Statistica Sinica **14**(2004), 103-125

Γ-MINIMAX WAVELET SHRINKAGE: A ROBUST INCORPORATION OF INFORMATION ABOUT ENERGY OF A SIGNAL IN DENOISING APPLICATIONS

Claudia Angelini and Brani Vidakovic

Istituto per le Applicazioni del Calcolo -Sezione di Napoli
and Georgia Institute of Technology

Abstract: In this paper we propose a method for wavelet-filtering of noisy signals when prior information about the L^2-energy of the signal of interest is available. Assuming the independence model, according to which the wavelet coefficients are treated individually, we propose a level dependent shrinkage rule that turns out to be the Γ-minimax rule for a suitable class, say Γ, of realistic priors on the wavelet coefficients.

The proposed methodology is particularly well suited for denoising tasks where signal-to-noise ratio is low, and it is illustrated on a battery of standard test functions. Performance comparisons with some others methods existing in the literature are provided. An example in atomic force microscopy (AFM) is also discussed.

Key words and phrases: Atomic force microscopy, bounded normal mean, Γ-minimaxity, shrinkage, wavelet regression.

1. Introduction

1.1. Γ-minimax theory

Γ-minimax theory, originally proposed in Robbins (1951), deals with the problem of selecting decision rules. Philosophically, the Γ-minimax criterion is situated in between the Bayes paradigm, which selects procedures that work well "on average", and the minimax paradigm, which guards against catastrophic outcomes, however unlikely. It has evolved from seminal papers in the fifties (Robbins (1951) and Good (1952)) and early sixties, through an extensive research on foundations and parametric families in the seventies, to a branch of Bayesian robustness theory, in the eighties and nineties. In this latter setup a comprehensive discussion of the Γ-minimax can be found in Berger (1984, 1985). A recent interest in the Γ-minimax theory has been pointed out, for example, in the work of Vidakovic (2000) and Noubiap and Seidel (2001).

The Γ-minimax paradigm involves incorporating the prior information about the statistical model, not via a single prior distribution, but rather by a family of plausible priors, say Γ. Such "family of priors" elicitations are often encountered

in practice. Given this family of priors, the decision maker looks for selecting an action that is optimal with respect to the least favorable prior in the family.

Inference of this kind is often interpreted in terms of a game. Suppose that the decision maker is Player II. Player I, an intelligent opponent to Player II, chooses the "least favorable" prior from the family Γ. Player II chooses an action that will minimize his loss, irrespective of what prior Player I has selected. The action of Player II is referred to as the Γ-minimax action.

The decision maker's actions are functions of observed data, and such functions are often called decision *rules.* In many models of interest, the exact Γ-minimax rules are intractable or, at best, computationally involved.

Formally, let $\mathcal{D}$ be the set of all decision rules and Γ be a family of prior distributions on the parameter space Θ. A rule δ^* is Γ-minimax if

$$\inf_{\delta\in\mathcal{D}}\sup_{\pi\in\Gamma} r(\pi,\delta) = \sup_{\pi\in\Gamma} r(\pi,\delta^*),$$

where $r(\pi,\delta) = E^\theta\left[E_\theta^{X|\theta}\mathcal{L}(\theta,\delta)\right] = E^\theta R(\theta,\delta)$ is the Bayes risk under the loss $\mathcal{L}(\theta,\delta)$. Note that when Γ is the set of all priors, the Γ-minimax rule coincides with minimax rule; when Γ is a singleton, then the Γ-minimax rule coincides with Bayes rule. The family Γ of plausible priors is usually given by parametric form or under generalized moment conditions. When the decision problem, viewed as a statistical game, has a value, then the Γ-minimax solution coincides with the Bayes rule with respect to the least favorable prior (see Berger (1985, Chap. 4)).

1.2. Wavelet shrinkage

In the present paper we consider a Γ-minimax approach to the classical nonparametric regression problem

$$Y_i = f(t_i) + \sigma\varepsilon_i, \quad i = 1,\dots,n, \tag{1}$$

where t_i, $i = 1,\dots,n$, is a deterministic equispaced design in $[0,1]$, the random errors ε_i are i.i.d. standard normal random variables and the noise level σ^2 may, or may not, be known. The interest is to recover the function f using the observations Y_i. Additionally, we assume that the unknown signal f has a bounded L^2-energy, hence it assumes values from a bounded interval. After applying a linear and orthogonal wavelet transformation, model (1) becomes

$$\begin{aligned} c_{J,k} &= \theta_{J,k} + \sigma\epsilon_{J,k}, \quad k = 0,\dots,2^J-1, \\ d_{j,k} &= \theta_{j,k} + \sigma\epsilon_{j,k}, \quad j = J,\dots,N-1,\ k = 0,\dots,2^j-1, \end{aligned} \tag{2}$$

where $d_{j,k}$ (c_{Jk}), $\theta_{j,k}$ and $\epsilon_{j,k}$ are the wavelet (scaling) coefficients (at resolution j and position k) of Y, f and ε, respectively; J and N are the coarsest and finest level of the wavelet decomposition. Since the orthogonality of the wavelet

transformation preserves the conditional independence of the wavelet coefficients, the stochastic structure of the noise and the bound on the energy, model (2) can be written as

$$[d|\theta] \sim \mathcal{N}(\theta, \sigma^2), \tag{3}$$

where, due to the independence of the coefficients, we have omitted the double indices j, k. The prior information on the energy bound implies that the unknown wavelet coefficient θ would assume its values from a bounded parameter space, say $\Theta = [-m, m]$.

Thresholding or shrinkage rules in the wavelet domain have been often proposed in the literature to estimate the location parameter θ in model (3) when no additional information on the parameter space Θ is available; see for example, Donoho and Johnstone (1994b and 1995) and related papers, Antoniadis (1997), Härdle, Kerkyacharian, Picard and Tsybakov (1998) (in the minimax setup), Abramovich, Sapatinas and Silverman (1998), Abramovich and Sapatinas (1999), Chipman, Kolaczyk and McCulloch (1997), Clyde, Parmigiani and Vidakovic (1998), Clyde and George (1999, 2000), Figueiredo and Novak (2001), Vidakovic (1998) and Vidakovic and Ruggeri (2001) (in the Bayesian setup). Recently wavelet methods based on block-shrinkage strategy have showed excellent performances, see Abramovich Besbeas and Sapatinas (2002), Cai (1999), Cai and Silverman (2001), Hall, Kerkyacharian and Picard (1998, 1999), and Hall, Penev, Kerkyacharian and Picard (1997). For a comprehensive discussion of the status of the art in wavelet methods for nonparametric regression problems we refer to Antoniadis, Bigot and Sapatinas (2001) where most of the methods are described and numerically compared.

1.3. Bayesian model in the wavelet domain

Over the last decade Bayesian methods in the wavelet domain have received considerable attentions, an extensive review can be found in the book by Vidakovic (1999). Informally speaking, a shrinkage rule in the wavelet domain replaces the observed empirical wavelet coefficients d with their shrunken versions $\hat{\theta} = \delta(d)$. The form of the particular rule $\delta(\cdot)$ characterizes the performance of the estimate.

Bayesian models in the wavelet domain have showed to be capable of incorporating prior information about the unknown signal such as smoothness, periodicity, sparseness, self-similarity and, for some particular basis (Haar), monotonicity (see for example Abramovich, Sapatinas and Silverman (1998), Berliner, Wikle and Milliff (1999)). This is usually achieved by eliciting a single prior distribution π on the space of parameters Θ, and then choosing the estimator $\hat{\theta} = \delta(d)$ that minimizes the Bayes risk with respect to the given prior.

It is well known that most of the noiseless signals encountered in practical applications have (for each resolution level) empirical distributions of wavelet

coefficients centered around zero and peaked at zero. A realistic Bayesian model that takes into account this prior knowledge should consider a prior distribution, say π, that produces a reasonable agreement with observations. A realistic prior distribution on the wavelet coefficient θ is given by

$$\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)\xi(\theta), \tag{4}$$

where δ_0 is a point mass at zero, ξ is a symmetric and unimodal distribution on the parameter space Θ and ϵ_0 is a fixed parameter in $[0, 1]$, usually level dependent, that regulates the amount of shrinkage for values of d close to 0. Prior models for wavelet coefficients as (4) have been indicated in the early 1990's by Berger and Müller (personal communication), considered in Abramovich, Sapatinas and Silverman (1998), Abramovich and Sapatinas (1999), Vidakovic (1998) and Vidakovic and Ruggeri (2001), among others.

It is however clear that specifying a single prior distribution π on the parameter space Θ can never be done exactly. Indeed the prior knowledge of real phenomena always contains some kind of approximation such that several types of distributions can match the prior belief, meaning that on the basis of the partial knowledge about the signal, it is possible to elicit only a family, say Γ, of plausible priors. In a robust Bayesian point of view the choice of a particular rule δ should not be influenced by the choice of a particular prior, as long as it is in agreement with our prior belief. Several approaches have been considered for measuring the robustness of a specific rule, Γ-minimax being one compromise.

In this paper we would like to incorporate prior belief on the boundedness of the energy of the signal (the L_2-norm of the regression function). The prior information on the energy bound often exists in real life problems and, as observed in Section 1.2, it can be modelled by the assumption that the parameter space Θ is bounded. Estimation of a bounded normal mean has been considered in Bickel (1981), Casella and Strawderman (1981), Donoho, Liu and MacGibbon (1990), Gatsonis, MacGibbon and Strawderman (1987), Marchand and Perron (2001), Miyasawa (1953) (in the minimax setup) and in Vidakovic and DasGupta (1996) (in the Γ-minimax setup). It is however well known that estimating a bounded normal mean represents a difficult task. In our context, if the structure of the prior (4) can be supported by the analysis of the empirical distribution of the wavelet coefficients, the precise elicitation of the distribution ξ cannot be done without some kind of approximation. Of course, when prior knowledge on the energy bound is available, then any symmetric and unimodal distribution supported on the bounded set, say $[-m, m]$, can be a possible candidate for $\xi(\theta)$.

Let Γ denote the family

$$\Gamma = \{\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)q(\theta), q(\theta) \in \Gamma_{SU[-m,m]}\}, \tag{5}$$

where $\Gamma_{SU[-m,m]}$ is the class of all symmetric and unimodal distributions supported on $[-m, m]$ and δ_0 is point mass at zero. One way to handle the incomplete specification of the prior is through the following model

$$\begin{cases} d|\theta & \sim \mathcal{N}(\theta, 1) \\ \theta & \sim \pi(\theta) \in \Gamma \\ \mathcal{L}(\theta, \delta) = (\theta - \delta)^2 & \text{Squared Error Loss.} \end{cases} \tag{6}$$

We stress that no generality is lost by assuming that $\sigma^2 = 1$.

The paper is organized as follows. Section 2 contains mathematical aspects and results concerning the Γ-minimax rule and its application to model (2). An exact risk analysis of the rule is discussed in Section 3. Section 4 proposes a sensible elicitation of hyper-parameters defining the model. Performance of the shrinkage rule in the wavelet domain and application to a data set are given in Section 5. In Section 6 we summarize the results and provide discussion on possible extensions. Proofs are deferred to the Appendix.

2. Γ-minimax Shrinkage

In this section we extend the result of Vidakovic and DasGupta (1996) to the class of priors defined in (5). We show that for m small the least favorable distribution is the uniform on $[-m, m]$ contaminated by a prior mass at zero. The corresponding Γ-minimax rule is a shrinkage rule that is applied in the context of wavelet regression.

Theorem 2.1. *Under* (6), *where* Γ *is defined in* (5), *we have*

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}} r(\pi, \delta).$$

The associated Γ*-minimax rule is the Bayes rule with respect to the least favorable prior*

$$\pi(\theta) = (\epsilon_0 + (1 - \epsilon_0)\alpha_0)\delta_0 + (1 - \epsilon_0) \sum_{k=1}^{p} \alpha_k \mathcal{U}[-m_k, m_k], \tag{7}$$

where $\alpha_k = \alpha_k(\epsilon_0) \geq 0$, $\sum_{k=0}^{p} \alpha_k = 1$, $m_k = m_k(\epsilon_0)$ *s.t.* $0 < m_1 < m2 < \ldots < m_p = m$. *The corresponding* Γ*-minimax rule is given by*

$$\delta_\pi(d) = d - \frac{(\epsilon_0 + (1 - \epsilon_0)\alpha_0) d\phi(d) - (1 - \epsilon_0) \sum_{k=1}^{p} \frac{\alpha_k}{2m_k} \left(\phi(d + m_k) - \phi(d - m_k)\right)}{(\epsilon_0 + (1 - \epsilon_0)\alpha_0)\phi(d) + (1 - \epsilon_0) \sum_{k=1}^{p} \frac{\alpha_k}{2m_k} \left(\Phi(d + m_k) - \Phi(d - m_k)\right)}, \tag{8}$$

where ϕ *and* Φ *denote the density and the cumulative distribution function of the standard normal random variable and* $\mathcal{U}$ *denotes the uniform distribution.*

Moreover, for any ϵ_0 there exists $m^* = m^*(\epsilon_0)$ such that, for any $m \leq m^*$, the least favorable prior is of the form

$$\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)\mathcal{U}[-m, m] \tag{9}$$

and the Γ-minimax rule is given by

$$\delta_\pi(d) = d - \frac{\epsilon_0 d\phi(d) - \frac{1-\epsilon_0}{2m}\left(\phi(d+m) - \phi(d-m)\right)}{\epsilon_0 \phi(d) + \frac{1-\epsilon_0}{2m}\left(\Phi(d+m) - \Phi(d-m)\right)}. \tag{10}$$

Remark 2.1. The value of $m^*(\epsilon_0)$ such that (9) holds is the largest value of m for which the maximum of $(1/z)\int_0^z R(v, \delta_\pi)dv$ is achieved at $z = m$, where $R(\cdot, \delta_\pi)$ represents the frequentist risk of the rule δ_π.

Table 1. Values of $m^*(\epsilon_0)$ for which the least favorable distribution in Γ is given by (9), and the corresponding values $m_1(\epsilon_0)$ of the support of the second uniform distribution when $m = m^*$ and $\sigma = 1$.

ϵ_0	0.	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$m^*(\epsilon_0)$	2.5323	2.862	3.606	4.171	4.346	4.416	4.442	4.446	4.446	4.447	4.448
$m_1(\epsilon_0)$	0.	0.	1.595	2.784	3.166	3.395	3.573	3.733	3.888	4.053	4.284

For m exceeding m^*, (9) is no longer the least favorable prior; the least favorable prior will contain other uniform distributions supported on $[-m_k, m_k]$ as in (7). Numerical work, analogous to that developed in Vidakovic and DasGupta (1996), can give an accurate approximation of the parameters α_k and m_k in (7), for any given ϵ_0 and m, however the exact values of the parameters are still unknown. Table 1 shows values of $m^*(\epsilon_0)$ for several choices of ϵ_0, and the corresponding values of $m_1(\epsilon_0)$ at which the additional uniform component of the prior is supported. The MATHEMATICA package was used in computing the values in Table 1, and in finding, for each ϵ_0, the largest m for which $\bar{R}$ defined in (17) is maximized at $z = m$. Computations show that m^* increases with ϵ_0 according to the analogous result obtained in DasGupta and Delampady (1994). However, when ϵ_0 is larger than about 0.7, the computed value of $m^*(\epsilon_0)$ tends to be less accurate since the risk function $\bar{R}$ becomes very flat in the neighborhood of m. Comparing the results obtained here for the family Γ in (5) with related results obtained in Vidakovic and DasGupta (1996), where the case $\epsilon_0 = 0$ has been considered, we see that for ϵ_0 exceeding 0.1 no additional point mass at zero is added with increasing m, i.e., α_0 in (8) is zero. However, for small values of ϵ_0 (for example $\epsilon_0 = 0.05$), α_0 is not zero. Indeed Vidakovic and DasGupta (1996) prove that, for the class of symmetric and unimodal distributions, the point mass at zero appears in the least favorable distribution when m increases, but with weight about 0.07.

When m is larger than m^* then the number of uniform distributions in the least favorable prior (7) increases, and the rule (8) can be only numerically evaluated. However the computational cost of computing optimal parameters in (8) for m large is high because of the multiplicity of parameters and because they depend on ϵ_0. For this reason (8) is not well suited for unrestricted applications.

In Bickel (1981), the problem of estimating a bounded mean of a normal distribution is considered within the minimax setup. It has been proved that when m increases, the weak limit of the least favorable priors (when taking the supremum of the risk with respect to the class of all priors), rescaled to the interval $[-1, 1]$, is $g_1(\theta) = \cos^2(\pi\theta/2)\mathbf{1}(|\theta| \leq 1)$. This fact implies that when m is large the least favorable prior in $\Gamma_{SU[-m,m]}$ is close to $g_m(\theta) = (1/m)\cos^2(\pi\theta/2m)\mathbf{1}(|\theta| \leq m)$. Applying this result in our context, we have that the least favorable prior in Γ is close to

$$\pi(\theta) = \epsilon_0\delta_0 + (1-\epsilon_0)\frac{1}{m}\cos^2(\frac{\pi\theta}{2m})\mathbf{1}(|\theta| \leq m). \tag{11}$$

The corresponding Bayes rule does not have a simple analytical form, since it involves the evaluation of the marginal distribution of d, and needs to be numerically computed. In this case also the prohibitive computational cost of would make the use of this rule not competitive in practical application. Figure 1 shows rule (10) (solid line) and the Bayes rule corresponding to Bickel's prior (11) (dashed line). Both rules are evaluated for $m = 20$, $\epsilon_0 = 0.8$ and $\sigma^2 = 1$. We observe that, since the value $m = 20$ exceeds m^* in Table 1, then (10) is no longer the exact Γ-minimax rule. Nevertheless, the two rules depicted in Figure 1 are very close. More discussion on this issue is deferred to Section 6.

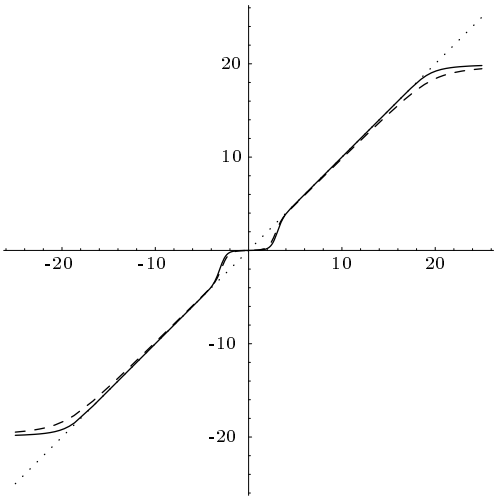

Figure 1. Bayes shrinkage rule (10) (solid line) and the Bayes rule corresponding to the prior (11) (dashed line), evaluated for $m = 20$, $\epsilon_0 = 0.8$ and $\sigma^2 = 1$. The 45° dotted line is superimposed as reference.

2.1. Γ-minimax shrinkage in the wavelet domain

After a proper rescaling, (10) can be applied to model (2) as rationale for the Bayesian wavelet shrinkage rule

$$\hat{\theta}_{jk} = d_{jk} - \sigma \frac{\epsilon_0 \frac{d_{jk}}{\sigma}\phi\left(\frac{d_{jk}}{\sigma}\right) - \frac{(1-\epsilon_0)\sigma}{2m}\left[\phi\left(\frac{d_{jk}+m}{\sigma}\right) - \phi\left(\frac{d_{jk}-m}{\sigma}\right)\right]}{\epsilon_0\phi\left(\frac{d_{jk}}{\sigma}\right) + \frac{(1-\epsilon_0)\sigma}{2m}\left[\Phi\left(\frac{d_{jk}+m}{\sigma}\right) - \Phi\left(\frac{d_{jk}-m}{\sigma}\right)\right]} \quad (12)$$

which is, from Theorem 2.1, the Γ-minimax rule under the assumption that m/σ is 'small' (Table 1 shows the bounds for what is meant by 'small'). For practical applications, we recommend using (12) for any value of m. We stress that for m/σ exceeding the critical values in Table 1, the rule (12) is not an exact Γ-minimax rule, however it remains close to the Γ-minimax rule and is computationally inexpensive. We observe that in order to apply (12), two hyper-parameters (ϵ_0 and m) can be conveniently chosen. The elicitation of such parameters is discussed in Section 4.

In order to complete model (2), we can place a diffuse prior on a scaling coefficient, i.e., $\theta_{Jk} \sim N(0, \tau^2)$, with $\tau^2 \to \infty$. When combining with the stochastic model (2), the resulting estimate $\hat{\theta}_{Jk} = c_{Jk}$ leaves the scaling coefficients unchanged.

Throughout the paper we estimate, according to Donoho and Johnstone (1994b), the noise level σ by

$$\hat{\sigma} = \frac{\text{median}(\{|\hat{d}_{N-1,k}| : k = 0, 1, \ldots, 2^{N-1} - 1\})}{0.6745}, \quad (13)$$

where $N = \log_2 n$ represent the finest level of wavelets detail, and we choose the primary resolution level J as

$$J(n) = \text{floor}(\log_2(\log(n))) + 1 \quad (14)$$

according to the asymptotic considerations given in Chapter 10 of the book by Härdle et al. (1998).

3. Risk Analysis of the Rule

Exact risk analysis of any proposed rule has received considerable attention since it allows for comparison of different wavelet-shrinkage methods. When the rule is given in a simple form, then the exact risk analysis can be carried out explicitly. For instance, Donoho and Johnstone (1994a) and Bruce and Gao (1996) provide exact risk analyses for hard and soft thresholding under squared error loss. Gao and Bruce (1997) give the rationale for introducing the "firm" or

"semi-soft" thresholding utilizing exact risk analysis. In our context the form of shrinkage rule (10) is more complex and the exact risk analysis had to be done numerically. The goal of our analysis is to explore robustness in risk, bias, and variance when the prior hyper-parameters change.

Computations performed in the software package MATHEMATICA produced Figures 2 and 3. We briefly describe the numerical findings expressed in the figures. As depicted in Figure 2(a), for $m = 3$, the shrinkage rules follow a desirable shrinkage pattern, differing for small values of d. For large values of ϵ_0, rules heavily shrink small values of d. The rules generally remain close to d for intermediate values of d. When $|d|$ exceeds m, rules remain bounded by $\pm m$, reflecting the prior knowledge that the signal energy is bounded. The parameter m controls the largest amplitude allowed in the wavelet coefficient corresponding to the signal and is directly proportional to the energy bound. We observe that, given m, the amount of shrinkage essentially depends on the choice of ϵ_0.

In Figure 2(b) the risks of rules in Figure 2(a) are presented. One can notice an obvious trade-off in the risk performance for small and large values of θ. When ϵ_0 is large the risk remains close to 0, for θ small; the flattest risk curve in Figure 2(b) corresponds to $\epsilon_0 = 0.1$.

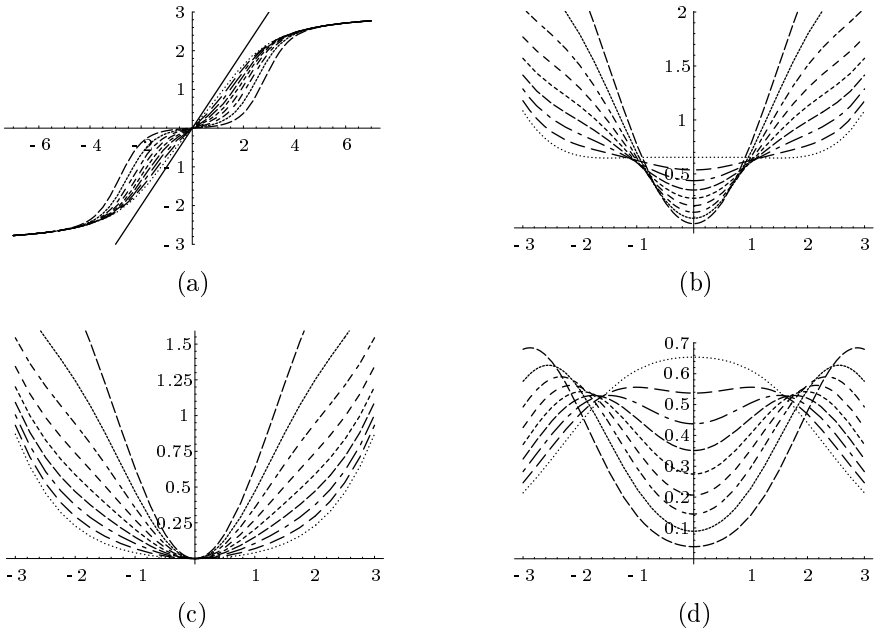

Figure 2. (a) Γ-minimax rules (10) for $m = 3$ and ϵ_0 ranging from 0.1 (upper envelope function) to 0.9 (lower envelope function); (b) Exact risks for rules in (a); (c) Bias^2 for rules in (a); (d) Variances for rules in (a).

The bias-squared, depicted in Figure 2(c) is uniformly (in θ) increasing when ϵ_0 increases. The lowest bias-squared curve correspond to $\epsilon_0 = 0.1$.

The variance functions, like the risks, exhibit a trade-off behavior for different values of ϵ_0. More precisely, large values of ϵ_0 produce variance values close to 0. On the other hand the variances increase, for the same value of ϵ_0, when θ increases. This behavior is illustrated in Figure 2(d).

Figure 3 describes the shape and the risk behavior of Γ-minimax rules for ϵ_0 fixed ($\epsilon_0 = 0.8$) and m ranging from 0.5 to 4 with the step 0.5. The panel (a) depicts forms of shrinkage rules. Note an overall heavy shrinkage for small values of m, and a "two-fold" shrinkage for m large (the curve closest to d). Inspection of this figure implies that the elicitation of m should be carefully considered since it can substantially influence the performance of the estimator. The exact risk behavior is similar for all rules at small values of θ, but the risk rapidly increases if m is small.

The bias-squared exhibits uniform monotonicity with respect to m. When m is increasing, the squared biases decrease in θ.

Finally, the variances depicted in Figure 3(d) indicate that values of θ close to m produce the most variability. The shape of variance functions is similar; the largest function corresponds to $m = 4$.

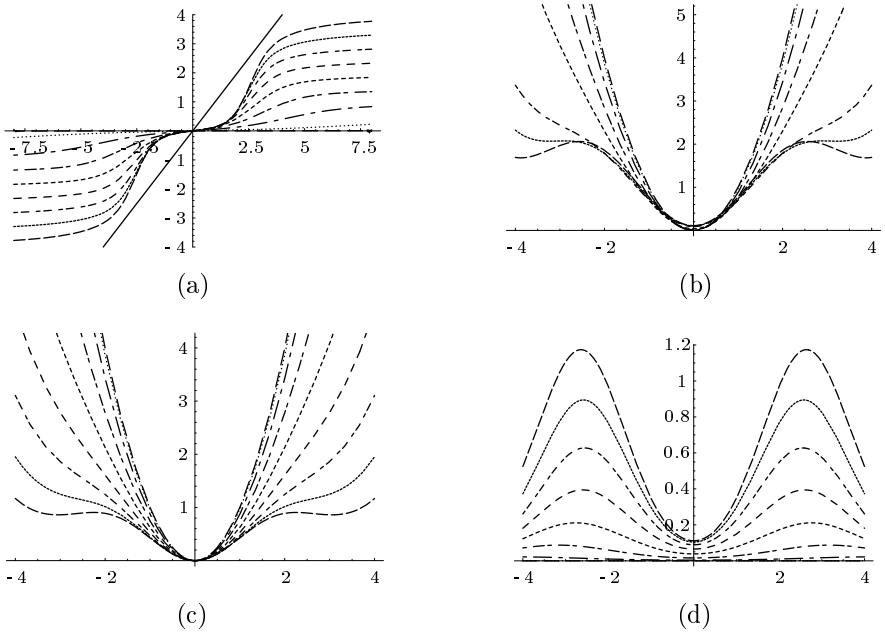

Figure 3. (a) Γ-minimax rules (10) for $\epsilon_0 = 0.8$ and m ranging from 0.5 to 4; (b) Exact risks for rules in (a); (c) Bias^2 for rules in (a); (d) Variances for rules in (a).

4. Elicitation of Parameters

The shrinkage rule (12) depends on the choice of the hyper-parameters ϵ_0 and m that should be carefully elicited in order to achieve good performance. Usually, the elicitation of hyper-parameters is one of the major issues in the Bayesian analysis and is carried out by taking into account available prior information. In our case such information concerns the smoothness, the sparseness, periodicity, self-similarity, as well as the energy, of the unknown signal. In this Section we propose a level dependent choice of hyper-parameters that is guided by considerations on the exact risk properties and on the shape of the shrinkage rule and that turns to be effective in our simulations.

It has been demonstrated that ϵ_0, the weight of the point mass at zero in the class Γ, regulates the amount of shrinkage at zero. This weight should depend on the prior information about the smoothness. It should be close to 1 at the finest level of detail and close to zero at coarse levels. However, the analysis of the exact risk shows that the shrinkage rule (10) is robust with respect to the choice of ϵ_0, at least for the values of ϵ_0 between 0.6 and 0.95. For practical purpose we propose an automatic choice of ϵ_0, one that is considered in Vidakovic and Ruggeri (2001). Level-dependent values of ϵ_0 are defined as

$$\epsilon_0(j) = 1 - \frac{1}{(j - J + 1)^\gamma}, \quad J \leq j \leq N - 1, \tag{15}$$

where J represents the coarsest level in the wavelet transformation and, in absence of additional information, the default value $\gamma = 2$ is considered. Since for smoother functions the wavelet coefficients decay more rapidly, the hyperparameter γ should increase with an increase in smoothness.

The sensitivity analysis on the performance of the estimator respect to the parameter γ is discussed in Section 5.

The elicitation of the hyper-parameter m requires more detailed discussion since it has been noticed that the choice of m can substantially influence performance of the estimator. In fact the Γ-minimax rule δ_π in Theorem 2.1 is sensitive to m and considerably more so than with respect to ϵ_0 (since the number of uniforms in the least favorable priors depends on m, while ϵ_0 influences the values of the parameters α_k and m_k). First, we already mention in Section 2.1 that the shrinkage rule is given by (12) is the Γ-minimax rule under the assumption that $m/\sigma < m^*(\epsilon_0)$. However, for practical application, we can use (12) for any m/σ.

Second, the elicitation of m (also if we limit attention to the shrinkage rule (12)) bears more influence in the performance of the estimator. This can be seen by inspecting the exact risk as a function of m.

In this paper we propose a level-dependent choice of m that depends only on noisy data. In fact, we estimate the size of the support of the wavelet coefficient

at each level by

$$m(j) = \max_k(|d_{j,k}|). \tag{16}$$

We stress that such a choice of m gives raise to a scale invariant empirical Bayesian shrinkage rule, which is particular interesting to work with in practical applications as observed in Figueidero and Nowak (2001). When the prior information about the distribution of energy content among the scales is available, then this information can be incorporated by magnitudes of the ratios $m(j)/\sigma$. A small $m(j)/\sigma$ would reflect the prior belief that the level j contributes little to the energy of the signal, while larger values would convey the information that the level is significant, in the sense that energetic feartures of the signal "live" in that particular scale.

5. Applications

In order to investigate the finite sample size performance of the Γ-minimax estimator (12) we carried out an extensive simulation study. We compared the proposed estimator with various estimators existing in the literature. In particular we consider the term-by-term Bayesian estimator *Bams* of Vidakovic and Ruggeri (2001), the classical term-by-term estimators *VisuShrink* of Donoho and Johnstone (1994b) and *Hybrid-SureShrink* of Donoho and Johnstone (1995), the scale invariant term-by-term Bayesian *ABE* method of Figueiredo and Nowak (2001), the "leave-out-half" version of the *Cross-Validation* method of Nason (1996), the term-by-term *False Discovery Rate* method of Abramovich and Benjamini (1995), and finally *NeighCoeff* of Cai and Silverman (2001) and *BlockJS* of Cai (1999) which represent classical estimators that, for achieving a better performance, incorporate the blocking procedure. Note that, for excellent numerical performance, we consider the *VisuShrink* and the "leave-out-half" version of the *CrossValidation* methods with the hard threshold and the *BlockJS* with the option 'Agument' (see Antoniadis, Bigot and Sapatinas (2001)).

We considered several standard test functions (Wave, Blocks, HeaviSine, Doppler, Corner, Blip, Angles, Parabolas and Time-Shifted Sine) which represent typical examples of signals encountered in practical applications. For each test function, $M = 200$ samples were generated by adding independent random noise $\varepsilon \sim N(0, \sigma^2)$ to $n = 256$ (small sample size), 512 (moderate sample size) and 1024 (large sample size) equally spaced points on [0,1]. The value of the noise level σ was taken to correspond to the values 3 (high noise level), 5 (moderate noise level) and 7 (low noise level) for the root signal-to-noise ratio (RSNR)

$$\mathrm{RSNR}(f,\sigma) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^n (f(t_i) - \bar{f})^2}}{\sigma}, \quad \text{where} \quad \bar{f} = \frac{1}{n}\sum_{i=1}^n f(t_i).$$

The goodness-of-fit for an estimator $\hat{f}$ of f was measured by its average mean squared error (AMSE) from the M simulations, defined as

$$AMSE(f) = \frac{1}{nM} \sum_{m=1}^{M} \sum_{i=1}^{n} (\hat{f}_m(t_i) - f(t_i))^2.$$

Symmlet 8-tap filter was used for all signals, except for Blocks where the Haar filter was considered. The hyper-parameters of the Γ-minimax rule are chosen according to the criteria discussed in Section 4. To fairly compare the estimators, the AMSE were computed on the same set of simulated data. All computations were carried out using MATLAB, with WaveLab toolbox (see Buckheit, Chen, Donoho, Johnstone and Scargle (1995)) and GaussWaveDen toolbox (see Antoniadis, Bigot and Sapatinas (2001)).

The aim of the first simulation was to examine the influence of the parameter γ on the finite sample performance of the proposed method. For each test function, sample size, and RSNR we have computed the AMSE (averaged over $M = 200$ samples) for a wide range of γ. The analysis shows that for moderate or high RSNR, the performance of the Γ-minimax estimator in terms of AMSE is quite robust with respect to the choice of γ. Larger values of γ would provide an almost noise-free reconstruction, but at the price of oversmoothing the singularities. When RSNR decreases, AMSE is significantly influenced by γ, and relatively large values of γ perform better in most cases. The value of $\gamma = 2$ has been selected as a default value when no information on the true signal is available. A relatively larger value of the parameter γ is to be preferred for regular functions or when the sample size increases. The importance of increasing γ when the signal-to-noise ratio is small is clearly seen in the application described in Section 5.1.

In the second simulation the AMSE, computed using the default value $\gamma = 2$, has been compared with the AMSE computed for the various classical and empirical Bayes term-by-term or block wavelet schemes used in this simulation study. The results across the various combinations of test functions, sample sizes and RSNR show that the Γ-minimax method often outperforms well known methods such as VisuShrink, Hybrid-SureShrink, Cross-Validation, ABE and FDR and BlockJS methods, and performs as well as (sometimes even better) BAMS and NeighCoeff methods. The Γ-minimax method shows improved performance compared to the BAMS whenever the RSNR is low. For high RSNR the advantage is not significant, although the Γ-minimax would protect against least favorable signals. Figures 4, 5 and 6 show the boxplots of the AMSE computed for the nine test functions based on $n = 1,024$ design points. We also observe that the choice of the default parameter results is effective when no additional information is available on the unknown regression function. We also reiterate that an improvement in the performance of the estimator when applied to low RSNR can be also achieved by increasing γ.

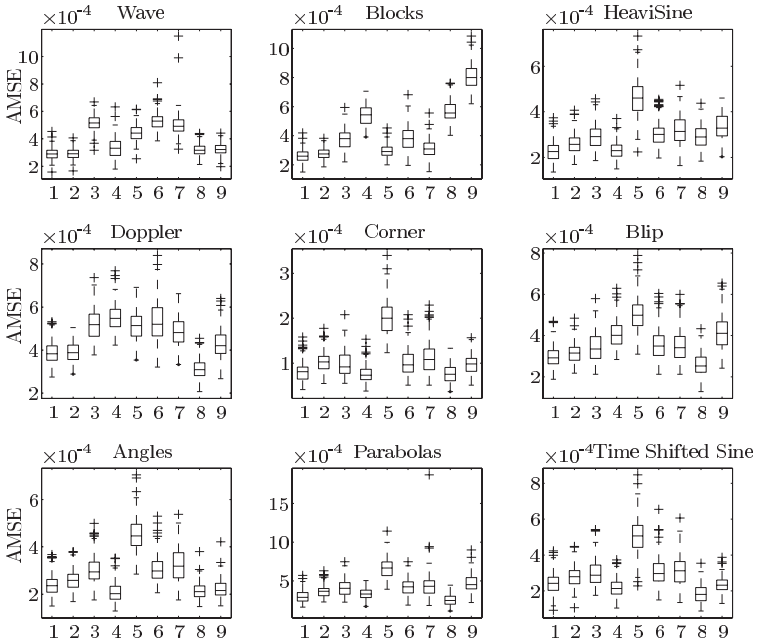

Figure 4. Boxplots of the AMSE for the nine methods considered: (1) Γ-minimax rule with $\gamma = 2$, (2) BAMS (3) VisuShrink with hard thresholding, (4) Hybrid-SureShrink, (5) ABE (6) "Leave-out-half" version of CrossValidation with hard thresholding, (7) FDR, (8) NeighCoeff and (9) BlockJS. The AMSE was computed on $n = 1,024$ design points and RSNR=3.

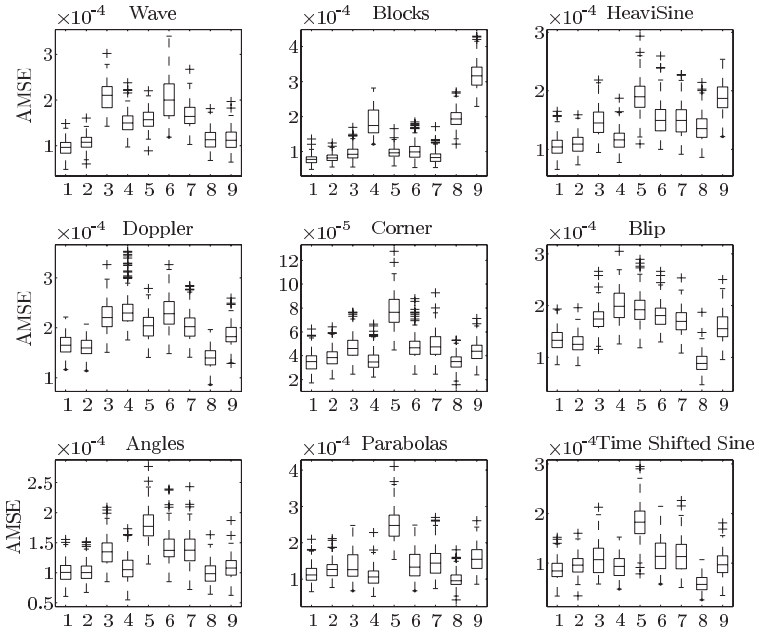

Figure 5. The same as in Figure 4, but with RSNR=5.

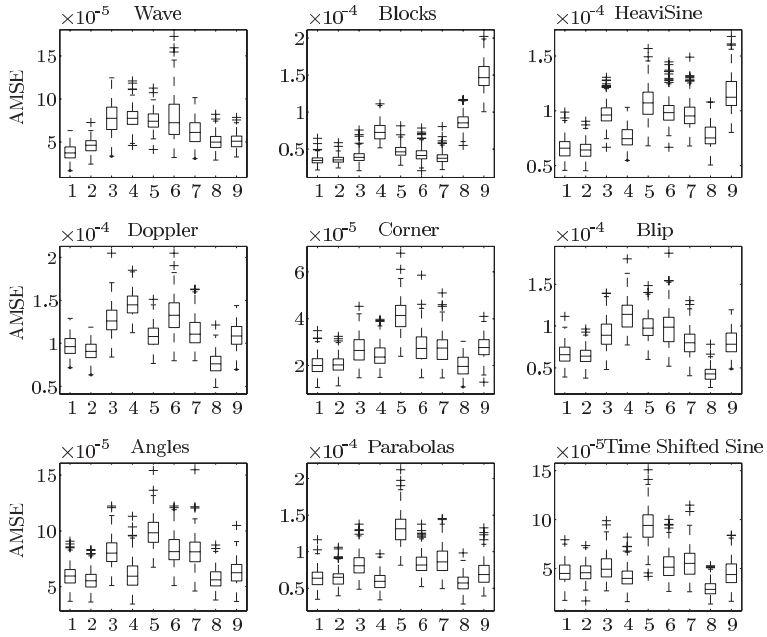

Figure 6. The same as in Figures 4 and 5, but with RSNR=7.

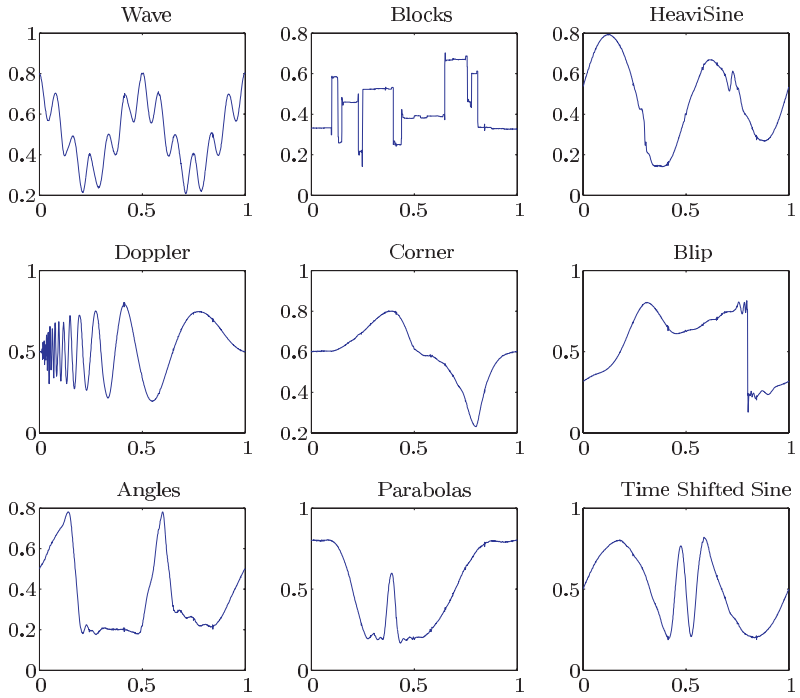

Figure 7. Estimates obtained using the Γ-minimax shrinkage with the default value $\gamma = 2$ for noisy samples of the nine test functions with RSNR=5 and $n = 1024$.

As an example of reconstruction, we show in Figure 6 the nine estimates obtained using the Γ-minimax method from $n = 1,024$ equispaced values with RSNR=5. The estimates usually show a good compromise in the smoothness of the function and the capability in preserving the singularities; smoother estimates can be obtained by increasing γ.

5.1. An example in atomic force microscopy

To illustrate features of the Γ-minimax shrinkage approach proposed here we used measurements in atomic force microscopy (AFM).

The AFM is a type of scanned proximity probe microscopy (SPM) that can measure the adhesion strength between two materials at the nanonewton scale (Binnig, Quate and Gerber (1986)). In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, and then the force required to separate the beam and sample is measured from the beam deflection. Beam vibration can be caused by factors such as thermal energy of the surrounding air or the footsteps of someone outside the laboratory. The vibration of a beam acts as noise on the deflection signal; in order for the data to be useful this noise must be removed.

The AFM data from the adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin was collected by Bryan Marshal from the BME Department at Georgia Institute of Technology. The technical description is provided in Marshall, McEver and Zhu (2001).

In Figure 8 the top panel shows the original noisy data. The middle panel shows the Γ-minimax estimate with the default parameter $\gamma = 2$, while the bottom panel shows Γ-minimax estimate with the parameter $\gamma = 8$. The sample size was $n = 2^{11}$ and Symmlet 8-tap filter was used to obtain the estimate. We observe that the latter estimate exhibits a smooth behavior, especially in the long-middle part without oversmoothing the bumps which are local features of interest.

6. Conclusions

In this paper we developed a method for wavelet-filtering of noisy signals when prior information about the L^2-energy of the signal is available. Assuming a Γ-minimax model, according to which the wavelet coefficients are treated individually, we propose a level dependent shrinkage rule. The proposed methodology was found well suited to denoise signals at any signal-to-noise ratio, however its advantage is of particular interest when the RSNR is low, which corresponds to the most problematic case. Dealing with ϵ-contaminated priors, a different approach has been recently considered in Angelini and Sapatinas (2002), where a wavelet thresholding rule based on the selection of the type II maximum likelihood prior is introduced. The resulting ML-IIThresh rule depends on only one hyper-parameter, but requires (for each wavelet coefficient) an iterative proce-

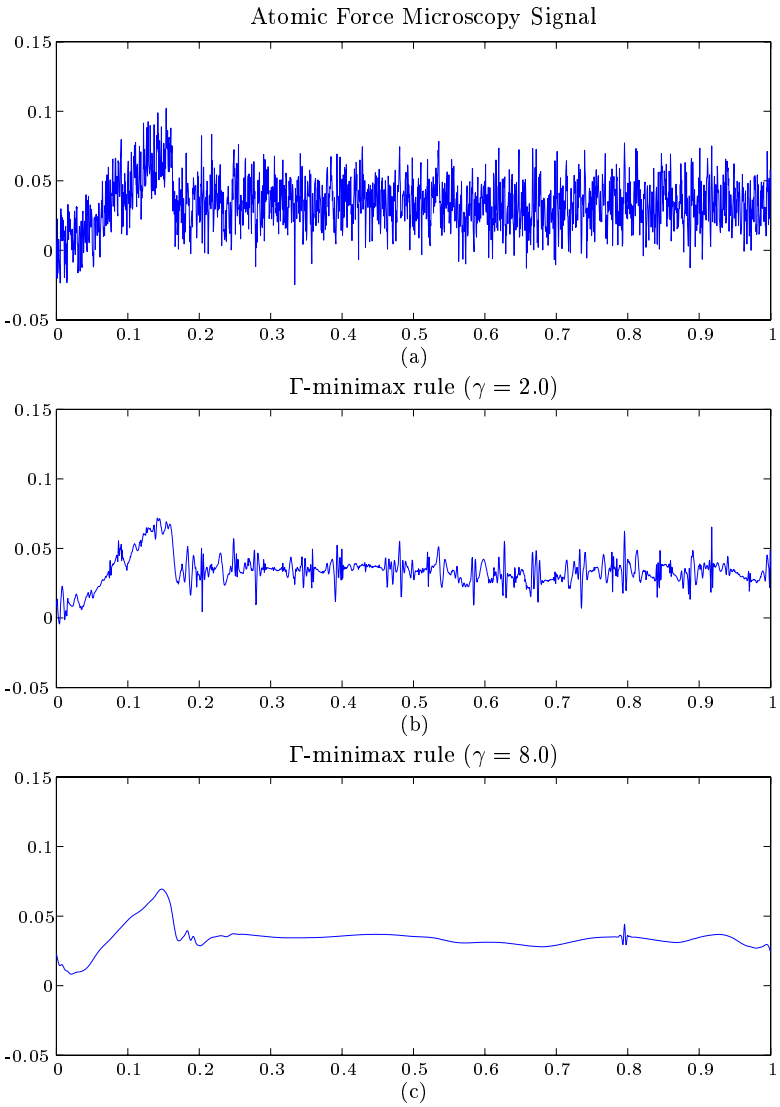

Figure 8. Top: Original AFM measurements; Middle: Γ-minimax estimator with the default parameter $\gamma = 2$; Bottom: Γ-minimax estimator with the parameter $\gamma = 8$.

dure to estimate the ML-II prior. Possible extensions of the method would use the restricted (linear, polynomial, etc.) Γ-minimax rules instead of the unrestricted. Such rules would provide additional simplicity with a minor expense in risk efficiency. Other extensions can be obtained by extending the proposed term-by-term shrinkage scheme to blocks of wavelet coefficients.

We already mentioned that the Bayes rule with respect to the Bickel prior is

not Γ-minimax. However, good approximations to Γ-minimax rules are possible. Theorem 2.2 of Bickel (1981) demonstrates how to construct an approximate minimax rule staring from $g_1(\theta)$. Bickel's construction easily adapts to our context and we hope to explore it elsewhere.

Appendix

Proof of Theorem 2.1. It is well known (Khintchine (1938), Dharmadhikari and Joag-dev (1988)) that any symmetric and unimodal random variable θ in $[-m, m]$ (with distribution $Q(\theta)$, and density $q(\theta)$) admits the representation $\theta = UZ$ where $U = \mathcal{U}[-m, m]$ and Z is a non negative random variable supported on $[0, m]$. Moreover, U and Z are independent. Indeed there is a "unique" correspondence between Q and the distribution function of Z, say F, up to a set of measure zero. The statistical game has a value, since

$$\begin{aligned} r(q,\delta) &= \int_0^m \int_{-1}^1 \frac{1}{2} R(uz,\delta)\,du\,dF(z) \\ &= \int_0^m \frac{1}{2z} \int_{-z}^z R(v,\delta)\,dv\,dF(z) \\ &= \int_0^m \frac{1}{z} \int_0^z R(v,\delta)\,dv\,dF(z) \stackrel{\text{def.}}{=} r'(F,\delta), \end{aligned}$$

$$\begin{aligned} \inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi,\delta) &= \inf_{\delta\in\mathcal{D}} \sup_{q\in\Gamma_{SU[-m,m]}} r(\epsilon_0\delta_0 + (1-\epsilon_0)q, \delta) \\ &= \inf_{\delta\in\mathcal{D}} \sup_{q\in\Gamma_{SU[-m,m]}} \{\epsilon_0 R(0,\delta) + (1-\epsilon_0) r(q,\delta)\} \\ &= \inf_{\delta\in\mathcal{D}} \sup_{F} \{\epsilon_0 R(0,\delta) + (1-\epsilon_0) r'(F,\delta)\} \\ &= \sup_{F} \inf_{\delta\in\mathcal{D}} \{\epsilon_0 R(0,\delta) + (1-\epsilon_0) r'(F,\delta)\} \\ &= \sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}} r(\pi,\delta). \end{aligned}$$

We have used the fact that, for any fixed value of m, the term of the risk coming from the point mass at zero, i.e., $r(\delta_0,\delta) = R(0,\delta) = \int_{-\infty}^{\infty} (\delta(d))^2\, d\Phi(d)$, does not depend on Q (hence on F), and that F is an arbitrary distribution in $[0, m]$.

Then, for any $z \in [0, m]$ we can define a new risk function

$$\bar{R}(z,\delta) = \epsilon_0 R(0,\delta) + (1-\epsilon_0)\frac{1}{z}\int_0^z R(v,\delta)\,dv \tag{17}$$

(and by continuity $\bar{R}(0,\delta) = R(0,\delta)$), and prove that it satisfies the five conditions on the risk given in Theorem 2.4 of Kempthorne (1987). Indeed,

(i) For any given distribution F on $[0, m]$ the Bayes rule with respect to $\int_0^m \bar{R}(z, \delta) dF(z)$ is unique, almost everywhere, due to the completeness of the normal model.

(ii) If F_n is any sequence of distributions which converges weakly to a distribution F, then the risk function (17) of the corresponding Bayes procedure converges uniformly on compacts to the risk function of the Bayes procedure corresponding to F:

$$\begin{aligned} &|\bar{R}(z, \delta_{F_n}) - \bar{R}(z, \delta_F)| \\ &\leq \epsilon_0 |R(0, \delta_{F_n}) - R(0, \delta_F)| + (1 - \epsilon_0) \frac{1}{z} \int_0^z |R(v, \delta_{F_n}) - R(v, \delta_F)| dv. \end{aligned}$$

Using the chain of implications

$$F_n \overset{weak}{\longrightarrow} F \Rightarrow \pi_n \overset{weak}{\longrightarrow} \pi \Rightarrow \delta_{F_n}(.) \overset{Uin(.)}{\longrightarrow} \delta_F(.) \Rightarrow R(\theta, \delta_{F_n}) \overset{U.in\theta}{\longrightarrow} R(\theta, \delta_F),$$

where $\overset{Uin(.)}{\longrightarrow}$ denotes uniform convergence, we get the result.

(iii) The parameter space is a compact and separable metric space since the support of F is $[0, m]$.

(iv)−(v) The risk function (17) is, for any decision rule, upper semi-continuous and analytic in the parameter space since it is true for the class of symmetric and unimodal distributions (the second term in (17)); the first term in (17) is a continuous, analytic function.

It follows that the least favorable distribution with respect to the risk $\bar{R}(z, \delta)$ is discrete (i.e., it is a linear combination of point masses at knots $m_i \in [0, 1]$ with probability α_i) $F^*(z) = \alpha_0 \mathbf{1}(z = 0) + \sum_{k=1}^p \alpha_k \mathbf{1}(z = m_k)$. Hence, the corresponding $q^* \in \Gamma_{SU[-m,m]}$ is $q^*(d) = \alpha_0 \delta_0 + \sum_{k=1}^p \alpha_k \mathcal{U}[-m_k, m_k]$, and the least favorable prior in Γ is the linear combination of uniforms and point mass at zero given in (7).

Finally, given the prior (7), it is easy to check that (8) is the Bayes rule. Indeed, since $d|\theta \sim N(\theta, 1)$, the Bayes rule will have the form $\delta_\pi(d) = d + (f'_\pi(d)/f_\pi(d))$, where $f_\pi(.)$ denotes the marginal distribution of d when the prior on θ is given by (7). By standard calculation we have

$$\begin{aligned} f_\pi(.) &= \int_\Theta \phi(d - \theta) \pi(\theta)\, d\theta = c_0 \phi(d) + \sum_{k=0}^p c_k \int_{-m_k}^{m_k} \frac{1}{2m_k} \phi(d - \theta)\, d\theta \\ &= c_0 \phi(d) + \sum_{k=0}^p \frac{c_k}{2m_k} \left[\Phi(d + m_k) - \Phi(d - m_k)\right], \end{aligned}$$

where $c_0 = \epsilon_0 + (1 - \epsilon_0)\alpha_0$ and $c_k = (1 - \epsilon_0)\alpha_k$, $k = 1, \ldots, p$. After taking the derivative of f_π with respect to d, we obtain (8).

To conclude, the case of small m (i.e., $m \leq m^*(\epsilon_0)$) has been considered in DasGupta and Delampady (1994) in a more general setup. Limiting the attention to model (6) we have that a distribution $\pi \in \Gamma$ is least favorable if

$$\sup_{\pi^* \in \Gamma} r(\pi^*, \delta_\pi) = r(\pi, \delta_\pi) = r(\pi) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta),$$

where π^* is any prior in Γ, δ is any rule in $\mathcal{D}$, and π and δ_π are defined in (9) and (10), respectively. For a fixed ϵ_0, $\sup_{q \in \Gamma_{SU[-m,m]}} r(q, \delta_\pi) = r(\mathcal{U}[-m, m], \delta_\pi)$, $\forall m \leq m^*$, where m^* is defined in Remark 2.1.

In fact, using the standard representation of a symmetric and unimodal random variable, we have

$$\begin{aligned}
\sup_{q \in \Gamma_{SU[-m,m]}} r(q, \delta) &= \sup_{q \in \Gamma_{SU[-m,m]}} E_q R(\theta, \delta) \\
&= \sup_F \int_{-1}^{1} \int_0^m R(uz, \delta) du dF(z) \\
&= \sup_{0 \leq z \leq m} \frac{1}{z} \int_0^z R(v, \delta) dv = r(\mathcal{U}[-m, m], \delta).
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\sup_{\pi^* \in \Gamma} r(\pi^*, \delta_\pi) &= \sup_{q \in \Gamma_{SU[-m,m]}} \{\epsilon_0 R(0, \delta_\pi) + (1 - \epsilon_0) r(q, \delta_\pi)\} \\
&= \epsilon_0 R(0, \delta_\pi) + (1 - \epsilon_0) \sup_{q \in \Gamma_{SU[-m,m]}} r(q, \delta_\pi) \\
&= \epsilon_0 R(0, \delta_\pi) + (1 - \epsilon_0) r(\mathcal{U}[-m, m], \delta_\pi) \\
&= r(\epsilon_0 \delta_0 + (1 - \epsilon_0)(\mathcal{U}[-m, m]) = r(\pi, \delta_\pi),
\end{aligned}$$

where δ_π is defined in (10) and π in (9).

Remark 7.1. We observe that the term $R(0, \delta_\pi)$ in (17) is a constant with respect to z. Hence $\bar{R}(z, \delta_\pi)$ reaches its maximum value at $z = m$ if and only if $(1/z) \int_0^z R(v, \delta_\pi) dv$ reaches its maximum at $z = m$. Since δ_π depends on ϵ_0, the value of m^* depends on ϵ_0, as well. Moreover, the parameters α_k and m_k, in the least favorable distribution (7), depend on ϵ_0.

Acknowledgements

The authors would like to thank an associate editor and two anonymous referees for their careful reading and insightful comments which significantly improved the paper. This work was done while the first author was visiting the Georgia Institute of Technology with support from University of Naples, Italy under "progetti giovani ricercatori" *DR n.* 4385/00 and Agenzia Spaziale Italiana. The work of the second author is supported in part by the NSF grant DMS 0004131 at GaTech. The MATLAB files and functions used in calculations are available upon request from the authors.

References

Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In *Wavelets and Statistics* Lect. Notes Statist. **103** (Edited by Antoniadis A. and Oppenheim G.), 5-14. Springer-Verlag, New York.

Abramovich, F, Besbeas, P. and Sapatinas, T. (2002). Empirical Bayes approach to block wavelet function estimation. *Comput. Statist. Data Anal.* **39**, 435-451.

Abramovich, F., Sapatinas T. and Silverman B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc. Ser. B* **60**, 725-749.

Abramovich, F. and Sapatinas, T. (1999). Bayesian approach to wavelet decomposition and shrinkage. In *Bayesian Inference in Wavelet Based Models* Lect. Notes Statist. **141** (Edited by P. Müller and B. Vidakovic), 33-50. Springer-Verlag, New York.

Angelini, C. and Sapatinas, T. (2002). Empirical Bayes approach to wavelet regression using epsilon-contaminated priors. Rapp. Tec., 239/02, Istituto per le Applicazioni del Calcolo - Sezione di Napoli, CNR, Italy.

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *J. Ital. Statist. Soc.* **6**, 97-144.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Soft.* **6**, 1-83.

Berger, J. O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analysis*, (Edited by J. Kadane), 63-124. Elsevier Science Publisher.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag. New York.

Berliner, L. M., Wikle, C. K. and Milliff, R. F. (1999). Multiresolution wavelet analyses in hierarchical Bayesian turbulence models. In *Bayesian Inference in Wavelet Based Models* Lect. Notes Statist. **141** (Edited by P. Müller and B. Vidakovic), 341-359. Springer-Verlag, New York.

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**, 1301-1309.

Binnig, G., Quate, C. F. and Gerber, Ch. (1986). Atomic force microscope. *Phys. Rev. Lett.* **56**, 930-933.

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

Buckheit, J. B., Chen, S., Donoho, D. L., Johnstone, I. M. and Scargle, J. (1995). About WaveLab. Technical Report, Department of Statistics, Stanford University, U.S.A.

Cai, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.

Cai, T. T. and Silverman, B. W. (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā B*, **63**, 127-148.

Casella, G. and Strawderman, W. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9**, 870-878.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Stat. Assoc.* **92**, 1413-1421.

Clyde, M. and George, E. I. (1999). Empirical Bayes estimation in wavelet nonparametric regression. In *Bayesian Inference in Wavelet Based Models* Lect. Notes Statist. **141** (Edited by P. Müller and B. Vidakovic), 309-322. Springer-Verlag, New York.

Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *J. Roy. Statist. Soc. Ser. B* **62**, 681-698.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85**, 391-401.

DasGupta, A. and Bose, A. (1988). Γ-minimax and restricted risk Bayes estimation of multiple Poisson means under ϵ-contaminations of the subjective prior. *Statist. Decisions* **6**, 311-341.

DasGupta, A. and Delampady, M. (1994). Γ-minimax estimation of vector parameters in restricted parameter spaces. Technical Report, Department of Statistics, Purdue University.

Dharmadhikari S. and Joag-dev, K. (1998). *Unimodality, Convexity, and Applications.* Academic Press Incorporated.

Donoho, D. L., Liu, R. and MacGibbon, B. (1990). Minimax risk over hyperrectangles. *Ann. Statist.* **18**, 1416-1437.

Donoho, D. L. and Johnstone I. M. (1994a). Minimax risk over l_p balls for l_q error. *Probab. Theory Related Fields* **99**, 277-304.

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

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200-1224.

Donoho, D. L., Johnstone I., Kerkyacharian G. and Picard D. (1995). Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B* **57**, 301-370.

Figueidero, M. A. T. and Nowak, R. D. (2001). Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior. *IEEE Trans. Image Process* **10**, 1322-1331.

Gao, H-Y. and Bruce, A. G. (1997). WaveShrink with firm shrinkage. *Statist. Sinica* **7**, 855-874.

Gatsonis, C., MacGibbon, B. and Strawderman W. E. (1987). On the estimation of a restricted normal mean. *Statist. Probab. Lett.* **6**, 21-30.

Good, I. J. (1952). Rational decisions. *J. Royal Statist. Soc. Ser. B* **14**, 107-114.

Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26**, 922-942.

Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica* **9**, 33-50.

Hall, P., Penev, S., Kerkyacharian, G. and Picard, D. (1997). Numerical performance of block thresholded wavelet estimators *Statist. Comput.* **7**, 115-124.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). Wavelets, approximation, and statistical applications. *Lect. Notes Statist.* **129**, Springer-Verlag, New York.

Kempthorne, P. J. (1987). Numerical specification of discrete least favorable prior distributions. *SIAM J. Sci. Statist. Comput.* **8**, 171-184.

Khintchine, A. Y. (1938). On unimodal distributions (in Russian). *Izv. Nauchno-Issled Inst. Math. Makh. Towsk. Gos., Univ.* **2**, 1-7.

Marchand E. and Perron F. (2001). Improving on the MLE of a bounded normal mean. *Ann. Statist.*, **29**, 1078-1093.

Marshall, B., McEver, R. and Zhu, C. (2001). Kinetic rates and their force dependence on the P-Selectin/PSGL-1 interaction measured by atomic force microscopy. *Proceedings of ASME* 2001, *Bioengineering Conference*, BED - **50**.

Miyasawa, K. (1953). On the minimax point estimations. *Bull. Math. Statist.* **5**, 1-17.

Nason, G. P. (1996). Wavelet shrinkage using cross-validation. *J. Roy. Statist. Soc. Ser B* **58**, 463-479.

Noubiap R. F. and Seidel W. (2001) An algorithm for calculating Γ-minimax decision rules under Generalized moment conditions. *Ann. Statist.* **29**, 1094-1116.

Robbins, H. (1951). Asymptotically sub-minimax solutions to compound statistical decision problems. In *Proc. Second Berkeley Symposium Math. Statist. and Prob.* **1**, 241-259. Berkeley University of California Press.

Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Stat. Soc. Ser. B* **45**, 173-179.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley, New York

Vidakovic, B. (2000). Γ-minimax: a paradigm for conservative robust Bayesians. In *Robust Bayesian Analysis* Lec. Notes Statist. **152** (Edited by D. Rios Insua and F. Ruggeri). Springer-Verlag, New York.

Vidakovic, B. and DasGupta, A. (1996). Efficiency of linear rule for estimating a bounded normal mean. *Sankhyā A* **58**, 81-100

Vidakovic, B. and Ruggeri, F. (2001). BAMS method: theory and simulations. *Sankhyā B* **63**, 234-249.

Istituto per le Applicazioni del Calcolo -Sezione di Napoli, CNR 80131, Naples, Italy.

E-mail:c.angelini@iac.cnr.it

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, U.S.A.

E-mail: brani@isye.gatech.edu

(Received August 2001; accepted May 2003)