On least squares estimators under Bernoulli-Laplacian mixture priors

Aleksandra Pižurica and Wilfried Philips
Ghent University
Image Processing and Interpretation Group
Sint-Pietersnieuwstraat 41, B9000 Ghent
Belgium
sanja@telin.ugent.be, philips@telin.ugent.be

Abstract

Several recent wavelet denoising methods use priors that are mixtures of two truncated Laplacian-shaped distributions, where a *Bernoulli* random variable controls switching between the central part of the distribution (describing the "insignificant" data) and its tails (describing the so-called "signal of interest"). The existing works combine this prior with heuristic shrinkage rules, and in this paper we derive the minimum mean squared error (MMSE) estimator. The derived expressions generalize previous related MMSE estimators under Laplacian priors and point out a new relationship between the asymptotic behavior of these estimators and the soft-thresholding. The new results also enable a comparison and demonstrate potentials of Bernoulli-Laplacian priors with respect to more common Bernoulli-Gaussian (normal-mixture) priors.

1 Introduction

A Bayesian estimator assumes a prior distribution for the quantity being estimated. For sparse signal representations, like the wavelet domain, effective priors include mixtures of two distributions where one distribution models the statistics of "significant" coefficients and the other distribution models the statistics of "non-significant" coefficients [1–7], with a *Bernoulli* random variable mixing the two distributions. Examples are the mixture of two normals [1], the mixture of a normal and point mass at zero [2–5] and the mixture of a Laplacian and point mass at zero [6, 7]. We address related, but somewhat different priors that are mixtures of two *truncated* (Laplacian) densities and that appear in several recent multiresolution denoising papers and books, like [8–11]. In this approach, the distribution of significant coefficients is defined by tails of a (generalized) Laplacian and the distribution of nonsignificant coefficients by the remaining central part of the distribution (scaled to integrate one). We call this prior *Bernoulli-Laplacian* (B-L). The existing works used B-L prior within a heuristic shrinkage rule only, which prevented its fair comparison to other, more common priors in Bayesian denoising.

This paper derives the minimum mean squared error (MMSE) solution under the B-L prior, and it also generalizes some of the previous results on MMSE estimation under Laplacian priors and Laplacian mixture priors from [6, 7]. The paper has two main objectives. First, it searches a justification for rather heuristical shrinkage rules adopted in [8–11] and also gives the expressions which enable direct upgrading (restating) these methods within the MMSE criterion. Secondly, this paper also shows

[1]A. Pižurica is a postdoctoral research fellow of FWO Flanders.

that the B-L model leads to a wavelet shrinkage rule which behaves as a sort of soft thresholding for large coefficients. From a practical point of view in image denoising, we show experimentally that the B-L mixture prior yields a better mean squared error performance than more common normal mixtures [1] and mixtures of a Laplacian and point mass at zero [7].

2 Problem outline

Consider a classical problem of estimating the unknown digital signal β_i $(i = 1...n)$ from the measurement noise ϵ_i, where the observations are

$$y_i = \beta_i + \epsilon_i, \quad i = 1, ..., n \tag{1}$$

and the noise samples ϵ_i are independent identically distributed normal random variables $\epsilon_i \sim N(0, \sigma^2)$. We assume the unknown noise-free signal is sparse and well modelled by a Laplacian distribution, which is a typical problem in wavelet domain image denoising [12]. If the input noise is additive white Gaussian, then (1) holds in each wavelet subband provided that the wavelet transform [13, 14] is orthogonal.

Let $\phi(\epsilon; \sigma^2)$ denote the zero-mean normal density of variance σ^2. The MMSE solution of (1), which minimizes the squared risk function, is the conditional mean [15]

$$E(\beta|y) = \frac{\int_{-\infty}^{\infty} \beta f(y|\beta) f(\beta) d\beta}{\int_{-\infty}^{\infty} f(y|\beta) f(\beta) d\beta} = \frac{\int_{-\infty}^{\infty} \beta \phi(y-\beta; \sigma^2) f(\beta) d\beta}{\int_{-\infty}^{\infty} \phi(y-\beta; \sigma^2) f(\beta) d\beta} \tag{2}$$

where $f(\beta)$ is the probability density function of β (to be called hereafter *density* and also *prior* for β) and $f(y|\beta)$ is the conditional density of y given β. Denote the hypothesis "signal component appears in the observed coefficient with significant energy" by H_1 and the opposite hypothesis by H_0. This unifies different mixture priors from [1–11] as $f(\beta) = P(H_0)f(\beta|H_0) + P(H_1)f(\beta|H_1)$ and the corresponding MMSE estimate of β is

$$E(\beta|y) = E(\beta|y, H_0)P(H_0|y) + E(\beta|y, H_1)P(H_1|y). \tag{3}$$

The Bayes' rule yields $P(H_1|y) = \mu\eta/(1 + \mu\eta)$ where $\mu = P(H_1)/P(H_0)$ is the *prior ratio* (i.e., prior odds) and $\eta = f(y|H_1)/f(y|H_0)$ is the *likelihood ratio*, also called *Bayes factor*. The mixture prior of [8–11] states

$$H_0 : |\beta| \leq T \quad \text{and} \quad H_1 : |\beta| > T \tag{4}$$

where T is a threshold that defines the signal of interest. The corresponding mixture densities $f(\beta|H_0)$ and $f(\beta|H_1)$ are truncated Laplacians (rescaled in order to integral one). This prior appeared previously with a heuristic estimator $\hat{\beta} = P(H_1|y)y$ and its spatially adaptive extensions. In [9, p.154], this rule that we call *ProbShrink*, has been described as posterior expected hard thresholding.

In this paper, we show that the *ProbShrink* rule can be interpreted as an approximative MMSE and we also derive the exact MMSE estimator under the corresponding prior. The noise standard deviation will be, as is conventional, estimated from the median of the absolute values of the coefficients at the highest level [7, 16]. We shall *not* consider the specification of the the threshold T and the prior probability $P(H_1)$ because these issues have already been addressed elsewhere. In particular, the threshold T was in [8–11] chosen as $T = \sigma$ motivated by the oracle reasoning by Donoho and collaborators. The prior probabilities of the two hypotheses, have been analyzed in different forms: fixed per subband [1, 11] or spatially varying, based on hidden Markov modelling (HMM) [17], Markov Random Field (MRF) modelling [8–10, 18] or local spatial/spectral activity indicators [11]. We shall present here denoising results for the case where $P(H_1)$ is fixed per subband, but notice that the derived expressions can be in the same way combined with more complex, spatially adaptive prior probabilities. For example, the expressions we derive here can be "plugged" into the spatially adaptive MRF-based approaches of [8,10] and spatially adaptive denoiser based on the same prior from [11] to restate these methods within the MMSE criterion.

3 The least squares estimator under B-L prior

As discussed above, our estimator is determined by the conditional means $E(\beta|y, H_0)$ and $E(\beta|y, H_1)$ and the Bayes factor η under the B-L prior. First, we normalize the noise standard deviation to $\sigma = 1$.

Proposition 1 Assume the wavelet coefficients in a given subband follow model (1) with $\sigma^2 = 1$, i.e., $y_i = \beta_i + \epsilon_i$, $i = 1, ..., n$, where ϵ_i are i.i.d. $N(0,1)$. Also, assume that the noise-free coefficients β_i are i.i.d. Laplacian random variables $f(\beta) = (\lambda/2)\exp(-\lambda|\beta|)$ and define the hypotheses H_0 and H_1 as in (4): $H_0 : |\beta| \leq T$ and $H_1 : |\beta| > T$. Let $\phi(y)$ and $\Phi(y)$ denote the probability density function and the cumulative distributions of the standard normal distribution $N(0,1)$ and define $\Psi(a;t) = \Phi(a + t) - \Phi(t)$.

The conditional mean of β given y under H_1 (the signal of interest *is* present) is

$$E(\beta|y, H_1) = y - \frac{\lambda r_-(y;\lambda;T) - e^{-(T^2/2+\lambda T)}(e^{Ty} - e^{-Ty})/\sqrt{2\pi}}{r_+(y;\lambda;T)} \quad (5)$$

where

$$r_+(y;\lambda;T) = e^{(y-\lambda)^2/2}\Phi(y-\lambda-T) + e^{(y+\lambda)^2/2}\Phi(-y-\lambda-T)$$
$$r_-(y;\lambda;T) = e^{(y-\lambda)^2/2}\Phi(y-\lambda-T) - e^{(y+\lambda)^2/2}\Phi(-y-\lambda-T).$$

The conditional mean of β given y under H_0 (the signal of interest is *not* present) is

$$E(\beta|y, H_0) = y - \frac{\lambda \rho_-(y;\lambda;T) + e^{-(T^2/2+\lambda T)}(e^{Ty} - e^{-Ty})/\sqrt{2\pi}}{\rho_+(y;\lambda;T)} \quad (6)$$

where

$$\rho_+(y;\lambda;T) = e^{(y-\lambda)^2/2}\Psi(\lambda-y;T) + e^{(y+\lambda)^2/2}\Psi(\lambda+y;T)$$
$$\rho_-(y;\lambda;T) = e^{(y-\lambda)^2/2}\Psi(\lambda-y;T) - e^{(y+\lambda)^2/2}\Psi(\lambda+y;T).$$

The conditional densities of noisy coefficients are $f(y|H_0) = A_0\sqrt{2\pi}\phi(y)\rho_+(y;\lambda;T)$ and $f(y|H_1) = A_1\sqrt{2\pi}\phi(y)r_+(y;\lambda;T)$, with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$, and the likelihood ratio is

$$\eta = \frac{f(y|H_1)}{f(y|H_0)} = (e^{\lambda T} - 1)\frac{r_+(y;\lambda,T)}{\rho_+(y;\lambda,T)}. \quad (7)$$

Proof: See Appendix.

A subband-adaptive formulation for $P(H_1)$ from [11] is $P(H_1) = 1 - \int_{-T}^{T} f(\beta)d\beta = \exp(-\lambda T)$. In this case, for $T = 0$ our estimator reduces to $E(\beta|y) = E(\beta|y, H_1) = y - \lambda r_-(y;\lambda, 0)/r_+(y;\lambda, 0)$, which is the MMSE estimate under the Laplacian prior from [6].

For arbitrary σ (see the bottom of Appendix) we have

$$E(\beta|y, H_1) = y - \frac{\sigma^2\lambda r_-(y/\sigma;\sigma\lambda;T) - \sigma e^{-(T^2/2+\sigma\lambda T)}(e^{Ty/\sigma} - e^{-Ty/\sigma})/\sqrt{2\pi}}{r_+(y/\sigma;\sigma\lambda;T)}. \quad (8)$$

and the conditional mean is $E(\beta|y) = y - \sigma^2\lambda r_-(y/\sigma;\sigma\lambda, 0)/r_+(y/\sigma;\sigma\lambda, 0)$.

4 Results and Discussion

The following analysis of the derived estimator yields two interesting results: (1) we establish a new relationship between the MMSE estimators under (Bernoulli-)Laplacian priors and the classical soft-thresholding and (2) we explicitly expose the approximations that relate the intuitive probabilistic shrinkers of [8–11] to the MMSE estimator under the same prior.

The conditional means $E(\beta|y,H_1)$ and $E(\beta|y,H_1)$ are illustrated in Fig. 1. For large $|y|$, $E(\beta|y,H_0)$ approaches $+T$ (when $y>0$) or $-T$ (when $y<0$). We observe that the values of $E(\beta|y,H_1)$ for large $|y|$ do not depend on the value of T and that $E(\beta|y,H_1)$ tends to the soft-thresholding estimates. It has been shown before that the *maximum a posteriori* (*MAP*) estimator under the Laplacian prior is equivalent to soft-thresholding with the threshold $\sigma^2\lambda$ [12, 19]. The following proposition describes a similar result for the *MMSE* estimator of *large magnitude* coefficients.

Proposition 2 For large $|y|$, $E(\beta|y,H_1)$ and $E(\beta|y)$ approach the MAP estimate of β, i.e.,

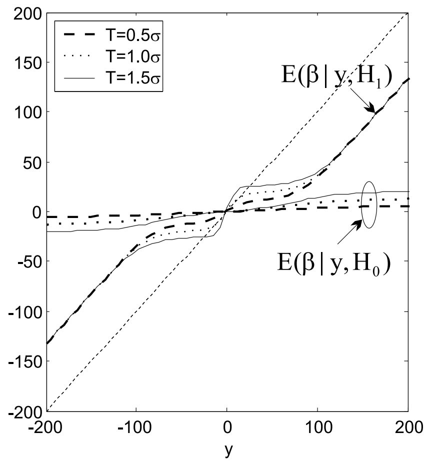
$$\lim_{|y|\to\infty} E(\beta|y,H_1) = \lim_{|y|\to\infty} E(\beta|y) = \mathrm{sgn}(y)(|y|-\sigma^2\lambda) \tag{9}$$

Proof: This follows from (8) if we note that $\lim_{y\to\infty}\Phi(y-\lambda-T)=\lim_{y\to-\infty}\Phi(-y-\lambda-T)=1$, and $\lim_{y\to-\infty}\Phi(y-\lambda-T)=\lim_{y\to\infty}\Phi(-y-\lambda-T)=-1$, which yields $\lim_{y\to\pm\infty} r_-(y;\lambda;T)/r_+(y;\lambda;T)=\pm1$ and $\lim_{y\to\pm\infty}(e^{Ty}-e^{-Ty})/r_+(y;\lambda;T)=0$. Notice that this asymptotic behavior has a practical importance because it is achieved already at moderately high coefficient values (see Fig. 1).

The derived MMSE solution and the illustration of the conditional means in Fig. 1 allow also a new interpretation of the so-called *ProbShrink* rule from [11]: $\hat{\beta}=P(H_1|y)y$ and the related shrinkers of [8–10]. Obviously, the *ProbShrink* estimator can be interpreted as an approximation of the MMSE estimate, which follows from the simplifications: $P(H_0|y)E(\beta|H_0,y)\cong 0$ and $E(\beta|H_1,y)\cong y$. Fig. 1) shows that $E(\beta|H_0,y)$ is close to zero for relatively small y. For large y, the product $P(H_0|y)E(\beta|H_0,y)$ again tends to zero because $E(\beta|H_0,y)$ is confined to finite values $\pm T$, while (7) yields $\lim_{|y|\to\infty} P(H_0|y)=\lim_{|y|\to\infty} 1/(1+\mu\eta)=0$ for any non-zero prior ratio μ. Fig. 1 also illustrates that the simplification, $E(\beta|H_1,y)\cong y$ is reasonable for relatively small y, while for larger magnitudes ($|y|>\sigma^2\lambda$) a more accurate linear approximation is $E(\beta|H_1,y)\cong \mathrm{sgn}(y)(|y|-\sigma^2\lambda)$.

The proposed prior has two parameters: threshold T and the scale parameter λ of the underlying Laplacian model. As explained in Section 2 we use $T=\sigma$ and the scale parameter is estimated in each subband [10]: $\lambda=[0.5(\sigma_y^2-\sigma^2)]^{-1/2}$, where σ_y^2 is the variance of the noisy coefficients.

Table 1 lists experimental peak-signal-to-noise ratio (PSNR) results for several shrinkers under two mixture priors: the analyzed Bernoulli-Laplacian (B-L) prior and the *two-normals* mixture of [1], that we denote here as Bernoulli-Gaussian (B-G): $f(\beta)=P(H_1)\phi(\beta;\sigma_1)+P(H_0)\phi(\beta;\sigma_2)$. A standard wavelet denoising procedure was applied: the noisy images were decomposed using a five-level orthogonal wavelet transform with the Daubechies' wavelet *symmlet* with eight vanishing moments [13]. In each wavelet subband the noise-free wavelet coefficients were estimated using the corresponding shrinkers for each of the analyzed methods. The inverse wavelet transform was applied to reconstruct the denoised image. The parameters of both priors were calculated adaptively for each subband. For the B-L prior we used the subband-adaptive formulation [11] $P(H_1)=1-(\lambda/2)\int_{-\sigma}^{\sigma}\exp(-\lambda|\beta|)d\beta=\exp(-\lambda\sigma)$. For the B-G prior, we used the default estimation of the hyperparameters from [1]. Several conclusions can be drawn from the results in Table 1. Compared to the MMSE estimator, *ProbShrink* under the same prior yields a minor performance loss (usually less than 0.1dB). For the MAP estimator this performance loss is up to 0.3dB. The results also demonstrate that



(a)

Figure 1: An illustration of $E(\beta|y, H_0)$ and $E(\beta|y, H_1)$ at the finest resolution scale for different T. For large y the conditional mean under H_1: $E(\beta|y, H_1)$ tends to the soft-thresholding function with the threshold $\sigma^2\lambda$.

the Bernoulli-Laplacian prior yields a significant improvement with respect to the two-normals mixture, which is often higher than 1dB. The difference in the performance of our method when using the estimated σ versus a priori known σ is negligible (on the tested images 0.01 to 0.1dB).

Table 2 compares the performance of the analyzed prior with the mixture of a Laplacian and point mass at zero (here denoted as B-L-PM prior) from [7]. The posterior mean and the posterior median estimators under this prior were shown to outperform a range of other estimators on one dimensional (1-D) signals and are currently used as benchmarks for 1-D signal denoising. We used *EBayesThresh* software available online from authors of [7] and ran simulations on four standard test signals: *Blocks* (blk), *Bumps* (bmp), *Doppler* (dop) *HeaviSine* (hea). The squared errors for each estimator were averaged over 100 replications of 512-point signals. Like in [7], for "low noise" the ratio of the standard deviations of noise and signal is 1/7; for "high noise" this ratio is 1/3. The posterior mean outperforms the posterior median which agrees with the conclusions of [7]. In all cases except *Blocks* with low noise level, the proposed estimator achieves the smallest error. The gain of the new estimator is bigger for the higher noise level. Notice that the B-L-PM prior can be seen as a special case of our prior, for $T = 0$.

5 Conclusion

We developed the MMSE estimator for the so-called *Bernoulli-Laplacian* mixture prior, which has been used in several earlier denoising methods, but combined with other, more intuitive wavelet shrinkers. Here derived results enable an interpretation and also a direct extension of these earlier methods within a well-established mean-squared error criterion. Our results also generalize some of the previous results on MMSE estimation with Laplacian priors, which can be seen as limiting cases in our approach. We also demonstrated a significant potential of Bernoulli-Laplacian priors as compared to the more common mixtures of two normals and mixtures of Laplacian and point mass at zero. It should be interesting to investigate, in a future work, polynomial approximations of the developed estimator.

6 Appendix: Proof of the proposition 1

Under the assumed Bernoulli-Laplacian prior, we have with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$. We write

$$E(\beta|y, H_1) = \frac{K_{H_1}(y; \lambda, T)}{m_{H_1}(y; \lambda, T)} \quad \text{and} \quad E(\beta|y, H_0) = \frac{K_{H_0}(y; \lambda, T)}{m_{H_0}(y; \lambda, T)} \tag{10}$$

$$f(\beta|y, H_1) = A_1 m_{H_1}(y; \lambda, T) \quad \text{and} \quad f(\beta|y, H_0) = A_1 m_{H_0}(y; \lambda, T) \tag{11}$$

where

$$\begin{aligned} m_{H_1}(y; \lambda, T) &= \int_{-\infty}^{-T} e^{\lambda\beta}\phi(y-\beta)d\beta \ + \ \int_{T}^{\infty} e^{-\lambda\beta}\phi(y-\beta)d\beta \\ K_{H_1}(y; \lambda, T) &= \int_{-\infty}^{-T} \beta e^{\lambda\beta}\phi(y-\beta)d\beta \ + \ \int_{T}^{\infty} \beta e^{-\lambda\beta}\phi(y-\beta)d\beta \\ m_{H_0}(y; \lambda, T) &= \int_{-T}^{0} e^{\lambda\beta}\phi(y-\beta)d\beta \ + \int_{0}^{T} e^{-\lambda\beta}\phi(y-\beta)d\beta \\ K_{H_0}(y; \lambda, T) &= \int_{-T}^{0} \beta e^{\lambda\beta}\phi(y-\beta)d\beta \ + \ \int_{0}^{T} \beta e^{-\lambda\beta}\phi(y-\beta)d\beta. \end{aligned} \tag{12}$$

On a nonnegative interval $[a, b]$, one can verify [6]

$$I(y; a; b, \lambda) = \int_a^b e^{-\lambda\beta}\phi(y-\beta)d\beta = \sqrt{2\pi}\phi(y)e^{(y-\lambda)^2/2}[\Phi(b-y+\lambda) - \Phi(a-y+\lambda)] \tag{13}$$

$$\begin{aligned} II(y; a; b, \lambda) = \int_a^b \beta e^{-\lambda\beta}\phi(y-\beta)d\beta = \sqrt{2\pi}\phi(y)e^{(y-\lambda)^2/2}\Big[\frac{1}{\sqrt{2\pi}}e^{-(a-y+\lambda)^2/2} - \\ \frac{1}{\sqrt{2\pi}}e^{-(b-y+\lambda)^2/2} + (y-\lambda)\Big(\Phi(b-y+\lambda) - \Phi(a-y+\lambda)\Big)\Big]. \end{aligned} \tag{14}$$

Using these identities, we find

$$m_{H_1}(y; \lambda, T) = I(y; T, \infty, \lambda) + I(-y; T, \infty, \lambda) = \sqrt{2\pi}\phi(y)r_+(y; \lambda; T) \tag{15}$$

and

$$\begin{aligned} K_{H_1}(y; \lambda, T) = II(y; T, \infty, \lambda) - II(-y; T, \infty, \lambda) = \\ \sqrt{2\pi}\phi(y)\Big[\frac{1}{\sqrt{2\pi}}e^{-\left(\frac{T^2}{2}+\lambda T\right)}\Big(e^{Ty} - e^{-Ty}\Big) + yr_+(y; \lambda; T) - \lambda r_-(y; \lambda; T)\Big] \end{aligned} \tag{16}$$

where

$$\begin{aligned} r_+(y; \lambda, T) &= e^{(y-\lambda)^2/2}\Phi(y-\lambda-T) + e^{(y+\lambda)^2/2}\Phi(-y-\lambda-T) \\ r_-(y; \lambda, T) &= e^{(y-\lambda)^2/2}\Phi(y-\lambda-T) - e^{(y+\lambda)^2/2}\Phi(-y-\lambda-T). \end{aligned} \tag{17}$$

Substituting (15) and (16) into (10) yields $E(\beta|y, H_1)$ from (5).

Table 1: Experimental PSNR[dB] results of several Bayesian wavelet shrinkers and two different mixture priors: *Bernoulli-Laplacian* (B-L) prior and *Bernoulli-Gaussian* (B-G) prior from [1], which is a mixture of two-normals. Wavelet type: Daubechies' least asymetrical wavelet with eight vanishing moments.

		Standard deviation of noise				Standard deviation of noise			
Prior	Estimator	10	15	20	25	10	15	20	25
		BOAT				LENA			
	noisy image	28.15	24.62	22.10	20.17	28.13	24.60	22.12	20.16
B-L	MAP	32.00	29.92	28.42	27.37	33.43	31.47	30.18	29.23
B-L	*ProbShrink* [11]	32.13	30.05	28.63	27.63	**33.69**	**31.70**	**30.39**	**29.41**
B-L	MMSE	**32.23**	**30.14**	**28.70**	**27.70**	33.62	31.65	**30.39**	**29.41**
B-G [1]	MMSE	31.01	29.04	27.66	26.68	32.84	30.97	29.66	28.74

Now we introduce $\Psi(a;T) = \Phi(a+T) - \Phi(a)$, to express

$$m_{H_0}(y;\lambda,T) = I(y;0,T,\lambda) + I(-y;0,T,\lambda) = \sqrt{2\pi}\phi(y)\rho_+(y;\lambda;T) \tag{18}$$

and

$$K_{H_0}(y;\lambda,T) = II(y;0,T,\lambda) - II(-y;0,T,\lambda) =$$
$$\sqrt{2\pi}\phi(y)\Big[-\frac{1}{\sqrt{2\pi}}e^{-\left(\frac{T^2}{2}+\lambda T\right)}\left(e^{Ty} - e^{-Ty}\right) + y\rho_+(y;\lambda;T) - \lambda\rho_-(y;\lambda;T)\Big] \tag{19}$$

where

$$\rho_+(y;\lambda,T) = e^{(y-\lambda)^2/2}\Psi(\lambda - y;T) + e^{(y+\lambda)^2/2}\Psi(\lambda + y;T)$$
$$\rho_-(y;\lambda,T) = e^{(y-\lambda)^2/2}\Psi(\lambda - y;T) - e^{(y+\lambda)^2/2}\Psi(\lambda + y;T). \tag{20}$$

Substituting (18) and (19) into (10) yields $E(\beta|y,H_0)$ from (6). Finally, the conditional densities of noisy wavelet coefficients are $f(y|H_0) = A_0 m_{H_0}(y;\lambda,T)$ and $f(y|H_1) = A_1 m_{H_1}(y;\lambda,T)$, with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$, which yields $\eta = f(y|H_1)/f(y|H_1)$ from (7). This completes the proof of Proposition 1. □ For arbitrary σ, we replace the integrals $I(y;a,b,\lambda)$ and $II(y;a,b,\lambda)$ from (13),(14) by $\mathcal{I}(y;a,b,\lambda,\sigma)$ and $\mathcal{II}(y;a,b,\lambda,\sigma)$, resp., where $\mathcal{I}(y;a,b,\lambda,\sigma) = \int_a^b e^{-\lambda\beta}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\beta)^2}{2\sigma^2}}d\beta$ and $\mathcal{II}(y;a,b,\lambda,\sigma) = \int_a^b \beta e^{-\lambda\beta}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\beta)^2}{2\sigma^2}}d\beta$ and use the change of variable $\beta' = \frac{\beta}{\sigma}$.

References

[1] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. of the Amer. Statist. Assoc*, vol. 92, pp. 1413–1421, 1997.

[2] F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a Bayesian approach," *J. of the Royal Statist. Society B*, vol. 60, pp. 725–749, 1998.

[3] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, no. 2, pp. 391–401, 1998.

Table 2: Sums of squared errors, averaged over 100 replications of 512 point signals. The wavelet transform is four-level orthogonal. The maximum standard deviation of entries given columnwise.

		LOW NOISE				HIGH NOISE			
Prior	Estimator	**blk**	**bmp**	**dop**	**hea**	**blk**	**bmp**	**dop**	**hea**
B-L	*ProbShrink* (MMSE)	16.03	2.59	0.37	12.82	72.47	11.35	1.36	48.32
B-L-PM [7]	*PostMean* (MMSE)	14.81	2.60	0.42	13.46	90.06	12.46	1.55	51.93
B-L-PM [7]	*PostMedian*	17.41	3.25	0.52	13.98	107.84	15.53	1.84	50.42
maximum standard deviation		3.68	0.60	0.09	3.14	19.30	2.15	0.26	9.27

[4] B. Vidakovic, "Wavelet-based nonparametric Bayes methods," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, ser. Lecture Notes in Statistics, D. D. Dey, P. Müller, and D. Sinha, Eds., vol. 133. Springer Verlag, New York, 1998, pp. 133–155.

[5] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Steins' principle," *IEEE Trans. Image Proc. (in press).*

[6] M. Hansen and B. Yu, "Wavelet thresholding via MDL for natural images," *IEEE Trans. Inform. Theory*, vol. 46, no. 5, pp. 1778–1788, Aug. 2000.

[7] I. Johnstone and B. Silverman, "Empirical Bayes selection of wavelet thresholds," *Annals of Statistics*, vol. 33, no. 4, pp.1700-1752, 2005.

[8] M. Jansen and A. Bultheel, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 629–639, 2001.

[9] M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer-Verlag, New York, 2001.

[10] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for wavelet based Bayesian image denoising," *IEEE Trans. Image Proc*, vol. 11, no. 5, pp. 545–557, May 2002.

[11] A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 654-665, March 2006.

[12] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, pp. 909–919, Apr. 1999.

[13] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

[14] S. Mallat, *A wavelet tour of signal processing.* Academic Press, London, 1998.

[15] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 1996.

[16] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[17] M. S. Crouse, R. D. Nowak, and R. G. Baranuik, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886–902, 1998.

[18] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. Image processing*, vol. 6, no. 4, pp. 549–565, 1997.

[19] M. Nikolova, "Estimées locales fortement homogènes," *Comp. Rendus Acad. Sci. Paris-Ser I*, vol. 325, pp. 665–670, 1997.