Minimum risk thresholds for data with heavy noise

Maarten Jansen
TU Eindhoven, dept. of Math. and Comp. Science
and K.U.Leuven, dept. of Comp. Science

September 2005

Abstract

In the estimation of data with many zeros (sparse data), such as wavelet coefficients, thresholding is a common technique. This paper investigates the behavior of the minimum risk threshold for large values of the noise standard deviation. It finds that the threshold depends quadratically on the noise standard deviation. The relevance of this result is situated in the context of both Bayesian and universal thresholding.

Keywords

Threshold, wavelet, Bayes

1 Introduction

Thresholding has been an intensively studied, yet straightforward procedure in nonlinear estimation of sparse sequences. The research on this subject has been engined by the popularity of wavelet theory and applications, but the problem of estimating sparse sequences also arises in other areas, such as the detection of gene expressions in micro array data.

Among the numerous threshold assessment procedures, the universal threshold takes a prominent position, as it offers many optimality properties [3, 5] the details of which are far beyond the scope of this paper. For a vector of N observations with constant variance σ^2, this threshold equals

$$\lambda_{\mathrm{univ}} = \sqrt{2 \log N}\sigma. \tag{1}$$

In practical applications, such as image denoising, the universal threshold is often found to be too conservative, i.e., it removes too much of the underlying data, thereby causing blur in the output. Nevertheless, the simplicity of expression (1) makes it a good starting point in many practical algorithms as well.

The expression (1) for the universal threshold has been derived as an asymptotic results on N in the first place. The expression suggests, however, that the threshold should depend linearly on the standard deviation of the noise. This seems to confirm the intuition that 'a good' (in some sense optimal?) threshold should be proportional to the amount of noise. On the other hand, some thresholds in Bayesian models [10, 2, 11] follow a quadratic rule:

$$\lambda_{\mathrm{bayes}} = C\sigma_{\mathrm{noise}}^2/\sigma_{\mathrm{signal}}, \tag{2}$$

with constant C depending on the used model. Depending on the exact Bayesian model and the Bayesian decision rule, this result is exact [11] or a numerical approximation of the asymptotic behavior for $\sigma_{\mathrm{signal}} \to \infty$, so not for the dependence on σ_{noise} in the first place [2]. Nevertheless, in any case, the Bayesian analysis seems to suggest that the threshold should be taken proportional to the noise variance, rather than to its standard deviation. This paper investigates the dependence of the minimum expected average squared error threshold on the standard deviation. It finds that the linear dependence holds for small amounts of noise, while even for moderate noise, the dependence becomes quadratic. This result holds regardless of the configuration of the coefficients, and hence for any prior model in Bayesian thresholding. The paper interprets this result and its significance with respect to the existing universal and Bayesian thresholds.

2 Definitions

Suppose we are given N observations in the following additive, i.i.d. (independently, identically distributed) normal model:

$$\boldsymbol{w} = \boldsymbol{v} + \boldsymbol{\omega}, \tag{3}$$

where $\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{\omega}$ are real vectors of length N, and $\boldsymbol{\omega}$ consists of independent, normally distributed random variables with zero mean and equal variance σ^2. The vector of noise-free variables $\boldsymbol{v}$ is typically sparse, meaning that it contains a lot of (near-)zeros, but no assumption in this sense is needed in the forthcoming results.

The noise-free variables are estimated by a pointwise threshold procedure, with one common threshold value λ:

$$\widehat{\boldsymbol{v}} = \boldsymbol{w}_{\boldsymbol{\lambda}},$$

where $\boldsymbol{w}_{\boldsymbol{\lambda}}$ stands for the output of a threshold operation applied to the vector $\boldsymbol{w}$. The precise definition of this threshold operation is subject to some degrees of freedom.

The hard threshold operation is defined as

$$\begin{aligned} \boldsymbol{w}_{\boldsymbol{\lambda}} = \mathrm{HT}_\lambda(\boldsymbol{v}) \Leftrightarrow \forall i = 1, \dots, N \quad : \quad & w_{\lambda i} = 0 \text{ if } |w_i| \leq \lambda \text{ and} \\ & w_{\lambda i} = w_i \text{ otherwise.} \end{aligned}$$

The soft threshold operation is defined as

$$\begin{aligned} \boldsymbol{w}_{\boldsymbol{\lambda}} = \mathrm{ST}_\lambda(\boldsymbol{v}) \Leftrightarrow \forall i = 1, \dots, N : w_{\lambda i} &= 0 \text{ if } |w_i| \leq \lambda \text{ and} \\ w_{\lambda i} &= \mathrm{sign}(w_i) \cdot (|w_i| - \lambda) \text{ otherwise.} \end{aligned}$$

In other words, soft thresholding not only sets small values to zero, but also shrinks the larger values by an amount equal to the threshold. As a consequence, the transition between small and large value operation is continuous (although not continuously differentiable). There exists numerous intermediate threshold operations that combine the continuity property of the soft threshold operation with the zero shrinkage property for large values in a hard threshold procedure [6, 1]. An important class of such threshold operations is defined in a Bayesian framework [9].

For any threshold operation (soft, hard or intermediate), the risk, or expected average squared error (EASE) is defined as

$$R(\lambda) = \mathrm{EASE}(\lambda) = \frac{1}{N} E \|\boldsymbol{w}_{\boldsymbol{\lambda}} - \boldsymbol{v}\|^2. \tag{4}$$

The minimization of the risk function is the objective in an important class of wavelet thresholds estimators, such as SURE (Stein's Unbiased Risk Estimator) [4] or (Generalized) Cross Validation [8]. Straightforward calculations (i.e., by taking the derivatives of the exact risk expressions found in the literature [3, Appendix 2]) show that the soft threshold risk in an additive, i.i.d. normal model is minimized if the threshold λ satisfies the following nonlinear equation:

$$\frac{\lambda}{\sigma} = \frac{\sum_{i=1}^{N} \phi\left(\frac{\lambda - v_i}{\sigma}\right) + \phi\left(\frac{\lambda + v_i}{\sigma}\right)}{\sum_{i=1}^{N} \left[1 - \Phi\left(\frac{\lambda - v_i}{\sigma}\right)\right] + \left[1 - \Phi\left(\frac{\lambda + v_i}{\sigma}\right)\right]}, \tag{5}$$

where $\phi(x)$ is the standard normal probability density function (pdf) and $\Phi(x)$ is the standard normal cumulative distribution function (cdf). For hard thresholding, the risk is minimized if

$$\lambda = \frac{2\sum_{i=1}^{N} v_i \left[\phi\left(\frac{\lambda - v_i}{\sigma}\right) - \phi\left(\frac{\lambda + v_i}{\sigma}\right)\right]}{\sum_{i=1}^{N} \phi\left(\frac{\lambda + v_i}{\sigma}\right) + \phi\left(\frac{\lambda - v_i}{\sigma}\right)}. \tag{6}$$

3 Main result

The main result of this paper is stated in the following proposition.

Proposition 1 *For a given vector $\boldsymbol{v}$ of fixed length N, and for increasing noise variance $\sigma^2 \to \infty$, the minimum risk soft threshold behaves as:*

$$\lambda \sim K \cdot \sigma^2, \tag{7}$$

where K is the solution of the nonlinear equation

$$K = \frac{\sum_{i=1}^{N} \left(e^{v_i K} + e^{-v_i K}\right)}{\sum_{i=1}^{N} v_i \left(e^{v_i K} - e^{-v_i K}\right)}. \tag{8}$$

Experiments show that this limit behavior is reached even at moderate noise levels, i.e., for values of σ within the range of the noise-free coefficients $\boldsymbol{v}$. Before proving this proposition, we also prove the following result for small noise levels.

Proposition 2 *Assume that a given vector $\boldsymbol{v}$ of fixed length N has $N_0 > 0$ zero entries, then for decreasing noise variance $\sigma^2 \to 0$, the minimum risk soft threshold behaves as:*

$$\lambda \sim k \cdot \sigma, \tag{9}$$

where k is the solution of the nonlinear equation

$$k = \frac{N_0 2\phi(k)}{N_1 + N_0 2[1 - \Phi(k)]}, \tag{10}$$

where $N_1 = N - N_0$.

Proof: Denote $T(\sigma) = \lambda(\sigma)/\sigma$, then Equation (5) becomes

$$T(\sigma) = \frac{\sum_{i=1}^{N} \phi\left(T(\sigma) - v_i/\sigma\right) + \phi\left(T(\sigma) + v_i/\sigma\right)}{\sum_{i=1}^{N} \left[1 - \Phi\left(T(\sigma) - v_i/\sigma\right)\right] + \left[1 - \Phi\left(T(\sigma) + v_i/\sigma\right)\right]}. \tag{11}$$

If all $|v_i| > 0$ (i.e., $N_0 = 0$), then taking the limit for $\sigma \to 0$ on both sides of (11) leads to $T(\sigma) \to 0$. Indeed, the numerator on the right hand side consists of terms that all tend to zero, as the arguments of the standard normal pdf all tend to infinity. The terms in the denominator on the other hand tend to $1 + 0$ or $0 + 1$.

If $N_0 > 0$, then taking the limits on both sides of (11) leads to:

$$k = \lim_{\sigma \to 0} T(\sigma) = \frac{N_0 2\phi\left(\lim_{\sigma\to 0} T(\sigma)\right)}{N_1 \cdot 1 + N_0 2\left[1 - \Phi\left(\lim_{\sigma \to 0} T(\sigma)\right)\right],}$$

which proves Proposition 2. □

For the proof of Proposition 1 we need the following lemma.

Lemma 1 *For $T(\sigma)$ as defined in (11), we have $\lim_{\sigma\to\infty} T(\sigma) = \infty$.*

Proof: Taking the limit on both sides of (11) leads to an equation for $L = \lim_{\sigma\to\infty} T(\sigma)$:

$$L = \frac{\phi(L)}{[1 - \Phi(L)]}.$$

This equation has no finite solution. □

Proof of Proposition 1:

Since $\lim_{\sigma\to\infty} T(\sigma) = \infty$, we can use the expansion $1 - \Phi(u) \approx \phi(u) \cdot \left[(1/u) - (1/u^3)\right]$ for approximation the equation (11). Although we omit the full analysis for reasons of space limits, it can be verified that all subsequent approximations leave the asymptotic behavior of the solution of (11) unchanged. Equation (11) can be written as:

$$\begin{aligned}
&\sum_{i=1}^{N} \phi(T + v_i/\sigma) + \phi(T - v_i/\sigma) = \\
&T\left[\sum_{i=1}^{N} \phi(T + v_i/\sigma)\left(\frac{1}{T + \frac{v_i}{\sigma}} - \frac{1}{(T + \frac{v_i}{\sigma})^3} \pm \dots\right)\right. \\
&\left. + \phi(T - v_i/\sigma)\left(\frac{1}{T - \frac{v_i}{\sigma}} - \frac{1}{(T - \frac{v_i}{\sigma})^3} \pm \dots\right)\right].
\end{aligned}$$

Using the fact that for $T\sigma \to \infty$, we have that

$$\frac{1}{1 + \frac{v_i}{T\sigma}} \approx 1 - \frac{v_i}{T\sigma} \text{ and } \frac{1}{\left(1 + \frac{v_i}{T\sigma}\right)^3} \approx 1 - 3\frac{v_i}{T\sigma},$$

this becomes

$$\begin{aligned}
&\sum_{i=1}^{N} \phi(T + v_i/\sigma)\left[\frac{v_i}{T\sigma} + \frac{1}{T^2}\left(1 - 3\frac{v_i}{T\sigma}\right)\right] \approx \\
&\quad \sum_{i=1}^{N} \phi(T - v_i/\sigma)\left[\frac{v_i}{T\sigma} - \frac{1}{T^2}\left(1 + 3\frac{v_i}{T\sigma}\right)\right]
\end{aligned}$$

Using the equality

$$\phi(T+v_i/\sigma) \pm \phi(T-v_i/\sigma) = \phi(T)\cdot e^{-v_i^2/2\sigma^2}(e^{-Tv_i/\sigma} \pm e^{Tv_i/\sigma}),$$

we can write the previous expression as:

$$\sum_{i=1}^{N} e^{-v_i^2/2\sigma^2} v_i \cdot (T/\sigma)\left(e^{Tv_i/\sigma} - e^{-Tv_i/\sigma}\right)$$
$$= \sum_{i=1}^{N} e^{-v_i^2/2\sigma^2}\left(e^{Tv_i/\sigma} + e^{-Tv_i/\sigma}\right).$$

For $\sigma \to \infty$, and fixed vector $\boldsymbol{v}$ the factors $e^{-v_i^2/2\sigma^2} \to 1$ rapidly. What remains is an equation in $K = T/\sigma$. Solving for T as a function of σ thus proceeds by solving for the constant $K = T/\sigma$ first, followed by letting $T(\sigma) = K\cdot\sigma$. This concludes the proof. □

4 The hard thresholding case

An argument, totally similar to the one for Proposition 1, leads to a result for hard thresholding:

Proposition 3 *For a given vector $\boldsymbol{v}$ of fixed length N, and for increasing noise variance $\sigma^2 \to \infty$, the minimum risk hard threshold behaves as:*

$$\lambda \sim K(\sigma)\cdot\sigma^2, \tag{12}$$

where $K(\sigma)$ is the solution of the nonlinear equation

$$K = \frac{2}{\sigma^2}\frac{\sum_{i=1}^{N} v_i\left(e^{v_iK} - e^{-v_iK}\right)}{\sum_{i=1}^{N}\left(e^{v_iK} + e^{-v_iK}\right)}. \tag{13}$$

It can be verified that $K(\sigma) \to \infty$, so the optimal hard threshold grows essentially faster than the optimal soft threshold.

5 Minimum risk thresholds in Bayesian models and Besov spaces

The discussion in this section concentrates on the links between the new results in this paper and existing observations in Bayesian thresholding. Next, we also discuss the link between the result and settings for the universal threshold. Suppose that the noise-free data vector is an instance of a multivariate random variable $\boldsymbol{V}$, with constant variance $\sigma_v^2 = \mathrm{Var}(V_i)$ and let $U_i = V_i/\sigma_v$, then $\mathrm{Var}(U_i) = 1$. Equation (8) can be written in the form $\sum_{i=1}^{N} G(K\cdot V_i) = 0$, where $G(x) = x(e^x - e^{-x}) - (e^x + e^{-x})$. If K_1 is the solution of $\sum_{i=1}^{N} G(K_1\cdot U_i) = 0$, then, obviously $K = K_1/\sigma_v$ is the solution of $\sum_{i=1}^{N} G(K\cdot\sigma_v U_i) = 0$. Taking the expected values, we find that $EK = EK_1/\sigma_v$, so the expected minimum risk threshold satisfies

$$E\lambda = EK_1\cdot\frac{\sigma}{\sigma_v}\cdot\sigma. \tag{14}$$

This threshold is not the same as the minimum expected risk threshold, studied in Bayesian shrinkage procedures [10, 2, 11]. Indeed, in those procedures, the expected value over random vector $\boldsymbol{V}$ is taken *before* it is minimized. Nevertheless, the results reported in this paper can be seen as an explanation for the behavior of the optimal thresholds as observed in the Bayesian setting. Note that Equation (14) has been derived from expressions for $\sigma \to \infty$. In the Bayesian setting, the expected minimum ASE threshold, $E\lambda$, is, however, a function of both σ and σ_v. If σ_v is large, the assumptions in the proof of Proposition 1 are no longer valid. We then have the following result.

Proposition 4 *If $\boldsymbol{v} = \sigma_v \boldsymbol{u}$, and suppose all u_i are different from zero, then for $\sigma_v \to \infty$ and for fixed noise standard deviation σ, the minimum risk soft threshold behaves as*

$$\lambda(\sigma_v) \sim \frac{2\sigma}{N} \sum_{i=1}^{N} \phi\left(\sigma_v \cdot \frac{u_i}{\sigma}\right). \tag{15}$$

If $\boldsymbol{u}$ has $N_0 > 0$ zero entries, then $\lim_{\sigma_v \to \infty} \lambda(\sigma_v) = \sigma L$, with L the solution of (10).

From this result, it follows immediately that in a Bayesian setting, the expected value of this threshold behaves as $E_U \lambda(\sigma_v) \sim 2\sigma E_U \phi(\sigma_v U/\sigma)$. If the prior on U is a double exponential (also known as the Laplacian distribution), then

$$E_U \lambda(\sigma_v) \sim \sqrt{2} \cdot e^{-(\sigma/\sigma_v)^2} \cdot 2\left[1 - \Phi(\sigma/\sigma_v)\right] \cdot \frac{\sigma^2}{\sigma_v}. \tag{16}$$

Up to the factor $[1 - \Phi(\sigma/\sigma_v)]e^{-(\sigma/\sigma_v)^2}$, which quickly approaches the constant $1/2$, this confirms the behavior of several Bayesian thresholds for $\sigma_v \to \infty$.
Proof of Proposition 4:
Equation (5) can be rewritten as

$$\begin{aligned} \frac{\lambda}{\sigma} \sum_{i=1}^{N} \left[1 - \Phi\left(\frac{\lambda - |u_i|\sigma_v}{\sigma}\right)\right] + \left[1 - \Phi\left(\frac{\lambda + |u_i|\sigma_v}{\sigma}\right)\right] \\ = \sum_{i=1}^{N} \phi\left(\frac{\lambda - |u_i|\sigma_v}{\sigma}\right) + \phi\left(\frac{\lambda + |u_i|\sigma_v}{\sigma}\right), \end{aligned} \tag{17}$$

where λ is now a function of σ_v. Let $L = \lim_{\sigma_v \to \infty} \lambda/\sigma$ and let $a_i = \lim_{\sigma_v \to \infty} (\lambda - |u_i|\sigma_v)/\sigma$, then taking limits on both sides of the equation, leads to

$$L \sum_{i=1}^{N} [1 - \Phi(a_i)] + L N_0 [1 - \Phi(L)] = \sum_{i=1}^{N} \phi(a_i) + N_0 \phi(L),$$

where N_0 is the number of zeros in $\boldsymbol{u}$. The left hand side is of the form $L \times R$, where R is finite for any values of L and a_i. The right hand side is finite for all values of L and a_i. As a consequence, L must be finite, and so, by definition of a_i, $a_i = -\infty$ if $u_i \neq 0$ and $a_i = L$ otherwise. The case $N_0 > 0$ now proceeds trivially to Equation (10). The case $N_0 = 0$ results in $L = 0$. As a consequence, for $\sigma_v \to \infty$, we have that in all arguments of ϕ and Φ, $\lambda \ll |u_i|\sigma_v$, leading immediately to Equation (15). □

In Equations (14) and (16) we have investigated the limit behavior of $E\lambda$ for $\sigma \to \infty$ and for $\sigma_v \to \infty$ respectively. Both equations are limit cases of the following general observation.

Proposition 5 *The expected minimum risk soft threshold $E\lambda(\sigma, \sigma_v)$ in a Bayesian model is of the form*

$$E\lambda(\sigma, \sigma_v) = S(\sigma/\sigma_v) \cdot \sigma. \tag{18}$$

Proof: From Equation (17), it follows that $E\lambda(\sigma, \sigma_v)$ can only depend on σ_v through the ratio σ/σ_v. Since, for reasons of symmetry, $E\lambda(k\sigma, k\sigma_v) = kE\lambda(\sigma, \sigma_v)$, this implies Eq. (18). □

For $S(\sigma/\sigma_v)$ to be a constant, like in the universal threshold, it is necessary that $E\lambda(k\sigma, \sigma_v) = kE\lambda(\sigma, \sigma_v) = E\lambda(k\sigma, k\sigma_v)$. This is only possible if $\sigma_v = \infty$, i.e., if the prior model is a heavy tailed distribution. As a matter of fact, this situation is intrinsically assumed in an asymptotic analysis of the minimum risk threshold for $N \to \infty$ [7]. Typical signals belong to Besov balls, i.e., function spaces whose members can be represented sparsely in a wavelet decomposition. Without going into the mathematical details, this sparsity ensures that if the number of observations from the signal increases, the dominant part of the information remains concentrated in a limited number of coefficients. That number of significant coefficients typically grows as $\log N$ if $N \to \infty$. As a consequence, those coefficients become more and more significant if $N \to \infty$. This can be seen as a case where $\sigma \to 0$, or, equivalently, $\sigma_v \to \infty$, and so, the minimum risk threshold is found to behave (up to a constant depending on the smoothness of the signal) as the universal threshold for $N \to \infty$. [7].

6 An illustration

As the results presented in this paper are asymptotic, it is interesting to have a look at the behavior of the minimum risk thresholds for finite values of the noise standard deviation. (By finite values, we mean intermediate values, i.e., values away from zero and infinity.) Figure 1 depicts a typical set of sparse data. It was generated as 2048 observations v_i from a model which is mixture of a point mass at zero and a Laplacian distribution with standard deviation equal to $10\sqrt{2}$. We can write $V_i \sim (1-p)\delta_0 + p$ Laplace. The mixture parameter p, controlling the sparsity, was set to $p = 1/50$. Figure 2(a) is a plot of the minimum risk threshold as a function of the noise standard deviation σ. Although the quadratic behavior is not very visible from this plot, it becomes much more prominent in a plot of the derivative of the minimum risk threshold, see Figure 2(b). The dashed line in this plot corresponds to the line $2K\sigma$, with K defined in (8). Although this asymptotic line is reached for large values of σ, an intermediate value of σ is already large compared to the many zeros in the noise-free data. For these zeros, the approximations made in the proof of Proposition 1 are valid. This results in a quadratic behavior, whose derivative is a line parallel to the asymptotic result.

7 Conclusion

This paper has investigated the behavior of minimum risk thresholds in the presence of intense noise, i.e., for $\sigma \to \infty$ and also in the presence of little noise, i.e., for $\sigma \to 0$. For $\sigma \to \infty$, the minimum risk threshold behaves as $\lambda \sim \sigma^2$ (Proposition 1), while for $\sigma \to 0$, we have that $\lambda \sim \sigma$ (Proposition 2). Proposition 3 proves that minimum risk hard thresholds behave essentially different from minimum risk soft thresholds.

The results have interesting interpretations with respect to existing thresholds: first, they explain the behavior of certain Bayesian thresholds. Secondly, they also can be brought in perspective to the

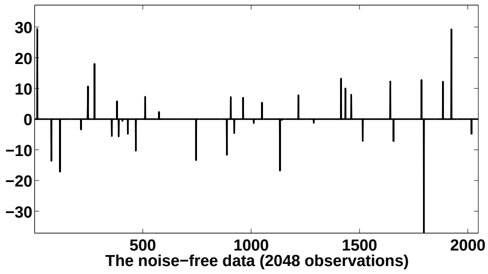

Figure 1: A sparse set of test data, generated as an instance from a mixture of a point mass at zero and a Laplacian distribution.

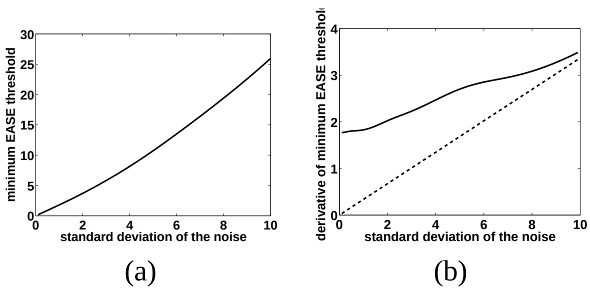

Figure 2: (a) Plot of minimum risk threshold λ as a function of the noise standard deviation σ. (b) Plot of the derivative $\lambda'(\sigma)$ of the function in Figure (a). The dashed line is $2K\sigma$, which is the asymptotic result from Proposition 1.

universal threshold and the assumption of sparsity in Besov spaces. As for the Bayesian framework, the elaboration of Proposition 4 with a double exponential prior leads to $E\lambda \sim \sqrt{2}\sigma^2/\sigma_v$ if the variance of the prior $\sigma_v^2 \to \infty$. A similar behavior follows from exact calculations in certain Bayesian frameworks and is taken as a heuristic rule in others. This heuristic approach is now confirmed by the results in this paper.

The results in this paper are limited to individual coefficient thresholding. More sophisticated thresholding, such as thresholds for joint neighborhoods [11] of coefficients, can obtain better quality. It is an interesting question if similar results hold for such a thresholding procedure.

References

[1] A.G. Bruce and H.-Y. Gao. Waveshrink with firm shrinkage. *Statistica Sinica*, 4:855–874, 1997.

[2] S. G. Chang, B. Yu, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000.

[3] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[4] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90:1200–1224, 1995.

[5] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26:879–921, 1998.

[6] H.-Y. Gao. Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4):469–488, December 1998.

[7] M. Jansen and A. Bultheel. Asymptotic behavior of the minimum mean squared error threshold for noisy wavelet coefficients of piecewise smooth signals. *IEEE Transactions on Signal Processing*, 49(6):1113–1118, June 2001.

[8] M. Jansen, M. Malfait, and A. Bultheel. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44, January 1997.

[9] I. M. Johnstone and B. W. Silverman. Needles and straw in haystacks: empirical bayes estimates of possibly sparse sequences. *Annals of Statistics*, 32(4):1594–1649, 2004.

[10] F. Ruggeri and B. Vidakovic. A Bayesian decision theoretic approach to the choice of thresholding parameter. *Statistica Sinica*, 9:183–197, 1999.

[11] L. Sendur and I. W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising. *IEEE Transactions on Signal Processing*, 50(11):2744–2756, 2002.