

Gamma-Minimax Wavelet Shrinkage for Signals with Low SNR

DIXON VIMALAJEEWA*, ANIRBAN DASGUPTA, FABRIZIO RUGGERI, AND
BRANI VIDA KOVIC

Abstract

In this paper, we propose a method for wavelet denoising of signals contaminated with Gaussian noise when prior information about the L^2 -energy of the signal is available. Assuming the independence model, according to which the wavelet coefficients are treated individually, we propose simple, level-dependent shrinkage rules that turn out to be Γ -minimax for a suitable class of priors.

The proposed methodology is particularly well suited in denoising tasks when the signal-to-noise ratio is low, which is illustrated by simulations on a battery of some standard test functions. Comparison to some commonly used wavelet shrinkage methods is provided.

KEYWORDS AND PHRASES: Wavelet Regression, Shrinkage, Bounded Normal Mean, Γ -minimaxity, Least Favorable Prior, Contamination Class, Low Signal-to-Noise Ratio.

1. INTRODUCTION

Function estimation by using shrinkage estimators in the wavelet domain was initiated in the works of David Donoho and Iain Johnstone in the early 1990's. Since then various shrinkage methods have been proposed. In this article, we propose and investigate a novel approach for estimation in the wavelet domain. This approach brings together the classic principle of Γ -minimaxity, Huber's famous ϵ -contamination class [23], mathematics of decision theory, aspects of multiple branches of pure mathematics, and then, also, detailed experimental and graphical analysis of the performance of our proposal. In these ways, this article is a new comprehensive analysis of the problem of estimation in the wavelet domain.

For the rest of this introductory section, we review fundamentals of Γ -minimax estimation, wavelet shrinkage, and Bayesian approaches to wavelet shrinkage. This section is mostly expository and gives background and the needed motivation.

1.1 Γ -Minimax Theory

The Γ -minimax paradigm, originally proposed by Robbins [29], deals with the problem of selecting decision rules in tasks of statistical inference. The Γ -minimax approach falls between the Bayes paradigm, which selects procedures that work well "on average a posteriori" for a specific prior, and the minimax paradigm, which guards against least favorable outcomes, however unlikely. This approach has evolved from seminal papers in the fifties [29, 21] 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. A comprehensive discussion of the Γ -minimax approach can be found in [5, 6]. Also see [8], a fundamental contribution.

The Γ -minimax paradigm incorporates the prior information about the statistical model by a family of plausible priors, denoted by Γ , rather than by a single prior. Elicitation of "prior families" is often encountered in practice. Given the family of priors, the decision maker selects a rule that is optimal with respect to the least favorable prior in the family.

We note that least favorable priors typically do not exist in unbounded parameter spaces. For example, there is no least favorable prior for estimation of unconstrained normal or Poisson means. There is a least favorable prior for estimating a binomial parameter, and this is a compact parameter space. But no amount of intuition can predict the least favorable prior for the binomial. And, as soon as the loss is changed to a normalized squared error, the least favorable prior changes. There are no links or intuition regarding least favorable priors that are portable. It is too delicate and problem specific. For example, in a quite general setup, Clarke and Barron in [13] show that Jeffreys' prior is asymptotically least favorable when the loss is Kullback-Leibler distance, thus providing theoretical justification for reference priors.

Inference of this kind is often interpreted in terms of game theory. Formally, if \mathcal{D} is a set of all decision rules and Γ is a family of prior distributions over the parameter space Θ , then a rule $\delta^* \in \mathcal{D}$ is Γ -minimax if

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta^*) = r(\pi^*, \delta^*), \quad (1.1)$$

*Corresponding author.

where $r(\pi, \delta) = E^\theta [E^{X|\theta} \mathcal{L}(\theta, \delta)] = E^\theta R(\theta, \delta)$ is the Bayes risk under the loss $\mathcal{L}(\theta, \delta)$. Here $R(\theta, \delta) = E_\theta^{X|\theta} \mathcal{L}(\theta, \delta)$ denotes the frequentist risk of δ , π^* is the least favorable prior, and $\mathcal{L}(\theta, \delta)$ is the loss function, often the squared error loss, $(\theta - \delta)^2$. Note that when Γ is the set of all priors, the Γ -minimax rule coincides with the minimax rule, if a minimax rule exists; when Γ contains a single prior, then the Γ -minimax rule coincides with Bayes' rule with respect to that prior. When the decision problem, viewed as a statistical game, *has a value*, that is, when $\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} \equiv \sup_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}}$, then the Γ -minimax solution coincides with the Bayes rule with respect to the least favorable prior. For the interplay between the Γ -minimax and Bayesian paradigms, see [6]. A review on Γ -minimax estimation can be found in [32].

In a nutshell, the Γ -minimax is a philosophical compromise between subjective Bayes with a single prior and the conservative approach of not having a prior at all.

1.2 Wavelet Shrinkage

We apply the Γ -minimax approach to the classical nonparametric regression problem

$$y_i = f(t_i) + \sigma \varepsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

where t_i , $i = 1, \dots, n$, are deterministic equispaced design points on $[0, 1]$, 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 from the observations y_i . Additionally, we assume that the unknown signal f has a bounded L^2 -energy, namely $\int_{[0,1]} f^2(t) dt$, and hence it assumes values from a bounded interval. After applying a linear and orthogonal wavelet transform, the model in (1.2) becomes (see, e.g., [16]),

$$\begin{aligned} c_{J_0, k} &= \theta_{J_0, k} + \sigma \varepsilon_{J_0, k}, \quad k = 0, \dots, 2^{J_0} - 1, \\ d_{j, k} &= \theta_{j, k} + \sigma \varepsilon_{j, k}, \quad j = J_0, \dots, J - 1, \\ & \quad k = 0, \dots, 2^j - 1, \end{aligned} \quad (1.3)$$

where $d_{j, k}$ ($c_{j, k}$), $\theta_{j, k}$ and $\varepsilon_{j, k}$ are the wavelet (scaling) coefficients (at resolution level j and location k) corresponding to y , f and ε , respectively; J_0 and $J - 1$ are the coarsest and finest levels of detail in the wavelet decomposition. If ε 's are i.i.d. standard normal, an arbitrary wavelet coefficient from (1.3) can be modeled as

$$|d|\theta \sim \mathcal{N}(\theta, \sigma^2), \quad (1.4)$$

where, due to (approximate) independence of the coefficients, we omitted the indices j , k . The prior information on the energy bound of the signal energy implies that a wavelet coefficient θ corresponding to the signal part in (1.2) assumes its values in a bounded interval, say $\Theta = [-m(j), m(j)]$, which depends on the level j .

Wavelet shrinkage rules have been extensively studied in the literature, but mostly when no additional information on

the parameter space Θ is available. For the implementation of wavelet methods in nonparametric regression problems, we also refer to [4], where some methods are described and numerically compared.

1.3 Bayesian Model in the Wavelet Domain

Bayesian shrinkage methods in wavelet domains have received considerable attention in recent years. Depending on a prior, Bayes' rules are shrinkage rules. The shrinkage process is defined as follows: A shrinkage rule is applied in the wavelet domain and the observed wavelet coefficients d are replaced by with their shrunken versions $\hat{\theta} = \delta(d)$. In the subsequent step, by the inverse wavelet transform, coefficients are transformed back to the domain of original data, resulting in a data smoothing. The shape of the particular rule $\delta(\cdot)$ influences the denoising performance.

Bayesian models on the wavelet coefficients have proved capable of incorporating some prior information about the unknown signal, such as smoothness, periodicity, sparseness, self-similarity and, for some particular bases (e.g., Haar), monotonicity.

The shrinkage is usually achieved by eliciting a single prior distribution π on the space of parameters Θ , and then choosing an estimator $\hat{\theta} = \delta(d)$ that minimizes the Bayes risk with respect to the adopted 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 for which the prior predictive distribution produces a reasonable agreement with observations. A realistic prior distribution on the wavelet coefficient θ is given by

$$\pi = \epsilon \delta_0 + (1 - \epsilon) \xi, \quad (1.5)$$

where δ_0 is a point mass at zero, ξ is a symmetric distribution on the parameter space Θ and ϵ is a fixed number in $[0, 1]$, usually level dependent, that regulates the amount of shrinkage for values of d close to 0. Priors for wavelet coefficients as in (1.5) have been considered by Vidakovic and Ruggeri in [33]. Prior literature on wavelet shrinkage by using Bayesian ideas includes important work by [12, 14, 30, 19], among others. A review by Reményi and Vidakovic in [28] overviews a range of Bayesian wavelet shrinkage methods and provides numerous references on Bayesian wavelet shrinkage strategies.

Our approach here is different from all the prior literature. Here is the motivation.

It is 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 uncertainty and multitude of prior 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 of

plausible priors, Γ . 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 ensuring the robustness of a specific rule, Γ -minimax being one compromise.

In this paper, we incorporate prior information 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 it can be modelled by the assumption that the parameter space Θ is bounded. Estimation of a bounded normal mean has been considered in numerous papers, e.g., [20, 7, 11, 15, 9, 17], and in the classic text of Ibragimov and Hasminskii [24].

It is well known that estimating a bounded normal mean represents a challenging problem. In our context, if the structure of the prior (1.5) can be supported by the analysis of the empirical distribution of the wavelet coefficients, even then precise elicitation of the distribution ξ cannot be done without some kind of approximation. Any symmetric distribution supported on the bounded set, say $[-m, m]$, can be a possible candidate for ξ .

Let Γ denote the family

$$\Gamma = \{\pi = \epsilon\delta_0 + (1 - \epsilon)\xi, \xi \in \Gamma_{S[-m, m]}\}, \quad (1.6)$$

where $\Gamma_{S[-m, m]}$ is the class of all symmetric distributions supported on $[-m, m]$, δ_0 is point mass at zero, and ϵ is a fixed constant between 0 and 1. We also require that the distribution ξ does not have an atom at 0.

We consider two models; both assume that wavelet coefficients follow normal distribution (which is a statement about the distribution of the noise), $d \sim \mathcal{N}(\theta, \sigma^2)$. In the Model I, the variance of the noise is assumed to be known, while in the Model II the variance is not known and is given a prior distribution. We have kept treatment of Model II limited. Detailed treatment of Model II would take too much additional space.

The rest of the paper is organized as follows. Section 2 contains mathematical aspects and results concerning the Γ -minimax rules. An exact risk analysis of the rule is discussed in Section 3. Section 4 proposes methods of elicitation of hyper-parameters defining the model. For example, one can imagine ‘‘estimating’’ ϵ and even m itself as hyperparameters by using simulated data. But it is going to depend on the run and none of the theoretical results of our paper would then stand. It can be the topic of an entirely different paper. Performance of the shrinkage rule in the wavelet domain and application to a data set are given in Section 5. In Section 6 we provide the conclusions. Proofs are mostly deferred to Appendix A and B.

2. CHARACTERIZING Γ -MINIMAX RULES

In this section, we discuss the existence and characterization of Γ -minimax shrinkage rules that are Bayes’ with

respect to least favorable priors on the interval $[-m, m]$. It will turn out that there exists a critical m^* such that for $m \leq m^*$, the least favorable prior is a three point prior, that is it assigns mass $1 - \epsilon$ at 0, and mass $\epsilon/2$ at the two boundary points $\pm m$. It is in this way that three point priors enter into our wavelet shrinkage analysis. We look at two scenarios, when the variance of the noise is known (Model I), and when it is not known (Model II). For reasons of space, we limit our treatment of Model II to just a short exposition.

We emphasize that we treat the case of Γ -minimax estimation in very general compact convex sets in general finite dimensional Euclidean spaces below. By giving a complete and self-contained derivation of the existence of a least favorable prior and the Γ -minimax estimator in such great generality when the parameter space is sufficiently small, we have given a useful unification of the Γ -minimax problem.

2.1 Model I

Theorem 1.

- (a) Let \mathcal{S}_0 be a compact convex set in R^p containing the origin $\mathbf{0}$ in its interior. Let $\mathbf{a} \in R^p$ and $b > 0$, define

$$\mathcal{S}_{\mathbf{a}, b} = b\mathcal{S}_0 + \mathbf{a},$$

and let $\partial\mathcal{S}_{\mathbf{a}, b}$ denote the boundary of $\mathcal{S}_{\mathbf{a}, b}$.

Let $\mathbf{X} \sim \mathbf{f}(\mathbf{x}|\boldsymbol{\theta})$, a density in the p -parameter exponential family, and consider estimation of $\boldsymbol{\theta}$ using squared error loss. Then, there exists a unique least favorable prior on $\mathcal{S}_{\mathbf{a}, b}$ as well as $\partial\mathcal{S}_{\mathbf{a}, b}$.

- (b) There exists $b^* > 0$ such that for $b \leq b^*$, the least favorable prior on $\partial\mathcal{S}_{\mathbf{a}, b}$ is also the least favorable prior on $\mathcal{S}_{\mathbf{a}, b}$.
- (c) If in particular, $\mathbf{X} \sim \mathbf{N}_{\mathbf{p}}(\boldsymbol{\theta}, \mathbf{I})$, and $\boldsymbol{\theta}$ belongs to a p -dimensional ball centered at $\mathbf{0}$ and of radius b , then for $b \leq b^*$, the least favorable prior in the ball is simply the uniform distribution on the boundary of the ball.
- (d) In the one-dimensional case, for the entire one parameter exponential family, if θ belongs to a compact interval $[u, v]$, the least favorable prior is necessarily finitely supported.
- (e) In the one dimensional case that $X \sim N(\theta, 1)$, let

$$\pi \in \Gamma = \{\epsilon\delta_0 + (1 - \epsilon)\xi\}, \quad (2.1)$$

where ϵ is fixed in $[0, 1]$, and ξ is any distribution on $[-m, m]$ without atoms at 0, that is, with no point-mass-at-zero. Then there exists $m^* > 0$ such that for $0 < m \leq m^*$ the least favorable prior is

$$[\theta] \sim \pi = \epsilon\delta_0 + \frac{1 - \epsilon}{2}(\delta_{-m} + \delta_m).$$

The Bayes rule with respect to this prior,

$$\delta_B(x) = \frac{m \sinh(mx)}{\cosh(mx) + \frac{\epsilon}{1 - \epsilon} e^{m^2/2}}, \quad (2.2)$$

is the Γ -minimax rule.

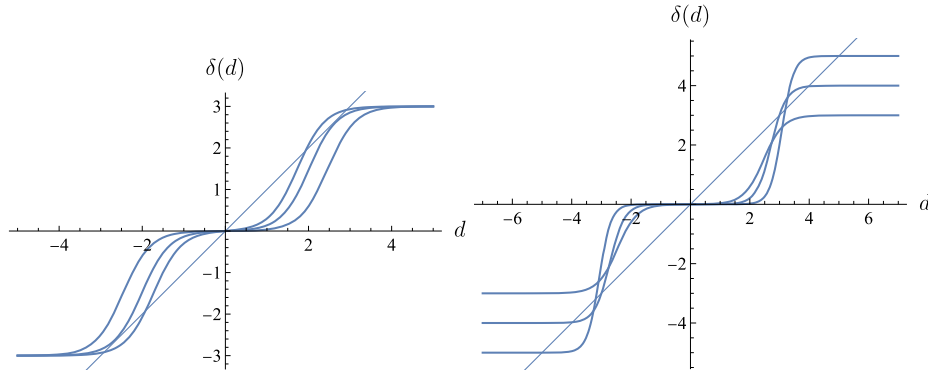


Figure 1: Γ -Minimax Rule for Model I. **Left:** Rules for $m = 3$ and $\epsilon = 0.5, 0.7,$ and 0.9 . **Right:** Rules for $\epsilon = 0.9$ and $m = 3, 4,$ and 5 .

Table 1. Values of m^* for both models for different values of ϵ .

ϵ	m^* (Model I)	m^* (Model II)
0.0	1.05674	0.81758
0.1	1.15020	0.91678
0.2	1.27739	1.05298
0.3	1.46988	1.25773
0.4	1.84922	1.52579
0.5	2.28384	1.74714
0.6	2.41918	1.91515
0.7	2.50918	2.05511
0.8	2.58807	2.19721
0.9	2.69942	2.40872
0.95	2.81605	2.63323
0.99	3.10039	3.24539

Remark. Figure 1 shows this shrinkage rule in (2.2) for selected values of the parameters ϵ and m . Note that the rules heavily shrink small coefficients, but unlike traditional shrinkage rules, remain bounded between $-m$ and m . The values of m^* are given in Table 1.

2.2 Model II

In Model II, the variance σ^2 is not known and is given an exponential prior. It is well-known that the exponential distribution is an entropy maximize in the class of all distributions supported on R^+ with a fixed first moment. This choice is non-informative, in a form of a maxent prior.

The model is:

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

$$[\sigma^2] \sim \mathcal{E}(\mu), \quad \left[f(\sigma^2) = \frac{1}{\mu} \exp \left\{ -\frac{\sigma^2}{\mu} \right\} \right]$$

The marginal likelihood is double exponential as an exponential scale mixture of normals,

$$[d|\theta] \sim \mathcal{DE} \left(\theta, \sqrt{\frac{\mu}{2}} \right)$$

$$\left[g(d|\theta, \mu) = \sqrt{\frac{1}{2\mu}} \exp \left\{ -\sqrt{\frac{2}{\mu}} |d - \theta| \right\} \right].$$

Theorem 2. If in Model II the family of priors on the location parameter is (2.1), the resulting Γ -minimax rule is:

$$\delta_B(d) = \frac{m \left(e^{-\sqrt{2/\mu} |d-m|} - e^{-\sqrt{2/\mu} |d+m|} \right)}{e^{-\sqrt{2/\mu} |d-m|} + \frac{2\epsilon}{1-\epsilon} e^{-\sqrt{2/\mu} |d|} + e^{-\sqrt{2/\mu} |d+m|}}, \tag{2.3}$$

which is Bayes with respect to the least favorable prior

$$[\theta] \sim \pi(\theta) = \epsilon \delta_0 + \frac{1-\epsilon}{2} (\delta_{-m} + \delta_m),$$

whenever $m \leq m^*$. Figure 2 shows the rule in (2.3) for selected values of parameters ϵ and m . The values of m^* depend on ϵ and are given in Table 1 for both models. We remark that the model in Theorem 2 is similar in spirit to those in BAMS wavelet shrinkage [33] and Bayesian Lasso [27]. Double exponential distributions in the marginal likelihood (for BAMS) and prior (for Bayesian Lasso) are obtained as exponential scale mixtures of normals.

Sketches of proofs of Theorems 1 and 2 are deferred to Appendix A and B.

3. RISK, BIAS, AND VARIANCE OF Γ -MINIMAX RULES

Frequentist risk of a rule δ , as a function of θ , can be decomposed as a sum of two functions, variance and bias-squared,

$$R(\delta, \theta) = E^{d|\theta} (\delta(d) - \theta)^2$$

$$= E^{d|\theta} (\delta(d) - E^{d|\theta} (\delta(d)))^2$$

$$+ (\theta - E^{d|\theta} (\delta(d)))^2.$$

To explore the behavior of the two risk components in the context of Models I and II, we selected the risk of Γ -minimax rule for $\epsilon = 0.8$ and $m = 2.197$. (Fig. 3). This

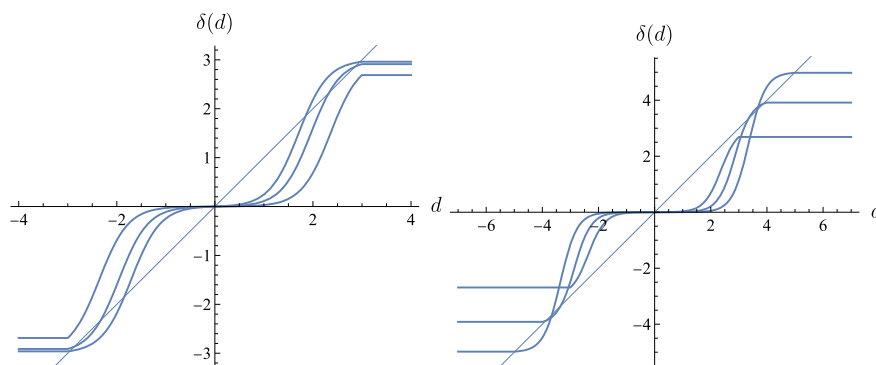


Figure 2: Γ -Minimax Rule for Model II. Left: Rules for $\epsilon = 0.5, 0.7,$ and 0.9 . Right: Rules for $m = 3, 4,$ and 5 .

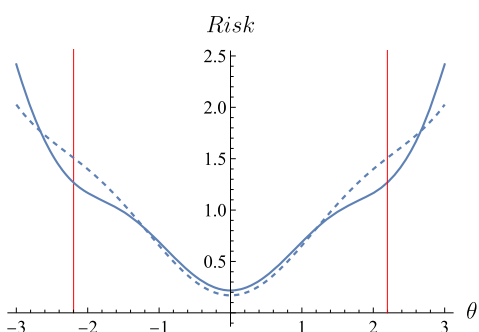


Figure 3: Risk of Γ -Minimax rule for $\epsilon = 0.8$ and $m = 2.197$. Dashed plots for Model I and solid for Model II.

particular value of m ensures that the rules are Γ -minimax and $m = m^*$ for model II. Note that δ in Model II shows a smaller risk for values of θ in the neighborhood of m , while for θ close to 0, the risk of the rule from Model I is smaller.

The most interesting finding is the behavior of the risk function of the rule. The risk function has a wavy shape with maxima at the locations depending on m . Typically the risk is maximized at $\theta = -m, m$. This is the most important finding in that now general minimax theory takes over and establishes that the proposed rule indeed is fully mini-

max and hence, also gamma-minimax. Similar, but less pronounced behavior is present in bias-squared function (Fig. 4 Left Panel). Compared to Model I, the variance of δ in Model II is significantly smaller for values of θ in the neighborhood of $\pm m$, and larger for θ in the neighborhood of zero. Preference in using either Model I or II depends on what size of signal part we are more interested in. If there is more uncertainty about signal bound m , the rule from Model II is preferable. However, Model I has a lower risk and both components of the risk are in the neighborhood of 0. This translates to a possibly more precise shrinkage of small wavelet coefficients.

4. ELICITATION OF PARAMETERS

The proposed Bayesian shrinkage procedures with three-point priors depend on three parameters, m, ϵ and μ , that need to be specified. The criteria used for selecting these parameters are critical for effective signal denoising. We propose these hyper-parameters to be elicited in an empirical Bayes fashion, that is, dependent on the observed wavelet coefficients.

- (1) **Elicitation of m :** The bound m in the domain of signal acquisition translates to level-dependent bounds on the parameter θ in the wavelet domain. Given a data

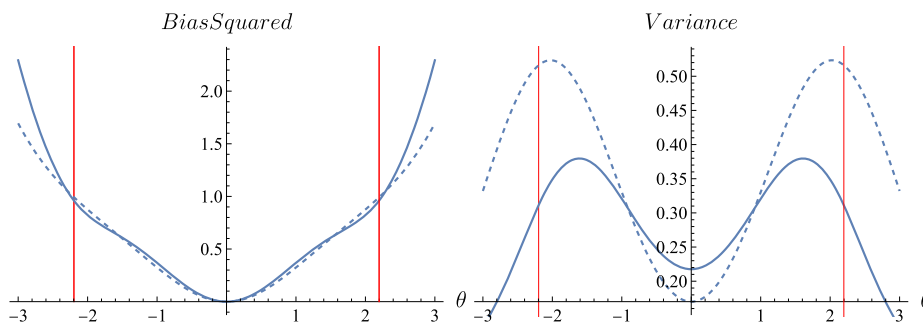


Figure 4: Risk components of Γ -minimax rule for $\epsilon = 0.8$ and $m = 2.197$. Dashed plots for Model I and solid for Model II. **Left:** Bias-squared; **Right:** Variance.

signal $y = (y_1, \dots, y_n)$, the hyper-parameter m at the j th multiresolution level is estimated as

$$m(j) = \hat{\sigma} \max(|y_i|) \left(\sqrt{2}\right)^{J-j}, \quad (4.1)$$

where $J = \log_2 n$ is the resolution level of the transformation, and $\hat{\sigma}$ an estimator of the noise size. The noise size is often estimated by a robust estimator of standard deviation that uses the wavelet coefficients at the finest multiresolution level,

$$\hat{\sigma} = \frac{\text{median}(|d_{J-1,\bullet} - \text{median}(d_{J-1,\bullet})|)}{0.6745}, \quad (4.2)$$

where $d_{J-1,\bullet}$ represents all detail coefficients in the level $J - 1$. The multiple $(\sqrt{2})^{J-j}$ in (4.1) reflects the increase of the extrema of the absolute value of wavelet coefficients within the decomposition level in j steps of the transform.

- (2) **Elicitation of ϵ :** This parameter controls the amount of shrinkage in the neighborhood of zero and overall shape of shrinkage rule. For the levels of fine details, this parameter should be close to 1. Based on the proposal for the ϵ given by Angelini and Vidakovic in [3] and Sousa et al. in [31], we suggest a level-dependent ϵ_0 as follows:

$$\epsilon_0(j) = 1 - \frac{1}{(j - J_0 + l)^k}, \quad (4.3)$$

where $J_0 \leq j \leq J - 1$ and k and l are positive constants. As we indicated, ϵ should be close to one at the multiresolution levels of fine details, and then be decreasing gradually for levels approaching the coarsest level [3]. When k and l are large, ϵ remains close to one over all levels. This results in an almost noise-free reconstruction, but could result in over-smoothing. On the other hand, $l > 1$ guarantees a certain level of shrinkage even at the coarsest level. Thus, the hyperparameters l and k should be selected with a care in order to achieve good performance for a wide range of signals. Numerical simulations guided us to suggest values $l \geq 6$ and $k = 2$ as reasonable choices. However, it is important to note that these parameters should depend on the smoothness of data signals and their size. We further discuss the selection of l and k for specific signals in Section 5.

- (3) **Elicitation of μ :** This parameter is needed only for Model II. Since the prior on the noise level σ^2 is exponential and the prior mean is μ , by simple moment matching we select μ as $\hat{\sigma}^2$. A possible choice for $\hat{\sigma}^2$ is a robust estimator as in (4.2).

5. SIMULATION STUDY

In the simulation study, we assessed the performance of the proposed shrinkage procedures on the

battery of standard test signals. We used nine different test signals (**step**, **wave**, **blip**, **blocks**, **bumps**, **heavisine**, **doppler**, **angles**, and **parabolas**), which are constructed to mimic a variety of signals encountered in applications (Fig. 5). As standardly done in literature, Haar and Daubechies six-tap (Daubechies 6) were used for Blocks and Bumps and Symmlet 8-tap filter was used for the remaining test signals. The shrinkage procedures are compared using the average mean square error (AMSE), as in (5.1). All simulations were performed using MATLAB software and toolbox *GaussianWaveDen* [4] that can be found at <http://www.mas.ucy.ac.cy/~fanis/links/software.html>.

We generated noisy data samples of the nine test signals by adding normal noise with zero mean and variance $\sigma^2 = 1$. The signals were rescaled so that $\sigma^2 = 1$ leads to a prescribed SNR. Each sample consisted of $n = 1024$ data points equally spaced on the interval $[0, 1]$. Figure 6 shows a noisy version of the nine test signals with SNR = 1/4. Each noisy signal was transformed into the wavelet domain. After the shrinkage was applied to the transformed signal, the inverse wavelet transform was performed on the processed coefficients to produce a smoothed version of a signal in the original domain. The AMSE was computed as

$$AMSE(f(t)) = \frac{1}{nN} \sum_{j=1}^N \sum_{i=1}^n \left(f(t_i) - \hat{f}_j(t_i)\right)^2, \quad (5.1)$$

where f denotes the original test signal and \hat{f}_j its estimator in the j -th iteration. To calculate the average mean square error this process was repeated $N = 100$ times.

The shrinkage procedure was applied to each test signal and the AMSE was computed for a range of parameter values of l and k . For example, Fig. 8 shows the average MSE obtained on the **heavisine** test signal when SNR = 1/5, and l and k vary in the range $l \in [2, 15]$ and $k \in [1.0, 3.5]$. As evident from Fig. 8, the estimator achieves its best performance for values $k \approx 2.4$ and $l \approx 5.8$. With these selected values of l and k , Fig. 7 shows that the estimator is sufficiently close to the original test signal, even though the SNR is quite small.

Based on our simulations, the optimal hyper-parameter values of l and k varied depending on the nature (e.g., smoothness) of the test signal. For larger values of k and l , the estimator performs better for smooth signals. This is because the corresponding wavelet coefficients rapidly decay with the increase in resolution. However, larger values of l and k may not detect localized features in signals (e.g., cusps, discontinuities, sharp peaks), resulting in over-smoothing. For low values of SNR, the three-point priors estimator is more sensitive to hyper-parameter values of l and k . When SNR increases, the estimation method shows better performance for most of the test signals with relatively small values of l and k . Moreover, higher values of parameters l and k are preferred when the sample size is large.

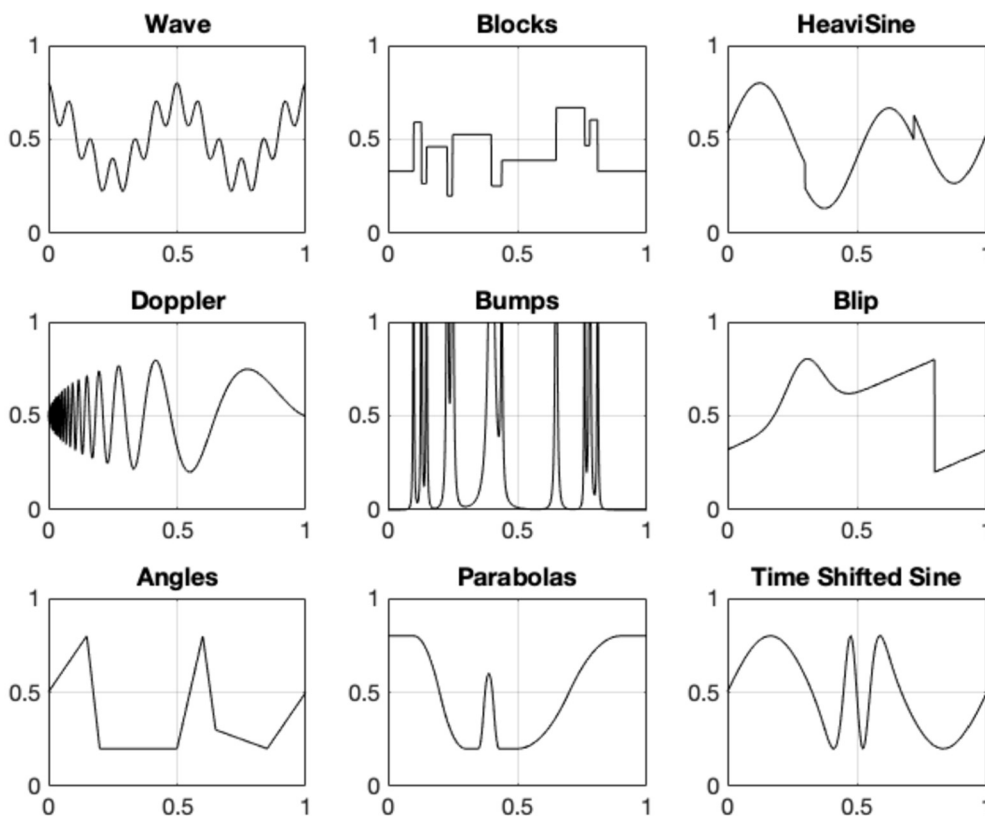


Figure 5: The nine test signals used in the simulation study.

In general, we suggest $k = 2.5$ and $l \geq 6$ as the most universal choice. The suggested values could, however, be adjusted depending on available information about the nature of signals.

5.1 Performance Comparison with Some Existing Methods

We compared the performance of the proposed three-point prior estimator with eight existing estimation techniques. The selected existing estimation techniques include: Bayesian adaptive multiresolution shrinker [33] (BAMS), Decomph [22], block-median and block-mean [1], hybrid version of the block-median procedure [1], blockJS [10], visushrink [16], and generalized cross validation [2]. The first five techniques are relying on a Bayesian procedure and they are based on level-dependent shrinkage. The blockJS method uses a level-dependent thresholding, while the visushrink and generalized cross-validation techniques use a global thresholding method. Readers can find more details about these techniques in [4].

In the simulation study, we computed the AMSE using the parameter values of $l = 6$ and $k = 2.5$ and compared with the AMSE computed for the selected estimation techniques. As can be seen in Fig. 9, the proposed estimator shows comparable and for some signals better performance when compared to the selected estimation methods.

In particular, for smooth signals (e.g., **wave**, **heavisine**), the three-point prior estimator shows better performance compared to non-smooth signals, such as **blip**, for instance. Moreover, when comparing the performance of the level-dependent estimation methods, the BAMS estimation method shows competitive (or better) performance for most of the cases. We also investigated the influence of SNR level and the sample size on the performance of proposed estimators, compared to the methods considered. For example, for higher SNR (Fig. 10), the three-point priors-based shrinkage procedure does not provide better performance except for **wave**, **angles**, and **time shifted sine**. In general, the Γ -minimax shrinkage shows comparable or better performance compared to other methods considered, when the SNR is low. Also, for larger sample sizes, the three-point prior shrinkage shows improved performance.

6. CONCLUSIONS

We proposed a method for wavelet denoising of signals when prior information about the size of the signal is available. Simple, level-dependent shrinkage rules that are Bayes with respect to a prior supported on three discrete points are at the same time Gamma-minimax, for a set of all priors with a bounded support symmetric about zero with a fixed point mass at zero. This statement is true when the

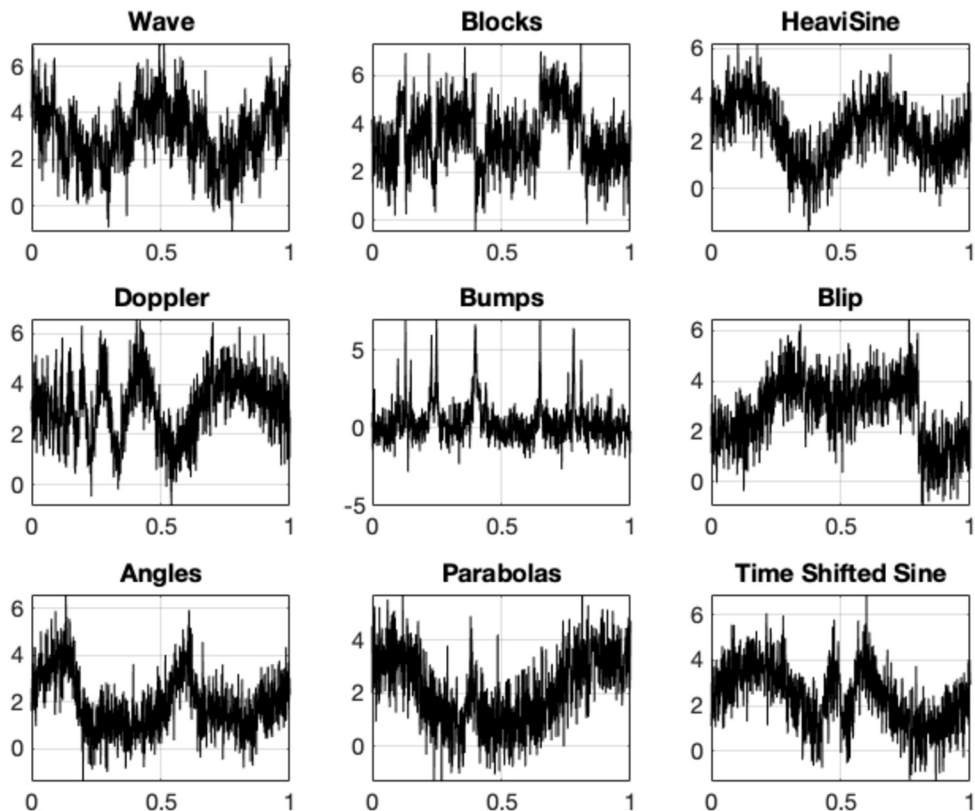


Figure 6: Noisy versions of the nine signals from Fig. 5.

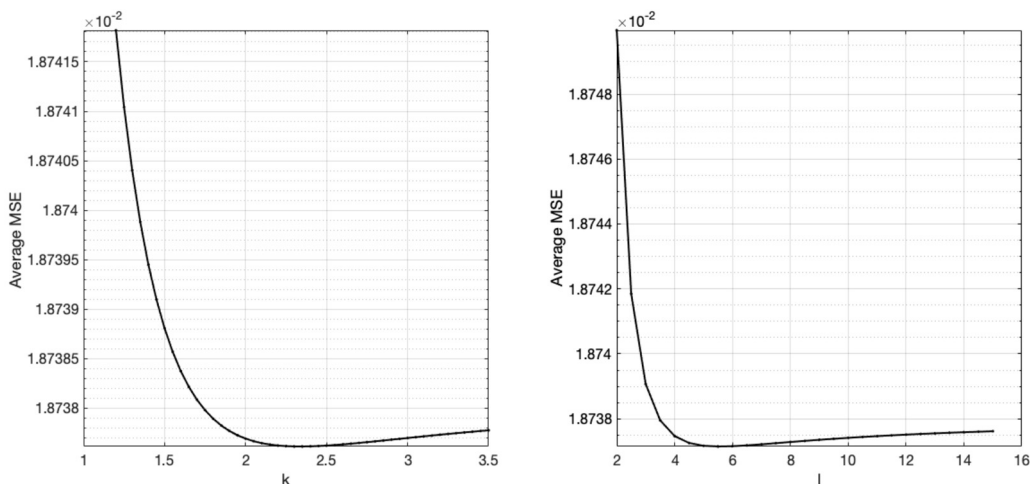


Figure 7: Change in average MSE(AMSE) as a function of hyper-parameters k (left) and l (right) exemplified on the `heavisine` test signal, $SNR = 1/5$, and $n = 1024$.

signal is bounded on $[-m, m]$, for small values of m , which in denoising applications translates to estimation of a signal with a low signal-to-noise ratio. For larger values of m the least favorable prior is still discrete with support on a finite number of points symmetrically distributed about 0, but finding exact locations and prior weights of these points

presents a challenging optimization problem. Thus, using the three point priors for high signal-to-noise ratio noisy signals is suboptimal from the standpoint of minimaxity and numerical experiments show that in this case, the three point prior shrinkage rules are underperforming compared to other standard shrinkage methods. We considered two estimation

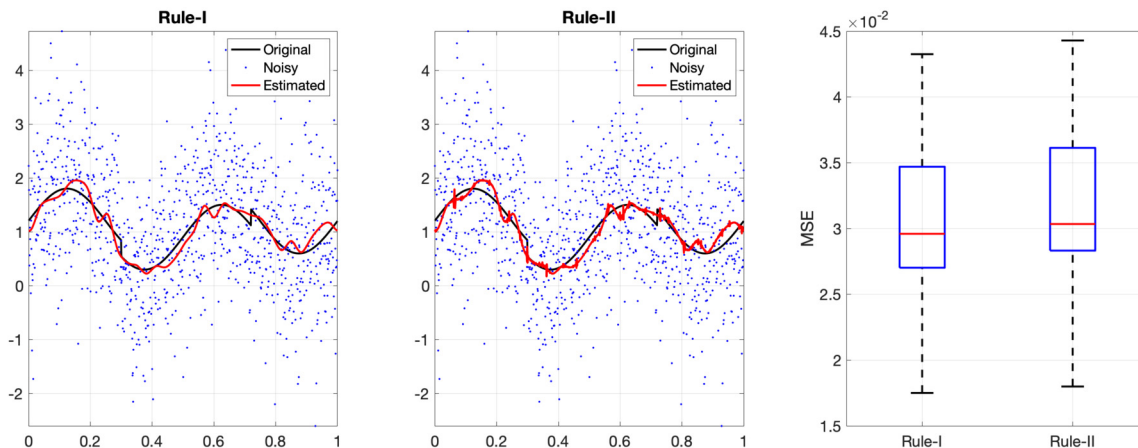


Figure 8: Estimation of heavisine test signal: Estimations obtained by three-point priors with $n = 1024$, $k = 2.4$, $l = 5.8$ and $SNR = 1/5$.

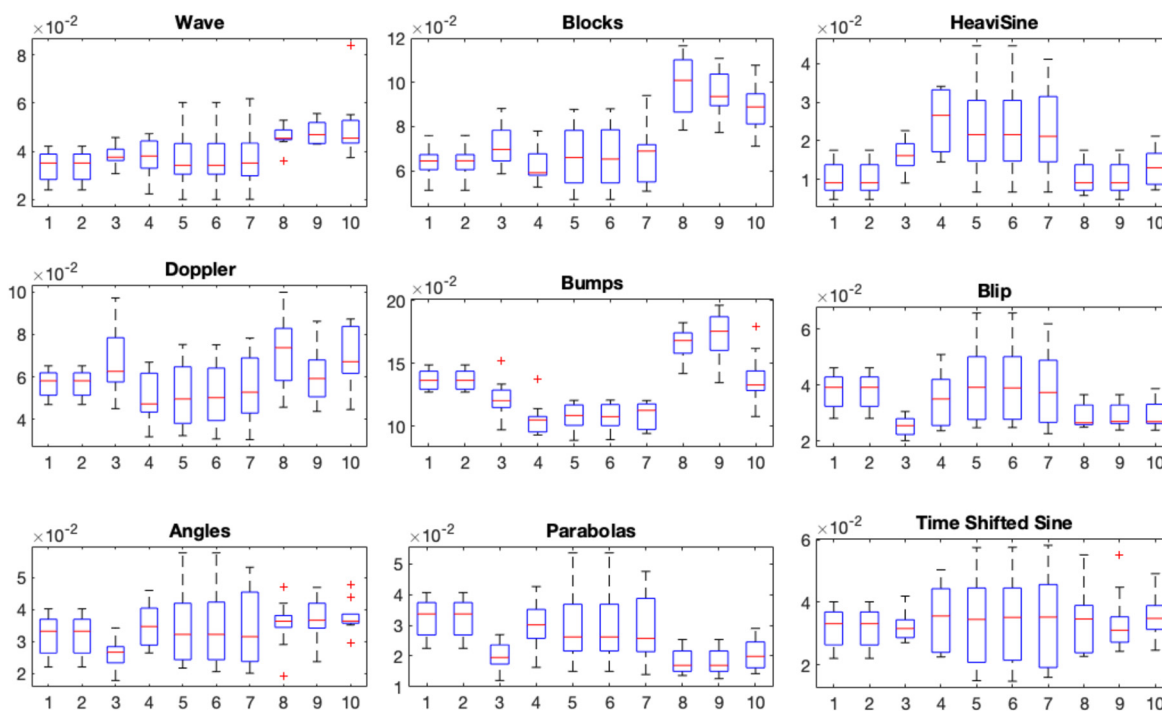


Figure 9: The box plots of MSE for the ten estimation methods: (1) Rule-I, (2) Rule-II, (3) Bayesian adaptive multiresolution shrinker (BAMS), (4) Decompsh, (5) Block-median, (6) Block-mean, (7) Hybrid version of the block-median procedure, (8) BlockJS, (9) Visu-Shrink, and (10) Generalized cross-validation. The MSE was computed by using $SNR = 1/5$, $k = 2.5$, $l = 6$, and $n = 1024$ data points.

scenarios: when the variance of the noise is known and when it is not known. In the first case a plug-in robust estimator of noise variance is used, while in the second the variance is given an exponential prior, and subsequently integrated out. The main theoretical result is given in Theorem 1, that describes the form of Gamma-minimax rules when parameter belongs to a very general convex set in finite dimensional

Euclidean space. The two estimation scenarios proposed in this paper are special cases of Theorem 1.

As demonstrated by simulations on a battery of standard test functions, the performance of the proposed shrinkage rules in terms of AMSE is comparable, and for some signals superior, to the state-of-art methods when SNR is low and the signals are smooth.

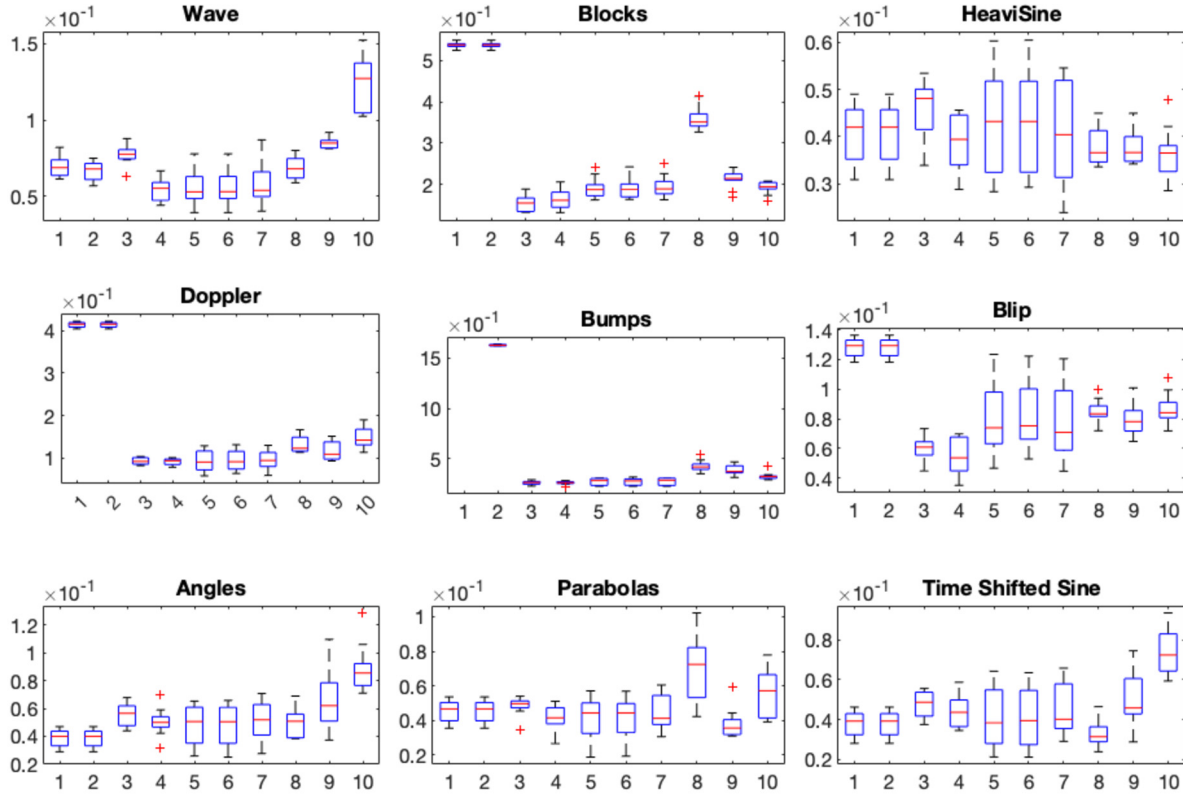


Figure 10: The same plot as in Fig. 8 with SNR = 3.

APPENDIX A. PROOF OF THEOREM 1

- (a) Let \mathcal{S}_0 be a fixed convex compact set in \mathcal{R}^p , $1 \leq p < \infty$ symmetric about $\mathbf{0}$ and containing $\mathbf{0}$ in its interior. Define

$$\mathcal{S}_{\mathbf{a},b} = b\mathcal{S}_0 + \mathbf{a},$$

where $\mathbf{a} \in \mathcal{R}^p$, and $b > 0$.

Define the family of prior distributions

$$\Gamma = \Gamma_\epsilon = \{\pi : \pi = \epsilon \delta_{\mathbf{a}} + (1 - \epsilon) \xi\},$$

where ξ is a symmetric measure on $\mathcal{S}_{\mathbf{a},b}$ with $\xi(\{\mathbf{a}\}) = \mathbf{0}$. Then, for any decision rule δ ,

$$\sup_{\pi \in \Gamma} r(\pi, \delta) = \epsilon R(\mathbf{a}, \delta) + (1 - \epsilon) \sup_{\xi} r(\xi, \delta).$$

Consider the classes \mathcal{G} of all probability measures and \mathcal{G}_s of all symmetric probability measures on $\mathcal{S}_{\mathbf{a},b}$. Let $\mathcal{G}_{s,0}$ be a class of all symmetric probability measures ξ on $\mathcal{S}_{\mathbf{a},b}$ with $\xi(\{\mathbf{a}\}) = \mathbf{0}$. By Prokhorov's theorem, \mathcal{G} is sequentially weakly compact, and since the weak topology on any finite dimensional linear space is metrizable, \mathcal{G} and therefore \mathcal{G}_s are also compact. Now we use the fact that the Bayes risk $r(\pi)$ considered as a functional of the prior π is a strictly concave functional due to the squared error loss being strictly convex [18, 26]. It also

being the case that $r(\pi)$ is upper semicontinuous [34], $\sup_{\pi \in \mathcal{G}}$ is attained.

This shows that a least favorable prior distribution exists for general $\mathcal{S}_{\mathbf{a},b}$, and also $\partial\mathcal{S}_{\mathbf{a},b}$, the boundary of $\mathcal{S}_{\mathbf{a},b}$.

- (b) Denote the least favorable prior on $\partial\mathcal{S}_{\mathbf{a},b}$ by $\mathcal{U}_{\mathbf{a},b}$. Let $\delta_{\mathbf{a},b}$ denote the Bayes estimator against this least favorable prior on $\partial\mathcal{S}_{\mathbf{a},b}$. The multivariate normal distribution being in the multiparameter exponential family, it follows in [15] that for sufficiently small $b > 0$, the function $R(\boldsymbol{\mu}, \delta_{\mathbf{a},b})$, as the risk function of $\delta_{\mathbf{a},b}$, is subharmonic on the whole of $\mathcal{S}_{\mathbf{a},b}$, and so, by the *maximum modulus theorem*, the maximum of $R(\boldsymbol{\mu}, \delta_{\mathbf{a},b})$ is attained on the boundary $\partial\mathcal{S}_{\mathbf{a},b}$ of $\mathcal{S}_{\mathbf{a},b}$, and $\delta_{\mathbf{a},b}$ is the unique minimax estimator of $\boldsymbol{\mu}$ when $\boldsymbol{\mu}$ belongs to $\mathcal{S}_{\mathbf{a},b}$. The main consequence is that when the parameter space $\mathcal{S}_{\mathbf{a},b}$ is sufficiently small, we have the sequence of equalities

$$\sup_{\pi \in \mathcal{G}} r(\pi) = \sup_{\pi \in \mathcal{G}_s} r(\pi) = \sup_{\pi \in \mathcal{G}_{s,0}} r(\pi) = r(\mathcal{U}_{\mathbf{a},b}).$$

- (c) Now consider the special case that $\mathcal{S}_{\mathbf{a},b}$ is a ball centered at \mathbf{a} with radius b . Without loss of generality, we may assume \mathbf{a} to be $\mathbf{0}$. Note now that our problem is invariant under the orthogonal group $\mathbf{X} \rightarrow \mathbf{P}\mathbf{X}$, $\boldsymbol{\mu} \rightarrow \mathbf{P}\boldsymbol{\mu}$, where \mathbf{P} is an orthogonal matrix. The invari-

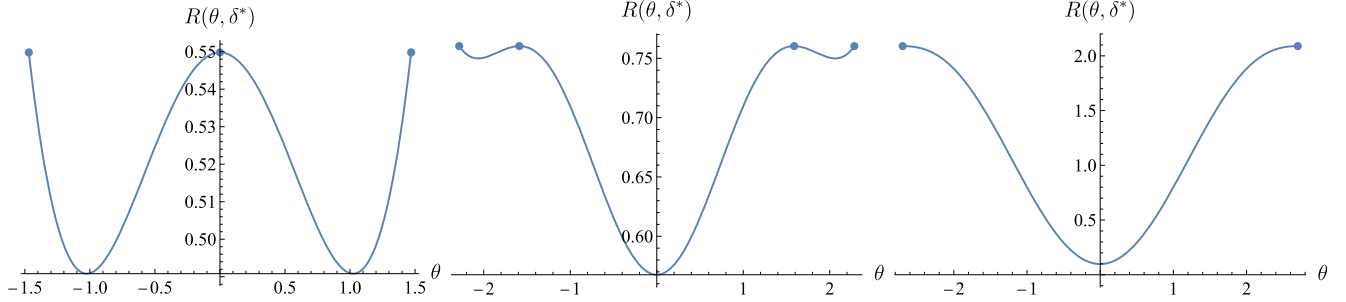


Figure 11: Model I: Frequentist risks of W-, VVV-, and V-shape, $R(\theta, \delta^*)$, for $\epsilon = 0.3$ (left), 0.5 (middle), and 0.9 (right).

ance property is true because the covariance matrix of \mathbf{X} is proportional to the identity, and $\mathcal{S}_{\mathbf{a},b}$ is a ball. By [25], $\mathcal{U}_{0,b}$ is spherically symmetric, and hence just the uniform distribution on the boundary of the ball $\mathcal{S}_{0,b}$. So, finally, for $b > 0$ sufficiently small, the estimator that is Bayes against the uniform prior on the boundary of the ball $\mathcal{S}_{0,b}$ is the unique minimax estimator of $\boldsymbol{\mu}$ when $\boldsymbol{\mu}$ belongs to $\mathcal{S}_{0,b}$, and $\mathcal{U}_{0,b}$ is the solution to the Γ -minimax problem of finding $\arg \max_{\pi \in \Gamma} r(\pi)$.

- (d) In the entire one parameter exponential family, the risk function $R(\theta, \delta)$ of any decision rule δ is a real analytic function of θ and hence has an analytic continuation to the complex plane \mathbb{C} . The Bayes rule against the least favorable prior has the constant risk equal to the minimax value on each point in the support of the least favorable prior. If the support is infinitely supported, then by the Bolzano-Weierstrass's theorem, the support has an accumulation point, and hence, by the analytic function theorem on the complex plane \mathbb{C} , the risk function of the minimax rule has a constant risk throughout the entire parameter space, which is a contradiction. This proves that the least favorable prior is finitely supported throughout the one parameter exponential family, whenever the parameter belongs to a compact interval.
- (e) In the setup of Model I, the point mass at 0 is a part of every prior; the second component $\xi(\theta)$ is as in Casella and Strawderman (1981), but non-atomic at 0, and for $m < m^*$ the least favorable prior is a three-point mass prior, with masses concentrated at $-m$, 0 , and m , with weights $(1 - \epsilon)/2$, ϵ , and $(1 - \epsilon)/2$. The threshold value m^* must be found numerically.

The corresponding Bayes rule δ^* is readily found by simplifying

$$\begin{aligned} \delta^*(d) &= \frac{\int \theta f(d|\theta) \pi(\theta) d\theta}{\int f(d|\theta) \pi(\theta) d\theta} \\ &= \frac{(1 - \epsilon)m/2 [\phi(d|m) - \phi(d|-m)]}{\epsilon\phi(d|0) + (1 - \epsilon)/2 [\phi(d|m) + \phi(d|-m)]}, \end{aligned}$$

where $\phi(d|\mu)$ is the pdf of normal $\mathcal{N}(\mu, 1)$ distribution. A simplified expression is given in (2.2).

To find values of m^* so that for $m \leq m^*$ the three point prior is least favorable, we analyze the frequentist risk, $R(\theta, \delta^*) = E^{d|\theta}(\theta - \delta^*)^2$, for a fixed ϵ , by varying m^* . Depending on ϵ , there are three possible shapes of the frequentist risk, which we denote as W, VVV, and V. Numerical work shows that values of ϵ that separate these three shapes are $\epsilon_1 \approx 0.45$ and $\epsilon_2 \approx 0.65$.

For small values of ϵ , ($< \epsilon_1$) the risk $R(\theta, \delta^*)$ is of W-shape, as in the left panel of Fig. 11.

This is a typical shape for a risk of the least favorable distribution in a class of all bounded on $[-m, m]$ distributions, for m small. The value m^* in this case is found from $R(0, \delta^*) = R(-m^*, \delta^*) = R(m^*, \delta^*)$. If $m > m^*$, the risk local maximum at 0 will exceed values at $\pm m$, and one could select a prior from Γ for which the payoff $r(\delta^*, \pi) > r(\delta^*, \pi_{m^*})$. Note that in increasing m in the search of this limiting m^* , the rule δ^* is simultaneously changing, since it depends on m , so the numerical work to find m^* is nontrivial.

For values of ϵ between ϵ_1 and ϵ_2 , the shape of frequentist risk is of VVV-type, as it is shown in middle panel of Fig. 11. In this case two local maxima for the frequentist risk appear at a pair of $\theta = \pm m_1$, $m_1 \leq m^*$. In the critical case that defines the m^* , $R(-m^*, \delta^*) = R(-m_1, \delta^*) = R(m_1, \delta^*) = R(m^*, \delta^*)$, and increasing m above such m^* will result in $R(\pm m^*, \delta_B) < R(\pm m_1, \delta_B)$. Placing more mass at $\pm m_1$ will result in higher payoff r and the three point prior is not the least favorable any longer.

The case when $\epsilon > \epsilon_2$ is most interesting since in the wavelet shrinkage, values of ϵ closer to 1 produce shrinkage rules of desirable shape. In this case, the frequentist risk is V-shaped, which flattens at the endpoints for $m = m^*$, that is, $\partial R / \partial \theta|_{\theta=m^*} = 0$ (the right panel in Fig. 11). In this case if we let $m = m^{**} > m^*$ the frequentist risk will start to decrease, so a prior with point masses that remain in points m^* , which are now inner to $[-m^{**}, m^{**}]$, will increase the payoff function r , and the three point prior with masses at $0, \pm m^{**}$ will not be the least favorable any longer.

APPENDIX B. MODEL II

If in Model II the normal likelihood is replaced by the marginal likelihood, after σ^2 is integrated out, then we get

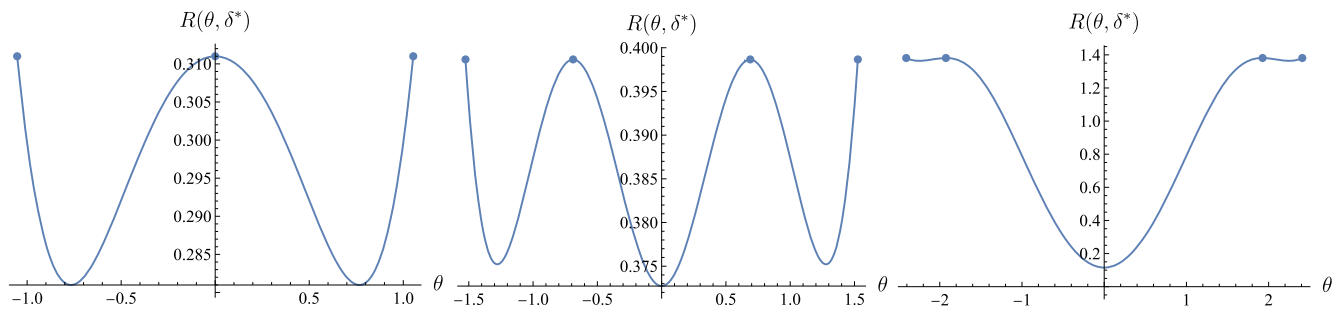


Figure 12: Model II: Frequentist risk of the Γ -minimax rule, $R(\theta, \delta^*)$, for $\epsilon = 0.2$ (Left), 0.4 (Middle), and 0.9 (Right).

a double exponential density. In Model II, there are two risk shapes, W and VVV , for $\epsilon \leq \epsilon_1$ and $\epsilon > \epsilon_1$ respectively, where ϵ_1 is between 0.3 and 0.4 . We show some plots below.

The argument is similar as in Model I, for the W -shape (Left panel in Fig. 12), a slight increase of m over m^* will make $R(\delta^*, 0) > R(\delta^*, \pm m)$ and one can choose $\pm m_1$ in the neighborhood of 0 , so that transferring some point mass from endpoints to $\pm m_1$ would increase the payoff. The argument for VVV -shaped risk (Middle and Right Panels in Fig. 12) is the same as in the Model I.

ACKNOWLEDGEMENTS

The authors thank the editor and the three anonymous referees for careful reading of the manuscript and constructive comments.

FUNDING

The work of Dixon Vimalajeewa was supported by H. O. Hartley endowment at Texas A&M University.

Accepted 15 June 2023

REFERENCES

- [1] ABRAMOVICH, F., BESBEAS, P. and SAPATINAS, T. (2002). Empirical Bayes approach to block wavelet function estimation. *Computational Statistics & Data Analysis* **39**(4) 435–451. [https://doi.org/10.1016/S0167-9473\(01\)00085-8](https://doi.org/10.1016/S0167-9473(01)00085-8). MR1910021
- [2] AMATO, U. and VUZA, D. T. (1997). Wavelet approximation of a function from samples affected by noise. *Rev. Roumanie & Mathematics with Applications* **36**(5) 101–111. [https://doi.org/10.1016/S0898-1221\(98\)00153-9](https://doi.org/10.1016/S0898-1221(98)00153-9). MR1639374
- [3] ANGELINI, C. and VIDAKOVIC, B. (2004). Γ -minimax wavelet shrinkage: A robust incorporation of information about energy of a signal in denoising applications. *Statistica Sinica* **14**(1) 103–125. MR2036764
- [4] ANTONIADIS, A., BIGOT, J. and SAPATINAS, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software* **6** 1–83. MR4024218
- [5] BERGER, J. O. (1984). The Robust Bayesian Viewpoint. In *Robustness of Bayesian Analysis* (J. K. Eds, ed.), 63–124. Elsevier Science Publisher, USA. MR0785367
- [6] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York. <https://doi.org/10.1007/978-1-4757-4286-2>. MR0804611
- [7] BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is Restricted. *The Annals of Statistics* **9**(6) 1301–1309. MR0630112
- [8] BICKEL, P. J. and COLLINS, J. R. (1983). Minimizing Fisher information over mixtures of distributions. *Sankhyā: The Indian Journal of Statistics, Series A (1961–2002)* **45**(1) 1–19. MR0749349
- [9] BISCHOFF, W. and FIEGER, W. (1992). Minimax estimators and Γ -minimax estimators for a bounded normal mean under the loss $\ell_p(\theta, d) = |\theta - d|^p$. *Metrika* **39** 185–197. <https://doi.org/10.1007/BF02614000>. MR1173577
- [10] CAI, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *The Annals of Statistics* **27**(3) 898–924. <https://doi.org/10.1214/aos/1018031262>. MR1724035
- [11] CASELLA, G. and STRAWDERMAN, W. E. (1981). Estimating a bounded normal mean. *The Annals of Statistics* **9**(4) 870–878. MR0619290
- [12] CHIPMAN, H. A., KOLACZYK, E. D. and MCCULLOCH, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association* **92**(440) 1413–1421.
- [13] CLARKE, B. S. and BARRON, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference* **41** 37–60. [https://doi.org/10.1016/0378-3758\(94\)90153-8](https://doi.org/10.1016/0378-3758(94)90153-8). MR1292146
- [14] CLYDE, M. and GEORGE, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4) 681–698. <https://doi.org/10.1111/1467-9868.00257>. MR1796285
- [15] DASGUPTA, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhyā: The Indian Journal of Statistics, Series A (1961–2002)* **47**(3) 326–332. MR0863726
- [16] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**(3) 425–455. <https://doi.org/10.1093/biomet/81.3.425>. MR1311089
- [17] DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *The Annals of Statistics* **18**(3) 1416–1437. <https://doi.org/10.1214/aos/1176347758>. MR1062717
- [18] FERGUSON, T. S. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York. MR0215390
- [19] GEORGE, E. I., LIANG, F. and XU, X. (2012). From minimax shrinkage estimation to minimax shrinkage prediction. *Statistical Science* **27**(1) 82–94. <https://doi.org/10.1214/11-STS383>. MR2953497
- [20] GHOSH, M. N. (1964). Uniform approximation of minimax point estimates. *The Annals of Mathematical Statistics* **35**(3) 1031–1047. <https://doi.org/10.1214/aoms/1177703262>. MR0164418
- [21] GOOD, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B (Methodological)* **14**(1) 107–114. MR0077033
- [22] HUANG, H. -C. and CRESSIE, N. (2000). Deterministic/stochas-

- tic wavelet decomposition for recovery of signal from noisy data. *Technometrics* **42**(3) 262–276. <https://doi.org/10.2307/1271081>. [MR1801033](#)
- [23] HUBER, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35**(1) 73–101. <https://doi.org/10.1214/aoms/1177703732>. [MR0161415](#)
- [24] IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981) *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York. [MR0620321](#)
- [25] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *The Annals of Mathematical Statistics* **28**(3) 573–601. <https://doi.org/10.1214/aoms/1177706874>. [MR0092325](#)
- [26] LEHMANN, E. L. (1959) *Testing Statistical Hypotheses*. Wiley, New York. [MR0107933](#)
- [27] PARK, T. and CASELLA, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**(482) 681–686. <https://doi.org/10.1198/016214508000000337>. [MR2524001](#)
- [28] REMÉNYI, N. and VIDA KOVIC, B. (2013). Bayesian wavelet shrinkage strategies: A review. In *Multiscale Signal Analysis and Modeling*, 317–346. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-4145-8_14. [MR3024475](#)
- [29] ROBBINS, H. E. (1951). Asymptotically subminimax solutions of compound statistical decision Problems. In *Proceedings of Second Berkeley Symposium Mathematics and Statistics and Probability*, 241–259. University of California Berkeley Press, California. [MR0044803](#)
- [30] SHANG, Z. (2011) *Bayesian variable selection: Theory and applications* **73**. Citeseer. [MR2982337](#)
- [31] SOUSA, A. R. D. S., GARCIA, N. L. and VIDA KOVIC, B. (2021). Bayesian wavelet shrinkage with beta priors. *Computational Statistics* **36** 1341–1363. <https://doi.org/10.1007/s00180-020-01048-1>. [MR4255812](#)
- [32] VIDA KOVIC, B. (2000). Γ -minimax: A paradigm for conservative robust Bayesians. In *Robust Bayesian Analysis* 241–259. Springer. https://doi.org/10.1007/978-1-4612-1306-2_13. [MR1795219](#)
- [33] VIDA KOVIC, B. and RUGGERI, F. (2001). BAMS method: Theory and simulations. *Sankhyā: The Indian Journal of Statistics, Series B* **63**(2) 234–249. [MR1895791](#)
- [34] WIJSMAN, R. A. (1970). Continuity of the Bayes risk. *The Annals of Mathematical Statistics* **41**(3) 1083–1085.
- Dixon Vimalajeewa. Texas A&M University, College Station, TX, USA. E-mail address: dixon.vimalajeewa@tamu.edu
- Anirban DasGupta. Purdue University, West Lafayette, IN, USA. E-mail address: dasgupta@purdue.edu
- Fabrizio Ruggeri. CNR–IMATI, Milan, Italy. E-mail address: fab.ruggeri@tiscali.it
- Brani Vidakovic. Texas A&M University, College Station, TX, USA. E-mail address: brani@stat.tamu.edu