Statistica Sinica **7**(1997), 855-874

WAVESHRINK WITH FIRM SHRINKAGE

Hong-Ye Gao and Andrew G. Bruce

MathSoft, Inc.

Abstract: Donoho and Johnstone's (1994) WaveShrink procedure has proven valuable for signal de-noising and non-parametric regression. WaveShrink has very broad asymptotic near-optimality properties. In this paper, we introduce a new shrinkage scheme, *firm*, which generalizes the hard and soft shrinkage proposed by Donoho and Johnstone (1994). We derive minimax thresholds and provide formulas for computing the pointwise variance, bias, and risk for WaveShrink with firm shrinkage. We study the properties of the shrinkage functions, and demonstrate that firm shrinkage offers advantages over both hard shrinkage (uniformly smaller risk and less sensitivity to small perturbations in the data) and soft shrinkage (smaller bias and overall L_2 risk). Software is provided to reproduce all results in this paper.

Key words and phrases: Bias estimation, firm shrinkage, minimax thresholds, non-parametric regression, signal de-noising, trend estimation, variance estimation, wavelet transform, WaveShrink.

1. Introduction

Suppose we observe data $\mathbf{y} = (y_1, \ldots, y_n)'$ given by

$$y_i = f_i + \sigma z_i \qquad i = 1, \ldots, n, \tag{1}$$

where $\{z_i\}$ are i.i.d. $N(0,\ 1)$. The goal is to estimate $f = (f_1, \ldots, f_n)'$ with small mean-square-error, i.e. to find an estimate $\hat{f}$ with small L_2 risk:

$$R(\hat{f},\ f) = \frac{1}{n}\sum_{i=1}^{n} E(\hat{f}_i - f_i)^2. \tag{2}$$

Donoho and Jonhstone (1994) have developed a powerful methodology called "WaveShrink" for estimating f; see also Donoho et al. (1995). WaveShrink has very broad asymptotic near-optimality properties. For example, WaveShrink achieves, within a factor of $\log n$, the optimal minimax risk over each functional class in a variety of smoothness classes and with respect to a variety of losses, including L_2 risk: (see Donoho et al. (1995)). WaveShrink is now well established as a technique for removing noise from signals and images.

WaveShrink is based on the principle of shrinking wavelet coefficients towards zero to remove noise. Let $\mathbf{w} = (w_1, \ldots, w_n)'$ be the empirical wavelet coefficients

and let W be the wavelet transform matrix. The WaveShrink estimate is obtained by the following 3 steps:

(i) Compute wavelet transform $\mathbf{w} = W\mathbf{y}$.

(ii) Apply a non-linear shrinkage rule $\delta_{\lambda_k}(\cdot)$ to the wavelet coefficients to obtain

$$\hat{w}_k = \delta_{\lambda_k \hat{\sigma}}(w_k), \tag{3}$$

where $\hat{\sigma}$ is an estimate of the noise standard deviation σ.

(iii) Invert the wavelet transform $\hat{\mathbf{f}}_\lambda = W^{-1}\hat{\mathbf{w}}$.

Donoho and Johnstone (1994) studied the behavior of WaveShrink using the *hard* and the *soft* shrinkage functions:

$$\delta_\lambda^H(x) = xI_{[|x|>\lambda]}$$
$$\delta_\lambda^S(x) = \operatorname{sgn}(x)(|x| - \lambda)_+,$$

where $\lambda \in [0, \infty)$. Bruce and Gao (1996) showed that hard shrink tends to have bigger variance (because of the discontinuity of the shrink function) and the soft shrink tends to have bigger bias (because of shrinking all big coefficients towards zero by λ). To remedy the drawbacks of hard shrink and soft shrink, we introduce a general *firm* (formerly called *semisoft*, the name firm was suggested by Professor David Donoho) shrinkage function $\delta_{\lambda_1,\lambda_2}$:

$$\delta_{\lambda_1,\lambda_2}(x) = \begin{cases} 0, & \text{if } |x| \le \lambda_1, \\ \operatorname{sgn}(x)\frac{\lambda_2(|x|-\lambda_1)}{\lambda_2-\lambda_1}, & \text{if } \lambda_1 < |x| \le \lambda_2, \\ x, & \text{if } |x| > \lambda_2. \end{cases} \tag{4}$$

Figure 1 displays the hard, soft and firm shrinkage functions. For values of x near the lower threshold λ_1, $\delta_{\lambda_1,\lambda_2}(x)$ behaves like $\delta_{\lambda_1}^S(x)$. For values of x above the upper threshold λ_2, $\delta_{\lambda_1,\lambda_2}(x) = \delta_{\lambda_2}^H(x) = x$. Note that hard shrinkage, with $\lambda_1 = \lambda_2$, and soft shrinkage, with $\lambda_2 = \infty$, are limiting cases of (4). A special case of firm shrinkage, with $\lambda_2 = 2\lambda_1$, was successfully used in spectral density estimation by Walden et al. (1995).

There is a parallel between the choice of shrinkage functions for WaveShrink and the choice of influence functions in robust statistics. The function $x - \delta_\lambda^H(x)$ is the influence function (ψ) for a trimmed mean estimator; $x - \delta_\lambda^S(x)$ is the influence function for a Huber estimator; and $x - \delta_{\lambda_1,\lambda_2}(x)$ is the influence function for a Hampel estimator (Hampel et al. (1986)).

The main objectives of this paper are to show the advantages of the firm shrinkage function over hard and soft shrinkage, and to further develop the use of WaveShrink with the firm shrinkage function. Towards this end, we extend the work of Donoho and Johnstone (1994) to derive minimax thresholds for firm

shrinkage. We also extend the work of Bruce and Gao (1996) to derive exact formulae for the finite sample bias and variance for WaveShrink under firm shrinkage. Finally, we illustrate some of the properties which make firm shrinkage advantageous over both hard and soft shrinkage. Following the principle of reproducible research as advocated by Buckheit and Donoho (1995), the software which produces all figures and tables in this paper are available by anonymous ftp.

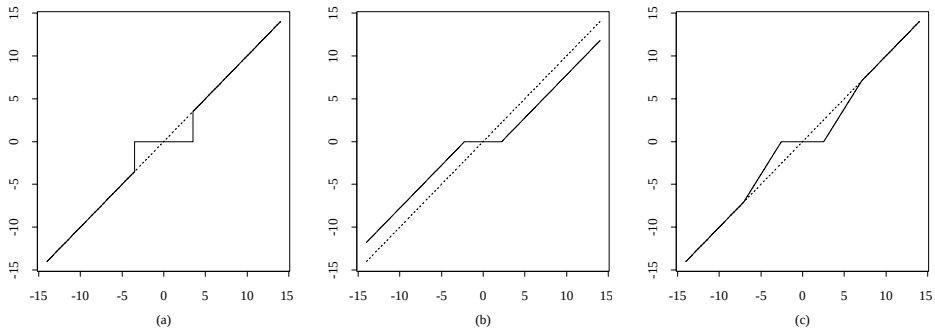

Figure 1. From left to right: hard, soft and firm shrinkage function (solid lines). Like hard shrinkage, firm is the identity function (dashed lines) for large enough $|x|$. Like soft shrinkage, firm shrinkage is continuous.

The paper is organized as follows. In Section 2, we give the formula for computing the L_2 risk for firm shrinkage of a Gaussian random variable. In Section 3, we use this formula to compute minimax thresholds for firm shrinkage. Finite sample bias and variance formulae are given in Section 4. Some advantages of firm shrinkage over hard and soft shrinkage are discussed in Section 5. Section 6 gives a summary and discusses on-going work. Proofs and technical details are given in Section 7. Computational and software details are given in Section 8.

2. L_2 Risk for Firm Shrinkage

In this Section, we study the L_2 risk of the firm shrinkage estimate of a Gaussian random variable $X \sim N(\theta,\ 1)$. The L_2 risk of the shrinkage function directly relates to the overall L_2 risk of the WaveShrink estimate, and is fundamental to the computation of minimax estimators: see Remark 1 and Section 3 below. The L_2 risk for firm shrinkage is defined as

$$R_{\lambda_1,\lambda_2}(\theta) = E\left\{\delta_{\lambda_1,\lambda_2}(X) - \theta\right\}^2. \tag{5}$$

The following theorem gives an analytical expression for $R_{\lambda_1,\lambda_2}(\theta)$.

Theorem 1. *For* $X \sim N(\theta,\ 1)$,

$$\begin{aligned} R_{\lambda_1,\lambda_2}(\theta) = \theta^2 \left\{\Phi(\lambda_1 - \theta) - \Phi(-\lambda_1 - \theta)\right\} + 1 - \Phi(\lambda_2 - \theta) + \Phi(-\lambda_2 - \theta) \\ + \Psi_0(\theta, \lambda_1, \lambda_2) + \Psi_0(-\theta, \lambda_1, \lambda_2) \\ + (\lambda_2 - \theta)\phi(\lambda_2 - \theta) + (\lambda_2 + \theta)\phi(\lambda_2 + \theta), \end{aligned}$$

where $r_1 = \lambda_1/(\lambda_2 - \lambda_1)$, $r_2 = \lambda_2/(\lambda_2 - \lambda_1)$, Φ *is the probability distribution function,* ϕ *is the probability density function for standard Gaussian random variable* $\Psi_0(\theta,\ \lambda_1,\ \lambda_2) = \Psi(\lambda_1 - \theta,\ \lambda_2 - \theta,\ r_2,\ r_1(\theta - \lambda_2))$ *and*

$$\begin{aligned} \Psi(a, b, c, d) &\equiv \int_a^b (cx + d)^2 \phi(x) dx \\ &= (c^2 + d^2)\Big(\Phi(b) - \Phi(a)\Big) + c\phi(a)(ac + 2d) - c\phi(b)(bc + 2d). \end{aligned}$$

The proof of the theorem follows by straightforward calculus. Formulas decomposing the risk into squared bias and variance components are given in Section 7.

Figure 2 displays the L_2 risk of the firm shrinkage estimates for different choices of (λ_1, λ_2). Figure 2 also displays the L_2 risk split into squared bias and variance. From Figure 2, it is evident that firm shrinkage offers a continuum with soft shrinkage at one extreme (high bias, low variance) and hard shrinkage at the other extreme (low bias, high variance).

Remark 1. For orthogonal wavelets, $W^{-1} = W'$. Therefore,

$$R(\hat{f}_\lambda,\ f) = \frac{1}{n} \sum_{k=1}^{n} E\left(\hat{w}_k - \theta_k\right)^2, \tag{6}$$

where $\theta = W\mathbf{f}$ are the true wavelet coefficients. If σ^2 is known, then the global risk for the WaveShrink estimate is

$$R(\hat{f}_\lambda,\ f) = \frac{\sigma^2}{n} \sum_{k=1}^{n} R_{\lambda_1,\lambda_2}\left(\frac{\theta_k}{\sigma}\right). \tag{7}$$

Hence, the form of $R_{\lambda_1,\lambda_2}(\theta)$ directly relates to the L_2 risk for the WaveShrink estimate.

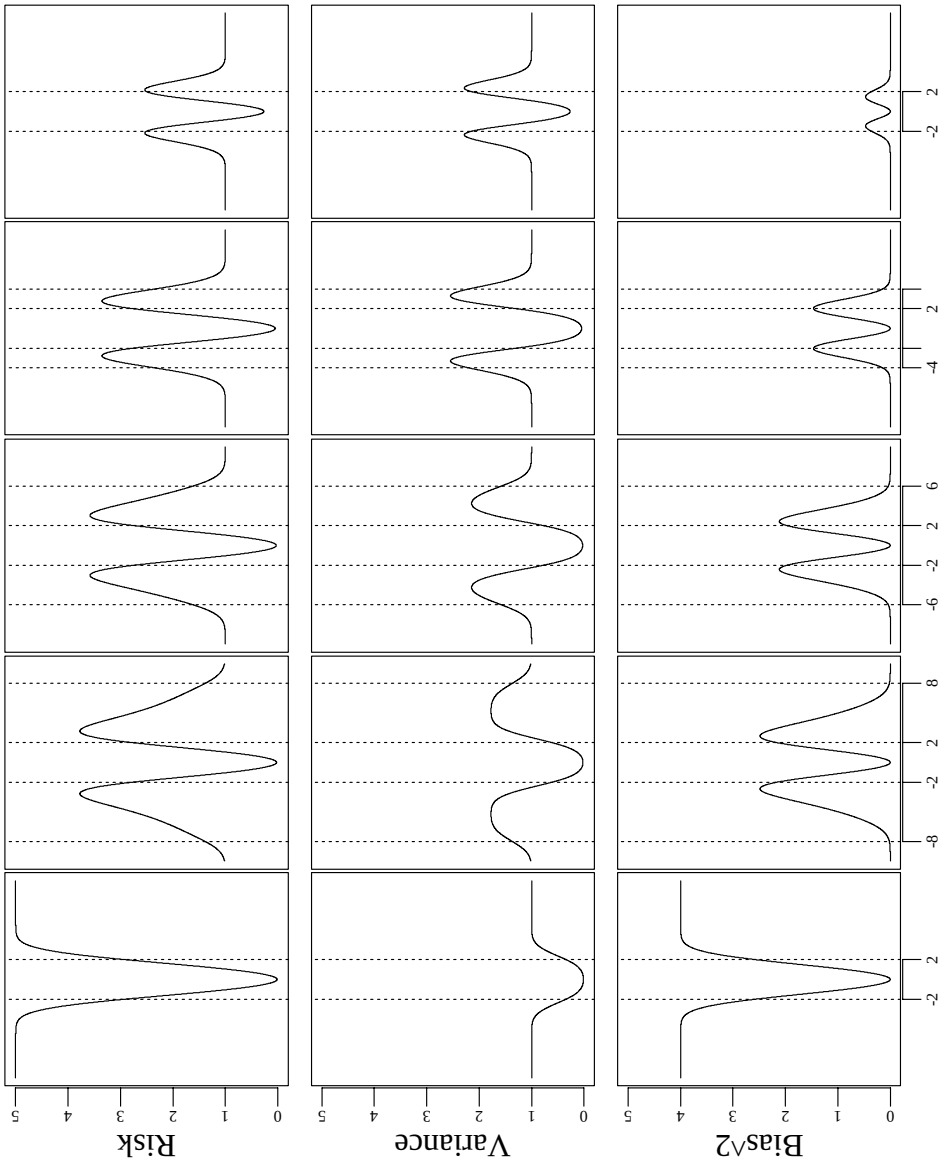

Figure 2. First row: L_2 risks vs. θ for firm shrinkage; middle row: variances vs. θ for firm shrinkage; bottom row: squared biases vs. θ for firm shrinkage. Dotted vertical lines indicate the thresholds. Thresholds are (from left to right): $\lambda_1 \equiv 2$ and $\lambda_2 = \infty$, 8, 6, 4, 2.

Remark 2. Visually, in Figure 2, it may appear that hard shrinkage has the smallest overall risk. Note, however, that hard shrinkage has much higher risk at the origin. Since the wavelet coefficients are typically near zero, the risk of

the WaveShrink estimate is often higher. For further elaboration, see Bruce and Gao (1996) and Section 4.

Remark 3. We can exploit the extra flexibility of firm shrinkage to achieve WaveShrink estimates with smaller L_2 risk. In fact, firm shrinkage can be made to have uniformly smaller L_2 risk than hard shrinkage. For each threshold λ, we can find many pairs of thresholds $\lambda_1 < \lambda < \lambda_2$ such that

$$R_{\lambda_1,\lambda_2}(\theta) < R^H_\lambda(\theta) \qquad \text{for all } \theta. \tag{8}$$

Figure 3 displays the region for the values (λ_1, λ_2) such that (8) holds when $\lambda = 3.33$.

A similar result does not hold for soft shrinkage. This is because soft shrinkage dominates at the origin.

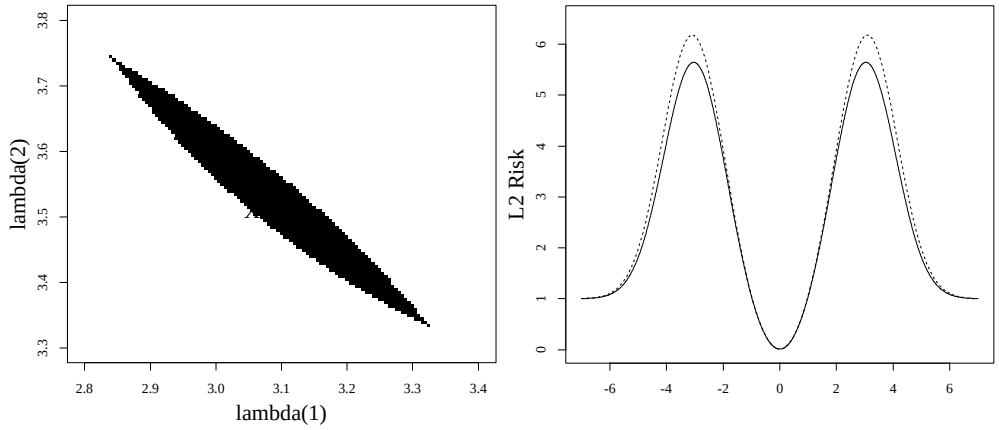

Figure 3. Left: the pairs of thresholds (λ_1, λ_2) for which firm shrinkage has uniformly smaller risk than hard shrinkage for $\lambda = 3.33$. Right: the risk curves under firm shrinkage with $\lambda_1 = 3.055$ and $\lambda_2 = 3.51$ (solid line) and under hard shrinkage with $\lambda = 3.33$ (dotted line). This pair of thresholds gives the greatest minimum risk improvement over hard shrinkage over the grid of thresholds searched.

3. Threshold Selection for Firm Shrinkage

One of the attractive aspects of the original Donoho and Johnstone (1994)-hereafter referred to as DJ94 — WaveShrink scheme is its computational and conceptual simplicity. Using a simple, data independent, threshold, one can achieve nearly optimal estimates in a minimax sense for a broad class of functionals in a variety of smoothness classes. In this section, we extend these results to firm shrinkage by computing several types of minimax thresholds. These thresholds

achieve a tighter upper bound on the L_2 minimax risk than both hard and soft shrinkage.

3.1. Minimax thresholds

Following DJ94, define the minimax quantity

$$\Lambda_n^{**} \equiv \inf_{\lambda_1 \le \lambda_2} \sup_{\theta} \Big\{ \frac{R_{\lambda_1,\lambda_2}(\theta)}{n^{-1} + \min(\theta^2, 1)} \Big\}. \tag{9}$$

The term $\min(\theta^2, 1)$ in the denominator is the ideal risk under the diagonal linear projection oracle: an oracle who tells you whether to "keep" or "kill" each coefficient, §2.1 of DJ94. Hence, from (7), we can bound the L_2 risk by

$$R(\hat{f}_{\lambda^*},\ f) \le \frac{\Lambda_n^{**}}{n} \Big\{ \sigma^2 + \sum_{k=1}^{n} \min(\theta_k^2, \sigma^2) \Big\}. \tag{10}$$

DJ94 show that $\Lambda_n^{**} \sim 2 \log n$ for hard and soft shrinkage, therefore $\hat{f}_{\lambda^*}$ is within a factor of $\log n$ of the optimal bound. Since firm shrinkage is guaranteed to have smaller minimax risk than soft and hard shrinkage, the nearly optimal asymptotic results hold for firm shrinkage as well.

The minimax thresholds $(\lambda_1^*, \lambda_2^*)$ can be derived by choosing $(\lambda_1^*, \lambda_2^*)$ to attain (9): (see Section 8 for computational details). Table 1 gives the minimax thresholds for selected values of n. From Table 1, we can see that the soft minimax quantities are improved by roughly 6% –19% and the hard minimax quantities are improved by roughly 19%-39%. As n increases, $\lambda_{n,1}^*$ also increases while $\lambda_{n,2}^*$ decreases. In other words, firm shrinkage moves from emulating soft shrinkage for small n to emulating hard shrinkage for large n. This corresponds to the findings of Bruce and Gao (1996), who determined that ideal soft shrinkage has lower risk for small n but ideal hard shrinkage has lower risk for large n.

Table 1. Minimax thresholds and minimax risk bounds for the firm, soft and hard shrinkages.

	firm			soft		hard	
n	$\lambda_{n,1}^*$	$\lambda_{n,2}^*$	Λ_n^{**}	λ_n^*	Λ_n^*	λ_n^*	Λ_n^*
64	1.678	8.004	2.931	1.474	3.124	2.697	4.078
128	1.893	7.980	3.464	1.669	3.755	2.913	4.735
256	2.116	7.549	4.033	1.859	4.439	3.117	5.409
512	2.331	7.259	4.634	2.045	5.172	3.312	6.098
1024	2.538	7.069	5.264	2.226	5.950	3.497	6.805
2048	2.737	6.939	5.921	2.403	6.771	3.674	7.529
4096	2.930	6.848	6.603	2.575	7.629	3.844	8.268
8192	3.116	6.799	7.307	2.743	8.522	4.008	9.022
16384	3.296	6.760	8.032	2.906	9.447	4.166	9.791
32768	3.471	6.748	8.776	3.066	10.399	4.319	10.572
65536	3.640	6.748	9.537	3.221	11.376	4.467	11.367

Define the supremum risk by

$$\Lambda_n(\lambda_1, \lambda_2) \equiv \sup_{\theta} \Big\{ \frac{R_{\lambda_1,\lambda_2}(\theta)}{n^{-1} + \min(\theta^2, 1)} \Big\}. \tag{11}$$

In Figure 4, we plot $\Lambda(\lambda_1, \lambda_2)$ vs. (λ_1, λ_2) for $n = 256$. The nearly minimax risk can be achieved with a broad range of threshold pairs (λ_1, λ_2). The supremum risk function $\Lambda(\lambda_1, \lambda_2)$ is particularly flat relative to the upper threshold λ_2.

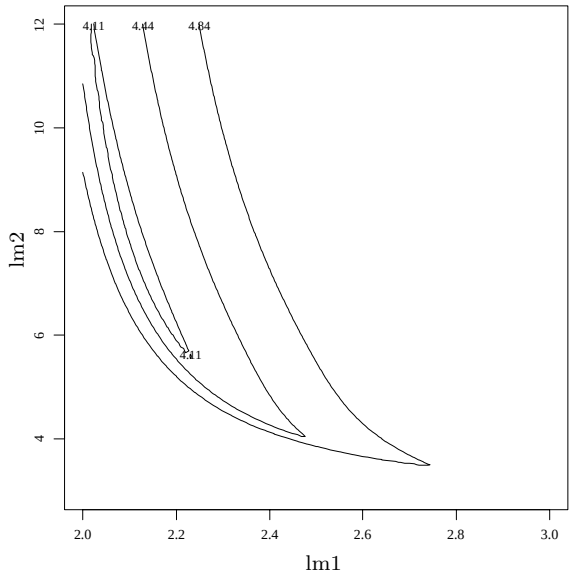

Figure 4. The supremum risk $\Lambda_n(\lambda_1, \lambda_2)$ defined by (11) is plotted over (λ_1, λ_2) with contour lines: $1.02\Lambda_n^{**}$, $1.10\Lambda_n^{**}$ and $1.20\Lambda_n^{**}$, where Λ_n^{**} is defined in (9).

3.2. Fix the upper threshold

The simplest rule for setting the WaveShrink threshold is the universal threshold $\sqrt{2\log n}$, referred to as *VisuShrink* in DJ94. The universal threshold ensures that

$$P\left\{ \max_{1\le i\le n} |X_i| > \sqrt{2\log n} \right\} \to 0, \qquad \text{as } n \to \infty,$$

where $X_1, \ldots, X_n$ are i.i.d. $N(0,\ 1)$ random variables. In fact, it can be shown that

$$P\left\{ \max_{1\le i\le n} |X_i| > \sqrt{2\log n} \right\} \sim \frac{1}{\sqrt{\pi \log n}}, \qquad \text{as } n \to \infty. \tag{12}$$

In general,

$$P\left\{ \max_{1\le i\le n} |X_i| > \sqrt{c\log n} \right\} \sim \frac{\sqrt{2}}{n^{c/2-1}\sqrt{c\pi \log n}}, \qquad \text{as } n \to \infty. \tag{13}$$

We can define other universal thresholds such that (13) converges to zero faster (i.e., to suppress noise more thoroughly). For example, threshold $\sqrt{4\log n}$ makes the above convergence rate $1/n\sqrt{2\pi\log n}$ and threshold $\sqrt{6\log n}$ makes the above convergence rate $1/n^2\sqrt{3\pi\log n}$.

Table 2. Minimax lower thresholds $\lambda^*_{n,1}$ for the upper thresholds set to the universal thresholds $\lambda^{(j)}_{n,2} = \sqrt{2j\log n}$ for $j = 1, 2, 3$. The minimax risk bound Λ^*_n is also given.

n	$\lambda^*_{n,1}$	$\lambda^{(1)}_{n,2}$	Λ^*_n	$\lambda^*_{n,1}$	$\lambda^{(2)}_{n,2}$	Λ^*_n	$\lambda^*_{n,1}$	$\lambda^{(3)}_{n,2}$	Λ^*_n
64	2.471	2.884	3.805	1.940	4.079	3.194	1.823	4.995	3.047
128	2.669	3.115	4.393	2.137	4.405	3.694	2.022	5.396	3.544
256	2.860	3.330	4.999	2.327	4.710	4.223	2.213	5.768	4.082
512	3.045	3.532	5.626	2.511	4.995	4.785	2.398	6.118	4.658
1024	3.223	3.723	6.272	2.689	5.266	5.379	2.576	6.449	5.272
2048	3.394	3.905	6.938	2.861	5.523	6.004	2.749	6.764	5.922
4096	3.560	4.079	7.622	3.027	5.768	6.658	2.916	7.064	6.604
8192	3.721	4.245	8.324	3.189	6.004	7.340	3.079	7.353	7.317
16384	3.876	4.405	9.043	3.346	6.230	8.049	3.237	7.630	8.058
32768	4.028	4.560	9.778	3.499	6.449	8.782	3.391	7.898	8.825
65536	4.175	4.710	10.528	3.648	6.660	9.538	3.541	8.157	9.616

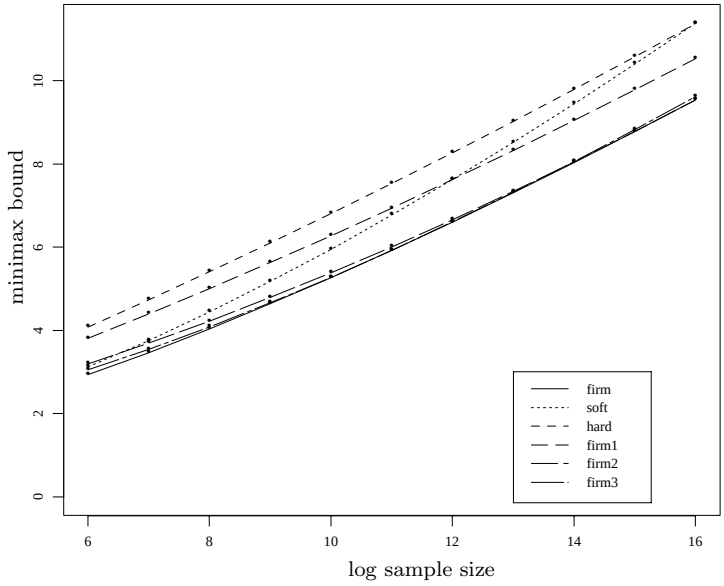

Figure 5. The minimax bounds Λ^{**}_n defined in (9) vs. $\log_2(n)$ for soft, hard, firm (all listed in Table 1) and firm with fixed upper thresholds (listed in Table 2 and labelled as "firm1", "firm2" and "firm3" in the legend).

These universal thresholds give, in some sense, an upper bound on the thresholds needed to suppress noise. Table 2 gives an alternative set of minimax thresholds for firm shrinkage with the upper threshold $\lambda_{n,2}$ set to one of the universal thresholds. Table 2 lists the minimax thresholds $\lambda^*_{n,1}$ and the minimax risk bounds Λ^*_n for $\lambda^{(1)}_{n,2} = \sqrt{2\log n}$, $\lambda^{(2)}_{n,2} = \sqrt{4\log n}$, $\lambda^{(3)}_{n,2} = \sqrt{6\log n}$ and some selected n. Figure 5 plots the minimax bounds for fixed upper thresholds, along with the minimax bounds for soft, hard and firm derived from previous section. From Table 2 and Figure 5, we can see that the minimax bounds with fixed upper thresholds are tighter than those for the hard shrinkage for all sample sizes, and tighter than those for the soft shrinkage for large sample sizes ($n \geq 4096$).

3.3. Fix the lower threshold

In some cases, it may be desirable to fix the lower threshold and optimize over the upper threshold. For example, a small threshold may be desired to preserve locally non-smooth features, such as jumps (to avoid overshrinking the coefficients associated with the jumps). Section 8 describes available software for computing the minimax upper threshold for a given lower threshold.

3.4. Use different oracle

Instead of the "keep" or "kill" oracle, we can also consider a "keep", "shrink" or "kill" (KSK) oracle. Let $X \sim N(\theta,\ \sigma^2)$, choose constants a, b, c such that $0 < a < 1 < b,\ 0 < c < 1$, and define

$$\hat{\theta} = \begin{cases} 0 & |\theta| \leq a\sigma & \text{kill,} \\ cX & a\sigma < |\theta| \leq b\sigma & \text{shrink,} \\ X & |\theta| > b\sigma & \text{keep.} \end{cases}$$

Then the ideal risk is

$$R(\theta) \equiv E\left\{(\hat{\theta} - \theta)^2\right\} = \begin{cases} \theta^2 & |\theta| \leq a\sigma, \\ c^2\sigma^2 + (c-1)^2\theta^2 & a\sigma < |\theta| \leq b\sigma, \\ \sigma^2 & |\theta| > b\sigma. \end{cases}$$

To make $R(\theta)$ continuous, choose

$$b = \sqrt{\frac{1+3a^2}{1-a^2}} \qquad c = \frac{2a^2}{1+a^2}.$$

In particular, for $a = 1/\sqrt{3}$, $b = \sqrt{3}$ and $c = 1/2$.

Software and tables for computing $(\lambda^*_1, \lambda^*_2)$ are available: see Section 8. The minimax thresholds using the KSK oracle are very close to those in Table 1.

4. Variance and Bias in Firm WaveShrink

Not only does firm shrinkage produce small minimax risk estimates, but it also can reduce the risk in many finite sample situations. In this section, we compute the exact pointwise bias and variance for WaveShrink using firm shrinkage, and show that firm can lead to substantially lower risk than both hard and soft shrinkage.

We can get access to the pointwise variance and bias of WaveShrink estimate using the expression in WaveShrink algorithm step (iii). Let

$$W^{-1} = \begin{pmatrix} \tilde{c}_{11} \cdots \tilde{c}_{n1} \\ \cdots \\ \tilde{c}_{1n} \cdots \tilde{c}_{nn} \end{pmatrix}$$

be the corresponding matrix for inverse wavelet transform. Then under model (1), $\mathbf{w} \sim N\left(\theta,\ \sigma^{\mathbf{2}}\mathbf{W}\mathbf{W}'\right)$ and the WaveShrink estimate can be written as

$$\hat{f}_i = \sum_k \tilde{c}_{ki}\hat{w}_k \qquad i = 1, \ldots, n, \tag{14}$$

where $\hat{w}_k$ are defined in (3).

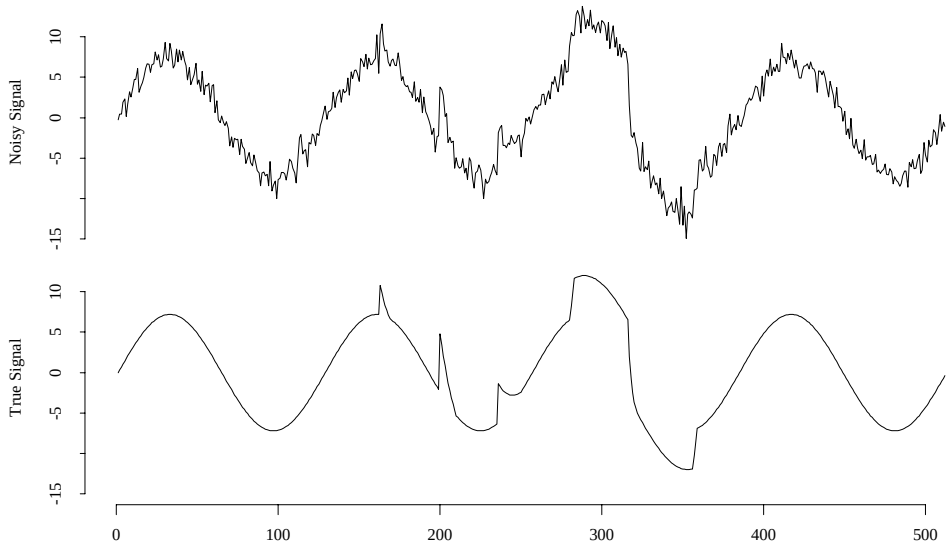

Figure 6. A synthetic "jumps" signal used to compare the variance, bias and risk of the various shrinkage estimates. The top panel displays the noisy signal and the bottom panel displays the original noise free signal. The noise is additive white Gaussian with the variance set so that the signal to noise ratio is 6.

Theorem 2. *If σ^2 is known, then for $i = 1, \ldots, n$,*

$$E(\hat{f}_i) = \sigma \sum_k \tilde{c}_{ki}\ M_\lambda\Big(\frac{\theta_k}{\sigma}\Big) \tag{15}$$

$$\mathrm{Var}\,(\hat{f}_i) = \sigma^2 \sum_{k,\ell} \tilde{c}_{ki}\tilde{c}_{\ell i}\ C_\lambda\Big(\frac{\theta_k}{\sigma}, \frac{\theta_\ell}{\sigma};\ \rho_{k,\ell}\Big), \tag{16}$$

where $M_\lambda(\cdot)$ and $C_\lambda(\cdot)$ are the mean and covariance functions of the firm shrinkage function defined in the Section 7 and $\rho_{k,\ell} = corr(w_k, w_\ell)$.

The proof follows from Bruce and Gao (1996), which gives a similar formula for hard and soft shrinkage.

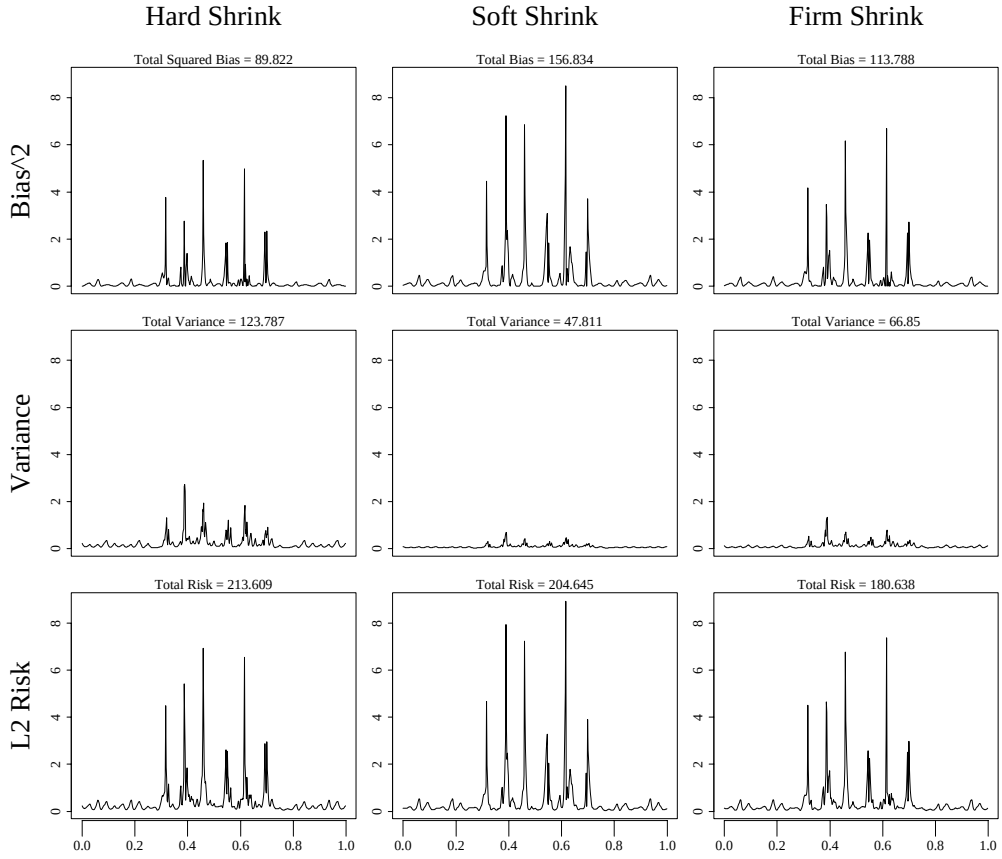

Figure 7. Bias, standard error, and L_2 risk for the signal of Figure 6. First row, from left to right: bias for hard, soft, and firm shrinkage. Second row, from left to right: standard error for hard, soft, and firm shrinkage. Third row, from left to right: L_2 risk for hard, soft, and firm shrinkage. The overall squared bias, variance and L_2 risk is given above the plots. Hard shrinkage has the smallest bias, soft shrinkage has the smallest variance, and firm shrinkage has smallest L_2 risk.

We use the results of Theorem 2 to present an example of how firm can produce estimates with lower risk. Figure 6 shows the synthetic "jumps" signal, which is a sinusoid punctuated by a variety of jumps and ramps. The signal contains additive Gaussian white noise and has a signal-to-noise ratio (SNR) of 6. The SNR is defined by $\text{SNR} = \dfrac{\text{SD(signal)}}{\text{SD(noise)}}$. In Figure 7, we plot the bias, standard error, and L_2 risk of WaveShrink using hard, soft, and firm shrinkage for the signal displayed in the top panel of Figure 7. Note that these values depend only on the underlying true signal and noise variance and not on the actual realization. The minimax thresholds are used for all three shrinkage functions. In this example, hard shrinkage has the smallest bias, soft shrinkage has the smallest variance, and firm shrinkage has smallest L_2 risk. This example is quite typical, and reflects very general behavior of hard, soft, and firm shrinkage. See also Bruce and Gao (1996).

Remark 1. Under certain conditions, Brillinger (1995) shows that, for each i, $\hat{f}_i$ is asymptotically Gaussian distributed. Hence, approximate confidence intervals can be constructed from the estimated pointwise standard errors

$$s_i = \hat{\sigma}\Big\{ \sum_{k,\ell} \tilde{c}_{ki}\tilde{c}_{\ell i}\, C_\lambda\Big(\frac{\hat{\theta}_k}{\hat{\sigma}}, \frac{\hat{\theta}_\ell}{\hat{\sigma}};\ \rho_{k,\ell}\Big)\Big\}^{1/2}. \tag{17}$$

Remark 2. The summations in (15)-(16) involve only P non-zero terms c_{ki} where P depends on the wavelet filter length and maximum resolution level J. In practice, if we shrink only a few levels, then the effort required to compute the variance and bias is not much greater than for the WaveShrink estimate itself. (See Section 8 for information on how to obtain an efficient software implementation of the above formulas.)

Remark 3. In the above example, the calculations are done with *periodized* wavelets. The formulas, however, are valid for more general boundary treatment methods (e.g., the "interval" wavelets of Cohen et al. (1993)).

Remark 4. For orthogonal wavelets, the variance formula in Theorem 2 simplifies considerably since $\rho_{k,\ell} = 0$ for all $k \neq \ell$. Note also that $\tilde{c}_{ij} = c_{ij}$.

5. Local Shift Sensitivity and Bias Reduction

Hard shrinkage is sensitive to small fluctuations in the data. This sensitivity is a result of the discontinuity in the shrinkage function at $x = \pm\lambda$. When values of wavelet coefficients are changed slightly (e.g. from barely below λ to barely above λ, or vice versa), this has a measurable effect on the WaveShrink estimate.

Following §2.1c of Hampel et al. (1986), define *local-shift sensitivity* by

$$\alpha_\lambda = \sup_{x \neq y} \Big| \frac{\delta_\lambda(y) - \delta_\lambda(x)}{y - x} \Big|. \tag{18}$$

It is easy to see that $\alpha_\lambda^H = \infty$, $\alpha_\lambda^S = 1$ and $\alpha_{\lambda_1,\lambda_2} = \lambda_2/(\lambda_2 - \lambda_1)$. Soft shrinkage and firm shrinkage are less sensitive to small perturbations in the data than hard shrinkage.

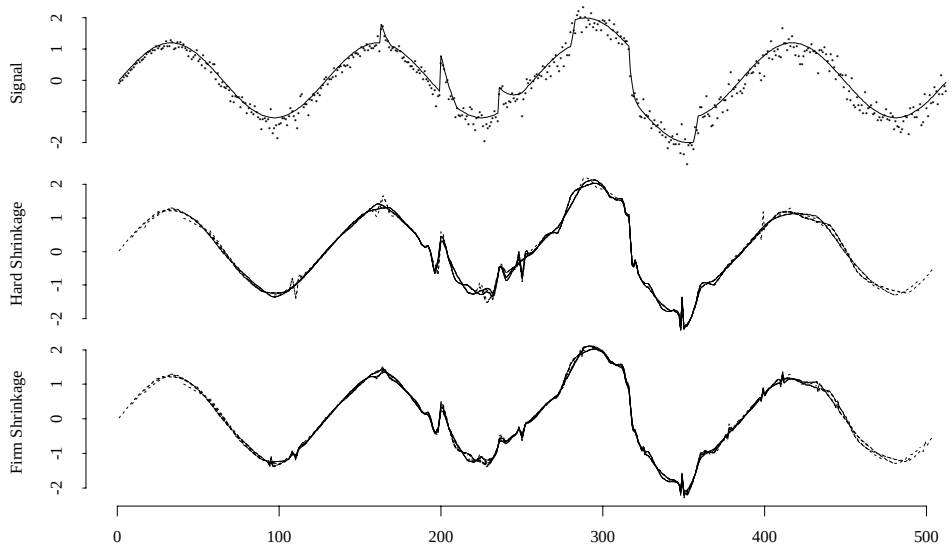

Figure 8. Hard shrinkage is more sensitive to small perturbations in the data than soft or firm shrinkage. The top panel displays a signal of length 512 with additive Gaussian white noise (signal to noise ratio of 5). The middle panel displays the estimates of the signal obtained using the WaveShrink procedure with hard shrinkage applied to 16 overlapping windows of length 384. The windows are obtained by successively dropping 8 samples from the beginning of the signal and adding 8 samples to the end of the signal. The bottom panel displays the sliding window WaveShrink estimates for firm shrinkage. Minimax thresholds are used for both hard and firm shrinkage and the scale of the noise is re-estimated for each window.

Figure 8 illustrates some of the problems associated with the local shift sensitivity of hard shrinkage. The top panel displays a signal of length 512 with additive Gaussian white noise. This signal is similar to that of Figures 6 and 7 except that the variance of the noise changes. The signal-to-noise ratio is 5. The middle panel displays the estimates of the signal obtained using the WaveShrink procedure with hard shrinkage applied to 16 overlapping windows of length 384.

The windows are obtained by successively dropping 8 samples from the beginning of the signal and adding 8 samples to the end of the signal. The bottom panel displays the 16 windowed WaveShrink estimates for firm shrinkage. Minimax thresholds are used for both hard and firm shrinkage and the scale of the noise is re-estimated for each window. The hard shrinkage estimates are visibly more volatile, particularly near the discontinuities.

In this example, the data hasn't been perturbed — only the estimated scale factor $\hat{\sigma}$ changes for each window. If we perturb the data as well, then we can expect hard shrinkage to exhibit even greater volatility.

Soft shrinkage almost always introduces bias. To measure this, define

$$\beta_\lambda = \inf\{r > 0;\ \delta_\lambda(x) = x,\ \text{when } |x| > r\}. \tag{19}$$

The shrinkage function does not generate bias for x larger than β_λ. It is easy to see that $\beta^S_\lambda = \infty$, $\beta^H_\lambda = \lambda$ and $\beta_{\lambda_1,\lambda_2} = \lambda_2$. Note that (19) is similar to the idea of *rejection point* in robust statistics literature: (see Hampel et al. (1986)).

The following example illustrates the bias sensitivity of soft shrinkage. Let $f_v(x) = v\ I_{[0\le x<0.5]}$, i.e. $f_v(x)$ is a step function with a jump of height v. The bigger the v, the bigger the signal-to-noise ratio, and the easier to estimate f_v. Table 3 shows that no matter how big v is, the soft shrinkage estimate is always biased. By contrast, the biases for both hard shrinkage and firm shrinkage (with the Haar wavelet) go to zero as $v \to \infty$.

Table 3. Bias comparison of soft, hard, and firm shrinkage with the Haar wavelet for $f_v(x) = vI_{[0\le x\le 0.5]}$, sample size $n = 1024$, and noise variance $\sigma^2 = 1$. Minimax thresholds are used for all shrinkage estimates. The first three columns compare the maximum bias $\max\{|E\hat{f}_i - f_i|\}$ and the last three columns compare the average squared bias $\sum_{i=1}^n (E\hat{f}_i - f_i)^2$.

	$\max\{\lvert E\hat{f}_i - f_i\rvert\}$			$\sum_{i=1}^n\{\lvert E\hat{f}_i - f_i\rvert^2\}$		
v	soft	hard	firm	soft	hard	firm
1	0.9502	0.9730	0.9597	0.9172	0.9617	0.9357
2	1.8425	1.9004	1.8637	3.4505	3.6699	3.5303
5	3.5824	2.9574	3.3368	13.4064	10.3365	12.0012
10	4.3657	1.4655	2.2923	22.2400	9.2453	12.2984
50	4.7023	0.0012	0.0670	29.7304	0.0001	0.2745
500	4.7023	0.0000	0.0000	29.7305	0.0000	0.0000
∞	4.7023	0.0000	0.0000	29.7305	0.0000	0.0000

6. Discussion

In this paper, we introduce the firm shrinkage function for use with the WaveShrink procedure. Firm shrinkage gives uniformly lower risk than hard

shrinkage, a lower minimax risk bound than both hard and soft shrinkage, and lower risk in many finite sample situations. In addition, firm shrinkage overcomes the sensitivity of hard shrinkage to small perturbations in the data and avoids the bias artifacts of soft shrinkage. We derive formulas and tables for L_2 risk, minimax thresholds, and pointwise variance and bias. Software is provided to reproduce all results in this paper.

The firm shrink function discussed in this paper is a piece-wise linear function and is not differentiable. An alternative is to replace the linear function in the interval $[\lambda_1, \lambda_2]$ with a higher order polynomial. This will enable one to get a differentiable shrink function, ensuring that the shrunken random variable has density when the original variable has one. For example, one could define

$$\delta_{\lambda_1,\lambda_2}(x) = \begin{cases} 0, & \text{if } |x| \le \lambda_1, \\ \operatorname{sgn}(x)(r_2 - r_1|x|)(|x| - \lambda_1)^2, & \text{if } \lambda_1 < |x| \le \lambda_2, \\ x, & \text{if } |x| > \lambda_2, \end{cases}$$

where $r_1 = (\lambda_1 + \lambda_2)/(\lambda_2 - \lambda_1)^3$ and $r_2 = 2\lambda_2^2/(\lambda_2 - \lambda_1)^3$.

7. Technical Details

Bruce and Gao (1996) give formulas for the mean, variance and covariance function of the hard and soft shrink functions. In this section, we give parallel results for the firm shrink function.

7.1. Mean and variance of shrinkage estimate

The relationships between the shrinkage functions:

$$\delta_{\lambda_1,\lambda_2}(X) = \delta^H_{\lambda_2}(X) + \frac{\lambda_2}{\lambda_2 - \lambda_1}\delta^S_{\lambda_1}(XI_{[|X|\le\lambda_2]}).$$

Notice also $\hat{w}_k = \delta_{\lambda\hat{\sigma}}(w_k) = \hat{\sigma}\delta_\lambda(w_k/\hat{\sigma}) = \sigma\delta_{\lambda\hat{\sigma}/\sigma}(w_k/\sigma)$.

Lemma 1. (*Mean*) *Suppose* $X \sim N(\theta, 1)$ *is a Gaussian random variable with mean* θ *and unit variance, then*

$$\begin{aligned} M_{\lambda_1,\lambda_2}(\theta) &\equiv E\left\{\delta_{\lambda_1,\lambda_2}(X)\right\} \\ &= M^H_{\lambda_2}(\theta) + \frac{\lambda_2}{\lambda_2 - \lambda_1}\left\{m_1(\lambda_1,\lambda_2,\theta) - m_1(\lambda_1,\lambda_2,-\theta)\right\}, \end{aligned}$$

where

$$\begin{aligned} M^H_\lambda(\theta) &= \theta + \theta\left[1 - \Phi(\lambda - \theta) - \Phi(\lambda + \theta)\right] + \phi(\lambda - \theta) - \phi(\lambda + \theta) \\ m_1(\lambda_1,\lambda_2,\theta) &\equiv \int_{\lambda_1-\theta}^{\lambda_2-\theta}(x + \theta - \lambda_1)\phi(x)dx \\ &= (\theta - \lambda_1)(\Phi(\lambda_2 - \theta) - \Phi(\lambda_1 - \theta)) - \phi(\lambda_2 - \theta) + \phi(\lambda_1 - \theta). \end{aligned}$$

It is easy to see that the mean function is odd, i.e. $M_\lambda(-\theta) = -M_\lambda(\theta)$.

Lemma 2. (*Variance*) *Suppose* $X \sim N(\theta, 1)$ *is a Gaussian random variable with mean* θ *and unit variance, then*

$$\begin{aligned} V_{\lambda_1,\lambda_2}(\theta) &\equiv \mathrm{Var}\left(\delta_{\lambda_1,\lambda_2}(X)\right) \\ &= V^H_{\lambda_2}(\theta) - \frac{\lambda_2}{\lambda_2-\lambda_1}\Big\{v_2(\lambda_1,\lambda_2,\theta) + v_2(\lambda_1,\lambda_2,-\theta)\Big\}, \end{aligned}$$

where

$$\begin{aligned} V^H_\lambda(\theta) &= (\theta^2+1)(2-\Phi(\lambda-\theta)-\Phi(\lambda+\theta)) \\ &\quad + (\lambda+\theta)\phi(\lambda-\theta) + (\lambda-\theta)\phi(\lambda+\theta) - M^H_\lambda(\theta)^2 \\ v_2(\lambda_1,\lambda_2,\theta) &= m_1(\lambda_1,\lambda_2,\theta)\Big\{2M^H_{\lambda_2}(\theta) + \frac{\lambda_2}{\lambda_2-\lambda_1}(m_1(\lambda_1,\lambda_2,\theta) - m_1(\lambda_1,\lambda_2,-\theta))\Big\} \\ &\quad - \frac{\lambda_2}{\lambda_2-\lambda_1} m_2(\lambda_1,\lambda_2,\theta) \\ m_2(\lambda_1,\lambda_2,\theta) &\equiv \int_{\lambda_1-\theta}^{\lambda_2-\theta} (x+\theta-\lambda_1)^2\phi(x)dx \\ &= (1+(\theta-\lambda_1)^2)\left\{\Phi(\lambda_2-\theta) - \Phi(\lambda_1-\theta)\right\} \\ &\quad - (\lambda_2-2\lambda_1+\theta)\phi(\lambda_2-\theta) - (\lambda_1-\theta)\phi(\lambda_1-\theta). \end{aligned}$$

The proofs of these lemmas are based on straight forward calculus, e.g.

$$E\left(XI_{[|X|\le\lambda]}\right) = \int_{-\lambda-\theta}^{\lambda-\theta} (x+\theta)\phi(x)dx.$$

Figure 2 displays the L_2 risk, variance and bias for the firm shrinkage estimates.

7.2. Covariance of shrinkage estimate

It is much more complex for the covariance function as it involves the bivariate Gaussian distribution.

Lemma 3. (*Covariance*) *Suppose*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

then

$$E\Big\{\delta_\lambda(X_1)\delta_\lambda(X_2)\Big\} = \sqrt{1-\rho^2}\int_{-\infty}^{\infty} \delta_\lambda(x+\theta_1)\phi(x) M_{\frac{\lambda}{\sqrt{1-\rho^2}}}\Big(\frac{\rho x+\theta_2}{\sqrt{1-\rho^2}}\Big)dx,$$

where $M_t(\cdot)$ *is the mean function for shrinkage* $\delta_t(\cdot)$ *as in Lemma* 1. *Then the covariance function* $C_\lambda(\theta_1,\theta_2;\ \rho)$ *can be computed by*

$$C_\lambda(\theta_1,\theta_2;\ \rho) = E\delta_\lambda(X_1)\delta_\lambda(X_2) - M_\lambda(\theta_1)M_\lambda(\theta_2).$$

For the firm shrinkage,

$$
\begin{aligned}
&E\Big\{\delta_{\lambda_1,\lambda_2}(X_1)\delta_{\lambda_1,\lambda_2}(X_2)\Big\}\\
&=\rho_1\Big\{G_1(\lambda_1-\theta_1,\theta_1,\rho',\theta_2',\lambda_2')-G_1(\lambda_1+\theta_1,-\theta_1,-\rho',\theta_2',\ \lambda_2')\Big\}\\
&\quad+r_2\rho_1\Big\{G_3(\lambda_2-\theta_1,\theta_1,\lambda_1',\lambda_2',\rho',\theta_2')-G_3(\lambda_2+\theta_1,-\theta_1,\lambda_1',\lambda_2',-\rho',\theta_2')\\
&\quad-G_3(\lambda_2-\theta_1,\theta_1,\lambda_1',\lambda_2',-\rho',-\theta_2')+G_3(\lambda_2+\theta_1,-\theta_1,\lambda_1',\lambda_2',\rho',-\theta_2')\\
&\quad+G_1(\lambda_1-\theta_1,\theta_1-\lambda_1,\rho',\theta_2',\lambda_2')-G_1(\lambda_2-\theta_1,\theta_1-\lambda_1,\rho',\theta_2',\lambda_2')\\
&\quad-G_1(\lambda_1+\theta_1,-\theta_1-\lambda_1,-\rho',\theta_2',\lambda_2')+G_1(\lambda_2+\theta_1,-\theta_1-\lambda_1,-\rho',\theta_2',\lambda_2')\Big\}\\
&\quad+r_2^2\rho_1\Big\{G_3(\lambda_1-\theta_1,\theta_1-\lambda_1,\lambda_1',\lambda_2',\rho',\theta_2')-G_3(\lambda_2-\theta_1,\theta_1-\lambda_1,\lambda_1',\lambda_2',\rho',\theta_2')\\
&\quad-G_3(\lambda_1+\theta_1,-\theta_1-\lambda_1,\lambda_1',\lambda_2',-\rho',\theta_2')+G_3(\lambda_2+\theta_1,-\theta_1-\lambda_1,\lambda_1',\lambda_2',-\rho',\theta_2')\\
&\quad-G_3(\lambda_1-\theta_1,\theta_1-\lambda_1,\lambda_1',\lambda_2',-\rho',-\theta_2')+G_3(\lambda_2-\theta_1,\theta_1-\lambda_1,\lambda_1',\lambda_2',-\rho',-\theta_2')\\
&\quad+G_3(\lambda_1+\theta_1,-\theta_1-\lambda_1,\lambda_1',\lambda_2',\rho',-\theta_2')-G_3(\lambda_2+\theta_1,-\theta_1-\lambda_1,\lambda_1',\lambda_2',\rho',-\theta_2')\Big\},
\end{aligned}
$$

where

$$
\begin{aligned}
H_1(a,b,c,d)&\equiv\int_a^\infty(x+b)(cx+d)\phi(x)dx=(ac+bc+d)\phi(a)+(bd+c)(1-\Phi(a))\\
H_2(a,b,c,d)&\equiv\int_a^\infty(x+b)\phi(x)\phi(cx+d)dx=A_1(a,d,c)+bA_0(a,d,c)\\
A_0(a,b,c)&\equiv\int_a^\infty\phi(x)\phi(cx+b)dx=\frac{\phi(b')}{\sqrt{1+c^2}}\Big(1-\Phi\Big(a\sqrt{1+c^2}+b'c\Big)\Big)\\
A_1(a,b,c)&\equiv\int_a^\infty x\phi(x)\phi(cx+b)dx=\frac{\phi(a)\phi(ac+b)}{1+c^2}-b'c\ A_0(a,b,c)\\
B_0(a,b,c)&\equiv\int_a^\infty\phi(x)\Phi(cx+b)dx=1-\Phi(a)-\Phi_2(a,\ b';-c')\\
B_1(a,b,c)&\equiv\int_a^\infty x\phi(x)\Phi(cx+b)dx=\phi(a)\Phi(ac+b)+cA_0(a,b,c)\\
B_2(a,b,c)&\equiv\int_a^\infty x^2\phi(x)\Phi(cx+b)dx=a\phi(a)\Phi(ac+b)+B_0(a,b,c)+cA_1(a,b,c)
\end{aligned}
$$

$$
\begin{aligned}
G_1(a,b,c,d,e)&\equiv\int_a^\infty(x+b)\phi(x)M_e^H(cx+d)dx\\
&=H_1(a,b,c,d)+H_2(a,b,c,d-e)-H_2(a,b,c,d+e)\\
&\quad+G_2(a,b,c,d,d-e)-G_2(a,b,c,d,d+e)\\
G_2(a,b,c,d,e,f)&\equiv\int_a^\infty(x+b)(cx+d)\phi(x)\Phi(ex+f)dx\\
&=bdB_0(a,f,e)+(bc+d)B_1(a,f,e)+cB_2(a,f,e)\\
G_3(a,b,c,d,e,f)&\equiv\int_a^\infty(x+b)\phi(x)m_1(c,d,ex+f)dx
\end{aligned}
$$

$$= G_2(a, b, e, f - c, e, f - c) - G_2(a, b, e, f - c, e, f - d)$$
$$+ H_2(a, b, e, f - c) - H_2(a, b, e, f - d)$$

$\rho_1 = \sqrt{1 - \rho^2}$, $r_2 = \lambda_2/(\lambda_2 - \lambda_1)$, $\lambda_1' = \lambda_1/\rho_1$, $\lambda_2' = \lambda_2/\rho_1$, $\rho' = \rho/\rho_1$, $\theta_2' = \theta_2/\rho_1$, $b' = b/\sqrt{1 + c^2}$, $c' = c/\sqrt{1 + c^2}$ *and*

$$\Phi_2(a, b;\ \rho) = \int_a^\infty \int_b^\infty n_2(x, y; \rho) dx dy,$$

where $n_2(x, y; \rho)$ *is the probability density function of a bivariate Gaussian* (X_1, X_2).

8. Software and Computational Details

All plots and calculations are done using extensions to the S+WAVELETS software toolkit version1.1 (Bruce et al. (1996)). S+WAVELETS is a module in the S-Plus software system (Statistical Sciences (1993)). The figures and tables are reproducible through software which can be obtained by anonymous ftp to `ftp.statsci.com` in the directory `pub/WAVELETS`.

The formulas for the mean, variance, L_2 risk, and covariances for the firm shrinkage function are given in Theorem 1 of Section 2 and Lemmas 1-3 of Section 7. These formulas are implemented by the functions `wv.shrink.mean`, `wv.shrink.var`, `wv.shrink.l2`, and `wv.shrink.cov`.

The values in Table 1 were computed by the following steps:

(1) The supremum risk $\Lambda_n(\lambda_1, \lambda_2)$ defined in (11), is computed for a given pair (λ_1, λ_2) using a quasi-Newton optimization with numerical derivatives (Dennis and Mei (1979)).
(2) Apply step (1) to a grid of (λ_1, λ_2) with increments $(\Delta_{\lambda_1}, \Delta_{\lambda_2}) = (0.01, 0.01)$.
(3) Compute the minimum and find a region for (λ_1, λ_2) such that $\Lambda_n(\lambda_1, \lambda_2)$ are within 1% of the minimum.
(4) Apply a quasi-Newton optimization procedure with numerical derivatives (Dennis and Mei (1979)) using initial values from the region obtained in step (3).

The values in Table 2 were computed using a grid search over λ_1 with increments $\Delta_{\lambda_1} = 0.0001$. At each grid point, the supremum over θ was computed using a quasi-Newton optimization with numerical derivatives (Dennis and Mei (1979)).

These (and other) minimax thresholds can be computed using the functions `minimax.hard` (keep or kill and shrink oracles), `minimax.soft1` (keep or kill oracle), `minimax.soft2` (shrink oracle), `minimax.firm` (keep or kill oracle), and `minimax.firm1` (keep, shrink, or kill oracle),

The pointwise variances, given by (16), are computed by the function `var.waveshrink`. For orthogonal wavelets, `var.waveshrink` computes the

variances using either a matrix formulation or more efficient implementation taking advantage of the sparsity in the matrix implied by the wavelet filters. For biorthogonal wavelets, which requires considerably greater computational effort, only the more efficient implementation is used. (Details are described in Bruce et al. (1996).)

Acknowledgements

This work was supported by ONR contract N00014-93-C-0106, by NSF Grant DMI-94-61370, and by ARMY contract DAABll-95-C-0026. The authors are indebted to Professor David Donoho for his comments and suggestions, in particular the name of the new shrinkage function "firm". Two anonymous referees contributed numerous improvements to the manuscript.

References

Brillinger, D. R (1995). Some uses of cumulants in wavelet analysis. *J. Nonparametr. Statist.* 4.

Bruce, A. G. and Gao, H.-Y. (1996). Understanding WaveShrink: Variance and bias estimation. *Biometrika* **83**, 727-745.

Bruce, A. G., Gao, H.-Y. and Percival, D. B. (1996). S+WAVELETS: Version 1.1 release notes. Technical report, 1700 Westlake Ave. N, Seattle, WA 98109-9891.

Buckheit, J. B. and Donoho, D. L. (1995). Wavelab and reproducible research. In *Wavelets and Statistics* (Edited by Antoniadis, A. and Oppenheim G), 55-82. Springer-Verlag Lecture Notes.

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Computational Harmonic Analysis* **1**, 54-81.

Dennis, J. E. and Mei, H. H. W. (1979). Two new unconstrained optimization algorithms which use function and gradient values. *J. Optim. Theory and Appl.* **28**, 453-493.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser.B* **57**, 301-369. (with discussion).

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via Wavelet shrinkage. *Biometrika* **81**, 425-455.

Hampel, F. R., Rousseeuw, P.J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics: the Approach Based on Influence Function.* John wiley, New York.

Statistical Sciences (1993). *S-PLUS User's Manual, Version* 3.2. StatSci, A Division of MathSoft, Inc., 1700 Westlake Ave., Suite 500, Seattle, WA 98109-9891.

Walden, A. T., McCoy, E. J. and Percival, D. B. (1995). Spectrum estimation by wavelet thresholding of multitaper estimators. Technical report, Imperial College of Science, Technology and Medicine.

MathSoft Inc., 1700 Westlake Ave., N, Suite 500, Seattle, WA 98109-9891, U.S.A.

E-mail: gao@statsci.com

E-mail: andrew@statsci.com

(Received October 1995; accepted October 1996)