Γ-Minimax Wavelet Shrinkage: A Robust Incorporation of Information about Energy of a Signal in Denoising Applications

Claudia Angelini

Istituto per Applicazioni della Matematica, CNR

Naples, Italy

Brani Vidakovic

Georgia Institute of Technology

Atlanta, GA 30332-0205, USA

Abstract

In this paper we propose a method for wavelet-filtering of noisy signals when prior information about the 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 Γ of realistic priors on the wavelet coefficients.

The proposed methodology, particularly applicable to noisy signals with a low signal-to-noise ratio, is illustrated on a battery of standard test functions. A real-life example in atomic force microscopy (AFM) is also discussed.

KEY WORDS: Wavelet Regression; Shrinkage; Bounded Normal Mean, Γ-minimaxity, Atomic Force Microscopy.

SHORT RUNNING TITLE: Γ-Minimax Wavelet Shrinkage.

1 Introduction

Gamma-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; 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).

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, Γ. Such "family of priors" elicitations are often encountered in practice. Given this family of priors, the decision maker is 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.

A 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 Γ-minimax rule coincides with minimax rule; when Γ

is a singleton then Γ-minimax rule coincides with Bayes rule. 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.

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

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

where t_i, $i = 1, \ldots, n$, is a deterministic equispaced design in $[0, 1]$, the random errors ε_i are i.i.d. centered normal random variables with common variance σ^2 and the interest is to recover the function f using the observations Y_i. Additionally, we assume that the unknown signal f has a bounded energy, hence it assumes values from a bounded interval. After applying a linear and orthogonal wavelet transformation, model (1) becomes

$$d_{j,k} = \theta_{j,k} + \epsilon_{j,k}$$

where $d_{j,k}$, $\theta_{j,k}$ and $\epsilon_{j,k}$ are the wavelet coefficients (at resolution j and position k) of Y, f and ε respectively. However, the orthogonality of the wavelet transformation preserves the independence of the wavelet coefficients, the stochastic structure of the noise and the bound on the energy. Due to the independence of the coefficients, in the sequel we will omit the double indices j, k and will work with a generic wavelet coefficient d. Therefore the model is

$$d = \theta + \epsilon \tag{2}$$

where $[d|\theta] \sim \mathcal{N}(\theta, \sigma^2)$ (σ^2 is assumed known).

Shrinkage rule in the wavelet domain have been often proposed in the literature to estimate the location parameter θ in model (2), for example in Donoho et al. (1995) and related papers in the minimax setup; Abramovich et al. (1998), Chipman et al (1997), Vidakovic (1998) and Vidakovic and Ruggeri (2001) in the Bayesian setup. Informally speaking, a shrinkage rule in the wavelet domain replaces the observed empirical wavelet coefficients d with their shrunk version $\hat{\theta} = \delta(d)$. 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) also the monotonicity. 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 signals encountered in practical applications have (for each resolution level) empirical distributions of detail wavelet coefficients centered around zero and peaked at zero. A realistic Bayesian model that takes into account this prior knowledge should consider prior distribution, π, that produces a reasonable agreement with observations. Hence a realistic prior distribution on the wavelet coefficient θ is if type

$$\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)\xi(\theta) \tag{3}$$

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 amount of shrinkage for values of d close to 0.

Prior models of this type have been indicated in the early 1990's by Berger and Müller (personal communication), considered in Abramovich et al (1998), 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, 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 of the compromising possibilities.

In this paper we would like to incorporate prior belief on the boundedness of energy of

the signal. *A priori* information on the energy bound often exists in real life problems and 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), Miyasawa (1953), Donoho et al (1990) in the minimax context and in Vidakovic and DasGupta (1996) in the Γ-minimax setup. It is however well know than estimating a bounded normal mean represents a difficult task. Thus, if the structure of the prior (3) 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 a bounded set can be a possible candidate for ξ.

Let Γ denote the family of priors of interest

$$\Gamma = \{\pi(\theta) = \epsilon_0\delta_0 + (1-\epsilon_0)q(\theta), q(\theta) \in \Gamma_{SU[-m,m]}\}, \tag{4}$$

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

$$\begin{cases} d|\theta & \sim & \mathcal{N}(\theta,1) \\ \theta & \sim & \pi(\theta) \in \Gamma \\ \mathcal{L}(\theta,\delta) & = & (\theta-\delta)^2 \ \text{ Squared Error Loss} \end{cases} \tag{5}$$

We stress that no generality is lost by assuming that $\Theta = [-m,m]$ (m depends on the resolution level of the wavelet coefficients) and that $\sigma^2 = 1$.

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

2 Model

In this section we extend the result of Vidakovic and DasGupta (1996) to the class of priors defined in (4). 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.

We first give the theoretical results.

Theorem 2.1 *Under the statistical model (5) where Γ is defined in (4), 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 distribution π in Γ.

The least favorable distribution π in Γ is of the form

$$\pi(\theta) = (\epsilon_0 + (1-\epsilon_0)\alpha_0)\delta_0 + (1-\epsilon_0)\sum_{k=1}^{p}\alpha_k\mathcal{U}[-m_k, m_k] \tag{6}$$

$$\text{where}\quad \alpha_k = \alpha_k(\epsilon_0) \geq 0;\quad \sum_{k=0}^{p}\alpha_k = 1;$$

$$m_k = m_k(\epsilon_0)\quad s.t.\quad 0 < m_1 < m_2 < \ldots < m_p = m,$$

and the corresponding Bayes rule is given by

$$\delta_\pi(d) = d - \frac{(\epsilon_0 + (1-\epsilon_0)\alpha_0)d\phi(d) - (1-\epsilon_0)\sum_{k=1}^{p}\frac{\alpha_k}{2m_k}\left(\phi(d+m_k) - \phi(d-m_k)\right)}{(\epsilon_0 + (1-\epsilon_0)\alpha_0)\phi(d) + (1-\epsilon_0)\sum_{k=1}^{p}\frac{\alpha_k}{2m_k}\left(\Phi(d+m_k) - \Phi(d-m_k)\right)}, \tag{7}$$

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

$$\pi(\theta) = \epsilon_0\delta_0 + (1-\epsilon_0)\mathcal{U}[-m, m] \tag{8}$$

and the Γ-minimax rule has the form

$$\delta_\pi(d) = d - \frac{\epsilon_0 d\phi(d) - \frac{1-\epsilon_0}{2m}\left(\phi(d+m) - \phi(d-m)\right)}{\epsilon_0\phi(d) + \frac{1-\epsilon_0}{2m}\left(\Phi(d+m) - \Phi(d-m)\right)}. \tag{9}$$

Remark 2.1 *The value of $m^*(\epsilon_0)$ such that (8) holds is the largest value of m for which the maximum of $\frac{1}{z}\int_0^z R(v,\delta_\pi)dv$ is achieved at $z = m$. $R(\bullet, \delta_\pi)$ represents the frequentist risk of the rule δ_π.*

Proof of Theorem 2.1. It is well known 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, F, up to a set of measure zero. Thus, the statistical game has a value, since from

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

it follows

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

In the above 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

$$\bar{R}(z, \delta) = \epsilon_0 R(0, \delta) + (1 - \epsilon_0)\frac{1}{z}\int_0^z R(v, \delta)dv \tag{10}$$

(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.

This condition follows from 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 (10) of the corresponding Bayes procedure converges uniformly on compacts to the risk function of the Bayes procedure corresponding to F.

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

Using the following chain of implications

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

where $\xrightarrow{Uin(.)}$ 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 (10) is, for any decision rule, upper semi-continuous and analytic in the parameter space.

This condition is satisfied since it is true for the class of symmetric and unimodal distributions [the second term in (10)]; the first term in (10) 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 (6).

Finally, given the prior (6) it is easy to check that (7) is the Bayes rule. Indeed, since $d|\theta \sim N(\theta,1)$ then the Bayes rule will have the form

$$\delta_\pi(d) = d + \frac{f'_\pi(d)}{f_\pi(d)},$$

where $f_\pi(.)$ denotes the marginal distribution of d when the prior on θ is given by (6). 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 the rule (7).

To conclude, the case of small m (i.e., $m \le m^*(\epsilon_0)$) has been considered in DasGupta and Delampady (1994) in a more general setup. Limiting the attention to model (5) we have that a distribution $\pi \in \Gamma$ is least favorable if

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

where π^* is any prior in Γ, δ is any rule in $\mathcal{D}$, and π and δ_π are defined in (8) and (9), respectively. For a fixed ϵ_0,

$$\sup_{q \in \Gamma_{SU[-m,m]}} r(q, \delta_\pi) = r(\mathcal{U}[-m,m], \delta_\pi),\ \forall m \le m^*, \tag{11}$$

where m^* is defined in Remark 2.1.

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

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

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 (9) and π in (8). □

Remark 2.2 *We observe that the term $R(0,\delta_\pi)$ in (10) is a constant with respect to z. Hence $\bar{R}(z,\delta_\pi)$ reaches its maximum value at $z=m$ if only if $\frac{1}{z}\int_0^z R(v,\delta_\pi)dv$ reaches the 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 (6), depend on ϵ_0.*

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

Table 1: Values of $m^*(\epsilon_0)$ for which the least favorable distribution in Γ is given by (8) and the corresponding values $m_1(\epsilon_0)$ of the support of the second uniform distribution when $m=m^*$.

For m exceeding m^*, (8) is no longer the least favorable prior; the least favorable prior will contain other uniform distributions supported on $[-m_k,m_k]$ as in (6). Numerical work,

analogous to the one developed in Vidakovic and DasGupta (1996), can give an accurate approximation of the parameters α_k and m_k in (6), 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. MATHEMATICA package was used in computing the values in Table 1, and in finding, for each ϵ_0, the largest m for which $\bar{R}$ 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 (4) 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 (7) is zero). However, for small values of ϵ_0 (for example $\epsilon_0 = 0.05$) α_0 is not zero. Indeed in Vidakovic and DasGupta (1996) has been proved 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 the weight of about 0.07.

When m is large the number of uniform distributions in the least favorable prior (6) increases, and the rule (7) can be only numerically evaluated. In Bickel (1981) 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 $g_1(\theta) = \cos^2(\frac{\pi\theta}{2})\mathbf{1}(|\theta| \leq 1)$. This fact implies that when m is large the least favorable prior in $\Gamma_{SU[-m,m]}$ is close to $g_m(\theta) = \frac{1}{m}\cos^2(\frac{\pi\theta}{2m})\mathbf{1}(|\theta| \leq m)$. Hence, the least favorable prior in Γ is close to

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

However, finding the corresponding Bayes rule cannot be done analytically and is beyond the scope of this paper. In the rest of the paper we limit our attention to the case of small m, where the shrinkage rule is (9).

3 Risk Analysis of the Rule

As we argued in Introduction, shrinkage rules of form (9) had been discussed in the literature on Bayesian wavelet shrinkage. In this section we explore their exact risk, bias-squared, and variance properties, when applied in the wavelet domain.

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 (1994) and Bruce and Gao (1996) provide exact risk analyses for hard and soft thresholding under the squared error loss. Bruce and Gao (1997) give the rationale for introducing the "firm" or "semi-soft" thresholding utilizing exact risk analysis. In our context the form of shrinkage rule (9) is more complex and the exact risk analysis had to be carried numerically. The goal of our analysis is to explore robustness in risk, bias, and variance when the prior hyperparameters change.

The computations performed in the software package MATHEMATICA produced Figures 1 and 2. Next, we briefly describe the numerical findings expressed in the figures.

As depicted in Fig.1(a), for $m = 3$, the shrinkage rules follow a desirable shrinkage pattern – for small values of d the rules behave differently. 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 Fig. 1(b) the risks of rules in Fig. 1(a) are presented. One can notice an obvious trade-off in the risk performance for small and large values of θ, respectively. When ϵ_0 is large the risk remains close to 0, for θ small; the risk is almost constant (the flattest risk curve in Fig. 1(b) corresponds to $\epsilon_0 = 0.1$).

The bias-squared, depicted in Fig. 1(c) is uniformly (in θ) increasing when ϵ_0 increases.

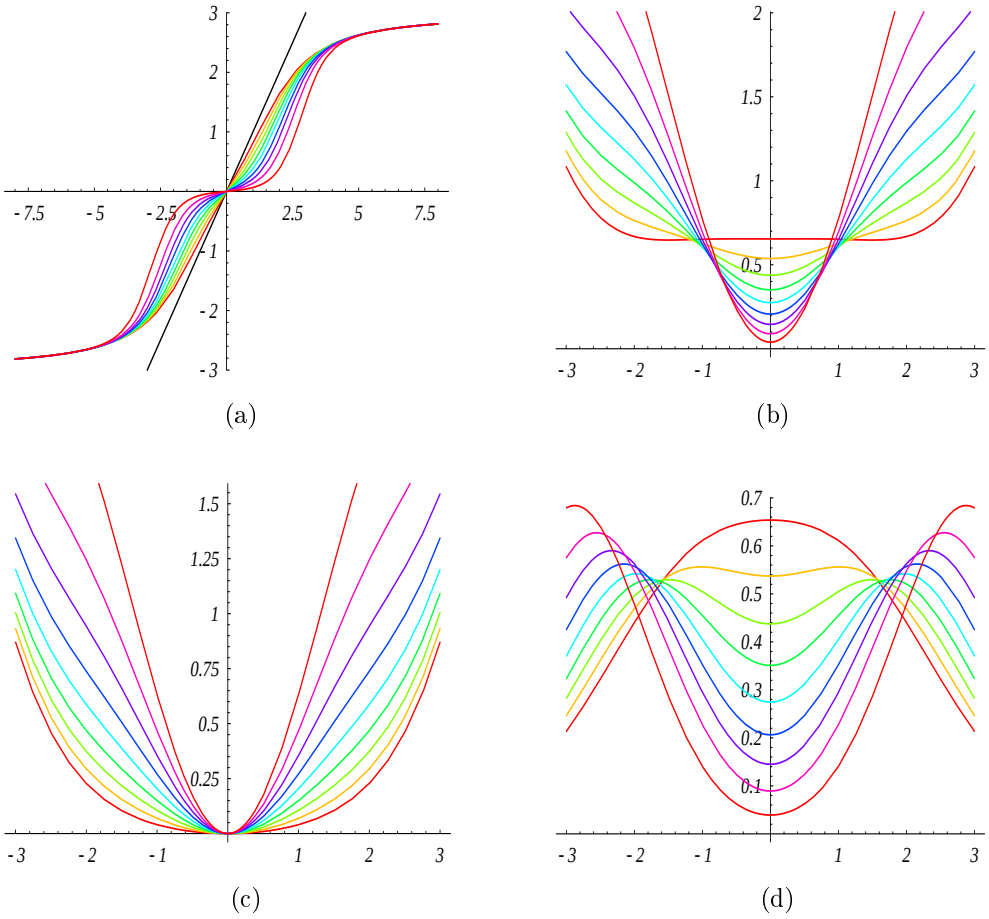

Figure 1: (a) Γ-minimax rules (9) for $m = 3$ and ϵ_0 ranging from 0.1 (upper envelope function) to 0.9 (lower envelope function); (b) Exact risks for rules in (a); (c) Bias^2 for rules in (a); (d) Variances for rules in (a).

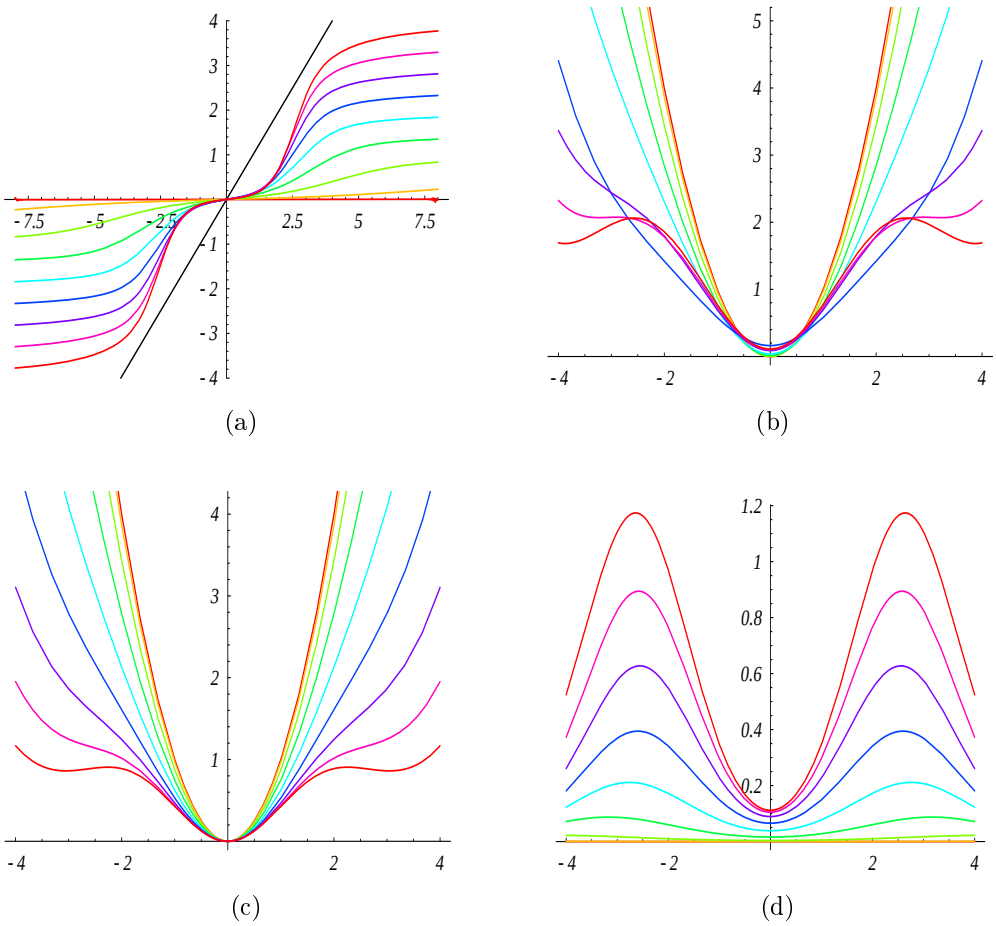

Figure 2: (a) Γ-minimax rules (9) for $\epsilon_0 = 0.8$ and m ranging from 0.5 to 4; (b) Exact risks for rules in (a); (c) Bias^2 for rules in (a); (d) Variances for rules in (a).

The lowest bias-squared curve correspond to $\epsilon_0 = 0.1$.

The variance functions, like the risks, exhibit an 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 Fig. 1(d).

Figure 2 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 biases-squared decrease in θ.

Finally, the variances depicted in Fig. 2(d) indicate that values of θ close to m produce most variability. The shape of variance functions is similar; the largest function corresponds to $m = 4$.

4 Elicitation of Parameters

The statistical model (5) depends on the choice of the hyperparameters ϵ_0 and m that should be carefully elicited in order to have an effective shrinkage rule. Usually, the elicitation of hyperparameters is a major issue 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, selfsimilarity, as well as the energy, of the unknown signal. We propose a level dependent choice of hyperparameters that is guided by considerations on the exact risk properties and on the shape of the shrinkage rule.

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 (9) 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

$$\epsilon_0(j) = \frac{1}{(j - J + 1)^\gamma}, \quad J \leq j \leq \log_2 n - 1,$$

where J represents the coarsest level in the wavelet transformation and $\gamma = 1.5$ is empirically chosen. The choice of the coarsest level is done according to $J = \text{floor}(\log_2(\log(\text{n}))) + 1$.

The elicitation of the hyperparameter m requires more detailed discussion since it has been noticed that the choice of m can substantially influence performance of the estimator. First, the Γ-minimax rule δ_π in Theorem 2.1 is sensitive with respect to m considerably more 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). For the sake of simplicity, we limit our attention to the case $m < m^*(\epsilon_0)$ where the shrinkage rule is explicitly given by (9). For such values of m, the resulting Bayes rules are particularly suited to noisy signals with low signal-to-noise ratio (SNR), i.e., to those situations for which most of other methods fail. When the SNR increases (or, equivalently, when m increases), for large value of ϵ_0, the Γ-minimax rule (7) becomes close to a thresholding rule curtailed at $-m$ and m. On the other hand, by increasing m, the Γ-minimax model (5) becomes close to the classical minimax problem, for which the thresholding rules have been found to be effective. Second, the elicitation of m 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 takes into account prior belief about the smoothness of the function:

$$m(j) = \min\left(p\sqrt{\left(\sigma_{d_j}^2 - \sigma_\epsilon^2\right)_+}, \max_k(|d_{j,k}|)\right), \tag{12}$$

where $\sigma^2_{d_j}$ is the variance of the empirical wavelet coefficients at scale j, and σ^2_ϵ is the variance of the noise. The variance of the noise can be estimated using a robust estimator at the finest level of detail. The standard rationale for (12) is that, since the transformation is orthogonal, the level of the noise is constant at each level, while the magnitude of the wavelet coefficients θ increases from the finest level to the coarsest level. When the resolution level j approaches the coarsest level, the estimate of $\sigma^2_{d_j}$ becomes inaccurate, since the number of coefficients within the level is small. Thus, the factor $\max_k(|d_{j,k}|)$ is taken into account. The parameter p is a real number ranging from 1 to 4, with values chosen such that $1/p$ is related to the smoothness of the function. In other words, the influence of p is significant at finest level details. High value of p, for example $p = 4$, allows the parameter space Θ wider at the finest level. This fact lead to preserving singularities, at the expense of still leaving some noise. Small values of p, for example $p = 1$, imply a narrow parameter space Θ at the finest level, which leads us to a smooth, almost noisy-free, reconstruction at the price of potentially distorting some of the singularities.

5 Applications

In order to analyze the performance of the Γ-minimax procedure we considered the standard test functions (`blocks, bumps, doppler,` and `heavisine`) and compared the average mean square error (AMSE), squared bias and variance with those of VisuShrink and SureShrink methods. If N denotes the number of simulational runs, the AMSE of the estimator $\hat{f} = (\hat{f}_1, \ldots, \hat{f}_n)$ is defined as

$$\frac{1}{Nn} \sum_{j=1}^{N} \sum_{i=0}^{n} (\hat{f}_{i,j} - f_i)^2,$$

where f_i are the components to be estimated and $\hat{f}_{i,j}$ are corresponding estimates in the j-th simulation run. In the first simulation the four test functions were rescaled so that an added standard normal i.i.d. noise produced a signal-to-noise-ratio (SNR) of 1. The sample size was $n = 1024$ and the wavelets used were: Symmlet 8-tap filter for `doppler` and `heavisine`, Haar for `blocks` and Daubechies 6-tap filter for `bumps`. The hyperparameters Γ-minimax

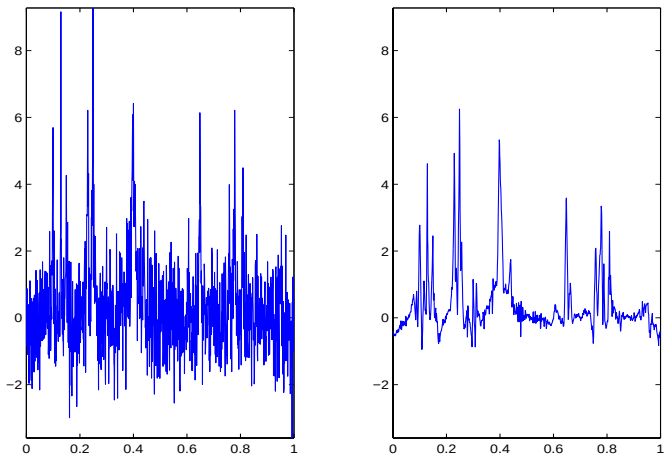

Figure 3: Left: noisy `bumps` [SNR=1, $n = 1024$, $\sigma^2 = 1$]; Right: The reconstructed signal using the Γ-minimax rule.

rule are chosen according to the criteria discussed in Section 4, in particular, the choice of p in (12) was 4 for `blocks` and `bumps`, 3 for `doppler` and 1 for `heavisine`. The coarsest level in wavelet decomposition used for computing VisuShrink and SureShrink estimators was $J = 3$, the same as for the Γ-rule. Moreover, we added results for VisuShrink and SureShrink rules with coarse level $J = 5$, as done in Chipman, Kolaczyk and McCulloch (1997). All computations have been carried out using MATLAB and WaveLab. The results for $N = 1000$ simulations are summarized in Table 2. An analysis of Table 2 yields that at low SNR the Γ-minimax rule outperforms both VisuShrink and SureShrink methods, since the restriction on size of θ allows for reduction of the components of the noise that produce most energetic wavelet coefficients. In the second simulation we computed the AMSE of the Γ-minimax rule (with its bias squared and variance components) on $N = 1000$ runs, for a variety of SNR's (0.5,1,1.5,2) and selection of sample sizes n (512,1024,2048,4096). The results are summarized in Table 3. As an illustration on application of the method on the test signals, Fig. 3 depicts a noisy and the reconstructed `bumps` signal from $n = 1024$ observations with SNR=1.

	BLOCKS	BUMPS
Γ-RULE	0.1094 (0.0573+0.0521)	0.1919 (0.1002+0.0917)
VISUSHRINK	0.2772 (0.2605+0.0167)	0.5326 (0.5075+0.0251)
SURESHRINK	0.2337 (0.0270+0.0207)	0.2390 (0.0771+0.1619)
VISUSHRINK (5)	0.1870 (0.1531+0.0339)	0.4041 (0.3626+0.0415)
SURESHRINK (5)	0.1738 (0.0832+0.0906)	0.2381 (0.0726+0.1655)
	HEAVISINE	DOPPLER
Γ-RULE	0.0168 (0.0081+0.0087)	0.0670 (0.0330+0.0340)
VISUSHRINK	0.0178 (0.0097+0.0081)	0.1890 (0.1735+0.0155)
SURESHRINK	0.0355 (0.0088+0.0267)	0.1427 (0.0221+0.1216)
VISUSHRINK (5)	0.0348 (0.0031+0.0317)	0.1244 (0.0898+0.0346)
SURESHRINK (5)	0.0383 (0.0030+0.0353)	0.1420 (0.0193+0.1227)

Table 2: AMSE, the Bias^2 and the variance components, obtained with 1000 simulational runs for the Γ-minimax rule (9), VisuShrink, and SureSrink. The sample size was $n = 1024$ and SNR=1.

5.1 An Example in Atomic Force Microscopy

To illustrate features of the Γ-minimax shrinkage approach proposed here we used the 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 cantalevar 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.

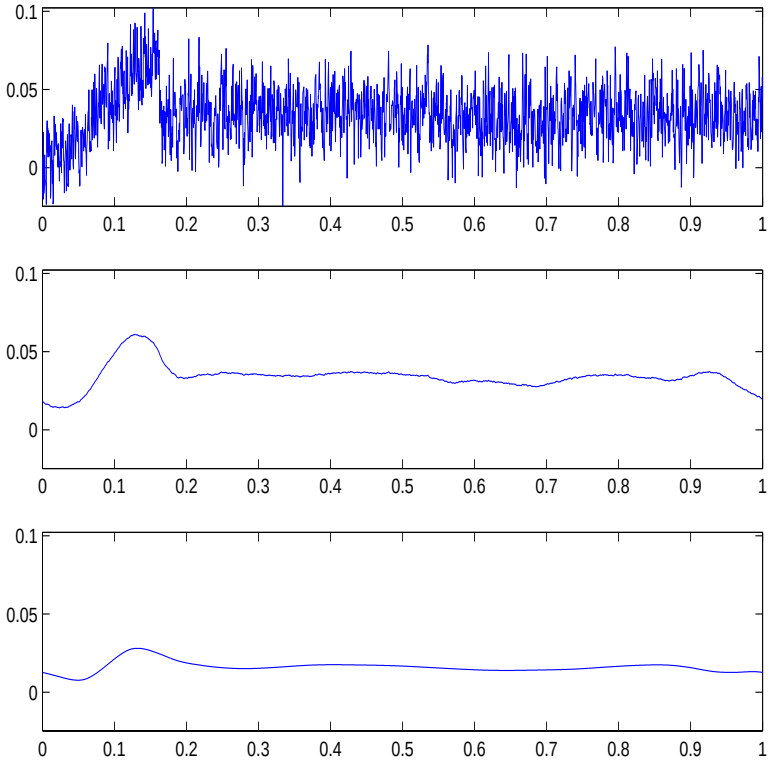

Figure 4: Top: Original AFM measurements; Middle: Γ-minimax estimator; Bottom: Universal thresholding estimator.

The AFM data from the adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin is 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 Fig. 4 the top panel shows the original noisy data. The middle panel shows the Γ-minimax estimator, while the bottom panel shows universal hard thresholding estimator. The sample size was $n = 2^{11}$.

6 Conclusions

In this paper we developed a method for wavelet-filtering of noisy signals when prior information about the 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 is well suited to noisy signals with a low signal-to-noise ratio. Applications include denoising of standard test functions and a real-life example in atomic force microscopy.

Possible extensions of the method can be in utilizing 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.

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

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

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis. Spriger 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(9), 930–933.

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

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

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.

DasGupta, A., and Bose, A. (1988) Γ-minimax and restricted risk Bayes estimation of multiple Poisson means under ϵ-contaminations of the subjective prior. Stat. and 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.

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., (1994) Minimax risk over l_p balls for l_q error. Prob. Theory relat. Fields, 277-304.

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

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

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

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 - Vol. 50.

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

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

Vidakovic, B., (1998). Non linear wavelet shrinkage with Bayes rules and Bayes factors. J. Amer. Stat. Soc., **45** B, pp. 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 (D. Rios Insua and F. Ruggeri, eds.) Springer-Verlag, New York.

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

Vidakovic, B. and Ruggeri, F. (2001). BAMS Method: Theory and Simulations. Sankhya, (Ser. A) **63**, in Press.

SNR	FUNCTION	n	AMSE	BIAS^2	VAR	SNR	FUNCTION	n	AMSE	BIAS^2	VAR
	BLOCKS	512	0.0934	0.0389	0.0545		BLOCKS	512	0.1980	0.0925	0.1085
		1024	0.0641	0.0200	0.0341			1024	0.1430	0.0856	0.0574
		2048	0.0433	0.0219	0.0214			2048	0.0901	0.0556	0.0345
		4096	0.0300	0.0128	0.0172			4096	0.0615	0.0372	0.0243
	BUMPS	512	0.1542	0.0844	0.0698		BUMPS	512	0.4081	0.2313	0.1768
		1024	0.1102	0.0590	0.0512			1024	0.2437	0.1295	0.1442
		2048	0.0756	0.0397	0.0359			2048	0.1488	0.0796	0.0691
0.5		4096	0.0484	0.0215	0.0269	1.5		4096	0.0891	0.0471	0.0420
	HEAVISINE	512	0.0185	0.0022	0.0163		HEAVISINE	512	0.0354	0.0184	0.0170
		1024	0.0105	0.0021	0.0084			1024	0.0262	0.0169	0.0093
		2048	0.0063	0.0021	0.0042			2048	0.0201	0.0150	0.0051
		4096	0.0052	0.0012	0.0040			4096	0.0138	0.0092	0.0045
	DOPPLER	512	0.0737	0.0266	0.0471		DOPPLER	512	0.1573	0.0839	0.0734
		1024	0.0465	0.0184	0.0281			1024	0.0905	0.0531	0.0374
		2048	0.0298	0.0137	0.0161			2048	0.0612	0.0367	0.0245
		4096	0.0211	0.0086	0.0125			4096	0.0471	0.0315	0.0156
	BLOCKS	512	0.1600	0.0759	0.0841		BLOCKS	512	0.2230	0.1141	0.1089
		1024	0.1094	0.0573	0.0521			1024	0.1793	0.1196	0.0597
		2048	0.0698	0.0387	0.0311			2048	0.1189	0.0731	0.0388
		4096	0.0460	0.0250	0.0210			4096	0.0790	0.0533	0.0257
	BUMPS	512	0.2964	0.1636	0.1328		BUMPS	512	0.4964	0.3001	0.1963
		1024	0.1919	0.1002	0.0917			1024	0.2864	0.1573	0.1290
		2048	0.1198	0.0614	0.0584			2048	0.1800	0.1024	0.0776
1		4096	0.0687	0.0324	0.0363	2		4096	0.1105	0.0635	0.0470
	HEAVISINE	512	0.0250	0.0085	0.0165		HEAVISINE	512	0.0490	0.0312	0.0178
		1024	0.0168	0.0081	0.0087			1024	0.0378	0.0277	0.0101
		2048	0.0120	0.0075	0.0045			2048	0.0298	0.0241	0.0057
		4096	0.0086	0.0044	0.0042			4096	0.0203	0.0153	0.0050
	DOPPLER	512	0.1212	0.0609	0.0603		DOPPLER	512	0.1854	0.1040	0.0814
		1024	0.0670	0.0330	0.0340			1024	0.1172	0.0745	0.0427
		2048	0.0490	0.0287	0.0203			2048	0.0726	0.0467	0.0259
		4096	0.0327	0.0184	0.0143			4096	0.0641	0.0466	0.0175

Table 3: AMSE, Bias squared and Variance components, obtained in 1000 simulational runs for a selection of sample sizes and SNR's. The test signals are rescaled such that the noise variance σ^2 remains 1.