THRESHOLD SELECTION FOR WAVELET SHRINKAGE OF NOISY DATA*

David L. Donoho
and
Iain M. Johnstone
Department of Statistics
Stanford University
Stanford, CA 94305

ABSTRACT

Methods based on thresholding and shrinking empirical wavelet coefficients hold promise for recovering and/or denoising signals observed in noise. For this talk we review and compare various proposals for the choice of thresholds. These include soft and hard thresholding, and thresholds that are fixed in advance or chosen level by level from an empricial optimality criterion. We present results from simulations and real data examples.

Key Words: Wavelet Shrinkage, Thresholding, Noisy Data, Denoising

1 INTRODUCTION

For simplicity, we concentrate on the case of a regularly sampled signal in the presence of near white Gaussian noise. Extensions to other data formats are possible: see [1] and the references therein. Thus, suppose (after a rescaling of the time interval) that the data consist of $n = 2^{J+1}$ pairs $(i/n, y_i), 1 \le i \le n$, with

$$y_i = f(i/n) + z_i. \tag{1}$$

The aim is to recover the signal f. For the derivations of the methods below, the noise $\{z_i\}$ is assumed to be independent and identically distributed Gaussian with mean zero and (known) variance σ^2: in practice the methods tolerate at least small departures from these assumptions.

1. Apply an empirical wavelet transform yielding coarse level scaling coefficients $\{v_k, k = 1, \ldots, 2^L\}$ at a prechosen level L, and wavelet coefficients $\{w_{j,k}, \quad k = 1, \ldots, 2^j, j = L, \ldots, J\}$, for a total of n transform coefficients in all.

2. To each individual wavelet coefficient, apply a threshold rule, either *soft* (a continuous function of the data which shrinks each observation) or *hard* (which retains unchanged only large observations):

$$\eta_S(w,t) = \begin{cases} w - t & w \ge t \\ 0 & |w| < t, \\ w + t & w \le -t \end{cases}$$

$$\eta_H(w,t) = \begin{cases} w & |w| \ge t \\ 0 & |w| < t. \end{cases}$$

Form

$$\tilde{w}_{j,k} = \eta(w_{j,k}, \sigma n^{-1/2} t_j)$$

The scaling coefficients $\{v_k, k = 1, \ldots, 2^L\}$ are left unchanged in order that gross structure is not lost.

3. Invert the empirical wavelet transform on $\{v_k, \tilde{w}_{j,k}\}$ to obtain an estimated curve $\hat{y}_i = \hat{f}(i/n)$.

Estimated Scale The noise variance σ^2 may be estimated, for example, by $(m/.6745)^2$ where m is the median absolute deviation of $\{w_{J,k}, k = 1, \ldots, 2^J\}$, the wavelet coefficients at the finest level.

2 CHOICE OF THRESHOLDS

Single threshold. A) *VisuShrink:* $t_j = \sqrt{2 \log n}$.

In conjunction with soft thresholding η_S and Gaussian white noise, this threshold choice generally produces 'noise-free' reconstructions (cf. [2]), sometimes at cost of shrinkage of genuine features. Hard thresholding preserves features (e.g. peak heights) better, but can yield less smooth fits.

B) *RiskShrink:* $t_j = \lambda_n$ (with soft thresholding).

The constants λ_n are defined and tabulated in [3], where an analog for hard thresholding is also described. They are defined to optimize a criterion based on mean squared error (MSE)

$$\sum_{i=1}^{n} E[\hat{f}(i/n) - f(i/n)]^2$$

*This work supported in part by NSF Grant DMS92-09130 and NIH Grant GM 21215

and are always less than $\sqrt{2\log n}$. For example, λ_n increases from 1.86 at $n = 256$ to 2.59 at $n = 4096$. Reconstructions typically have smaller MSE than those of VisuShrink, but have visible fine scale roughness.

Level dependent thresholds. These offer greater flexibility, which may be of critical importance for certain data types (e.g. inverse problems, see [4]) in which the noise depends on resolution level.
C) *SureShrink* [6] is an 'automatic' procedure for the model (1) in which $\hat{t_j}$ is estimated from data $(w_{j,k}; k = 1, \ldots, 2^j)$ at level j. Let $\epsilon^2 = \sigma^2/n$.

a) Choose $\tilde{t}_j$ to minimize the unbiased estimate of MSE:

$$S_j(t) = 2^j\epsilon^2 + \sum_{k=1}^{2^j} \min(w_{j,k}^2, t^2\epsilon^2) - 2\epsilon^2 I\{w_{j,k}^2 \leq t^2\epsilon^2\}$$

over the range $0 \leq t \leq t_j^F = \sqrt{2\log 2^j}$.

b) Let $s_j^2 = 2^{-j}\sum_{k=1}^{2^j}(\epsilon^{-2}w_{j,k}^2 - 1)$ and set

$$\hat{t}_j = \begin{cases} \tilde{t}_j & s_j^2 \geq 3j2^{-(j+1)/2} \\ t_j^F & \text{otherwise.} \end{cases}$$

Step b) is a preliminary test for size of the signal coefficients at level j: if the signal is deemed too small, then no attempt to estimate t_j is made and the fixed threshold t_j is used instead.

3 EXAMPLE

Figure 1 shows an ESCA spectrum based on millions of counts aggregated into 1024 bins (data courtesy Jean-Paul Bibérian). To simulate a noisier "low statistics" situation, a simulated spectrum was constructed using panel (a) as a "true" intensity function, and drawing Poisson samples with maximum 4000 counts/bin. The simulated spectrum is transformed by the Anscombe variance-stabilizing transformation for Poisson data, $y_i = 2\sqrt{N_i + .375}$, and then processed exactly as if the data obeyed the white noise model. For visual clarity, the figures show a only the subset of the reconstructions from bins 400 to 900. In this case soft thresholding obscures the peak at about bin 620, perhaps due to excessive shrinkage of wavelet coefficients. One may alternatively use hard thresholding, or *SURE Shrink*) – these both retain the peak at about bin 620, although SureShrink has some residual high-frequency roughness, for example in bins 100 to 450.

References

[1] D.L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In Ingrid Daubechies, editor, *Different Perspectives on Wavelets*, pages 173–205. American Mathematical Society, Providence, R.I., 1993.

[2] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series B*, 1994. with discussion, to appear.

[3] D. L. Donoho and I. M Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 1994. to appear.

[4] D.L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Technical Report 403, Department of Statistics, Stanford University, 1992.

[5] D. L. Donoho and I. M Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 1995. to appear.

Figure Caption. (a) Maximum cell count is 50,011 (before normalization), but treated as true intensity spectrum $I_i, i = 1, \ldots, 1024$. (b) Simulated spectrum with $N_i \sim Poisson(4000 I_i/\max(I_i))$. (c-e) Wavelet transform calculated using Daubechies filters of order 3, combined with the corresponding edge filters of Cohen et. al. Thresholding of wavelets done using soft (d) and hard(e) thresholding at $\sqrt{2\log n}$ and using SUREShrink (c).