Sankhyā : The Indian Journal of Statistics
Special issue on Wavelets
2001, Volume 63, Series B, Pt. 2, pp. 234–249

BAMS METHOD: THEORY AND SIMULATIONS

By BRANI VIDAKOVIC
Georgia Institute of Technology, Atlanta
and
FABRIZIO RUGGERI
CNR-IAMI, Milano

SUMMARY. In this paper we address the problem of model-induced wavelet shrinkage. Assuming the *independence* model according to which the wavelet coefficients are treated individually, we discuss a level-adaptive Bayesian model in the wavelet domain that has two important properties: (i) it realistically describes empirical properties of signals and images in the wavelet domain, and (ii) it results in simple optimal shrinkage rules to be used in fast wavelet denoising. The proposed denoising paradigm **BAMS** (short for Bayesian Adaptive Multiresolution Shrinker) is illustrated on an array of Donoho and Johnstone's standard test functions and is compared to some standard wavelet-based smoothing methods.

1. Introduction

Wavelet based statistical shrinkage procedures proposed by Donoho, Johnstone, and their co-authors in the early 1990's (Donoho and Johnstone, 1994, Donoho *et al.*, 1995, and references in Donoho, 1997) have proven to be an excellent data analytic tool in function/density estimation problems. The shrinkage was performed by thresholding-type rules and the optimal thresholds had been established either by an asymptotic minimax, exact risk, or penalized mean square error calculation.

Subsequent vibrant research focussed on several generalizations of the original Donoho-Johnstone paradigm; see Vidakovic (1999) for a general overview.

Paper received. July 2000.
AMS (2000) *subject classification.* Primary 65T60; secondary 62F15.
Key words and phrases. Wavelet regression, shrinkage, adaptivity, denoising.

A variety of methods, based on Bayes estimators of the "signal part" in an observed wavelet coefficient, have shown to be capable of incorporating prior information about the unknown signal (energy content, smoothness, periodicity, and self-similarity, for instance). Many aspects regarding this Bayesian approach to wavelet shrinkage can be found in contributed chapters in the volume by Müller and Vidakovic (1999).

In this paper we propose a shrinkage method, based on a Bayesian model in the wavelet domain, that addresses some of the shortcomings of the previously proposed Bayesian methods. We discuss these shortcomings and present our new model in Section 2. Section 2 also addresses the important issue of tuning the model hyperparameters with the goal of general usability of the proposed method. Section 3 contains the simulational study performed on the standard array of test-functions and comparisons with some other methods. The proposed **BAMS** (**B**ayesian **A**daptive **M**ultiresolution **S**hrinker) method is implemented in MATLAB and made available to the public[1].

2. The Model

We now describe the statistical model, the rationale behind it, and the resulting optimal shrinkage rules.

Suppose the observed data $\mathbf{y}$ represent the sum of an unknown signal $\mathbf{f}$ and random noise ϵ. Coordinate-wise, $\mathbf{y}_i = \mathbf{f}_i + \epsilon'_i,\ i = 1, \ldots, n$. In the wavelet domain (after applying a linear and orthogonal wavelet transformation W to the observed data), expression (1) becomes $d_{jk} = \theta_{jk} + \epsilon_{jk},\ i = 1, \ldots, n$, where d_{jk}, θ_{jk}, and ϵ_{jk} are the j,k-th coordinates in the traditional scale/shift wavelet-enumeration of vectors $W\mathbf{y}, W\mathbf{f}$ and $W\epsilon'$, respectively. Our assumption is that the coefficients d_{jk} can be considered independently, since the wavelet transformations are decorrelating. When modelling in practice, such an assumption prove to be very reasonable. In the exposition that follows, we omit the double index jk and work with a "typical" wavelet coefficient, d. Therefore, our model is

$$d = \theta + \epsilon. \tag{1}$$

It is now standard practice in wavelet shrinkage to specify a location model on the wavelet coefficients, elicit the prior on their locations (the signal

[1]All program codes, images, and simulations can be found at the web-site `www.isye.gatech.edu/~brani/wavelet.html`. This is in the spirit of Donoho's initiative for reproducible research; see Buckheit and Donoho (1995).

part in wavelet coefficients), exhibit the Bayes estimator for the locations and, if the resulting Bayes rules are shrinkage estimators, apply the inverse wavelet transformation to the estimators.

In considering this model-induced shrinkage, the main concern was, of course, performance of the induced shrinkage rules, measured by the realized mean square error, while no attention was paid to the match between models and data in the wavelet domain. It is certainly desirable for selected models to well-describe our empirical observations, for the majority of signals and images. At the same time, the calculation of shrinkage rules should remain inexpensive. Our experience is that the realistic but complicated models, for which the rules are obtained by extensive MCMC simulations, are seldom accepted by practitioners, despite their reportedly good performance.

We demonstrate that two desirable goals, *simplicity and reality*, can be achieved simultaneously. We adopt a paradigmatic location model in which the wavelet coefficients d are modelled via a density $f(d-\theta)$ where θ is the signal part. The same model can be used for all coefficients, or, with slight modifications, to be localized with respect to scale and/or time, depending on the nature of our observations.

We discuss the model building in stages: the likelihood, the priors, the calculation of the Bayes rule and selection of the hyperparameters.

2.1 *The likelihood.* We assume, as is commonly done, that the errors are normal, and thus $[d|\theta,\sigma^2]$ is distributed as $\mathcal{N}(\theta,\sigma^2)$.

Many researchers have discussed their findings about the empirical distribution of the wavelet coefficients; see, for example, Buccigrossi and Simoncelli (1999), Leporini *et al.* (1999), Mallat (1989), Ruggeri (1999), Simoncelli (1999), and Vidakovic (1998). Their opinions can be summarized in the following statement:

"For most of the signals and images encountered in practice, the empirical distribution of a typical detail wavelet coefficient is notably centered about zero and peaked at it."

Mallat (1989) used the exponential power distribution family $\mathcal{EPD}$ [2] as a

---

[2]The $\mathcal{EPD}(\alpha,\beta)$ family has density given by

$$f(d) = C\cdot e^{-(|d|/\alpha)^\beta},\ \alpha,\beta>0.$$

The normalizing constant is $C=\frac{\beta}{2\alpha\Gamma(1/\beta)}$. Since

$$E|D| = \alpha\frac{\Gamma(2/\beta)}{\Gamma(1/\beta)} \quad \text{and} \quad ED^2 = \alpha^2\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)},$$

the parameters α and β can be estimated from the data.

model to fit empirical distributions of wavelet coefficients. After estimating the parameters of the $\mathcal{EPD}$ from the coefficients, quantile based thresholding rules could be established, which originally was the reason for considering the $\mathcal{EPD}$ model.

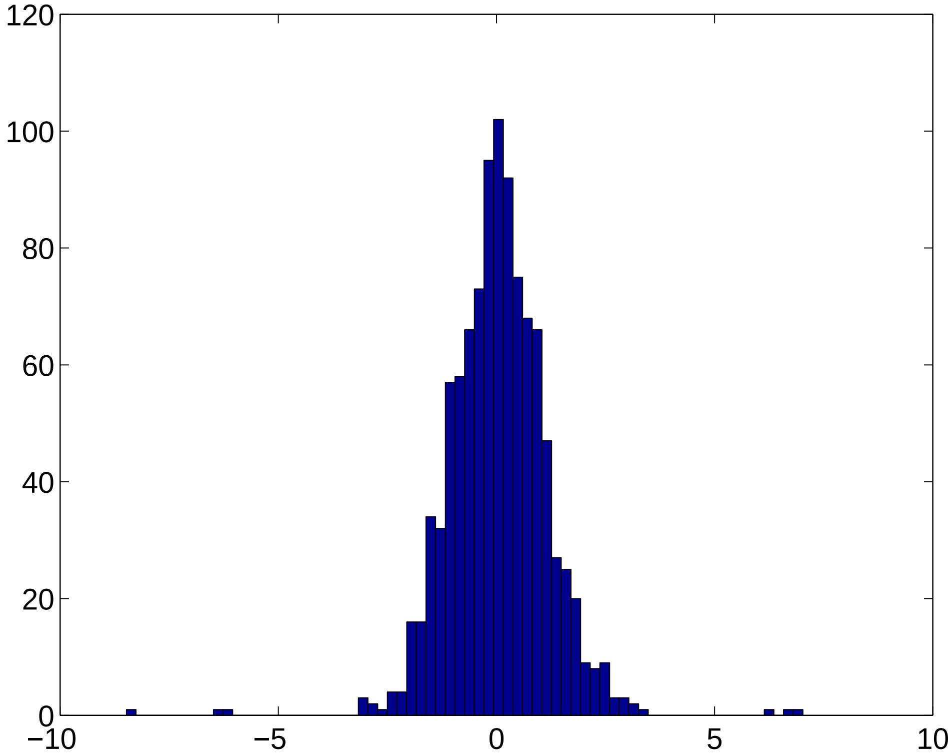


Figure 1. Empirical distribution of coefficients from 10th detail level (15th level is the finest, $n = 2^{16}$) of a noisy `doppler` signal.

A Bayesian model is realistic if the marginal (predictive) distribution of the observations "agrees" with the observations. In our approach, we require the marginal distribution of the wavelet coefficients to be close to the model in Mallat (1989).

Such marginal "matching" modelling is impossible if plug-in estimators of σ^2 are used. The argument is as follows:

Suppose that in the normal $\mathcal{N}(\theta, \sigma^2)$ model on $[d|\theta, \sigma^2]$ one replaces σ^2 with an estimator $\hat{\sigma}^2$. Then, the marginal likelihood for the location remains normal, and the corresponding overall marginal is a location mixture of normals. It is a well-known fact that location mixtures of normals (convolutions) result in densities which are flatter than the normal distribution itself. Since the family $\mathcal{EPD}$ is a scale mixture of normals and the exponential mixing distribution is supported on $[0, \infty)$; it follows that the $\mathcal{EPD}$ cannot possibly be a normal convolution because of the following result due to DasGupta (1992).

THEOREM **1** *Let the density f be a scale mixture of normals, given by*

$$f(x) = \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x}{2\sigma^2}} \nu(d\sigma^2),$$

for some scale-mixing measure ν. Then f is a normal convolution (location mixture of normals) iff $\nu([0, 1]) = 0$, i.e., iff the mixing measure ν does not have support in $[0, 1]$.

In view of the previous facts, it is advisable to assume that the $\mathcal{EPD}$ marginal $f(d|\theta)$ is a scale mixture of normals. We propose the double exponential distribution $\mathcal{DE}$ (i.e. $\beta = 1$ in $\mathcal{EPD}$) as a realistic, Bayesian-justified, and mathematically manageable choice. Note that we will be using $\mathcal{DE}$ to denote both the double exponential law and its density function; it will be clear from context which one is being considered.

2.2 *Prior selection.* In the previous discussion we suggested the choice of a double exponential distribution as a partially integrated likelihood (marginal likelihood on location). The double exponential can be obtained by marginalizing a normal likelihood by adopting exponential prior on its variance.

It is well-known that the exponential distribution is an entropy maximizer in the class of all distributions supported on $(0, \infty)$ with a fixed first moment. Thus our choice is a *maxent* prior. That is, it is the noninformative one in the above class.

Our choice of prior on the signal part is based on inspecting the empirical realizations of coefficients of pure signals (noiseless data). It is worth mentioning that the choice of a normal model for d and a double exponential on θ was recommended by Ruggeri and Vidakovic (1999) in a different context – when looking for pairs (model/prior) ensuring the existence of an optimal thresholding rule according to the Bayesian decision theoretic paradigm. Also, Vidakovic (1998) uses the same marginal likelihood as suggested here but with general priors on θ without point mass at zero; resulting in shrinkage rules that are not simple.

The scale in $\xi(\theta)$ is assumed fixed, although it could depend on level j in the decomposition, as was suggested by Clyde *et al.* (1998) in a different modelling setup. Clyde *et al.* (1998) propose Inverse-Gamma prior on σ^2. Due to this choice and additional level of hierarchy in the model, their method does not produce an effective Bayes rule. However, simple approximations are available.

In the wavelet context, Müller (1994) first suggested the priors on θ that contain a point mass at zero; that is, the ϵ–contamination

$$\pi(\theta) = \epsilon\delta(0) + (1-\epsilon)\xi(\theta). \tag{2}$$

This additional point mass at zero ensures the enhanced shrinkage in the resulting Bayes rule, as shown in Vidakovic and Ruggeri (1999). The point mass at zero, δ_0, is the shrinkage inducer whereas $\xi(\theta)$ (in our case $\mathcal{DE}$) is a spread distribution that models wavelet coefficients with large energies (squared magnitudes).

Besides, adequate changes in ϵ make it possible to adapt the shrinkage rules level-wise. In addition, when ϵ increases, the full marginal on d will look increasingly like the "spiky" marginal likelihood. Several functions $\epsilon(j)$ have already been proposed (Clyde *et al.*, 1998). We will discuss our proposal in Section 2.4.

2.3 *Derivation of the shrinkage rule.* In this section, we give a summary of our model and the analytic form of the Bayesian shrinkage rule, postponing its computation to the Appendix.

Starting with $[d|\theta,\sigma^2] \sim N(\theta,\sigma^2)$ and the prior $\sigma^2 \sim \mathcal{E}(\mu)$, $\mu > 0$, with density $f(\sigma^2|\mu) = \mu e^{-\mu\sigma^2}$, we obtain the marginal likelihood

$$d|\theta \sim \mathcal{DE}\left(\theta, \frac{1}{\sqrt{2\mu}}\right), \qquad \text{with density } f(d|\theta) = \frac{1}{2}\sqrt{2\mu}e^{-\sqrt{2\mu}|d-\theta|}.$$

If the prior on θ is

$$[\theta] \sim \mathcal{DE}(0,\tau),$$

then the predictive distribution of d is

$$[d] \sim m(d) = \frac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}}e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu},$$

and the corresponding Bayes rule with respect to the squared error loss is

$$\delta(d) = \frac{\tau(\tau^2 - 1/(2\mu))de^{-|d|/\tau} + \tau^2(e^{-|d|\sqrt{2\mu}} - e^{-|d|/\tau})/\mu}{(\tau^2 - 1/(2\mu))(\tau e^{-|d|/\tau} - (1/\sqrt{2\mu})e^{-|d|\sqrt{2\mu}})}. \tag{3}$$

Figure 2(a) depicts the rule in (3). Notice that this is a shrinkage rule, but it is close to a linear shrinkage rule, known to be under-performing in wavelet-based methods.

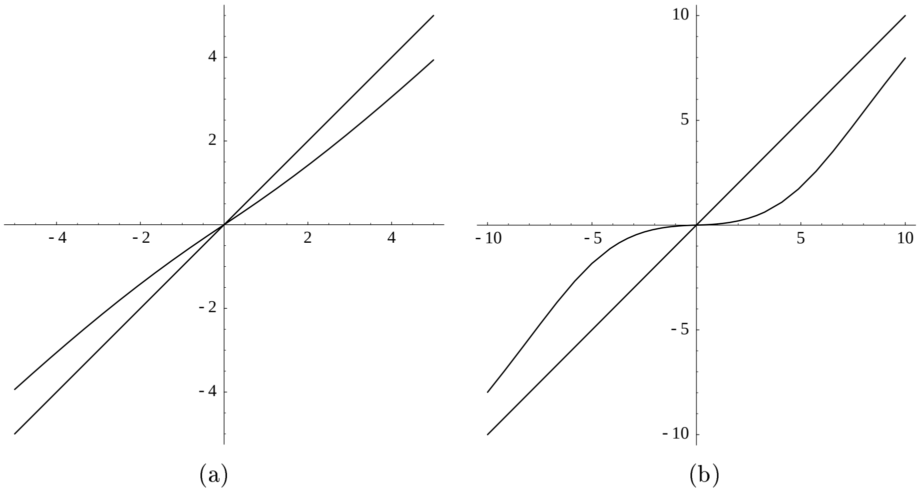


Figure 2. Bayes rules: (a) No point mass, $\tau = 2$, $\mu = 1/2$. (b) $\epsilon = 0.9, \tau = 2$, $\mu = 1/2$.

Rules with a more desirable shape result from the ϵ-contaminated priors (2). When

$$[\theta|\epsilon] \sim \epsilon\delta_0 + (1-\epsilon)\mathcal{DE}(0,\tau), \tag{4}$$

the marginal is

$$d \sim m^*(d) = \epsilon\mathcal{DE}(0, \frac{1}{\sqrt{2\mu}}) + (1-\epsilon)m(d)$$

and the Bayes rule is:

$$\delta^*(d) = \frac{(1-\epsilon)\ m(d)\ \delta(d)}{(1-\epsilon)\ m(d) + \epsilon\ \mathcal{DE}\left(0, \frac{1}{\sqrt{2\mu}}\right)}. \tag{5}$$

Figure 2(b) depicts the rule in (5). Notice that this rule is close to a thresholding rule: it heavily shrinks small-in-magnitude arguments while the large ones are shrunk slightly. As presented in Figure 3(a), the Bayes rule (5) falls between comparable hard and soft thresholding rules. For comparison purposes, the exact risk of the three rules is shown in Figure 3(b). Notice that the risk ($R(\theta, \delta) = E^{d|\theta}(\theta - \delta(d))^2$) for the Bayes rule is smaller than the risk for the thresholding rules for small values of argument θ.

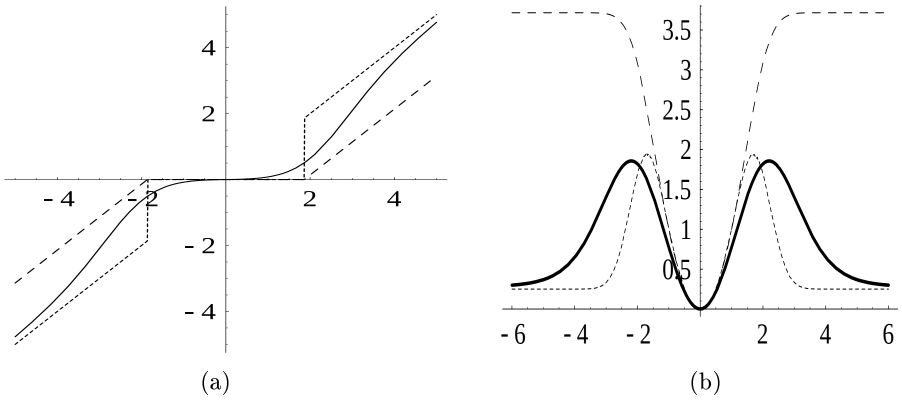


Figure 3. (a) Bayes rule (3) and comparable hard and soft thresholding rules. (b) Exact risk plots for the rules from (a). Notice that the risk for the Bayes rule generally falls between risks for the thresholding rules. For small values of θ, the risk of the Bayes rule is smaller than the risk of the thresholding rules.

2.4 *Tuning the model parameters.* One of the main issues in any Bayesian analysis is the eliciting of the hyperparameters specifying the statistical model. Often the hyperparameters may have their own priors, leading to hierarchical models.

The described model depends on 3 hyperparameters that have to be specified. Purely subjective eliciting of priors is difficult since, in general, users may lack intuition and interpretation of wavelet-domain priors. Subjective priors can be easily elicited only when the prior information concerns smoothness or self-similarity. A variety of default solutions is available (see Yang and Berger, 1997), but default choices do not seem to be very suitable in function estimation, since observations can vary tremendously and some degree of informativeness and/or data dependence should be exploited.

We propose an empirical (moment matching) parameter specification that works well for standard test cases and emphasizes that the nature of the data may call for different parameter values.

Data dependent specifications of the three parameters via Maximum-likelihood II (ML-II) procedure may provide alternative procedures. The ML-II procedure is a standard Empirical Bayes tool which mimics the maximum likelihood procedure of estimating parameters, but with predictive distributions in place of likelihoods.

1. μ is the reciprocal of the mean for the prior on σ^2, or, equivalently, the square root of the precision for σ^2. We first estimate σ by a robust Tukey's $\texttt{pseudos} = (Q_1 - Q_3)/C$, where Q_1 and Q_3 are the first and the third quartiles of the finest level of details in the decomposition and $1.3 \leq C \leq 1.5$. We propose $\frac{1}{\texttt{pseudos}}$ as a default value for μ. According to the Law of Large Numbers, this ratio should be close to the "true" μ.

2. ϵ is the weight of the point mass at zero in the prior on θ and should depend on level j. Depending on our prior information about smoothness, ϵ should be close to 1 at the finest level of detail and close to 0 at the coarsest level. We propose a hyperbolic decay in j,

$$\epsilon(j) = 1 - \frac{1}{(j - \texttt{coarsest} + 1)^{\gamma}}, \quad \texttt{coarsest} \leq j \leq \log_2 n,$$

where $\texttt{coarsest}$ is the coarsest level subjected to shrinkage. For the standard test functions, various sample sizes, and various signal-to-noise ratios, values $\texttt{coarsest} = 3$ and $\gamma = 1.5$ worked the best.

3. τ is the scale of the "spread part" in the prior (4). In the case of a double exponential prior, the variance of the signal part is $2\tau^2$. Because of the independence between the error and the signal parts, we have $\sigma_d^2 = 2(1-\epsilon(j))^2\tau^2 + 1/\mu$, where σ_d^2 is the variance of the observations d. Taking into account that ϵ depends on j, we adopted $2(1-\epsilon(j))^2 \approx 1$ as a mid-point of range of ϵ. Such a choice yields

$$\tau = \sqrt{\max\{(\sigma_d^2 - \frac{1}{\mu}), 0\}}\ .$$

Note when $\tau = 0$, the prior and the posterior put all their mass at 0, which results in $\delta(d) = 0$.

The simulations provided in the next section are performed with these default values for model hyperparameters.

3. Simulations

To assess the performance of the procedure we considered the standard test functions and compared the average mean square error AMSE, variance and squared bias of BAMS estimators to those listed in Chipman *et al.* (1997) [Table 1, page 1420; VisuShrink, SureShrink and ABWS methods]. If N is the number of simulations, the AMSE of the estimator $\hat{f} = (\hat{f}_1, \ldots, \hat{f}_n)$ is calculated as

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

where f_i are the components to be estimated and $\hat{f}_{ij}$ are corresponding estimates in jth simulational run.

VisuShrink (Donoho and Johnstone, 1994) is a procedure which, although not intended to minimize overall MSE, minimizes variance, retains asymptotic minimax optimality, and results in visually pleasing reconstructions. SureShrink (Donoho and Johnstone, 1995) is an adaptive procedure that uses different thresholding strategies levelwise depending on the energies of the coefficients in the levels. ABWS (Adaptive Bayesian Wavelet Shrinkage; Chipman *et al.*, 1997) is an effective, level-dependent Bayesian procedure in which the prior on the signal part is a mixture of two normals with different scales. One of the normal components in the mixture-prior has small variance, thus approximating standard point mass at zero component.

We should note that simple Bayes rules are achieved by Chipman *et al.* (1997). Their ABWS method matches the shape of the observed distribution of wavelet coefficients at prior-level by appropriate mixture of two normal distributions. The ABWS rules are almost thresholding and their performance is excellent.

Four standard test functions (`blocks, bumps, doppler` and `heavisine`) are rescaled so that an added standard normal noise produces a signal-to-noise ratio (SNR) of 7. The sample size is $n = 1024$ and the wavelets we used are: Symmlet 8 for `doppler` and `heavisine` (8 stands of number of taps in the wavelet filter), Haar for `blocks` and Daubechies 6 for `bumps`. The MSE is split into the variance and squared bias parts for comparison purpose. $N = 1000$ simulations are summarized in table 1. Since the shrinkage rule (5) is given in a closed form and it is fast to implement, the simulation of $N = 1000$ runs using MATLAB and WAVELAB takes several seconds on Ultra 80 SUN Station.

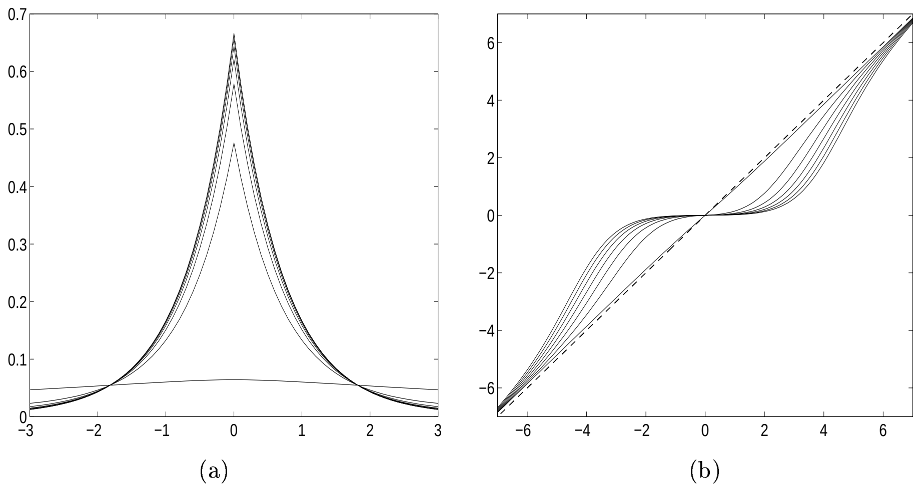
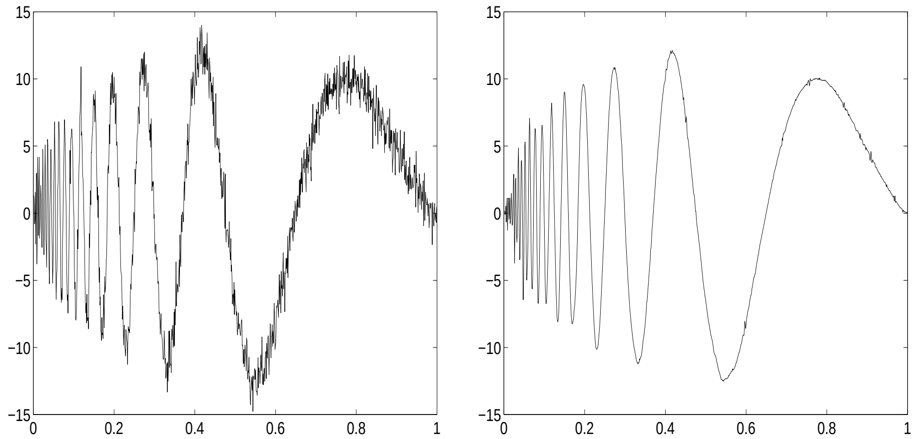


Figure 4. (a) Level-wise marginal distributions for the wavelet coefficients superimposed; (b) Shrinkage Bayes rules corresponding to marginals in panel (a).

Table 1. MSE (Variance+Bias^2) for VisuShrink, SureShrink, ABWS, and BAMS on standard test functions. Test signals are rescaled so that SNR equals 7 and the noise variance σ^2 equals 1.

	`blocks`	`bumps`
VisuShrink	0.6840 (0.0719 + 0.6122)	1.5707 (0.1165 + 1.4543)
SureShrink	0.2225 (0.1369 + 0.0856)	0.6827 (0.2660 + 0.4167)
ABWS	0.0995 (0.0874 + 0.0121)	0.3495 (0.2228 + 0.1267)
BAMS	0.1107 (0.0965 + 0.0142)	0.3404 (0.1976 + 0.1428)
BAMS (100,000 runs)	0.1108 (0.0962 + 0.0146)	0.3378 (0.1967 + 0.1411)
	`doppler`	`heavisine`
VisuShrink	0.4850 (0.0523 + 0.4327)	0.1204 (0.0339 + 0.0864)
SureShrink	0.2285 (0.0946 + 0.1340)	0.0949 (0.0416 + 0.0534)
ABWS	0.1646 (0.1006 + 0.0640)	0.0874 (0.0442 + 0.0433)
BAMS	0.1482 (0.0899 + 0.0584)	0.0815 (0.0511 + 0.0304)
BAMS (100,000 runs)	0.1474 (0.0883 + 0.0591)	0.0805 (0.0504 + 0.0301)

We repeated the simulation with $N = 100,000$ and reported the results in Table 1; we found only minor differences from the numbers obtained in $N = 1,000$ simulations. Levelwise marginal distributions for the coefficients are superimposed at the panel (a) in Figure 4. Figure 4 (b) shows the shrinkage rules corresponding to marginals in Figure 4(a). Note that the marginals support Mallat's findings and that the shrinkage rules can be thought as a continuous compromise between hard and soft tresholding rules. Clyde *et al.* (1998) report average mean-square errors comparable to both ABWS and BAMS. For a particular selection of prior on ϵ and $n = 1024$, they report **blocks**: 0.106, **bums**: 0.326, **doppler**: 0.151, and **heavisine**: 0.096.



(a) (b)

Figure 5. (a) A noisy `doppler` signal [SNR=7, n=1024, noise variance $\sigma^2 = 1$]. (b) Signal reconstructed by BAMS.

Table 2 gives an extensive simulational study of the BAMS method only. For $N = 1000$ runs, we were interested in MSE for four different SNR's (3, 5, 7, 10) and five sample sizes n (256, 512, 1024, 2048, 4096).

4. **Conclusions**

The shrinkage procedure in the BAMS method is adaptive in scale, and with minor modifications can also be made adaptive in time. This additional time-adaptivity can be introduced by taking into account local energies from some coarse level of detail. This "energy adjusting" affects ϵ - the amount of the point mass at zero; see Vidakovic and Bielza (1998) for an implementation of such time-adaptivity.

Table 2. MSE FOR BAMS: A VARIETY OF SAMPLE SIZES AND SNR'S.

FUNCTION	n	SNR=3	SNR=5	SNR=7	SNR=10
BLOCKS	256	0.3343	0.2835	0.2412	0.2080
	512	0.2101	0.1943	0.1763	0.1567
	1024	0.1583	0.1311	0.1107	0.0942
	2048	0.0921	0.0788	0.0665	0.0549
	4096	0.0560	0.0458	0.0390	0.0343
BUMPS	256	0.6419	0.6996	0.7554	0.8607
	512	0.4834	0.5132	0.5573	0.6093
	1024	0.2969	0.3263	0.3404	0.3508
	2048	0.1823	0.1978	0.2049	0.2144
	4096	0.1009	0.1070	0.1114	0.1176
DOPPLER	256	0.3378	0.3821	0.3887	0.4114
	512	0.1954	0.2131	0.2264	0.2391
	1024	0.1180	0.1350	0.1482	0.1590
	2048	0.0687	0.0783	0.0868	0.0939
	4096	0.0484	0.0497	0.0487	0.0460
HEAVISINE	256	0.1462	0.1754	0.1985	0.2245
	512	0.0957	0.1185	0.1374	0.1584
	1024	0.0607	0.0707	0.0815	0.0958
	2048	0.0402	0.0471	0.0531	0.0611
	4096	0.0332	0.0351	0.0363	0.0382

At the expense of losing the closed-form of the shrinkage rule, it is possible to build additional levels of hierarchy in the model; for example,

$$[\epsilon|\alpha,\beta] \sim \mathcal{B}eta(\alpha,\beta)$$

where α and β are hyperparameters.

The hyperparameters could be found by an ML-II procedure; the product of predictive distributions [from a particular level] is maximized with respect to hyperparameters:

$$\text{argmax} \prod_{d_i \in \text{level } j} m^*(\epsilon,\mu,\tau|d_i).$$

The issues of identifiability of parameters and efficiency of the optimization procedure are yet to be resolved.

It should be also stressed that our proposal is probably only one in the array of reasonable models. The selection we made was dictated by simplicity of the rule and relation of the marginal to the observed data. A MATLAB program implementing BAMS is available at:

`http://www.isye.gatech.edu/~brani/wavelet.html`.

Also, BAMS method is implemented in the MATLAB toolbox of Antoniadis, Bigot, and Sapatinas (Antoniadis *et al.*, 2001) available at http://www-lmc.imag.fr/SMS/software/wavden/.

Acknowledgment. We thank the Editor and two anonymous referees for their insightful comments.

Appendix

In this Appendix we present the derivation of the shrinkage rule (3).

Consider $d|\theta \sim \mathcal{DE}\left(\theta, \frac{1}{\sqrt{2\mu}}\right)$, $\theta \sim \mathcal{DE}(0,\tau)$. It follows that the (prior) predictive density is

$$m(d) = \frac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}} e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu}$$

and the Bayes rule is

$$\delta(d) = \frac{\tau(\tau^2 - 1/(2\mu)) d e^{-|d|/\tau} + \tau^2(e^{-|d|\sqrt{2\mu}} - e^{-|d|/\tau})/\mu}{(\tau^2 - 1/(2\mu))(\tau e^{-|d|/\tau} - (1/\sqrt{2\mu}) e^{-|d|\sqrt{2\mu}})}.$$

PROOF. Assume $d > 0$.

$$\begin{aligned} m(d) &= \sqrt{2\mu}/(4\tau)\left\{ e^{-d\sqrt{2\mu}} \int_{-\infty}^{0} e^{\theta\sqrt{2\mu}+\theta/\tau} d\theta + e^{-d\sqrt{2\mu}} \int_{0}^{d} e^{\theta\sqrt{2\mu}-\theta/\tau} d\theta + \right. \\ &\quad \left. + e^{d\sqrt{2\mu}} \int_{d}^{\infty} e^{-\theta\sqrt{2\mu}-\theta/\tau} d\theta \right\} \\ &= \sqrt{2\mu}/(4\tau)\left\{ \frac{e^{-d\sqrt{2\mu}}}{\sqrt{2\mu}+1/\tau} + \frac{e^{-d\sqrt{2\mu}}}{\sqrt{2\mu}-1/\tau}\left(e^{d\sqrt{2\mu}-d/\tau} - 1\right) \right\} + \frac{e^{-d/\tau}}{\sqrt{2\mu}+1/\tau} \\ &= \frac{\tau e^{-d/\tau} - \frac{1}{\sqrt{2\mu}} e^{-\sqrt{2\mu} d}}{2\tau^2 - 1/\mu}. \end{aligned}$$

The Bayes rule can be expressed as $\delta(d) = N_0/m(d)$, where

$$\begin{aligned} N_0 &= \sqrt{2\mu}/(4\tau)\left\{ e^{-d\sqrt{2\mu}} \int_{-\infty}^{0} \theta e^{\theta\sqrt{2\mu}+\theta/\tau} d\theta + e^{-d\sqrt{2\mu}} \int_{0}^{d} \theta e^{\theta\sqrt{2\mu}-\theta/\tau} d\theta + \right. \\ &\quad \left. + e^{d\sqrt{2\mu}} \int_{d}^{\infty} \theta e^{-\theta\sqrt{2\mu}-\theta/\tau} d\theta \right\} \\ &= \sqrt{2\mu}/(4\tau)\left\{ \frac{-e^{-d\sqrt{2\mu}}}{(1/\tau+\sqrt{2\mu})^2} + \frac{d e^{-d/\tau}}{\sqrt{2\mu}-1/\tau} + \frac{e^{-d\sqrt{2\mu}}}{(1/\tau-\sqrt{2\mu})^2} - \right. \end{aligned}$$

$$
\begin{aligned}
& - \frac{e^{-d/\tau}}{(1/\tau - \sqrt{2\mu})^2} + \frac{de^{-d/\tau}}{\sqrt{2\mu} + 1/\tau} + \frac{e^{-d/\tau}}{(1/\tau + \sqrt{2\mu})^2} \Big\} \\
= & \frac{\tau(\tau^2 - 1/(2\mu))de^{-d/\tau} + \tau^2(e^{-d\sqrt{2\mu}} - e^{-d/\tau})/\mu}{\tau^2 - 1/(2\mu)},
\end{aligned}
$$

so expressions for $\delta(d)$ and $\delta^*(d)$ follow immediately.

References

ANTONIADIS, A., BIGOT, J. and SAPATINAS, T. (2001). Wavelet estimators in nonparametric regression: Description and simulative comparison. *Technical report*, IMAG, France. http://www-lmc.imag.fr/SMS/software/wavden/wavden.pdf.zip.

BUCCIGROSSI, R. and SIMONCELLI, E. (1999). Image compression via joint statistical characterization in the wavelet domain, *IEEE Transaction on Image Processing*, **8**, 1688–1701.

BUCKHEIT, J. and DONOHO, D. (1995). WaveLab and reproducible research. In *Wavelet and Statistics*, Eds. Antoniadis and Oppenheim, Lecture Notes in Statistics, **103**, 55–81, Springer-Verlag, New York.

CHIPMAN, H., KOLACZYK, E. and MCCULLOCH, R. (1997). Adaptive Bayesian wavelet shrinkage, *Journal of the American Statistical Association*, **92**, 1413–1421.

CLYDE, M., PARMIGIANI, G. and VIDAKOVIC, B. (1998). Multiple shrinkage and subset selection in wavelets, *Biometrika*, **85**, 391–402.

DASGUPTA, A. (1992). Distributions which are Gaussian convolutions, *Technical Report*, **92-34**, Department of Statistics, Purdue University, West Lafayette, IN.

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.

DONOHO, D. (1997). CART and best-ortho-basis: A connection. *Annals of Statistics*, **25**, 1870–1911.

DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425–455.

DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia?, *Journal of the Royal Statistical Society, Series B*, **57**, 301–369.

LEPORINI, D., PESQUET, J.-C. and KRIM, H. (1999). Best basis representations with prior statistical models. In *Bayesian Inference in Wavelet Based Models*, Eds. Müller and Vidakovic, 155–172, Springer-Verlag, New York.

MALLAT, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

MÜLLER, P. (1994). Private communication.

MÜLLER, P. and VIDAKOVIC, B. (1999). *Bayesian Inference in Wavelet Based Model*. Springer Verlag, New York.

RUGGERI, F. (1999). Robust Bayesian and Bayesian decision theoretic wavelet shrinkage. In *Bayesian Inference in Wavelet Based Models*, Eds. Müller and Vidakovic, 139–154, Springer-Verlag, New York.

RUGGERI, F. and VIDAKOVIC, B. (1999). Bayesian decision theoretic approach to the choice of thresholding parameter, *Statistica Sinica*, **9**, 183-197.

SIMONCELLI, E. P. (1999). Bayesian denoising of visual images in the wavelet domain. In *Bayesian Inference in Wavelet Based Models*, Eds. Müller and Vidakovic, 291–308, Springer-Verlag, New York.

VIDAKOVIC, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors, *Journal of the American Statistical Association*, **93**, 173–179.

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

VIDAKOVIC, B. and BIELZA, C. (1998). Time-adaptive wavelet denoising. *IEEE Transactions on Signal Processing*, **46**, 2549–2554.

VIDAKOVIC, B. and RUGGERI, F. (1999). Expansion estimation by Bayes rules, *Journal of Statistical Planning and Inference*, **79**, 223–235.

YANG, R. and BERGER, J. (1997). A catalog of noninformative priors, *Discussion Paper*, **97-42**, ISDS, Duke University, Durham, NC.

BRANI VIDAKOVIC
SCHOOL OF SYSTEMS AND
INDUSTRIAL ENGINEERING
437 GROSECLOSE BUILDING
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA 30332-0205
E-mail: brani@isye.gatech.edu

FABRIZIO RUGGERI
ISTITUTO PER LE APPLICAZIONI DELLA
MATEMATICA E DELL'INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE
VIA AMPÉRE 56
I-20131 MILANO, ITALY
E-mail: fabrizio@iami.mi.cnr.it