previously known. Other groups have independently developed methods for de-noising 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. This agreement of diverse theoretical and empirical work is very encouraging, and suggests that wavelets will soon have a large impact on how scientists treat noisy data.

In this brief tour, I will only describe the mechanics of some wavelet shrinkage techniques and give examples. Software is available to compute all the displays presented in this paper; discussion I mention work which proves the various theoretical advantages of the new techniques.

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 [DJ92a] 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 (1992) [CDJV] 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] Setting all wavelet coefficients $\hat{\alpha}_{j,k} = 0$ for $j > J$, invert the wavelet transform, producing the estimate $\hat{f}(t)$, $t \in [0,1]$.

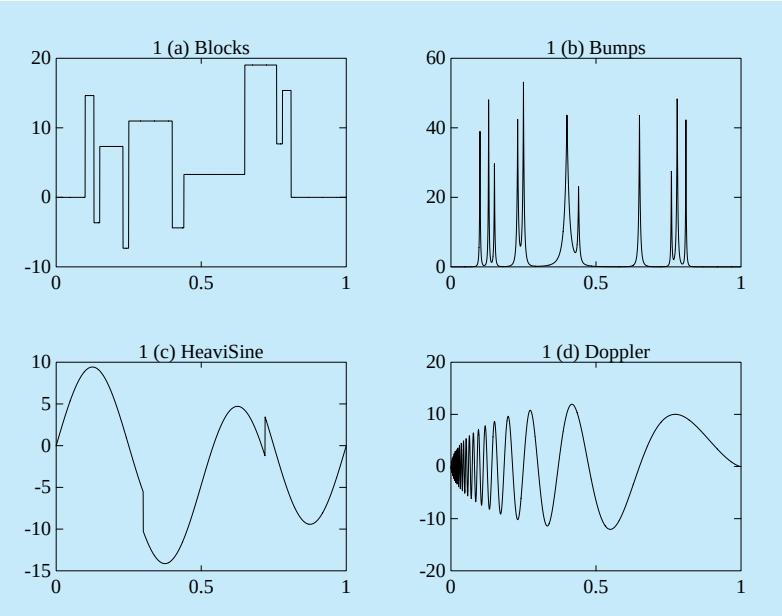

Figure 1: Four functions (*Blocks, Bumps, Heavisine* and *Doppler*)

This method shrinks the empirical wavelet coefficients towards zero. Statisticians consider this an example of multivariate shrinkage estimates, e.g. Efron & Morris (1975) [EM75], Stein (1981) [St81].

To see how this works, we take four functions, *Blocks*, *Bumps*, *HeaviSine*, and *Doppler*, illustrated in figure 1. Here $n = 2048 = 2^{11}$. Noisy versions are depicted in figure 2. Reconstructions by the method are depicted in figure 3. 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, figure 4, and 5, show the results of two state-of-the-art adaptive 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 [St81]). 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.

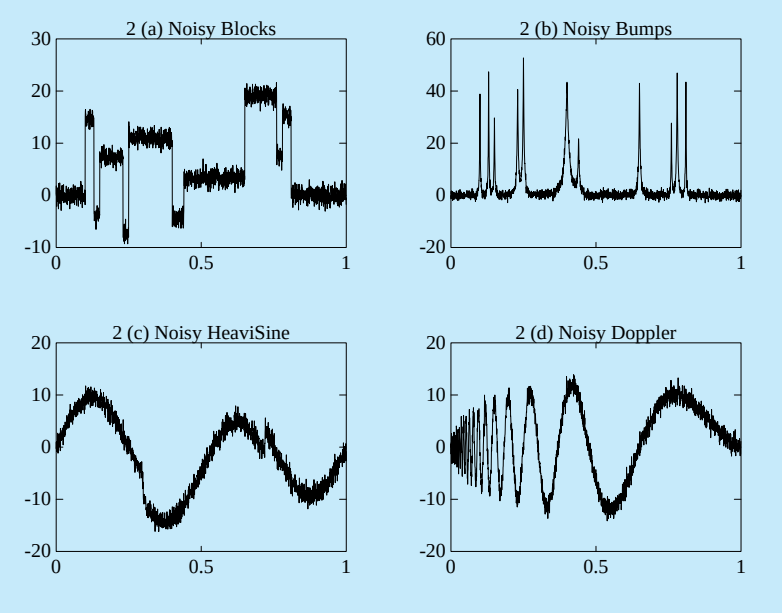

Figure 2: Four functions with noise

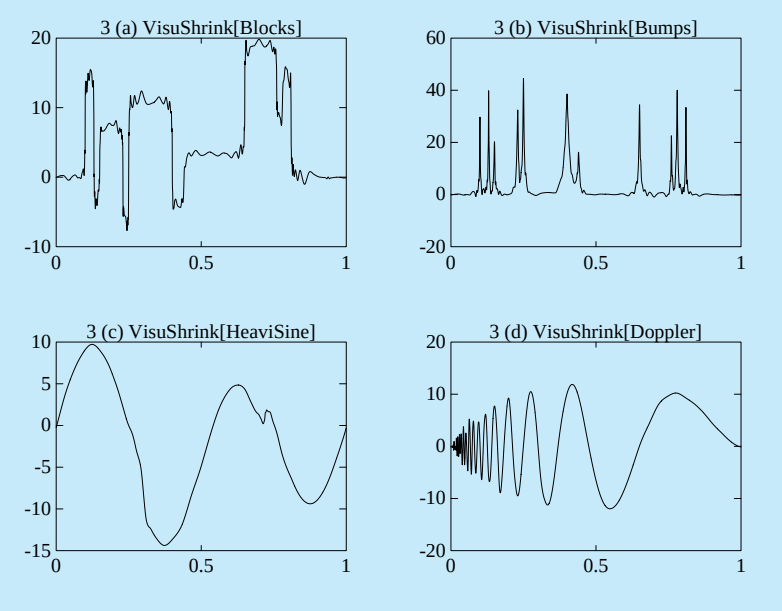

Figure 3: Reconstructed functions

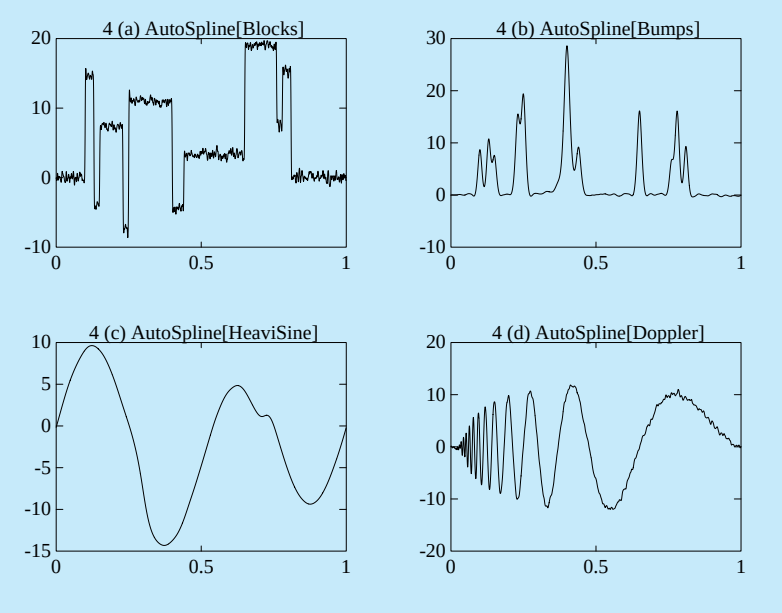

Figure 4: Functions smoothed by adaptive splines

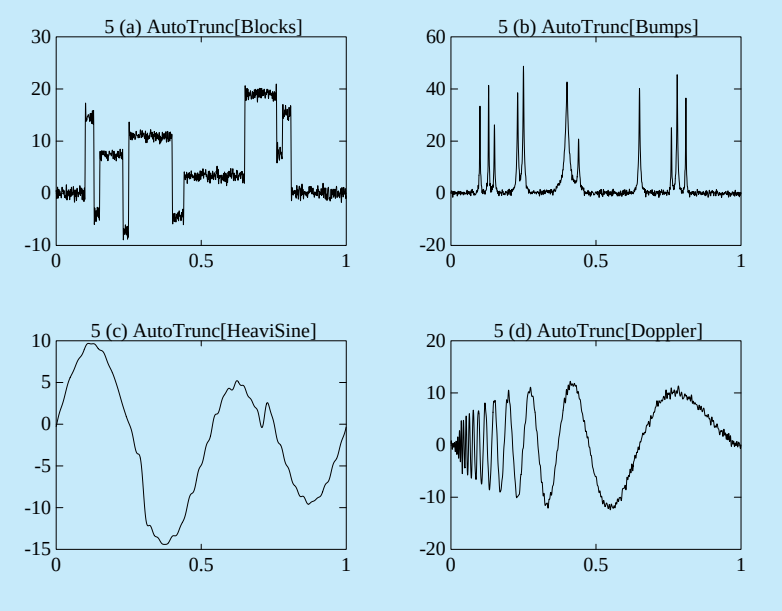

Figure 5: Functions smoothed by Stein's method [St81]

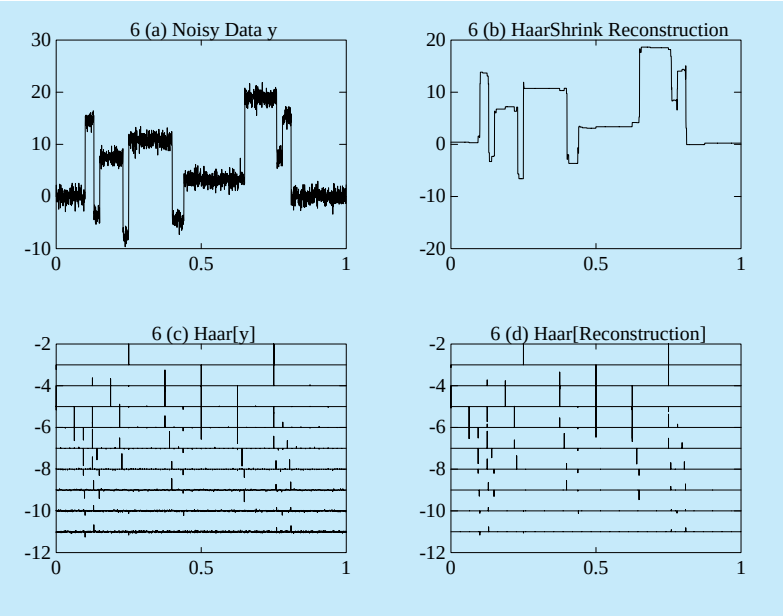

Figure 6: Wavelet shrinkage – Haar wavelets

3 Why it works: Data Compression

A depiction of the wavelet shrinkage method in operation is given in Figure 6. Here we use Haar-basis shrinkage on a noisy version of object *Blocks*. The figure shows the original, noisy data, the noisy Haar coefficients, the thresholded coefficients, and the reconstruction. 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 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

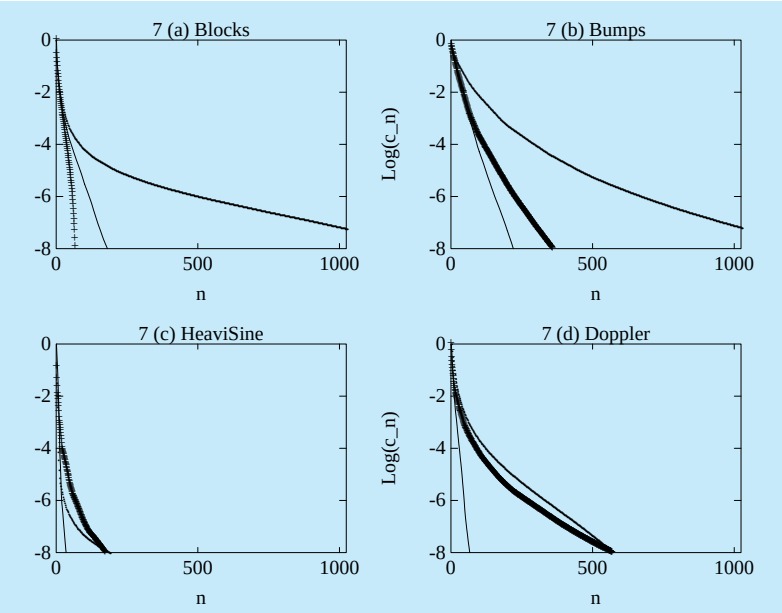

Figure 7: Compression numbers for Fourier (——), Harr (+++) and Daubechies-8 (——) bases

the vector θ can approximated by a vector with n nonzero entries. Setting $N(\epsilon) = \#\{i : |\theta_i| \geq \epsilon\}$,

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

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 [D92c]).

Figure 7 shows the extent to which the different orthogonal bases compress the objects. The logarithm of the compression numbers is shown, plotted against n. The heavy line shows compression numbers in the Fourier Basis; the '+' 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 Fourier basis. Traditional methods of smoothing are effectively little else than (non-ideal) diagonal projectors in the Fourier basis. At the same time, soft-thresholding closely mimicks an ideal diagonal projector in the wavelet basis [DJ92a]. The compression advantages of the wavelet basis are responsible for the mean-squared error advantages of wavelet shrinkage.

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 $n = m^2$.

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 $\sigma = 1$.

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". The results of doing this on a density f which is a mixture of normal densities are depicted in figure 8. 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

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

and apply the Wahba (1980) [W80] 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 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 9.

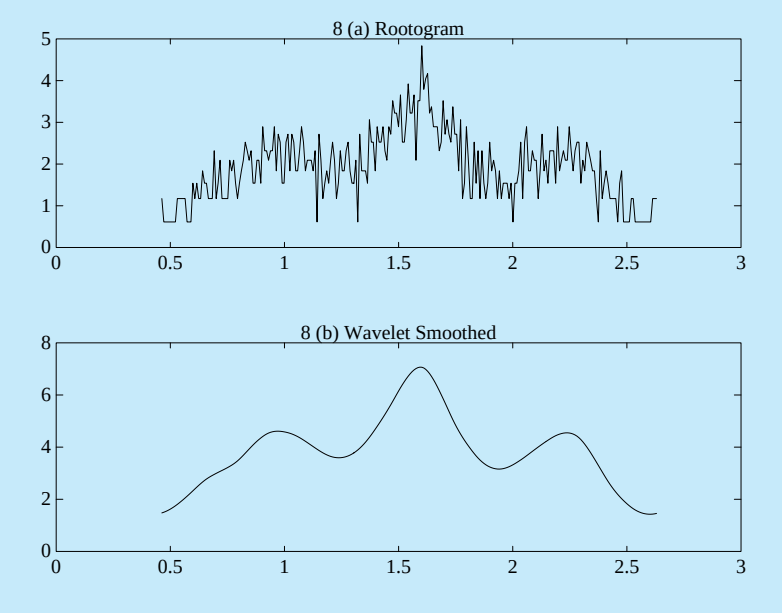

Figure 8: Smoothed rootogram

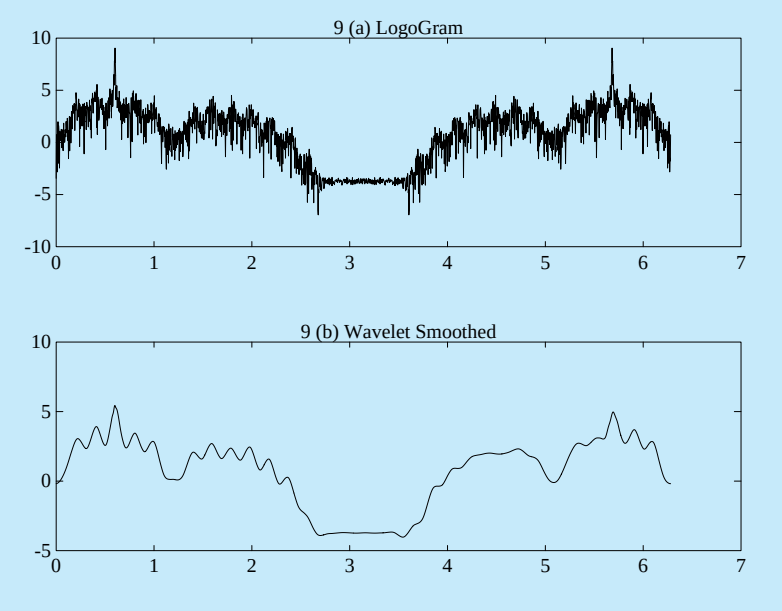

Figure 9: Smoothed logogram

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. Hong-Ye Gao at Berkeley is writing his Ph.D. thesis in part on a finer analysis of this question.

5 Discrete Inverse Problems

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

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

where Kf is a transformation of f. Examples include: Fourier transformation (magnetic resonance imaging), Laplace transformation (fluoresecence 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 problem of recovering wavelet coefficients in the presence of *non-white* noise. In this rapid tour I will 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:

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

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

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

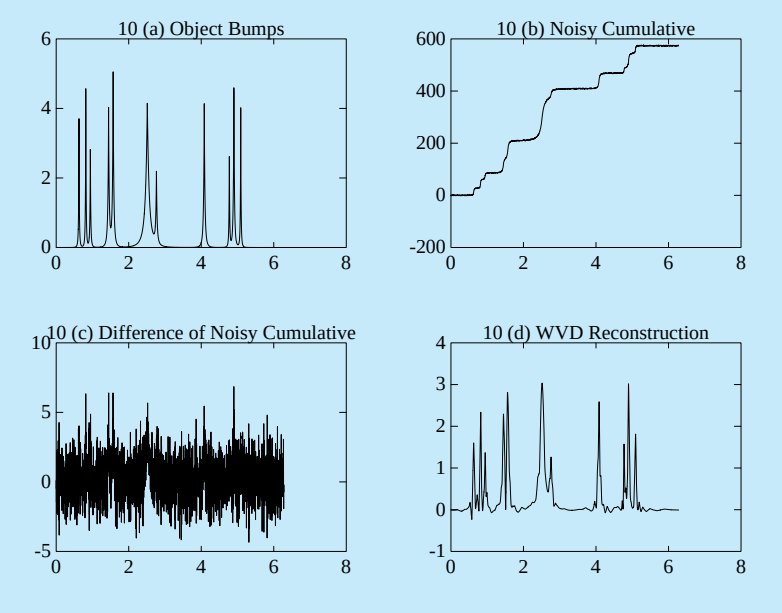

Figure 10: Smoothed by level dependent thresholds

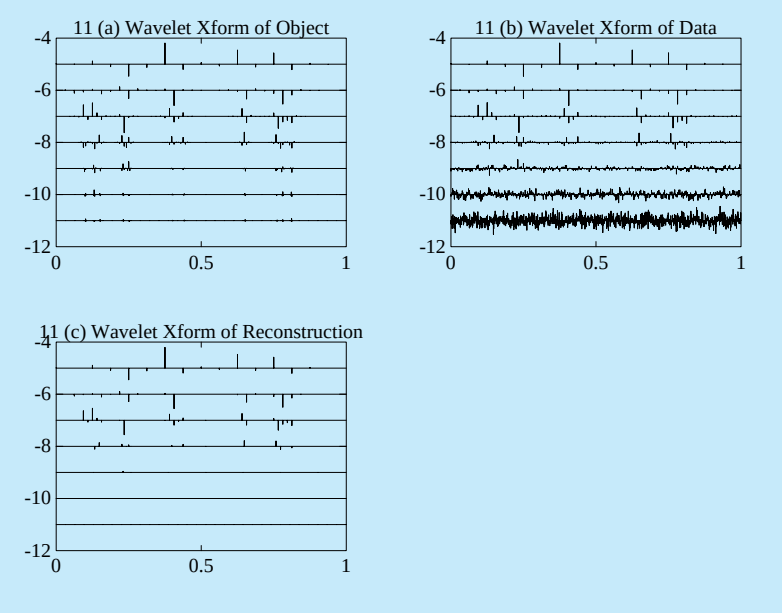

Figure 11: Wavelet amplitudes for level dependent thresholding

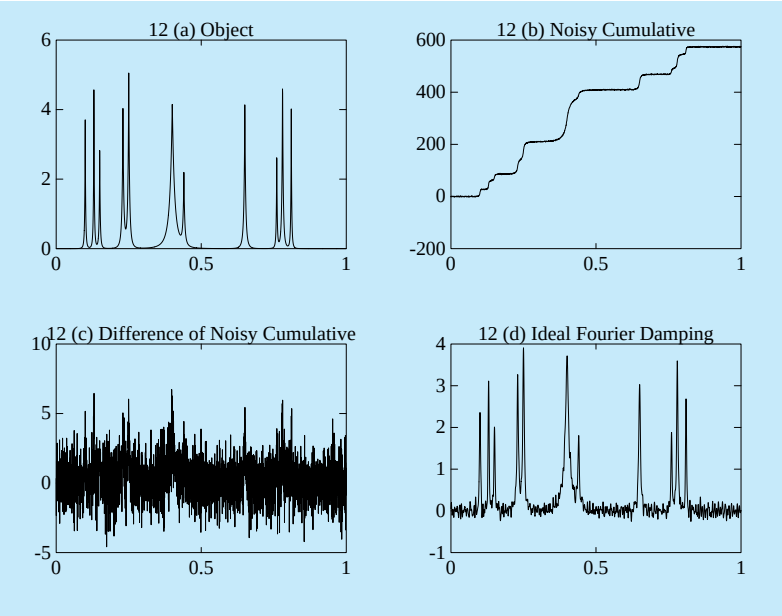

Figure 12: Ideal Fourier damping

this gives the reconstruction depicted in figure 10. The situation in the wavelet domain is depicted in figure 11. 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 instead try traditional approaches, we get the results in figure 12. 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 a formal convolution inverse k^{-1}; we may attempt to invert this relation, forming

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

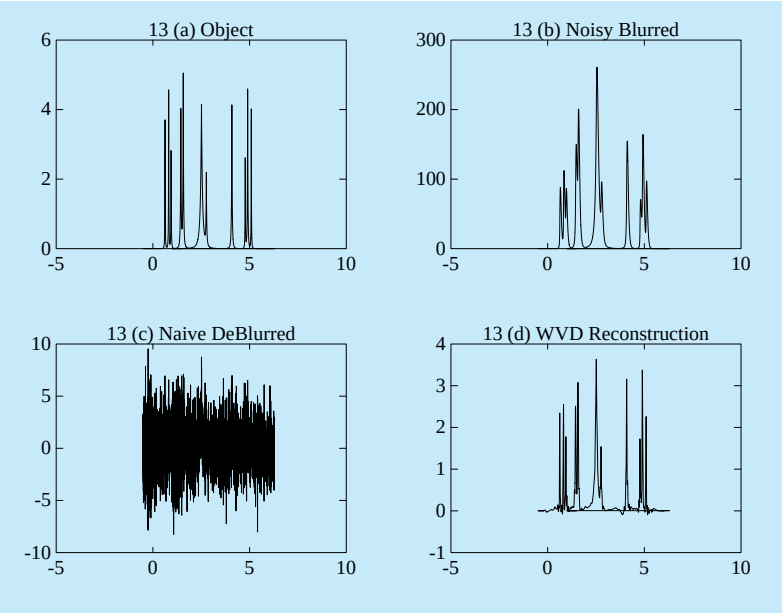

Figure 13: Resolution dependent thresholding

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

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

where $MAD((v_i)_i) = Median((|v_i|)_i)$.

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 depicted in figure 13. The situation in the wavelet domain is depicted in figure 14. 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. We 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. With this resolution-dependent thresholding, the noise is heavily damped, while the main structure in object *Bumps* persists.

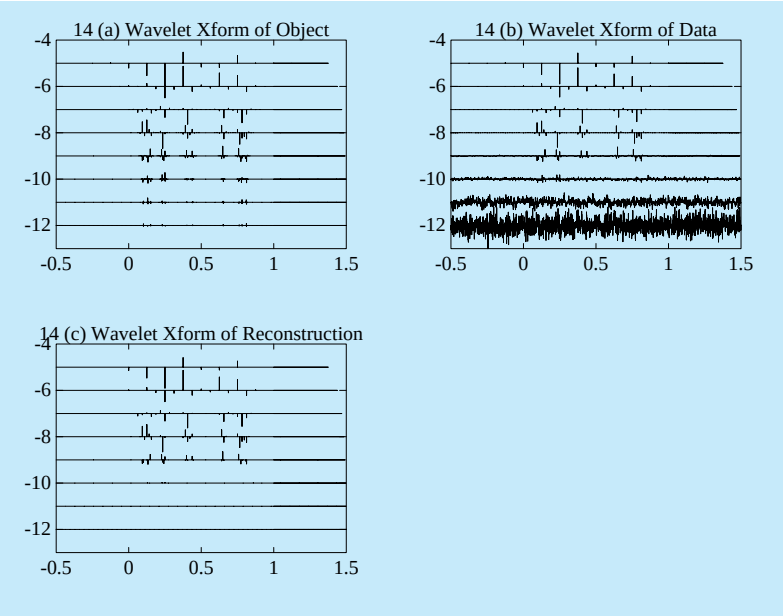

Figure 14: Wavelet amplitudes for resolution dependent thresholding

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

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

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

$$c_{j,k}^{(\epsilon)}(Kf) = \langle \psi_{j,k}^{(\epsilon)}, f \rangle$$

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 $c_{j,k}^{(\epsilon)}$ 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 15 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 reponsible for the fact that the representers effectively vanish for certain ranges of θ.

[D92a] 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.

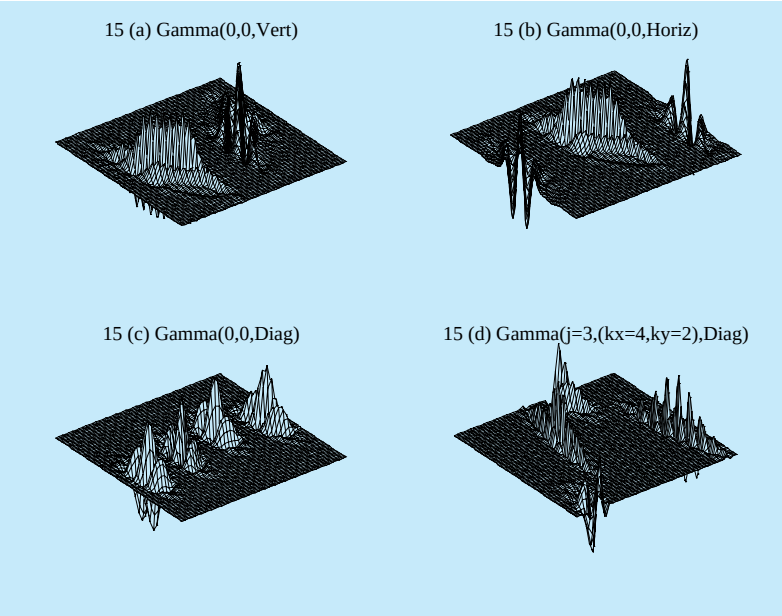

Figure 15: Radon transform

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

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

etc. 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

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

with threshold

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

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 Theory

The methods just described possess a disarming simplicity. In fact they achieve many goals simultaneously. For example, in the context of section 2, [DJ92c] shows that in estimating a function of unknown Hölder smoothness at a point, the estimator $\hat{f}(t_0)$ attains within a constant factor of the optimal behavior among all measurable procedures. [DJ92d] 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 limited by the wavelet employed. [DJ92b] shows that there is a way to remove these logarithmic factors by clever choice of thresholds; [DJ92d] shows how to use Stein's Unbiased estimate of Risk [St81] in order to do so in a practical way.

In [KP] and [JKP], 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 sought to smooth 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 [WW75]. 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. Older methods achieve the minimax rate only over special subsets of the full range of Besov and Triebel classes.

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 3, 4, 5) 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 [D92a]. 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.

References

[CDJV] Cohen, A., Daubechies, I., Jawerth, B., and Vial, P. (1992). Multiresolution analysis, wavelets, and fast algorithms on an interval. To appear, *Comptes Rendus Acad. Sci. Paris* (A).

[D92a] Donoho, D.L. (1992a) Nonlinear solution of linear inverse problems via Wavelet-Vaguelette Decomposition. Technical Report, Department of Statistics, Stanford University.

[D92b] Donoho, D.L. (1992b) Unconditional bases are optimal bases for data compression and for statistical estimation. Technical Report, Department of Statistics, Stanford University.

[D92c] Donoho, D.L. (1992c) De-Noising via Soft-Thresholding. Technical Report, Department of Statistics, Stanford University.

[DJ92a] D.L. Donoho & I.M. Johnstone, (1992a) Ideal Spatial Adaptation via Wavelet Shrinkage. Department of Statistics, Stanford University.

[DJ92b] D.L. Donoho & I.M. Johnstone, (1992b) Minimax Estimation via Wavelet Shrinkage. Department of Statistics, Stanford University.

[DJ92c] Donoho, D.L. & Johnstone, I.M. (1992c). New minimax theorems, thresholding, and adaptation. Manuscript.

[DJ92d] Donoho, D.L. & Johnstone, I.M. (1992d). Adapting to unknown smoothness by wavelet shrinkage. Manuscript.

[EM75] Efron, B. and Morris, C. (1975) Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assn.* **70**, 311-319.

[JKP] Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1992) Estimation d'une densité de probabilité par méthode d'ondelettes. To appear *Comptes Rendus Acad. Sciences Paris* (A).

[KP] Kerkyacharian, G. and Picard, D. (1992) Density estimation in Besov Spaces. *Statistics and Probability Letters* **13** 15-24

[St81] Stein, C. (1981) Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.

[WW75] Wahba, G. and Wold, S. (1975) A completely automatic French curve. *Commun. Statist.* **4** pp. 1-17.

[W80] Wahba, G. (1980) Automatic smoothing of the log periodogram. *J. Amer. Statist. Assn.* **75** 122-132.

Note: the articles of Donoho and Johnstone are available by "anonymous" FTP from `playfair.stanford.EDU`. After connecting, use commands `cd pub/donoho/reports`, then `ls` to get a list of files, followed by `get filename` to get copies of individual files. Files are available in LaTex, in DVI format, and in PostScript.