StatSci

Research Report
No. 39

WaveShrink with Semisoft Shrinkage

Hong-Ye Gao
Andrew G. Bruce

Last revised
September 28, 1995

Acknowledgements: This work was supported by ONR contract N00014-93-C-0106 and by NSF Grant DMI-94-61370.

StatSci Division of MathSoft, Inc.
1700 Westlake Ave. N, Suite 500
Seattle, WA 98109.9891, USA
Tel: (206) 283-8802
FAX: (206) 283-6310

E-mail: andrew@statsci.com
gao@statsci.com
ftp: ftp.statsci.com
(pub/WAVELETS)

WaveShrink and Semisoft Shrinkage

Hong-Ye Gao and Andrew G. Bruce

September 28, 1995

Abstract

Donoho and Johnstone's 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, *semisoft*, which generalizes the hard and soft shrinkage proposed by Donoho and Johnstone. We derive minimax thresholds and provide formulas for computing the pointwise variance, bias, and risk for WaveShrink with semisoft shrinkage. We study the properties of the shrinkage functions, and demonstrate that semisoft 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: Bias Estimation; Confidence Interval; Minimax Thresholds; Non-parametric Regression; Semisoft Shrinkage; Signal De-noising; Trend Estimation; Variance Estimation; Wavelet Transform; WaveShrink;

1 Introduction

Suppose we observe data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ given by

$$y_i = f_i + \sigma z_i \qquad i = 1, 2, ..., n \tag{1}$$

where $\{z_i\}$ are iid $N(0,\ 1)$. The goal is to estimate f 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 Johnstone (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 $\mathbf{w} = (w_1, w_2, \ldots, w_n)'$ be the empirical wavelet coefficients and let W be the wavelet transform matrix. The WaveShrink estimate is obtained by following 3 steps:

[1] Compute wavelet transform $\mathbf{w} = W\mathbf{y}$.

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

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

where $\hat{\sigma}$ is an estimate of the scale σ.

[3] 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:

$$\delta_\lambda^H(x) = xI_{[|x|>\lambda]} \tag{4}$$

$$\delta_\lambda^S(x) = \operatorname{sgn}(x)(|x| - \lambda)_+ \tag{5}$$

where $\lambda \in [0, \infty)$. Bruce and Gao (1995b) 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 the general *semisoft* shrinkage function $\delta_{\lambda_1,\lambda_2}$:

$$\delta_{\lambda_1,\lambda_2}(x) = \begin{cases} 0 & \text{if } |x| \le \lambda_1 \\ \operatorname{sgn}(x)\frac{\lambda_2(|x|-\lambda_1)}{\lambda_2-\lambda_1} & \text{if } \lambda_1 < |x| \le \lambda_2 \\ x & \text{if } |x| > \lambda_2 \end{cases} \tag{6}$$

Figure 1 displays the hard, soft and semisoft shrinkage functions. For values x near the lower threshold λ_1, $\delta_{\lambda_1,\lambda_2}(x)$ behaves like $\delta^S_{\lambda_1}(x)$. For values x above the upper threshold λ_2, $\delta_{\lambda_1,\lambda_2}(x) = \delta^H_{\lambda_2}(x) = x$. Note that hard shrinkage, with $\lambda_1 = \lambda_2$, and soft shrinkage, with $\lambda_2 = \infty$, are limiting cases of (6).

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

The main objectives of this paper are to show the advantages of the semisoft shrinkage function over hard and soft shrinkage, and to further develop the use of WaveShrink with the semisoft shrinkage function. Towards this end, we extend the work of Donoho and Johnstone (1994) to derive minimax thresholds for semisoft shrinkage. We also extend the work of Bruce and Gao (1995b) to derive exact formula for the finite sample bias and variance for WaveShrink under semisoft shrinkage. Finally, we illustrate some of the properties which make semisoft 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.

The paper is organized as follows. In section 2, we give the formula for computing the L_2 risk for semisoft shrinkage of a Gaussian random variable. In section 3, we use this formula to compute minimax thresholds for semisoft shrinkage. Finite sample bias and variance formula are given in section 4. Some advantages of semisoft 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 Semisoft Shrinkage

In this section, we study the L_2 risk of the semisoft 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 semisoft shrinkage is defined as

$$R_{\lambda_1,\lambda_2}(\theta) = E\left(\delta_{\lambda_1,\lambda_2}(X) - \theta\right)^2 \tag{7}$$

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

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

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

$$
\begin{aligned}
&+ \Psi(\lambda_1 - \theta,\ \lambda_2 - \theta,\ r_2,\ r_1(\theta - \lambda_2)) + \Psi(\lambda_1 + \theta,\ \lambda_2 + \theta,\ r_2,\ -r_1(\theta + \lambda_2)) \\
&+ (\lambda_2 - \theta)\phi(\lambda_2 - \theta) + (\lambda_2 + \theta)\phi(\lambda_2 + \theta) \qquad (8)
\end{aligned}
$$

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 and*

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

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

Figure 2 displays the L_2 risk of the semisoft 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 semisoft 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,

$$R(\hat{f}_\lambda,\ f) = \frac{1}{n} \sum_{k=1}^{n} E\left(\hat{w}_k - \theta_k\right)^2 \qquad (9)$$

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

$$R(\hat{f}_\lambda,\ f) = \frac{\sigma^2}{n} \sum_{k=1}^{n} R_{\lambda_1, \lambda_2}\left(\frac{\theta_k}{\sigma}\right). \qquad (10)$$

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

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 of the wavelet coefficients are typically near zero, the risk of the WaveShrink estimate is often higher. For further elaboration, see Bruce and Gao (1995b) and section 4.

Remark 3: We can exploit the extra flexibility of semisoft shrinkage to achieve WaveShrink estimates with smaller L_2 risk. In fact, semisoft 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

$$R_{\lambda_1, \lambda_2}(\theta) < R_\lambda^H(\theta) \qquad \text{for all } \theta. \qquad (11)$$

Figure 3 displays the region for the values (λ_1, λ_2) such that (11) holds when $\lambda = 3.33$.

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

3 Threshold Selection for Semisoft 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 semisoft 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

$$\Lambda_n^{**} \equiv \inf_{\lambda_1 \leq \lambda_2} \sup_{\theta} \left\{ \frac{R_{\lambda_1,\lambda_2}(\theta)}{n^{-1} + \min(\theta^2, 1)} \right\} \tag{12}$$

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 (10), we can bound the L_2 risk by

$$R(\hat{f}_{\lambda^*},\ f) \leq \frac{\Lambda_n^{**}}{n} \left\{ \sigma^2 + \sum_{k=1}^{n} \min(\theta_k^2, \sigma^2) \right\}. \tag{13}$$

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 semisoft shrinkage is guaranteed to have smaller minimax risk than soft and hard shrinkage, the nearly optimal asymptotic results hold for semisoft shrinkage as well.

The minimax thresholds $(\lambda_1^*, \lambda_2^*)$ can be derived by choosing $(\lambda_1^*, \lambda_2^*)$ to attain (12): 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, semisoft 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 (1995b), who determined that ideal soft shrinkage has lower risk for small n but ideal hard shrinkage has lower risk for large n.

Define the supremum risk by

$$\Lambda_n(\lambda_1, \lambda_2) \equiv \sup_{\theta} \left\{ \frac{R_{\lambda_1,\lambda_2}(\theta)}{n^{-1} + \min(\theta^2, 1)} \right\} \tag{14}$$

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.

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

$$P\left\{\max_{1\le i\le n}|X_i| > \sqrt{2\log n}\right\} \sim \frac{1}{\sqrt{\pi\log n}}, \qquad \text{as } n\to\infty. \tag{15}$$

In general,

$$P\left\{\max_{1\le i\le n}|X_i| > \sqrt{c\log n}\right\} \sim \frac{\sqrt{2}}{n^{c/2-1}\sqrt{c\pi\log n}}, \qquad \text{as } n\to\infty. \tag{16}$$

We can define other universal thresholds such that (16) converges to zero faster (i.e., to suppress noise more thoroughly). For example, threshold $\sqrt{4\log n}$ makes above convergence rate $1/n\sqrt{2\pi\log n}$ and threshold $\sqrt{6\log n}$ makes above convergence rate $1/n^2\sqrt{3\pi\log n}$.

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 semisoft 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. Even with the upper threshold fixed, table 2 shows that the minimax risk bounds are tighter than for soft or hard shrinkage.

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

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

Then the ideal risk is

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

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 Semisoft WaveShrink

Not only does semisoft 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 semisoft shrinkage, and show that semisoft 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 [3]. 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

$$\hat{f}_i = \sum_k \tilde{c}_{ki}\hat{w}_k \qquad i = 1, 2, ..., n \tag{17}$$

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

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

$$E(\hat{f}_i) = \sigma \sum_k \tilde{c}_{ki}\ M_\lambda\left(\frac{\theta_k}{\sigma}\right) \tag{18}$$

$$var(\hat{f}_i) = \sigma^2 \sum_{k,\ell} \tilde{c}_{ki}\tilde{c}_{\ell i}\ C_\lambda\left(\frac{\theta_k}{\sigma}, \frac{\theta_\ell}{\sigma};\ \rho_{k,\ell}\right) \tag{19}$$

where $M()$ and $C()$ are the mean and covariance functions of the semisoft shrinkage function defined in the section 7 and $\rho_{k,\ell} = corr(w_k, w_\ell)$.

The proof follows from Bruce and Gao (1995b), which gives a similar formula for hard and soft shrinkage.

We use the results of theorem 2 to present an example of how semisoft can produce estimates with lower risk. Figure 5 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 of 6. In figure 6, we plot the bias, standard error, and L_2 risk of WaveShrink using hard, soft, and semisoft shrinkage for the signal displayed in the top panel of figure 5. 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 semisoft shrinkage has smallest L_2 risk. This example is quite typical, and reflects very general behaviour of hard, soft, and semisoft shrinkage (see also Bruce and Gao (1995b)).

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

$$s_i = \hat{\sigma}\sqrt{\sum_{k,\ell} \tilde{c}_{ki}\tilde{c}_{\ell i}\; C_\lambda\left(\frac{\hat{\theta}_k}{\hat{\sigma}}, \frac{\hat{\theta}_\ell}{\hat{\sigma}};\; \rho_{k,\ell}\right)} \tag{20}$$

Remark 2: The summations in (18)-(19) 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

$$\alpha_{\lambda} = \sup_{x \neq y} \left| \frac{\delta_{\lambda}(y) - \delta_{\lambda}(x)}{y - x} \right|. \tag{21}$$

It is easy to see that $\alpha^H_{\lambda} = \infty$, $\alpha^S_{\lambda} = 1$ and $\alpha_{\lambda_1,\lambda_2} = \frac{\lambda_2}{\lambda_2 - \lambda_1}$. Soft shrinkage and semisoft shrinkage are less sensitive to small perturbations in the data than hard shrinkage.

Figure 7 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 5 and 6 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 semisoft shrinkage. Minimax thresholds are used for both hard and semisoft 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

$$\beta_{\lambda} = \inf\{r > 0;\ \delta_{\lambda}(x) = x, \text{ when } |x| > r\}. \tag{22}$$

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 (22) 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 $f_v(x) = v\, I_{[0 \leq x < 0.5]}$, 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 semisoft shrinkage (with the Haar wavelet) go to zero as $v \to \infty$.

6 Discussion

In this paper, we introduce the semisoft shrinkage function for use with the WaveShrink procedure. Semisoft 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, semisoft 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.

We can improve further upon the semisoft shrinkage function by replacing the linear function in 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 $r_1 = \frac{\lambda_1+\lambda_2}{(\lambda_2-\lambda_1)^3}$ and $r_2 = \frac{2\lambda_2^2}{(\lambda_2-\lambda_1)^3}$. In practical terms, we expect that the relative gain over linear semisoft shrinkage will be small.

7 Technical Details

Bruce and Gao (1995b) 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 semisoft 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

$$\hat{w}_k = \delta_{\lambda\hat{\sigma}}(w_k) = \hat{\sigma}\delta_\lambda\left(\frac{w_k}{\hat{\sigma}}\right) = \sigma\delta_{\lambda\frac{\hat{\sigma}}{\sigma}}\left(\frac{w_k}{\sigma}\right).$$

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

$$M_{\lambda_1,\lambda_2}(\theta) \equiv E\left\{\delta_{\lambda_1,\lambda_2}(X)\right\} = M^H_{\lambda_2}(\theta) + \frac{\lambda_2}{\lambda_2 - \lambda_1}\left\{m_1(\lambda_1,\lambda_2,\theta) - m_1(\lambda_1,\lambda_2,-\theta)\right\} \qquad (23)$$

where

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

It is easy to see that the mean function is odd function, 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*

$$V_{\lambda_1,\lambda_2}(\theta) \equiv var(\delta_{\lambda_1,\lambda_2}(X)) = V_{\lambda_2}^H(\theta) - \frac{\lambda_2}{\lambda_2-\lambda_1}\left\{v_2(\lambda_1,\lambda_2,\theta) + v_2(\lambda_1,\lambda_2,-\theta)\right\} \tag{24}$$

where

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

The proofs of these lemmas are 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 semisoft shrinkage estimates.

7.2 Covariance of Shrinkage Estimate

It is much more complex for covariance function, it involves 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

$$E\left\{\delta_\lambda(X_1)\delta_\lambda(X_2)\right\} = \sqrt{1-\rho^2}\int_{-\infty}^{\infty}\delta_\lambda(x+\theta_1)\phi(x)M_{\frac{\lambda}{\sqrt{1-\rho^2}}}\left(\frac{\rho x+\theta_2}{\sqrt{1-\rho^2}}\right)dx \qquad (25)$$

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 semisoft shrinkage,

$$\begin{aligned}
&E\left\{\delta_{\lambda_1,\lambda_2}(X_1)\delta_{\lambda_1,\lambda_2}(X_2)\right\} \quad = \\
&\quad \sqrt{1-\rho^2}\left\{G_1\left(\lambda_1-\theta_1,\ \theta_1,\ \rho',\ \theta_2',\ \lambda_2'\right) - G_1\left(\lambda_1+\theta_1,\ -\theta_1,\ -\rho',\ \theta_2',\ \lambda_2'\right)\right\} + \\
&\quad + \frac{\lambda_2\sqrt{1-\rho^2}}{\lambda_2-\lambda_1}\left\{G_3\left(\lambda_2-\theta_1,\ \theta_1,\ \lambda_1',\ \lambda_2',\ \rho',\ \theta_2'\right) - G_3\left(\lambda_2+\theta_1,\ -\theta_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ \theta_2'\right)\right. \\
&\quad - G_3\left(\lambda_2-\theta_1,\ \theta_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ -\theta_2'\right) + G_3\left(\lambda_2+\theta_1,\ -\theta_1,\ \lambda_1',\ \lambda_2',\ \rho',\ -\theta_2'\right) + \\
&\quad + G_1\left(\lambda_1-\theta_1,\ \theta_1-\lambda_1,\ \rho',\ \theta_2',\ \lambda_2'\right) - G_1\left(\lambda_2-\theta_1,\ \theta_1-\lambda_1,\ \rho',\ \theta_2',\ \lambda_2'\right) - \\
&\quad \left. - G_1\left(\lambda_1+\theta_1,\ -\theta_1-\lambda_1,\ -\rho',\ \theta_2',\ \lambda_2'\right) + G_1\left(\lambda_2+\theta_1,\ -\theta_1-\lambda_1,\ -\rho',\ \theta_2',\ \lambda_2'\right)\right\} + \\
&\quad + \frac{\sqrt{1-\rho^2}\lambda_2^2}{(\lambda_2-\lambda_1)^2}\left\{G_3\left(\lambda_1-\theta_1,\ \theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ \rho',\ \theta_2'\right) - G_3\left(\lambda_2-\theta_1,\ \theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ \rho',\ \theta_2'\right)\right. \\
&\quad - G_3\left(\lambda_1+\theta_1,\ -\theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ \theta_2'\right) + G_3\left(\lambda_2+\theta_1,\ -\theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ \theta_2'\right) - \\
&\quad - G_3\left(\lambda_1-\theta_1,\ \theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ -\theta_2'\right) + G_3\left(\lambda_2-\theta_1,\ \theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ -\rho',\ -\theta_2'\right) + \\
&\quad \left. + G_3\left(\lambda_1+\theta_1,\ -\theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ \rho',\ -\theta_2'\right) - G_3\left(\lambda_2+\theta_1,\ -\theta_1-\lambda_1,\ \lambda_1',\ \lambda_2',\ \rho',\ -\theta_2'\right)\right\}
\end{aligned}$$

where

$$\begin{aligned}
G_1(a,b,c,d,e) &\equiv \int_a^\infty (x+b)\phi(x)M_e^H(cx+d)dx = H_1(a,b,c,d) + H_2(a,b,c,d-e) - \\
&\quad - H_2(a,b,c,d+e) + H_3(a,b,c,d,d-e) - H_3(a,b,c,d,d+e) \\
G_3(a,b,c,d,e,f) &\equiv \int_a^\infty (x+b)\phi(x)m_1(c,d,ex+f)dx = H_3(a,b,e,f-c,e,f-c) - \\
&\quad - H_3(a,b,e,f-c,e,f-d) + H_2(a,b,e,f-c) - H_2(a,b,e,f-d) \\
H_1(a,b,c,d) &\equiv \int_a^\infty (x+b)(cx+d)\phi(x)dx = (ac+bc+d)\phi(a) + (bd+c)(1-\Phi(a)) \\
H_2(a,b,c,d) &\equiv \int_a^\infty (x+b)\phi(x)\phi(cx+d)dx \ = \ A_1(a,d,c) + bA_0(a,d,c) \\
H_3(a,b,c,d,e,f) &\equiv \int_a^\infty (x+b)(cx+d)\phi(x)\Phi(ex+f)dx \\
&= bdB_0(a,f,e) + (bc+d)B_1(a,f,e) + cB_2(a,f,e) \\
A_0(a,b,c) &\equiv \int_a^\infty \phi(x)\phi(cx+b)dx = \frac{\phi(b')}{\sqrt{1+c^2}}\left(1-\Phi\left(a\sqrt{1+c^2}+b'c\right)\right)
\end{aligned}$$

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

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

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

where $n_2(x,y;\rho)$ is the probability density function of 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 (Bruce and Gao (1994)). 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 semisoft 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 (14), 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 $\Lambda_n(\lambda_1, \lambda_2)$ were 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.semisoft` (keep or kill oracle), and `minimax.semisoft1` (keep, shrink, or kill oracle),

The pointwise variances, given by (19), 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 and Gao (1995a).

List of Tables

List of Figures

	semisoft			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

Table 1: Minimax thresholds and minimax risk bounds for semisoft shrinkage, compared with soft and hard shrinkage.

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

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

	$\max\{\lvert E\hat{f}_i - f_i\rvert\}$			$\sum_{i=1}^{n}\{\lvert E\hat{f}_i - f_i\rvert^2\}$		
v	soft	hard	semisoft	soft	hard	semisoft
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

Table 3: Bias comparison of soft, hard, and semisoft shrinkage with the Haar wavelet for $f_v(x) = vI_{[0\leq x\leq 0.5]}$, 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 $\max\{\lvert E\hat{f}_i - f_i\rvert\}$ and the last three columns compare the average squared bias $\sum_{i=1}^{n}(E\hat{f}_i - f_i)^2$.

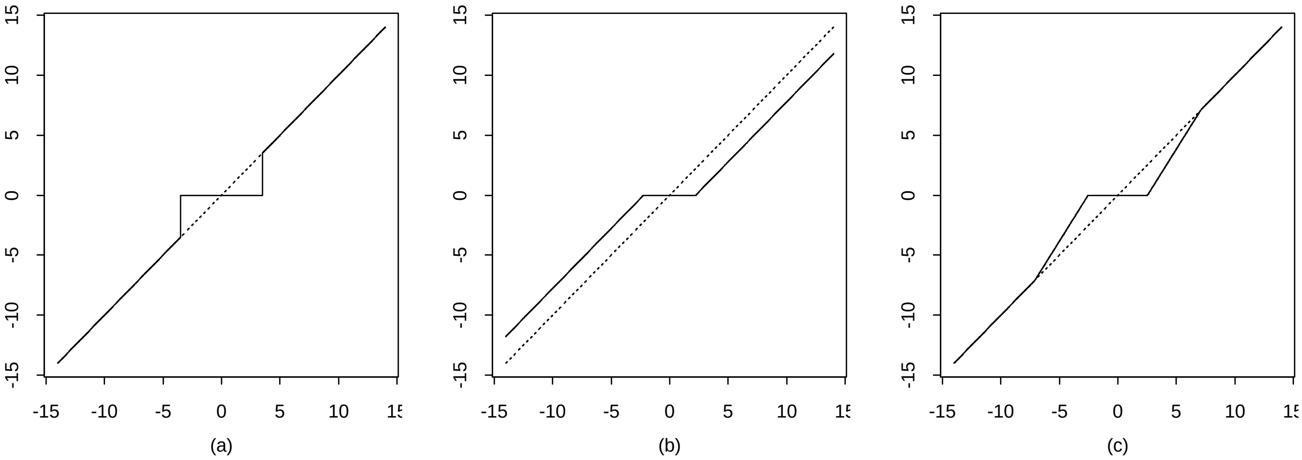

Figure 1: From left to right: hard, soft and semisoft shrinkage function (solid lines). Like hard shrinkage, semisoft is the identity function (dashed lines) for large enough $|x|$. Like soft shrinkage, semisoft shrinkage is continuous.

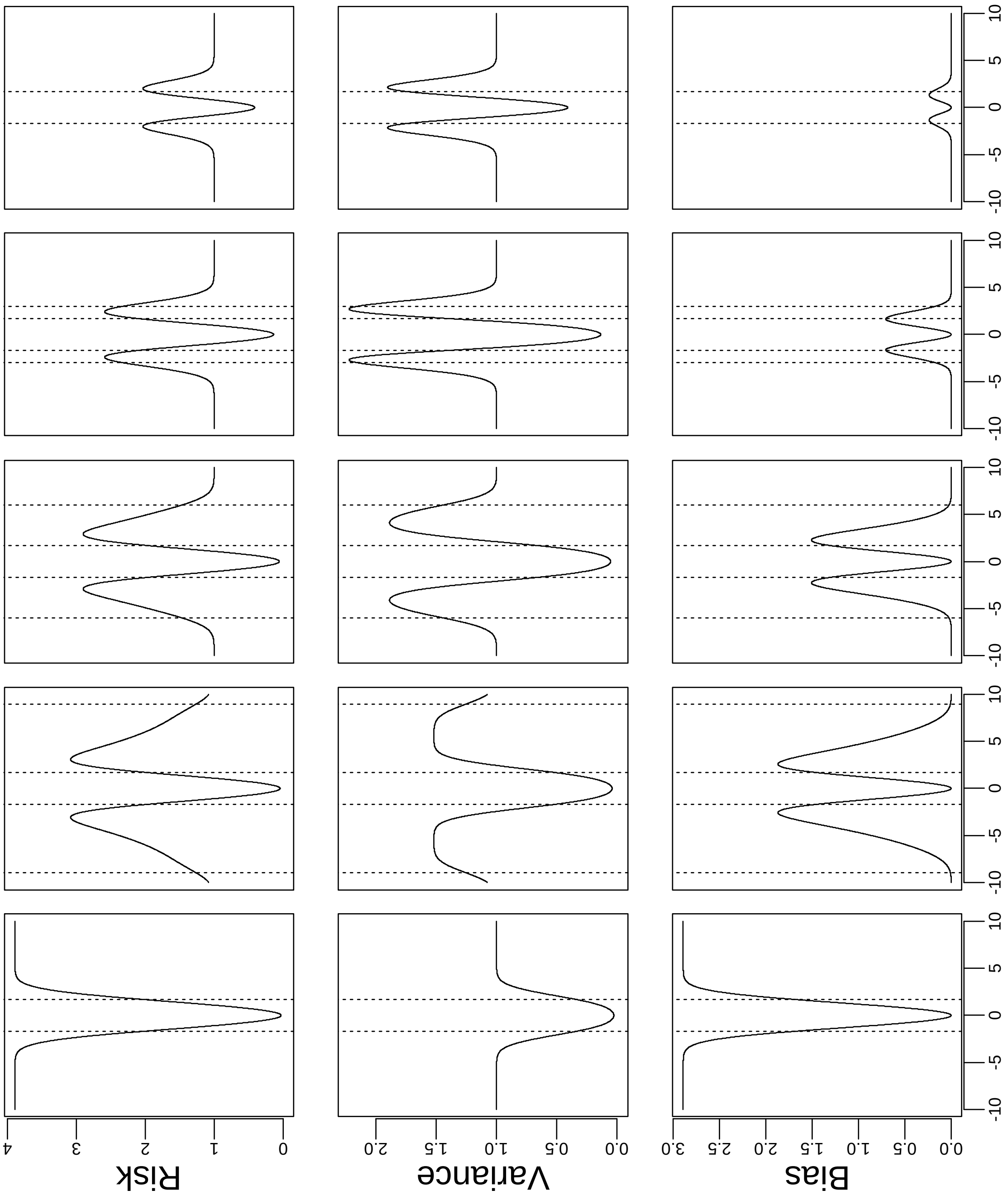

Figure 2: First row: L_2 risks vs. θ for semisoft shrinkage, with $\lambda_1 \equiv 1.7$ and $\lambda_2 = \infty, 9, 6, 3, 1.7$. Middle row: variances vs. θ for semisoft shrinkage, with $\lambda_1 \equiv 1.7$ and $\lambda_2 = \infty, 9, 6, 3, 1.7$. Bottom row: squared biases vs. θ for semisoft shrinkage, with $\lambda_1 \equiv 1.7$ and $\lambda_2 = \infty, 9, 6, 3, 1.7$. Dotted vertical lines indicate the thresholds.

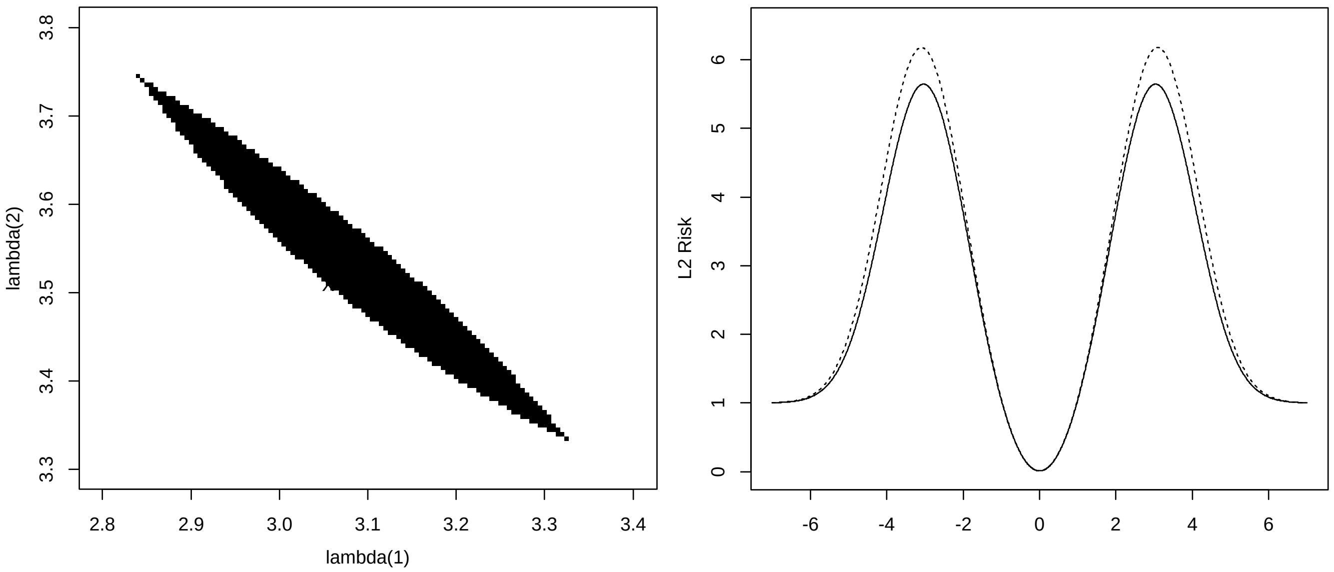

Figure 3: Left: the pairs of thresholds (λ_1, λ_2) for which semisoft shrinkage has uniformly smaller risk than hard shrinkage for $\lambda = 3.33$. Right: the risk curves under semisoft shrinkage with $\lambda_1 = 3.055$ and $\lambda_2 = 3.51$ (solid line) and under hard shrinkage with $\lambda = 3.33$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

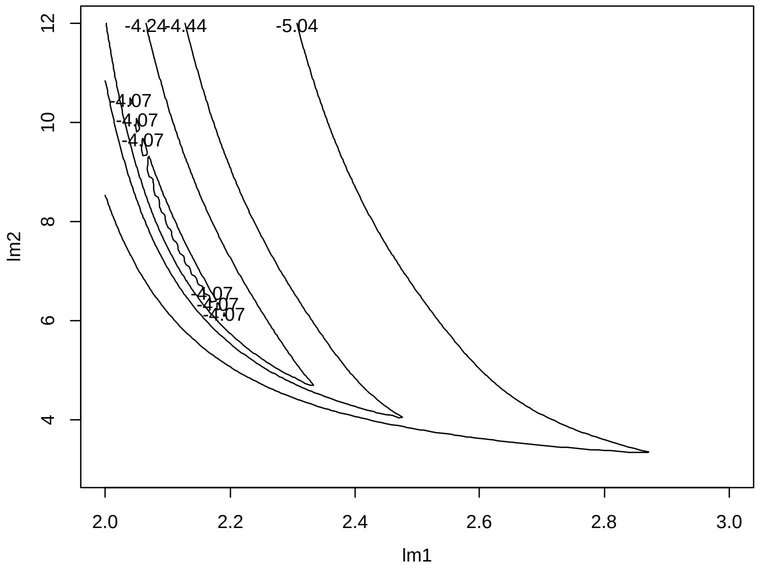

Figure 4: A contour plot of the supremum risk $\Lambda_n(\lambda_1, \lambda_2)$ defined by (14) over (λ_1, λ_2) with contour lines: 1%, 5%, 10% and 25% of the minimum.

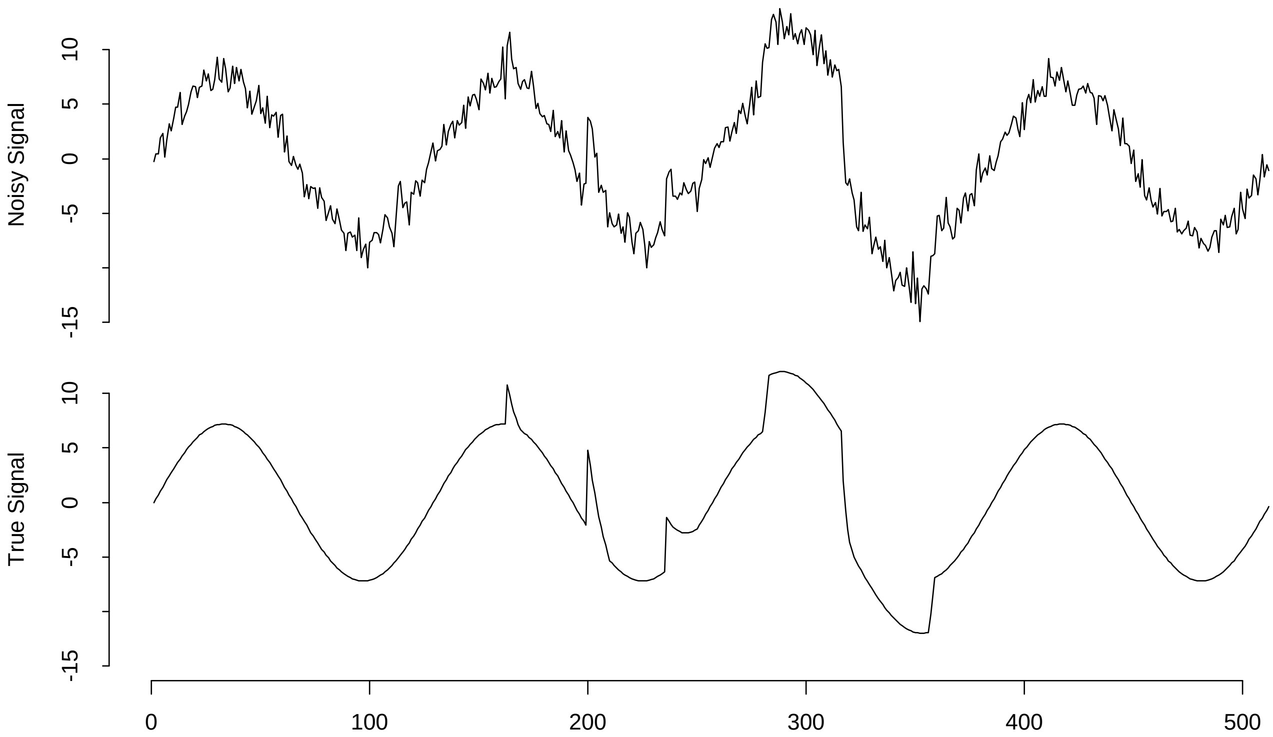

Figure 5: A synthetic "jumps" signal used to compare the variance, bias and risk of the various shrinkage estimates. The top panel displays the noisy signal and the bottom panel displays the original noise free signal. The noise is additive white Gaussian with the variance set so that the signal to noise ratio is 6.

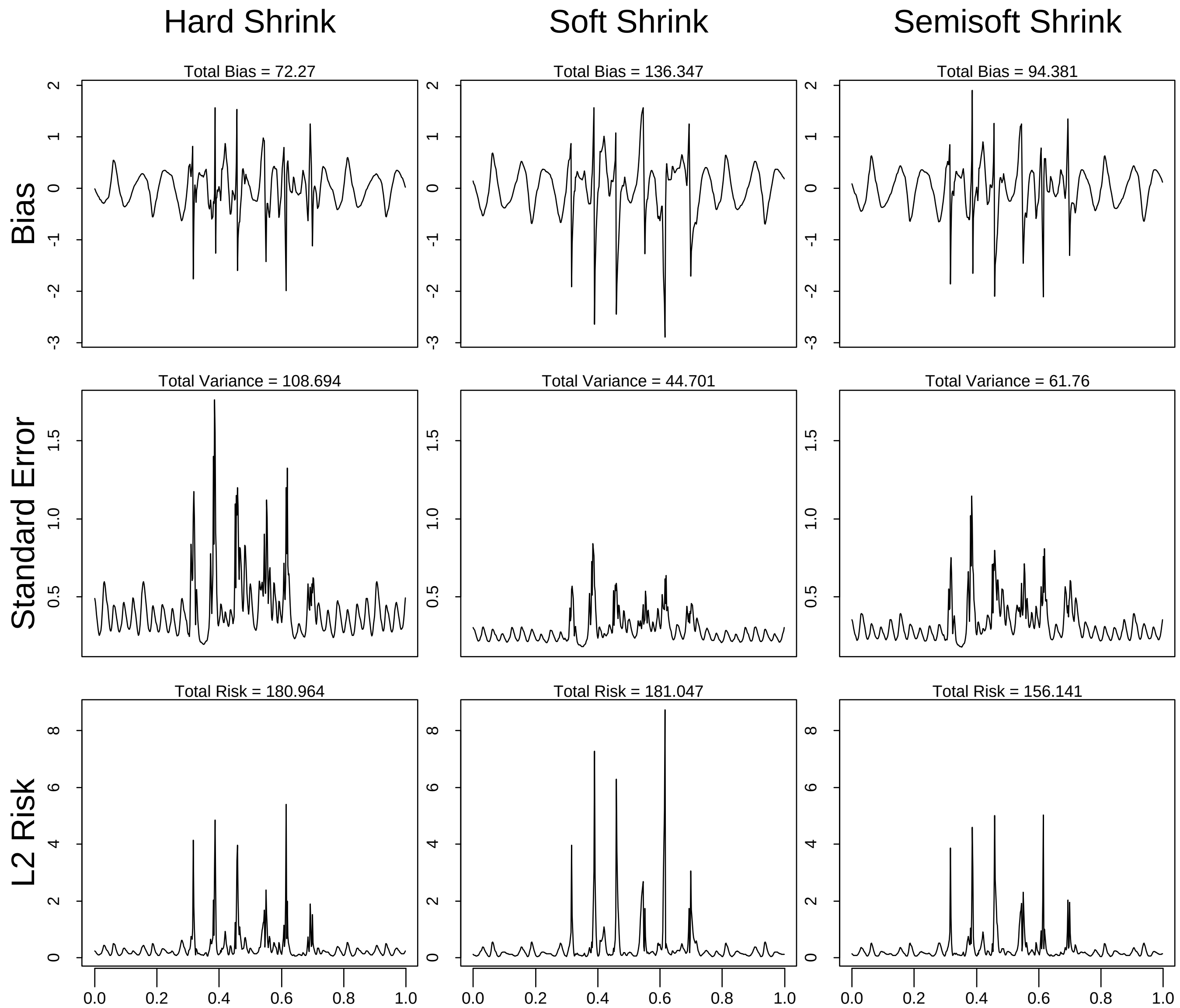

Figure 6: Bias, standard error, and L_2 risk for the signal of figure 5. First row, from left to right: bias for hard, soft, and semisoft shrinkage. Second row, from left to right: standard error for hard, soft, and semisoft shrinkage. Third row, from left to right: L_2 risk for hard, soft, and semisoft 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 semisoft shrinkage has smallest L_2 risk.

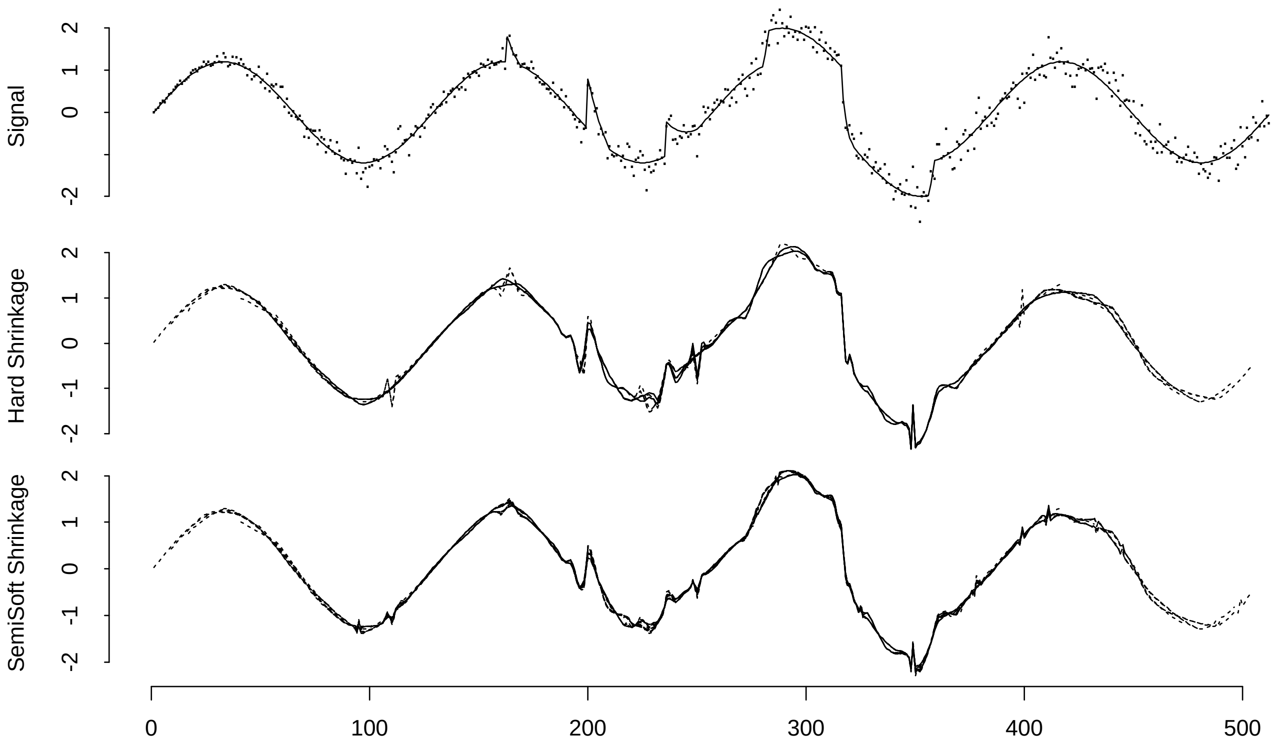

Figure 7: Hard shrinkage is more sensitive to small perturbations in the data than soft or semisoft shrinkage. The top panel displays a signal of length 512 with additive Gaussian white noise (signal to noise ratio of 5). The middle panel displays the estimates of the signal obtained using the WaveShrink 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 semisoft shrinkage. Minimax thresholds are used for both hard and semisoft shrinkage and the scale of the noise is re-estimated for each window.

References

Brillinger, D. R. (1995). Uses of Cumulants in Wavelet Analysis. Technical report, University of California, Berkeley.

Bruce, A. G. and Gao, H.-Y. (1994). *S+WAVELETS Users Manual*. StatSci Division of MathSoft, Inc., 1700 Westlake Ave. N, Seattle, WA 98109-9891.

Bruce, A. G. and Gao, H.-Y. (1995a). Fast Computation of WaveShrink Variances. Technical report, StatSci Division of MathSoft, Inc., 1700 Westlake Ave. N, Seattle, WA 98109-9891. In preparation.

Bruce, A. G. and Gao, H.-Y. (1995b). Understanding WaveShrink: Variance and Bias Estimation. Technical report, StatSci Division, MathSoft, Inc., 1700 Westlake Ave. N, Seattle, WA 98109-9891.

Buckheit, J. B. and Donoho, D. L. (1995). Wavelab and Reproducible Research. Technical Report 474, Stanford University.

Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the Interval and Fast Wavelet Transforms. *Applied and 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. *Journal of Optimization Theory and Applications*, 28(4):453–483.

Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1995). Wavelet Shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series 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 functions*. 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.