Adaptive Probabilistic Wavelet Shrinkage for Image Denoising

Aleksandra Pižurica and Wilfried Philips

Technical Report July 2004
Ghent University
Department Telecommunications and Information Processing
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

Adaptive Probabilistic Wavelet Shrinkage for Image Denoising

Aleksandra Pižurica and Wilfried Philips

Abstract

We study a Bayesian wavelet shrinkage approach for natural images based on a probability that a given coefficient contains a significant noise-free component, which we call "signal of interest". First we develop new subband adaptive wavelet shrinkage method of this kind for the generalized Laplacian prior for noise free coefficients. We compare the new shrinkage approach with other subband adaptive Bayesian shrinkage rules in terms of mean squared error performance. The results demonstrate that the new method outperforms existing Bayesian thresholding rules for natural images. We also extend the new shrinkage method to a spatially adaptive procedure. In the spatially adaptive version of the method, each coefficient is shrunk according to how probable it is that it presents a signal of interest, based on its value, based on a measurement from the local surrounding and based on the global statistical properties of the coefficients in a given subband. The procedure is fully automatic and fast. The new method yields the results that are among the best state-of-the-art ones and it outperforms much more complex recent related methods.

Keywords: Image denoising, wavelets, Bayesian estimation, generalized likelihood ratio

I. Introduction

In image denoising, where a trade-off between noise suppression and the preservation of actual image discontinuities must be made, solutions are sought which can "detect" important image details and accordingly adapt the degree of noise smoothing. The wavelet transform [1–5] facilitates the detection of important image details in the presence of noise: the wavelet coefficients representing the main image discontinuities are relatively large as compared to coefficients that only contain noise. Noise reduction in the wavelet domain is usually done by

A. Pižurica and W. Philips are with the Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. E-mail: Aleksandra.Pizurica@telin.rug.ac.be, philips@telin.rug.ac.be, Tel: +32 9 264 34 12, Fax: +32 9 264 42. 95. **Corresponding author:** Aleksandra Pižurica

shrinking the magnitudes of noisy wavelet coefficients. Ideally, the coefficients that contain primarily noise should be reduced to negligible values while the ones containing a "significant" noise-free component should be reduced less. A common shrinkage approach is thresholding [6, 7], where the coefficients with magnitudes below a certain threshold are treated as "non significant" and are set to zero. The remaining, "significant" coefficients are kept unmodified (hard-thresholding) or reduced in magnitude (soft-thresholding).

Shrinkage estimators can also result from Bayesian methods [8–33]. A Bayesian approach imposes a prior distribution of noise-free data. Common models for noise-free subband data include (generalized) Laplacian [2, 8, 15, 17], double stochastic (Gaussian scale mixture) priors [24–26] and mixtures of two distributions where one distribution models the statistics of "significant" coefficients and the other one the statistics of "insignificant" data [9–14]. Combined with these priors for marginal statistical distributions, Hidden Markov Tree (HMT) models are often employed to incorporate inter-scale dependencies [21–23] and Markov Random Fields (MRF) have been used to model intra-scale (spatial) dependencies [29–32].

Regardless of the particular prior, Bayesian wavelet domain denoising methods have been developed along the following two main lines. The first class of methods optimizes the threshold selection for hard- and soft-thresholding [8–11]. The soft thresholding method of [8] known as *BayesShrink*, employs a threshold that is optimal in terms of mean squared error (MSE) for the generalized Laplacian prior. The second class of methods replaces somewhat ad-hoc choice of hard- and soft- thresholding rules by shrinkage functions that result from minimizing a Bayesian risk, typically under a quadratic cost function (minimum mean squared error - MMSE estimation [13–16]) or under a delta cost function (maximum a posteriori - MAP estimation [17]). The above listed methods are *subband adaptive*: they are optimized with respect to the marginal subband statistics. The use of bivariate and joint statistics of wavelet coefficients is addressed in [18, 19], respectively. In practice, *spatially adaptive* Bayesian estimators are efective, where a given parameter of the marginal prior is refined with respect to the local spatial context. Such a spatially adaptive extension of the *BayesShrink* rule is reported in [20]

and different spatially adaptive versions of the MMSE estimators are developed in [21–28].[1]

This paper focuses on another class of Bayesian shrinkage methods [29–31] including our previous work [32, 33], where the shrinkage rule is: multiply empirical wavelet coefficient with the probability that it contains a significant noise-free component, given a set of measurements calculated from the empirical coefficients. While this "probabilistic shrinkage" rule is intuitively appealing, its motivation in terms of some well established optimization criteria was less clear so far. The effectiveness of the existing methods [29–32] is largely due to the use of powerful MRF priors for spatial context. While such priors could have been combined with the MMSE and MAP estimates as well, some important questions remain: why should one consider multiplying a coefficient with the probability that it contains a significant signal (instead of using some common Bayesian estimator, like MMSE or MAP)? How can we develop a simple subband adaptive[2] probabilistic shrinkage function and how would such a function compare with the well known Bayesian shrinkers, like MAP and MMSE estimates and *BayesShrink*?

Here we first address these questions and some other related aspects including the MMSE estimation with Laplacian mixture priors, by providing a thorough analysis and experimental evaluation. The goals of this analysis are to develop a new subband adaptive Bayesian shrinkage rule as a valuable alternative to the common ones and, at the same time, to provide a motivation for a class of methods that shrink the wavelet coefficients according to probability of signal presence, while building a new, simple and efficient representative of this class. First we develop new subband adaptive probabilistic wavelet shrinkage rule for the (generalized) Laplacian prior. The results demonstrate that the new subband adaptive shrinkage outperforms MAP and *BayesShrink* in terms of MSE, and that it yields a performance close to the MMSE estimation under the same priors, while it is less complex and offers a flexible framework for spatially adaptive extensions. We also develop a spatially adaptive version of the new denoising method which is simple to implement, fast and effective in practice. The new method outperforms

[1] In [21–23] a mixing parameter for a mixture of two normals is calculated adaptively for each spatial position using a HMT model, while in [24–28], the multiplier of a Gaussian scale mixture prior is calculated from the local surrounding of each spatial position.

[2] Subband adaptive here means adapted to the marginal staistics only, i.e., without using spatial context models.

some of the much more complex recent ones, which employ HMT and MRF models and yields results that are in terms of MSE and visually among the best published ones in the field.

The paper is organized as follows. Section II provides the background on minimum mean squared error estimation of image wavelet coefficients. The emphasis is on MMSE estimation with mixture priors that are superpositions of two distributions. We give an original, brief unifying description of the most relevant methods. In Section III, we develop and analyze new subband adaptive shrinkage functions for natural images. In Section IV, we present one possible approach to extend the developed shrinkage functions to a spatially adaptive procedure and we demonstrate and discuss the implementations with orthogonal and with redundant wavelet transforms. Finally, we conclude the paper in Section V.

II. Background and related work

Assume the input image is contaminated with additive white Gaussian noise of zero mean and variance σ^2. An orthogonal wavelet transformation [1–4] of the noisy input yields then an equivalent additive white noise model in each wavelet subband[3]

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

where β_i are noise-free wavelet coefficients, ϵ_i are independent identically distributed (i.i.d.) normal random variables $\epsilon_i \sim N(0, \sigma^2)$ and n is the number of coefficients in a subband. The MMSE solution is the conditional mean of the noise-free coefficient value [15]

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

where $\phi(\epsilon; \sigma^2)$ is the zero-mean normal density $N(0, \sigma^2)$, i.e., $\phi(\epsilon; \sigma^2) = (1/\sqrt{2\pi\sigma^2}) \exp(-\epsilon^2/2\sigma^2)$, $f(\beta)$ is the probability density function of the noise-free coefficients and $f(y|\beta)$ is the conditional probability density function (likelihood) of y given β. In the following, we use the term *density* for the probability density function, and also we call the density $f(\beta)$ "prior (model) for β". Below, several often used models are listed.

[3] As it is common in the related literature, for compactness we omit here the indices that denote the scale and the orientation and we denote the spatial position with a single index, like in a raster scanning.

A. MMSE estimation under generalized Laplacian priors

The histograms of noise-free wavelet coefficients of natural images, in each subband of the decomposition, are typically long-tailed and sharply peaked at zero. A generalized Laplacian distribution (also called generalized Gaussian distribution) is thus a widely accepted prior model for noise-free subband data (e.g., [2, 8, 15, 17]):

$$f(\beta) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})} \exp(-\lambda|\beta|^{\nu}), \tag{3}$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$ is the Gamma function, $\lambda > 0$ is the *scale* parameter and ν is the *shape* parameter. For natural images, ν is typically in the range $\nu \in [0, 1]$. The model parameters λ and ν are estimated from the noise-free histogram[4] as follows: if σ_β^2 is the sample variance and κ_β the kurtosis of the noise-free histogram, then [15]

$$\sigma_\beta^2 = \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}, \quad \kappa_\beta = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})}. \tag{4}$$

Simoncelli and Adelson [15] studied the MMSE approach (2) to image denoising under the generalized Laplacian prior (3). The resulting estimator performs a Bayesian "coring" of the coefficients in a manner similar to soft-thresholding.

A special case in the family (3), with $\nu = 1$ (the so-called Laplacian or double exponential) is often used because of its analytical tractability [16, 17] since it usually does not produce a noticeable degradation in performance. In [17] it was shown that the maximum a posteriori (MAP) estimate under the Laplacian prior is a soft-thresholding function with the threshold $\sqrt{2}\sigma^2/\sigma_\beta$. The MMSE estimate under the Laplacian prior (for the normalized noise standard deviation $\sigma = 1$) was derived in [16] and we shall generalize it in Section III-B.

B. MMSE estimation under mixture priors

Mixture priors are often used to reduce the computation complexity and/or to adapt a Bayesian estimator to the surrounding of each coefficient. For example, Mihçak *et al* [24], Strela *et al* [25], and Portilla *et al* [26] use *Gaussian scale mixture* models: each coefficient is

[4] The parameter estimation from the *noisy* coefficient histogram is explained in [15] and also later in this paper.

modeled as the product of two independent random variables: $\beta = \sqrt{z}u$, where z is a positive scalar (a function of the surrounding coefficient values, like the local variance in [24]), and u is an element of a Gaussian random field. The MMSE estimate (2) with such priors takes the form of a Wiener estimator.

Another common class of mixture priors, are superpositions of two distributions, where one of these models the statistics of "significant" ("high energy" or "important") coefficients and the other one models the statistics of "non-significant" coefficients and where the mixing parameter is usually a Bernoulli random variable [10–14, 21–23]. Within this framework, Chipman *et al* [13] proposed a *mixture of two normals*, which is also adopted in [21–23]. Abramovich *et al* [11] and Clyde et *al* [14] consider a related prior for β which is a mixture of a normal distribution and a point mass at zero. Vidakovic [10] gives a nice overview of these and related approaches. Hansen and Yu [16] consider a mixture of a point mass at zero and the Laplacian distribution. The analysis in Section III will start from a generalization of that model. At this point, it is useful to give a unifying description of these approaches. Above mentioned mixture priors can be seen as special cases of the following mixture model class:

$$f(\beta) = P(H_1)f(\beta|H_1) + P(H_0)f(\beta|H_0), \tag{5}$$

where H_1 denotes the hypothesis "signal component appears in the observed coefficient with significant energy" and H_0 denotes the opposite hypothesis. For the prior (5) the minimum mean squared error estimate of the noise-free coefficient value is

$$E(\beta|y) = E(\beta|y, H_0)P(H_0|y) + E(\beta|y, H_1)P(H_1|y). \tag{6}$$

Using the Bayes' rule one can show that

$$P(H_1|y) = \frac{\mu\eta}{1+\mu\eta}, \quad \text{where} \quad \mu = \frac{P(H_1)}{P(H_0)} \quad \text{and} \quad \eta = \frac{f(y|H_1)}{f(y|H_0)}, \tag{7}$$

where the product $\mu\eta$ is usually referred to as the generalized likelihood ratio [34]. We can express the conditional means $E(\beta|y, H_0)$ and $E(\beta|y, H_1)$ as

$$E(\beta|y, H_0) = \int_{-\infty}^{\infty} \beta f(\beta|y, H_0)d\beta = \frac{\int_{-\infty}^{\infty} \beta\phi(y-\beta;\sigma)f(\beta|H_0)d\beta}{\int_{-\infty}^{\infty} \phi(y-\beta;\sigma)f(\beta|H_0)d\beta}, \tag{8}$$

$$E(\beta|y,H_1)=\int_{-\infty}^{\infty}\beta f(\beta|y,H_1)d\beta=\frac{\int_{-\infty}^{\infty}\beta\phi(y-\beta;\sigma)f(\beta|H_1)d\beta}{\int_{-\infty}^{\infty}\phi(y-\beta;\sigma)f(\beta|H_1)d\beta}. \tag{9}$$

For the mixture of two normals of the form $f(\beta|H_0)=N(0,\sigma_0^2)$ and $f(\beta|H_1)=N(0,\sigma_1^2)$, (8) and (9) reduce to $E(\beta|y,H_0)=\sigma_0^2/(\sigma_0^2+\sigma_1^2)y$ and $E(\beta|y,H_1)=\sigma_1^2/(\sigma_0^2+\sigma_1^2)y$. In Section III-B, we derive the above expressions for a mixture prior that generalizes the one in [16], where it was proposed $H_0:\beta=0$ and $H_1:\beta\neq 0$, with a Laplacian prior for β.

III. Subband adaptive Bayesian wavelet shrinkage

Here we develop new Bayesian wavelet shrinkage functions, which are adapted to the marginal subband statistics (histograms) of image wavelet coefficients. We define a "*signal of interest*" as a noise-free coefficient component that exceeds a specific threshold T and formulate the following two hypotheses: H_0: "signal of interest is absent" and H_1: "signal of interest is present" (in a given coefficient) precisely as:

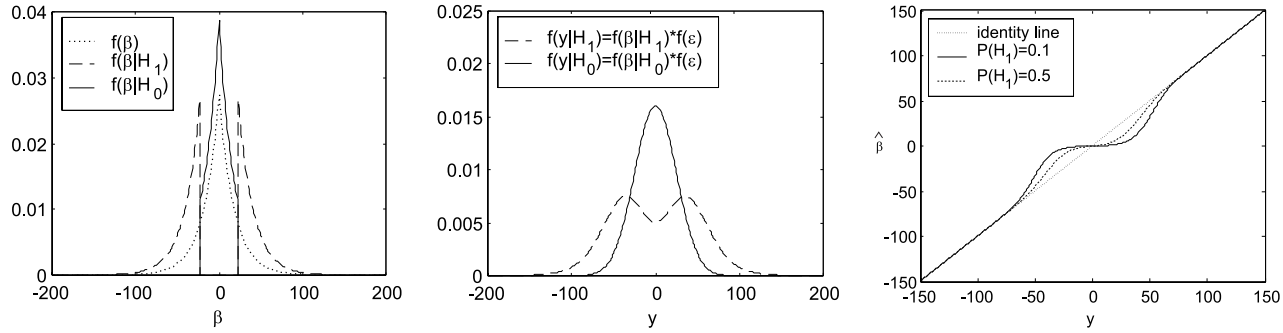
$$H_0:|\beta|\leq T \quad \text{and} \quad H_1:|\beta|>T. \tag{10}$$

For $T=0$, the resulting mixture prior (5) reduces to the one in [16]. As an alternative to the MMSE estimate (6), here we consider a simpler estimator

$$\hat{\beta}=P(H_1|y)y=\frac{\mu\eta}{1+\mu\eta}y, \tag{11}$$

where each coefficient is multiplied with the probability that it contains a signal of interest. We call this rule *ProbShrink* and develop the resulting shrinkage functions for the (generalized) Laplacian prior for noise free coefficients. We will show that this Bayesian shrinker for appropriately defined threshold T yields a mean squared error (MSE) close to the MMSE estimate (2) under the Laplacian prior for β, while it outperforms the corresponding MAP estimator and soft-thresholding rules with the MSE-optimum threshold.

Fig. 1(a) shows an example of the conditional densities of *noise-free* coefficients $f(\beta|H_0)$ and $f(\beta|H_1)$. The conditional densities of *noisy* coefficients (likelihoods) $f(y|H_0)$ and $f(y|H_1)$



(a) (b) (c)

Fig. 1. (a) An illustration of the probability density functions of noise-free coefficients: $f(\beta)$ (dotted), $f(\beta|H_0)$ (solid) and $f(\beta|H_1)$ (dashed). (b) The resulting conditional densities of noisy coefficients $f(y|H_0)$ (solid) and $f(y|H_1)$ (dashed). (c) *ProbShrink* rule $\hat{\beta} = P(H_1|y)y$, where $P(H_1)$ is a parameter.

are given by the following convolutions

$$f(y|H_0) = \int_{-\infty}^{\infty} \phi(y-\beta;\sigma) f(\beta|H_0) d\beta \ , \quad f(y|H_1) = \int_{-\infty}^{\infty} \phi(y-\beta;\sigma) f(\beta|H_1) d\beta \tag{12}$$

and are illustrated in Fig. 1(b). Fig. 1(c) shows the resulting *ProbShrink* rule (11), where $P(H_1)$ (i.e., the prior ratio μ) is left as a free parameter. Next we address the specification of the prior probabilities in a given subband.

A. Adapting the prior probabilities to the subband statistics

In this Section we consider a subband adaptive approach, meaning that the prior probability (that a signal of interest is present in a coefficient) $P(H_1)$ is the same for all the coefficients in a given subband. In related approaches it has been usually assumed $P(H_1) = P(H_0) = 0.5$ (e.g., [16, 29–32]) or $P(H_1)$ was estimated empirically as the fraction of the observed *noisy* coefficients that are above a certain threshold [13]. In our earlier approach [33], $P(H_1)$ was also estimated empirically using binary masks $\mathbf{x} = \{x_1, ..., x_n\}$ that indicate the positions of "significant" ($x_l = 1$) and "insignificant" ($x_l = 0$) coefficients for each subband. The mask detection was based on interscale correlations and the required probability was estimated by the fraction of labels "1": $P(H_1) = \sum_{l=1}^{n} x_l/n$.

Here we propose to derive the probability $P(H_1)$ from the prior model for the noise-free coefficients in a given subband. First, we note that from the definition $H_0 : |\beta| \leq T$ and

$H_1 : |\beta| > T$, and in case where $P(H_1)$ is constant per subband, it follows that

$$P(H_1) = \int_{-\infty}^{\infty} f(\beta|H_1)d\beta = 1 - \int_{-T}^{T} f(\beta)d\beta. \tag{13}$$

Next we develop this expression and analyze *ProbShrink* rule (11) in cases where the prior for noise-free coefficients is Laplacian or generalized Laplacian density.

B. ProbShrink rule for the Laplacian prior

For the Laplacian prior, the conditional densities of noise-free coefficients are

$$f(\beta|H_0) = \begin{cases} A_0 \exp(-\lambda|\beta|), & \text{if } |\beta| \le T \\ 0, & \text{if } |\beta| > T, \end{cases} \tag{14}$$

$$f(\beta|H_1) = \begin{cases} 0, & \text{if } |\beta| \le T, \\ A_1 \exp(-\lambda|\beta|), & \text{if } \beta > T, \end{cases} \tag{15}$$

where it is straightforward to show that

$$A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1) \quad \text{and} \quad A_1 = (\lambda/2)e^{\lambda T}. \tag{16}$$

From (13) it follows that $P(H_1) = 1 - \lambda \int_0^T \exp(-\lambda\beta)d\beta = 1 - \exp(-\lambda T)$ yielding

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{\exp(-\lambda T)}{1 - \exp(-\lambda T)}. \tag{17}$$

For the *ProbShrink* rule (11), we still need to derive the likelihood ration $\eta = f(y|H_1)/f(y|H_0)$. For completeness, we also derive the conditional means from (8) and (9). We start by normalizing the noise standard deviation to $\sigma = 1$ and later generalize to arbitrary σ.

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 (according to (3) with $\nu = 1$) and define the hypotheses H_0 and H_1 as in (10): $H_0 : |\beta| \le T$ 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

$$E(\beta|y,H_1) = y - \frac{\lambda r_-(y;\lambda;T) - e^{-(T^2/2+\lambda T)}(e^{Ty} - e^{-Ty})/\sqrt{2\pi}}{r_+(y;\lambda;T)}, \tag{18}$$

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

$$E(\beta|y,H_0) = y - \frac{\lambda \rho_-(y;\lambda;T) + e^{-(T^2/2+\lambda T)}(e^{Ty} - e^{-Ty})/\sqrt{2\pi}}{\rho_+(y;\lambda;T)}, \tag{19}$$

where

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

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

The conditional densities (likelihoods) of noisy coefficients are

$$f(y|H_0) = A_0\sqrt{2\pi}\phi(y)\rho_+(y;\lambda;T) \tag{20}$$

$$f(y|H_1) = A_1\sqrt{2\pi}\phi(y)r_+(y;\lambda;T) \tag{21}$$

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

$$\eta = \frac{f(y|H_1)}{f(y|H_0)} = (e^{\lambda T}-1)\frac{r_+(y;\lambda,T)}{\rho_+(y;\lambda,T)}. \tag{22}$$

Proof: See Appendix A.

Note that the expression for the conditional mean $E(\beta|y)$ derived by Hansen and Yu [16] is a special case of (18) corresponding to $T = 0$. Note also that in our approach for $T = 0$ one has $P(H_1) = 1$ and $E(\beta|y) = E(\beta|y,H_1) = y - \lambda r_-(y;\lambda,0)/r_+(y;\lambda,0)$, where $\sigma = 1$.

For an arbitrary value of σ one can show (see Appendix B) that the previous expressions generalize as follows:

$$E(\beta|y,H_1) = y - \frac{\sigma^2\lambda r_-(y/\sigma;\sigma\lambda;T) - \sigma e^{-(T^2/2+\sigma\lambda T)}(e^{Ty/\sigma} - e^{-Ty/\sigma})/\sqrt{2\pi}}{r_+(y/\sigma;\sigma\lambda;T)}, \tag{23}$$

$$E(\beta|y,H_0) = y - \frac{\sigma^2\lambda\rho_-(y/\sigma;\sigma\lambda;T) + \sigma e^{-(T^2/2+\sigma\lambda T)}(e^{Ty/\sigma} - e^{-Ty/\sigma})/\sqrt{2\pi}}{\rho_+(y/\sigma;\sigma\lambda;T)}, \qquad (24)$$

For arbitrary σ, the densities of noisy coefficients are $f(y|H_0) = (A_0/\sigma)\sqrt{2\pi}\phi(y/\sigma)\rho_+(y/\sigma;\sigma\lambda;T)$ and $f(y|H_1) = (A_1/\sigma)\sqrt{2\pi}\phi(y/\sigma)r_+(y/\sigma;\sigma\lambda;T)$ and the likelihood ratio is

$$\eta = \frac{f(y|H_1)}{f(y|H_0)} = (e^{\lambda T} - 1)\frac{r_+(y/\sigma;\sigma\lambda;T)}{\rho_+(y/\sigma;\sigma\lambda;T)}. \qquad (25)$$

The MMSE estimate $E(\beta|y)$ is the special case of $E(\beta|y,H_1)$ for $T = 0$:

$$E(\beta|y) = y - \sigma^2\lambda r_-(y/\sigma;\sigma\lambda,0)/r_+(y/\sigma;\sigma\lambda,0). \qquad (26)$$

Also, the probability density function of noisy wavelet coefficients $f(y)$ is the special case of $f(y|H_1)$, for $T = 0$: $f(y) = (\lambda/2\sigma)\sqrt{2\pi}\phi(y/\sigma)r_+(y/\sigma;\sigma\lambda;0)$. The special case of this expression, for $\sigma = 1$, was previously derived in [16].

Note that the mixture prior $f(\beta) = P(H_1)f(\beta|H_1) + P(H_0)f(\beta|H_0)$ with the conditional densities (14) and (15), and with subband-adaptive prior probability $P(H_1)$ defined in (13) reduces to the Laplacian density $f(\beta) = (\lambda/2)\exp(-\lambda|\beta|)$. Computing the conditional mean (6): $E(\beta|y) = E(\beta|y,H_0)P(H_0|y) + E(\beta|y,H_1)P(H_1|y)$ using (7), (17), (22)-(24) is thus equivalent to the classical MMSE estimate (2) under the Laplacian prior. This also means that the MMSE estimate $E(\beta|y) = E(\beta|y,H_0)P(H_0|y) + E(\beta|y,H_1)P(H_1|y)$ is a generalization of the MMSE estimate under the Laplacian prior, which reduces to this special case when $P(H_1)$ is constant per subband as defined in (13). The main importance of using the described mixture prior is that it provides a simple methodology for making the shrinkage spatially adaptive, by adapting $P(H_1)$ and $P(H_0)$ to the local spatial context, which can be achieved using e.g., HMT [21,22] or MRF [29,31,32] models or using local spatial activity indicators as in Section IV. A question is whether such a MMSE approach would be of practical importance, regarding the complexity of the involved expressions for $E(\beta|y,H_0)$ and $E(\beta|y,H_1)$. This is possibly a subject for further research. The conditional means $E(\beta|y)$, $E(\beta|y,H_1)$ and $E(\beta|y,H_1)$ are illustrated in Fig. 2. For large $|y|$, $E(\beta|y,H_0)$ approaches $+T$ (when $y > 0$) or $-T$ (when $y < 0$). This is a consequence of the formulation $H_0 : |\beta| \leq T$, which imposes $-T \leq E(\beta|y,H_0) \leq T$. The

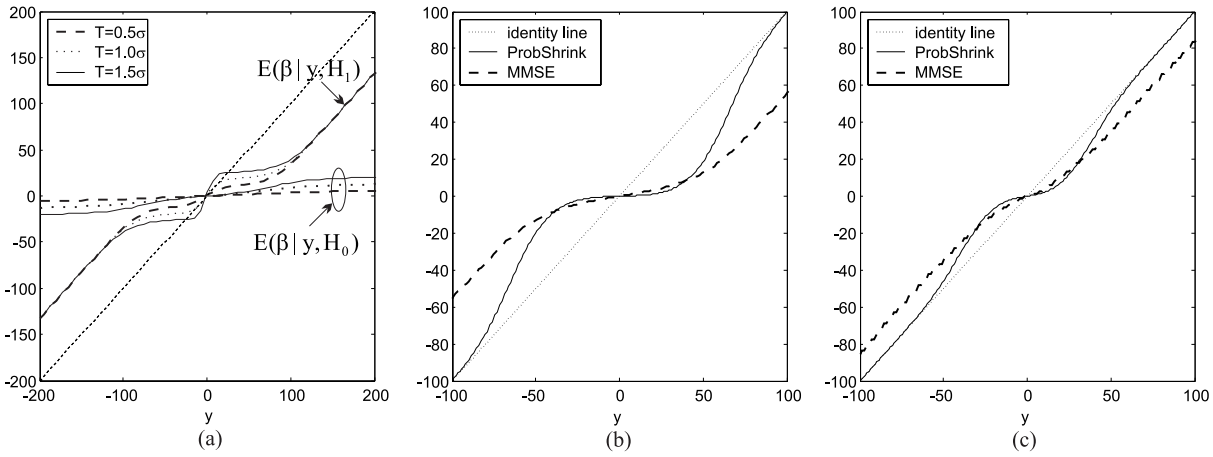


Fig. 2. (a) An illustration of $E(\beta|y, H_0)$ and $E(\beta|y, H_0)$ at the finest resolution scale for different T. (b) and (c) Comparison between the MMSE estimate $E(\beta|y)$ and the proposed *ProbShrink* rule for $T = \sigma$ at the finest and at the second finest scale, respectively.

values of $E(\beta|y, H_1)$ for large $|y|$ do not depend on the value of T. From (23) it can be shown[5] that when $|y|$ tends to infinity, $E(\beta|y, H_1)$ and $E(\beta|y)$ approach to the maximum a posteriori (MAP) estimate of β, i.e., $\lim_{|y|\to\infty} E(\beta|y, H_1) = \lim_{|y|\to\infty} E(\beta|y) = \mathrm{sgn}(y)(|y| - \sigma^2\lambda)$.

Fig. 2 also illustrates the *ProbShrink* rule in comparison to the MMSE estimate. While the MMSE estimate is analyzed here for completeness, we restrict the further analysis to *ProbShrink* rule and use the MMSE estimate for the performance comparison (Section III-D).

C. ProbShrink rule for the generalized Laplacian prior

When the noise-free first-order distribution is modelled by a generalized Laplacian (3), the conditional densities of noise-free coefficients are

$$f(\beta|H_0) = \begin{cases} B_0 \exp(-\lambda|\beta|^\nu), & \text{if } |\beta| \le T \\ 0, & \text{if } |\beta| > T, \end{cases} \tag{27}$$

and

$$f(\beta|H_1) = \begin{cases} 0, & \text{if } |\beta| \le T, \\ B_1 \exp(-\lambda|\beta|^\nu), & \text{if } \beta > T, \end{cases} \tag{28}$$

with the normalization constants (see Appendix C):

$$B_0 = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)} \quad \text{and} \quad B_1 = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Big[1 - \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)\Big]}. \tag{29}$$

[5] Note that $\lim_{y\to\infty} \Phi(y-\lambda-T) = \lim_{y\to-\infty} \Phi(-y-\lambda-T) = 1, \lim_{y\to-\infty} \Phi(y-\lambda-T) = \lim_{y\to\infty} \Phi(-y-\lambda-T) = -1$, which yields $\lim_{y\to\pm\infty} \frac{r_-(y;\lambda;T)}{r_+(y;\lambda;T)} = \pm 1$ and $\lim_{y\to\pm\infty} \frac{e^{Ty} - e^{-Ty}}{r_+(y;\lambda;T)} = 0$.

Fig. 3. Test images. **Top** left to right: *Airfield*, *Barbara*, *Boat*. **Bottom** left to right: *Couple*, *Goldhill*, *Lena*.

where $\Gamma_{inc}(x,a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$ is the *incomplete gamma function*. From (13) we have that (see also Appendix C): $P(H_1) = 1 - \int_{-\infty}^{\infty} f(\beta) d\beta = 1 - \Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)$ and thus

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{1 - \Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^{\nu}, \frac{1}{\nu}\Big)}. \tag{30}$$

The likelihood ratio $\eta = f(y|H_1)/f(y|H_0)$ is calculated using (12), where $f(y|H_0)$ and $f(y|H_1)$ result from convolving (27) and (28), respectively, with the normal density $N(0, \sigma^2)$.

D. Experimental performance evaluation

There are three goals of the experimental performance evaluation in this Section. Firstly, we wish to investigate the optimal (in terms of mean squared error) choice of the parameter T, which specifies the signal of interest in the proposed *ProbShrink* rule (11). Secondly, we wish to compare the performance of the resulting shrinkage rules under Laplacian and under generalized Laplacian priors. Finally, we wish to evaluate the performance of these new Bayesian shrinkers with respect to MAP and MMSE estimators and with respect to *BayesShrink* of [8]. Recall that the MAP estimate under the Laplacian prior is soft-thresholding with the threshold $\sigma^2\lambda = \sqrt{2}\sigma^2/\sigma_\beta$ and that *BayesShrink* is soft-thresholding with the threshold σ^2/σ_β that is for natural

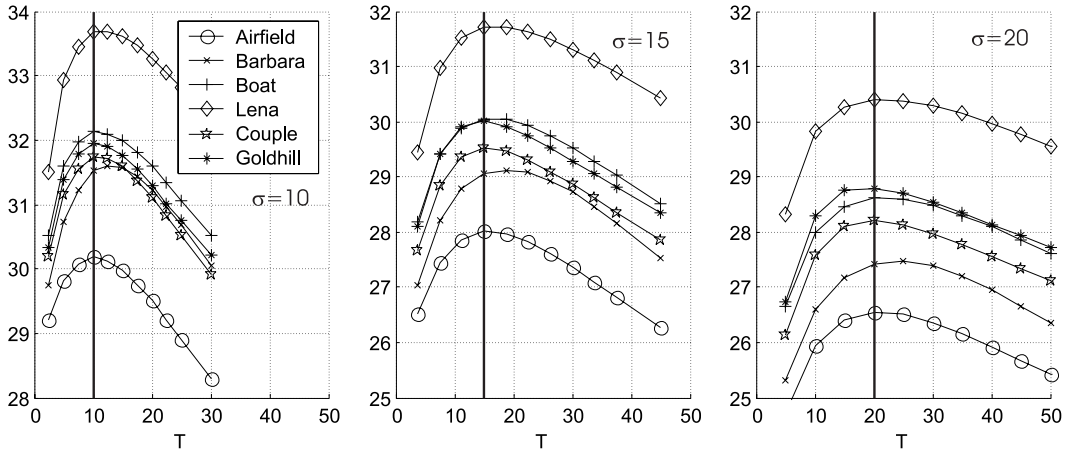


Fig. 4. Resulting PSNR[dB] values for the subband adaptive *ProbShrink* estimator under the Laplacian prior, as a function of the threshold T. From left to right the noise standard deviation is $\sigma = 10$, $\sigma = 15$ and $\sigma = 20$.

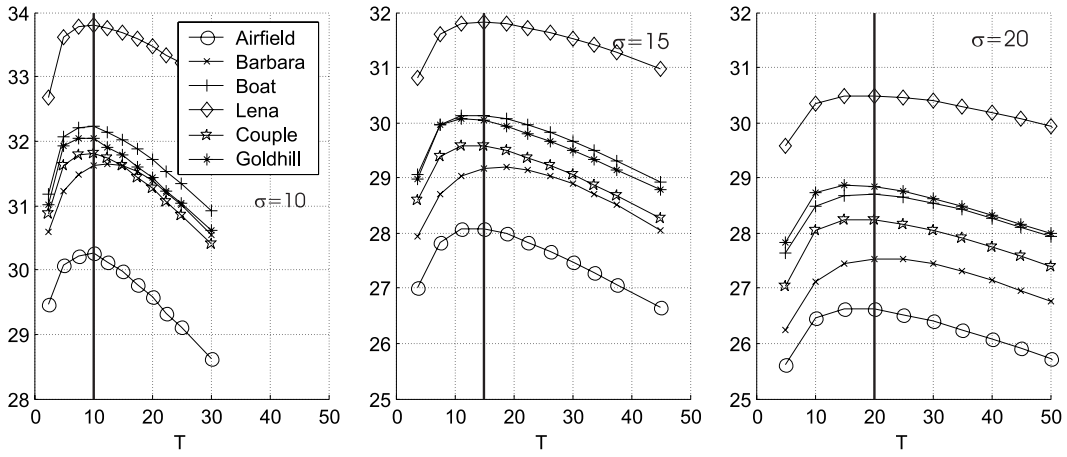


Fig. 5. Same as in Fig. 4, but for the generalized Laplacian prior.

images optimal in terms of mean squared error.

We used six representative natural images from Fig. 3, which were corrupted by various amounts of artificial additive white Gaussian noise. The experimental results are shown in Fig. 4, Fig. 5 and in Table I, where PSNR denotes the peak signal to noise ratio defined as: $\text{PSNR} = 10\log_{10}(255^2/\text{MSE})$, where MSE is the mean squared error. The diagrams in Fig. 4 and Fig. 5 demonstrate clearly that for the *ProbShrink* rule the optimal threshold value is a function of noise level. For both the Laplacian and the generalized Laplacian prior the PSNR peaks at $T \simeq \sigma$. All the following results in this paper correspond to this threshold selection.

TABLE I

EXPERIMENTAL PSNR[dB] RESULTS OF SEVERAL BAYESIAN WAVELET SHRINKERS UNDER LAPLACIAN (Lap.) AND GENERALIZED LAPLACIAN (gen. Lap) PRIORS, FOR THE DAUBECHIES' LEAST ASYMETRICAL WAVELET WITH EIGHT VANISHING MOMENTS.

		Standard deviation of noise			
Prior	Estimator	10	15	20	25
		AIRFIELD			
Lap.	noisy image	28.14	24.63	22.10	20.19
Lap.	MAP	30.18	27.86	26.33	25.25
Lap.	MMSE	**30.43**	**28.19**	**26.71**	**25.67**
Lap.	*ProbShrink*	30.19	28.01	26.55	25.52
gen. Lap.	*ProbShrink*	30.25	28.09	26.62	25.56
gen. Lap.	*BayesShrink* soft-thresholding	30.24	27.98	26.49	25.44
		BARBARA			
	noisy image	28.12	24.59	22.09	20.17
Lap.	MAP	31.33	28.96	27.36	26.08
Lap.	MMSE	31.35	28.93	27.34	26.18
Lap.	*ProbShrink*	31.52	29.05	27.41	26.21
gen. Lap.	*ProbShrink*	**31.62**	**29.17**	**27.54**	**26.32**
gen. Lap.	*BayesShrink* soft-thresholding	31.24	28.86	27.32	26.20
		BOAT			
	noisy image	28.15	24.62	22.10	20.17
Lap.	MAP	32.00	29.92	28.42	27.37
Lap.	MMSE	**32.23**	**30.14**	**28.70**	**27.70**
Lap.	*ProbShrink*	32.13	30.05	28.63	27.63
gen. Lap.	*ProbShrink*	**32.23**	30.13	**28.70**	27.69
gen. Lap.	*BayesShrink* soft-thresholding	32.01	29.98	28.55	27.54
		COUPLE			
	noisy image	28.15	24.60	22.11	20.18
Lap.	MAP	31.65	29.26	27.89	26.95
Lap.	MMSE	31.81	**29.61**	**28.29**	**27.32**
Lap.	*ProbShrink*	31.73	29.53	28.20	27.18
gen. Lap.	*ProbShrink*	**31.82**	29.60	28.24	27.24
gen. Lap.	*BayesShrink* soft-thresholding	31.70	29.46	28.08	27.09
		GOLDHILL			
	noisy image	28.13	24.63	22.11	20.20
Lap.	MAP	31.83	29.72	28.50	27.66
Lap.	MMSE	**32.16**	**30.16**	**28.95**	**28.09**
Lap.	*ProbShrink*	31.95	30.02	28.80	27.88
gen. Lap.	*ProbShrink*	32.04	30.05	28.85	27.94
gen. Lap.	*BayesShrink* soft-thresholding	31.94	29.95	28.71	27.82
		LENA			
	noisy image	28.13	24.60	22.12	20.16
Lap.	MAP	33.43	31.47	30.18	29.23
Lap.	MMSE	33.62	31.65	30.36	29.40
Lap.	*ProbShrink*	33.69	31.70	30.39	29.41
gen. Lap.	*ProbShrink*	**33.80**	**31.82**	**30.49**	**29.51**
gen. Lap.	*BayesShrink* soft-thresholding	33.47	31.53	30.26	29.30

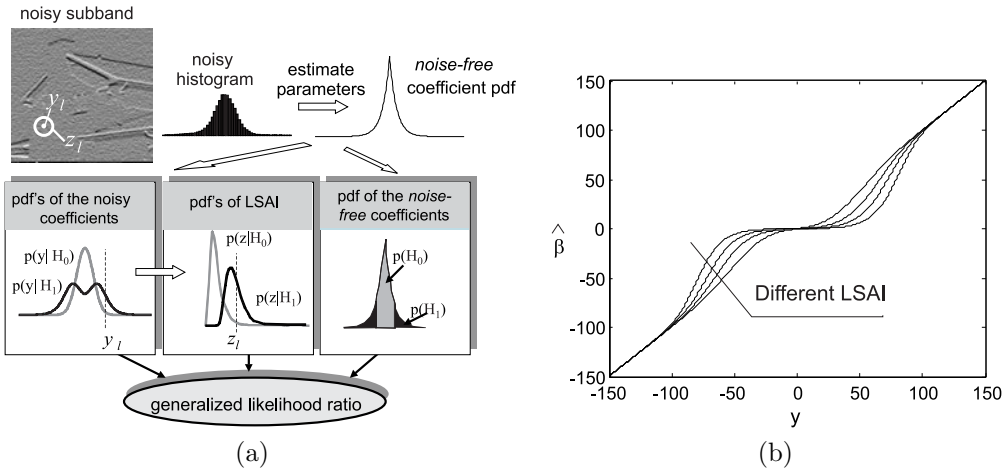
The results in Table I demonstrate that *ProbShrink* rule outperforms the MAP estimation in terms of mean squared error and that it yields similar MSE performance as the classical MMSE estimate $E(\beta|y)$ from (2). Next, the results demonstrate that *ProbShrink* with the generalized Laplacian prior yields slightly higher PSNR as compared to the same rule with the Laplacian prior. Finally, the results in Table I also demonstrate that the proposed *ProbShrink* rule outperforms *BayesShrink* rule on all tested images. Since *BayesShrink* is soft-thresholding with the MSE optimum threshold, we can deduce that *ProbShrink* (at least on the tested images) outperforms soft thresholding with any threshold that is constant per subband.

IV. Spatially adaptive Bayesian shrinkage

The shrinkage approach analyzed so far was subband-adaptive: if two noisy coefficients from the same subband were of equal magnitudes than they were shrunk by the same amount no matter their spatial position and no matter their local surrounding. Now we adapt the estimator to the local spatial context in the image. For the *ProbShrink* rule $\beta = P(H_1|y)y$, the most natural way to achieve such a spatial adaptation is to estimate the prior probability of signal presence $P(H_1)$ adaptively for each coefficient instead of fixing it per subband.

Similarly as in our related work [33], we achieve this by conditioning the hypothesis H_1 on a *local spatial activity indicator* (LSAI) such as the locally averaged coefficient magnitude or the local variance. In general, LSAI can be defined as the parent coefficient magnitude as in [18], the average of the (weighted) coefficient magnitudes [20, 23] or the local variance [24] within a small window, the estimate of the local regularity [5, 29, 32] at the corresponding position, etc. To estimate the probability that "signal of interest" is present at the position l we now take into account not only the global coefficient histogram but we also take into account the local spatial activity indicator. The probability of the hypothesis H_1 at the position l is thus estimated *given* LSAI z_l, i.e., it is conditioned on z_l. Starting from (11) and replacing the ratio of "global" probabilities $P(H_1)/P(H_0)$ with

$$\frac{P(H_1|z_l)}{P(H_0|z_l)} = \frac{f(z_l|H_1)}{f(z_l|H_0)} \frac{P(H_1)}{P(H_0)}, \tag{31}$$



(a) (b)

Fig. 6. (a) An illustration of the proposed denoising method, where pdf denotes the probability density function and where LSAI denotes the local spatial activity indicator. (b) The resulting shrinkage rule is a family of characteristics, which corespond to different values of LSAI.

we obtain a spatialy adaptive shrinkage estimator

$$\hat{\beta}_l = \frac{\eta_l \xi_l \mu}{1 + \eta_l \xi_l \mu} y_l, \tag{32}$$

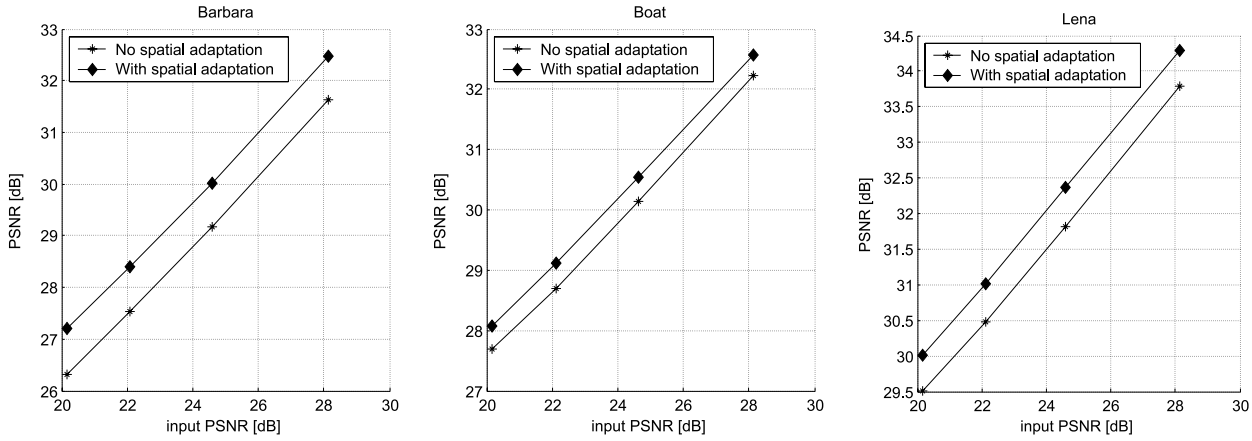
where

$$\eta_l = \frac{f(y_l|H_1)}{f(y_l|H_0)}, \quad \xi_l = \frac{f(z_l|H_1)}{f(z_l|H_0)} \quad \text{and} \quad \mu = \frac{P(H_1)}{P(H_0)}. \tag{33}$$

The characteristic parts of the method are illustrated in Fig. 6, where the generalized likelihood ratio denotes the product $\eta_l \xi_l \mu$. The proposed method has a nice heuristic explanation: each coefficient is shrunk according to how probable it is that it presents useful information, based on its value (via η_l), based on a measurement from the local surrounding (via ξ_l) and based on the global statistical properties of the coefficients in a given subband (via μ). The estimator (32) is of the same form as in our previous method [33]. The difference is that μ, η_l and ξ_l were in [33] estimated differently, based on a preliminary coefficient classification according to user-defined notion of the signal of interest, while now we develop an alternative analytical approach which does not involve any user-optimized parameters and binary classifications.

We define LSAI as the locally averaged magnitude of the coefficients in a relatively small square window $\delta(l)$ of a fixed size N, within the same subband:

$$z_l = \frac{1}{N} \sum_{k \in \delta(l)} \omega_l, \tag{34}$$



(a) (b) (c)

Fig. 7. Performance of the proposed *ProbShrink* method without spatial adaptation (subband adaptive shrinkage) and with spatial adaptation on three test images: (a) *Barbara*, (b) *Boat* and (c) *Lena*.

(a) (b) (c) (d)

Fig. 8. Visual performance of different versions of the proposed *ProbShrink* method. (a) Noisy *Barbara* image, $\sigma = 20$, PSNR=22.09dB. (b) Subband adaptive shrinkage in the orthogonal transform, PSNR=27.54dB. (c) Spatially adaptive shrinkage in the orthogonal transform PSNR=28.4dB. (d) Spatially adaptive shrinkage in the non-decimated transform PSNR=29.53dB.

where ω_l denotes the coefficient magnitude $\omega_l = |y_l|$. For practical reasons, we simplify the statistical characterization of z_l considerably assuming that all the coefficients within the small window are equally distributed and conditionally independent. With these simplifications, $f(Nz_l|H_1)$ is given by N convolutions of $f(\omega_l|H_1)$ with itself and $f(Nz_l|H_0)$ is given by N convolutions of $f(\omega_l|H_0)$ with itself. The resulting spatially adaptive estimator (32) yields a significant improvement with respect to the subband-adaptive estimator (11) as it is illustrated in Fig. 7). In all cases, the window size 3x3 was used, which was experimentally found optimal. Fig. 8 demonstrates that the visual improvement resulting from the spatial adaptation of the estimator is also evident.

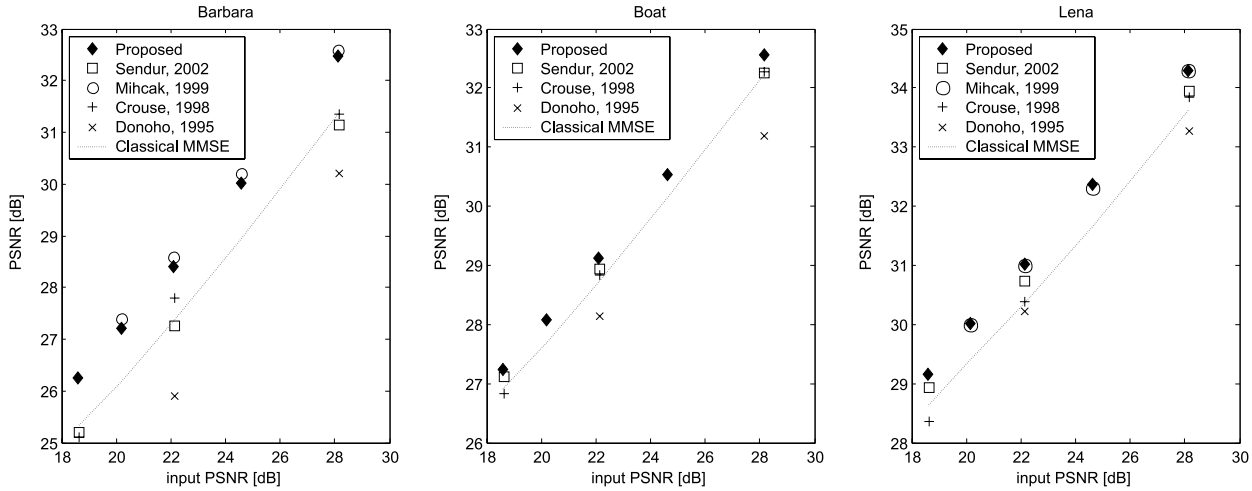


Fig. 9. Performance of the proposed spatially adaptive shrinkage in comparison with several recent methods using orthogonal wavelet transform: [7] (Donoho, 1995), [21] (Crouse, 1998), [24] (Mihcak, 1999) and [18] (Sendur, 2002). "Classical MMSE" denotes the MMSE estimation under the Laplacian prior.

A. Results in the orthogonal wavelet representation

Here we present the results of the proposed spatially adaptive algorithm with the orthogonal wavelet transform, where we used five decomposition levels, the least asymmetrical Daubechies wavelet *symmlet* with eight vanishing moments [1]. We used the generalized Laplacian prior and the fixed window size 3x3. Fig. 9 shows the results in comparison with several representative denoising methods that use the orthogonal transform: the recent bivariate shrinkage method of Şendur and Selesnick [18], the locally adaptive Wiener method of Mihcak et al [24], the Hidden Markov Tree (HMT) approach of Crouse *et al* [21] and the soft thresholding of Donoho and Johnstone [7] with the SURE threshold (that minimizes the Stein's unbiased risk estimate). We also show in Fig. 9 the results of the classical MMSE estimation (2) with the Laplacian prior for noise-free coefficients. The experimental results indicate that the proposed method offers the results that are among the best available ones where the orthogonal wavelet transform is used. On the three tested 512x512 images, *Lena*, *Barbara* and *Boat* the new method provides higher signal to noise ratio as compared to the bivariate shrinkage of [18] and the HMT approach of [21]. Our results also compare well with those of [24], which are the best reported ones with the orthogonal transform.

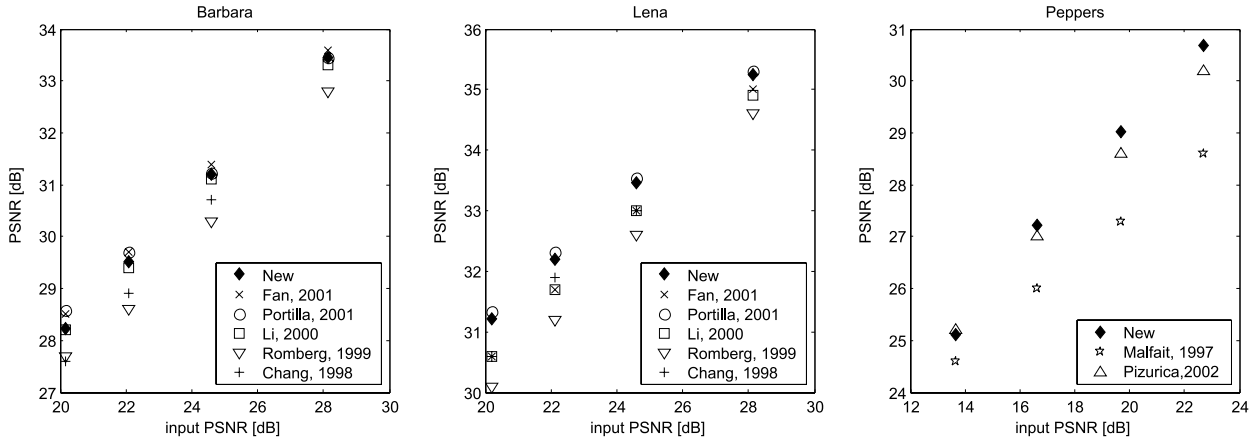


Fig. 10. Performance of the proposed method implemented with a non-decimated transform in comparison with several recent methods, which use a redundant wavelet transform: [29] (Malfait, 1997), [20] (Chang, 1998), [22] (Romberg, 1999), [28] (Li, 2000), [23] (Fan, 2001), [26] (Portilla, 2001) and [32] (Pizurica, 2002).

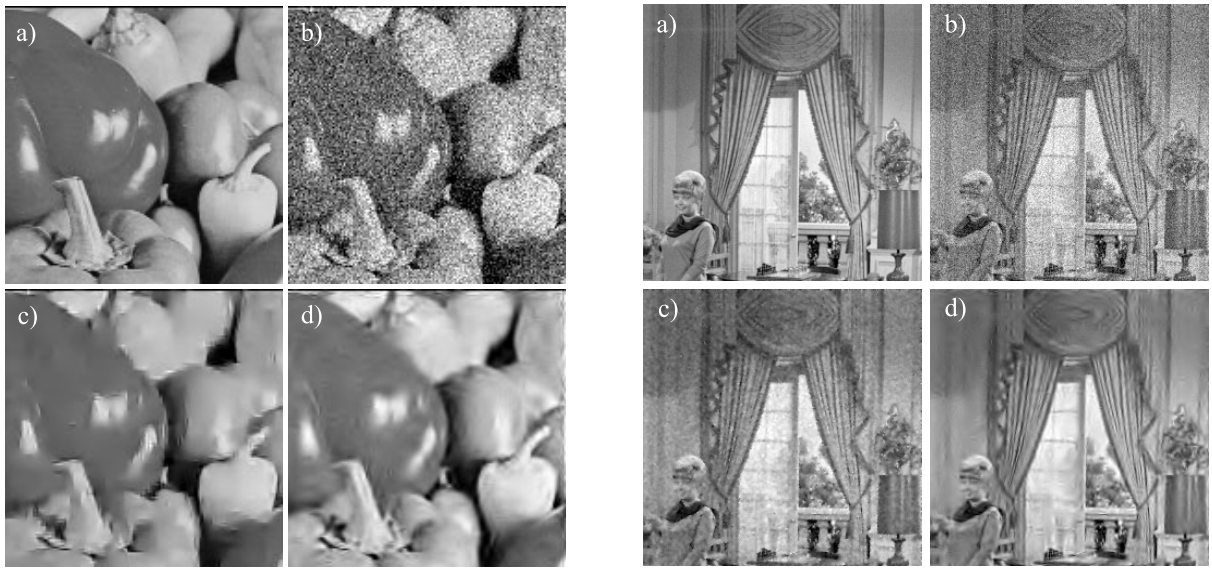


Fig. 11. **Left**: (a) noise-free part of the *Peppers* image, (b) noisy image, $\sigma = 37.5$, (c) the result of the MRF-based wavelet domain method of [32] and (d) the result of the new method. **Right**: noise-free part of the *Couple* image, (b) noisy image, $\sigma = 15$, (c) the result of the spatially adaptive Matlab's Wiener filter in the image domain and (d) the result of the new method.

B. Results in the redundant wavelet representation

In order to compare the performance of the proposed method with the best available results in the field, we need to use a redundant rather than orthogonal wavelet representation. In this respect, two approaches are common: (1) *cycle spining* [35]: (apply the orthogonal transform to several cyclically shifted image versions and average over unshifted denoising

results) and (2) denoising in a *non-decimated* (stationary) wavelet representation. Here we used the non-decimated transform, implemented with the algorithm *à trous* [4].

A non-decimated wavelet transform maps input white noise into correlated one which makes the modelling more complicated. A possible solution in this respect is to divide the coefficients in each subband at the scale 2^j into 2^{2^j} sets of non-correlated coefficients and to process these coefficient sets independently as in [20]. With these coefficient separations however the benefit of the non-decimated representation is not fully exploited (e.g., intrascale comparisons between the nearest-neighboring coefficients are no longer possible). Otherwise, if the estimation is not performed over sets of non-correlated coefficients certain model simplifications are needed, like, e.g., in [26, 29, 31, 32]. We performed experiments both with and without separating the coefficients into the sets of non-correlated elements. The resulting signal to noise ratio was on all tested images 0.1-0.2dB higher *without* the coefficient separation (window sizes were optimized in both cases).

In Fig. 10, the results are plotted in comparison with seven recently published methods, which use overcomplete wavelet transform: spatially adaptive thresholding of [20]; locally adaptive Wiener filtering combined with a directional filtering along the edges [28], MMSE estimation with a Gaussian scale mixture prior of [26], two different MMSE estimation variants with HMT models [22, 23] and two earlier probabilistic shrinkage methods based on MRF priors [29, 32]. The results are plotted for three test images: 512x512 *Lena* and *Barbara* images and 256x256 *Peppers* image. We experimented only with two types of orthogonal wavelets: Daubechies wavelets and symmlets [1, 4]. Among these, on the *Barbara* image the best results were obtained using the symmlet with eight vanishing moments, while for *Lena* and *Peppers* images the best results were obtained using the shortest Daubechies wavelet (i.e., the Haar wavelet). Visual quality of results is illustrated in Fig. 11. In all cases, we used four decomposition levels and the square window size 7x7, which was experimentally found optimal.

The results in Fig. 10 demonstrate that the new method outperforms more complex related ones that are based on HMT priors [22] and on MRF priors [29, 32]. Visual improvement is

illustrated in Fig. 11. The new, non-recursive method is much faster as compared to our previous MRF-based method from [32]. On a Pentium IV processor with 1.8GHz the new method takes 14s to process 512x512 image; the processing time of the MRF based method [32] is three times as long with the same processor. Here proposed method is also less complex as compared to the sophisticated MMSE approach of [26], while yielding a similar MSE performance (in [26], the reported processing time for a 512x512 image was 12.8 min on a Pentium III processor with 900 MHz). Fig. 10 also shows that in comparison with the approaches of similar complexity [20, 28], the new method yields a significant improvement.

C. Implementation details

In the proposed method, the parameters λ and ν of the generalized Laplacian prior for noise-free data are estimated from the noisy histogram in each subband, like in [8, 15]. In particular, we first solve the equation

$$\frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} = \frac{m_4 + 3\sigma_n^4 - 6\sigma^2\sigma_y^2}{(\sigma_y^2 - \sigma^2)^2}. \tag{35}$$

where σ_y^2 and m_4 denote the variance and fourth moment of the noisy histogram respectively. Then we compute the scale parameter ν as

$$\lambda = \left((\sigma_y^2 - \sigma^2)\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}\right)^{-1/2}. \tag{36}$$

For the Laplacian prior ($\nu = 1$) we have $\lambda = [0.5(\sigma_y^2 - \sigma^2)]^{-1/2}$. The results in this paper were obtained assuming that the noise standard deviation σ was known (as it is usual for reporting the results in case of artificially added noise). In practice the noise standard deviation is usually not known in advance, but its reliable estimate can be obtained as the median absolute deviation of the coefficients in the highest frequency subband devided by 0.6745 [7].

V. Conclusion

We developed a new wavelet domain denoising method based on probability that a given coefficient represents a significant noise-free component called here "signal of interest". The

new method differs significantly from the previous related ones that were based on Markov Random Fields and presents a complementary extension to our recent heuristic method [33].

The main conclusion of this paper is that a "probabilistic wavelet shrinkage" approach that we named here *ProbShrink* presents in terms of mean squared error performance a valuable alternative to more common Bayesian methods like MMSE and MAP estimators and the thresholding with the MSE optimum threshold. We demonstrated this both in case of estimators that are adapted to the marginal subband statistics only and in case where spatially adaptive versions of such estimators are used. This also presents a new theoretical and practical motivation for a class of previous related methods [29–33]. As side results of this analysis we also developed analytical expressions for MMSE estimation with Laplacian mixture priors, opening some new possibilities for further research in this field.

New spatially adaptive denoising method proposed here yields superior results as compared to some much more complex recent approaches based on HMT and MRF models. These results motivate strongly further development of the presented concept. We believe that main improvements of the current scheme should come from defining more effective local spatial activity indicators and from modelling accurately their conditional densities given the presence and given the absence of a signal of interest. Improvements can also result from replacing the classical wavelet transform with complex wavelets [36] or with a steerable pyramid as in [26,27].

VI. Appendix A

Proof of the proposition 1

When the noise-free coefficient histogram is modelled by a Laplacian density we have that

$$f(\beta|H_0) = \begin{cases} A_0 \exp(-\lambda|\beta|), & \text{if } |\beta| \le T \\ 0, & \text{if } |\beta| > T, \end{cases} \quad \text{and} \quad f(\beta|H_1) = \begin{cases} 0, & \text{if } |\beta| \le T, \\ A_1 \exp(-\lambda|\beta|), & \text{if } \beta > T, \end{cases}$$

with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$. The conditional means $E(\beta|y, H_1)$ and $E(\beta|y, H_0)$ from (8) and (9), and the likelihoods $f(\beta|y, H_0)$ and $f(\beta|y, H_1)$ from (12) can be

written as

$$E(\beta|y,H_1)=\frac{K_{H_1}(y;\lambda,T)}{m_{H_1}(y;\lambda,T)},\quad \text{and}\quad E(\beta|y,H_0)=\frac{K_{H_0}(y;\lambda,T)}{m_{H_0}(y;\lambda,T)}, \tag{37}$$

$$f(\beta|y,H_1)=A_1 m_{H_1}(y;\lambda,T)\quad \text{and}\quad f(\beta|y,H_0)=A_1 m_{H_0}(y;\lambda,T), \tag{38}$$

where

$$m_{H_1}(y;\lambda,T)=\int_{-\infty}^{-T} e^{\lambda\beta}\phi(y-\beta)d\beta \;+\; \int_{T}^{\infty} e^{-\lambda\beta}\phi(y-\beta)d\beta, \tag{39}$$

$$K_{H_1}(y;\lambda,T)=\int_{-\infty}^{-T} \beta e^{\lambda\beta}\phi(y-\beta)d\beta \;+\; \int_{T}^{\infty} \beta e^{-\lambda\beta}\phi(y-\beta)d\beta, \tag{40}$$

$$m_{H_0}(y;\lambda,T)=\int_{-T}^{0} e^{\lambda\beta}\phi(y-\beta)d\beta \;+\int_{0}^{T} e^{-\lambda\beta}\phi(y-\beta)d\beta, \tag{41}$$

$$K_{H_0}(y;\lambda,T)=\int_{-T}^{0} \beta e^{\lambda\beta}\phi(y-\beta)d\beta \;+\; \int_{0}^{T} \beta e^{-\lambda\beta}\phi(y-\beta)d\beta. \tag{42}$$

On a nonnegative interval $[a,b]$, one can verify the following two identities [16]

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

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

Using these identities, we find

$$m_{H_1}(y;\lambda,T)=I(y;T,\infty,\lambda)+I(-y;T,\infty,\lambda)=$$
$$\sqrt{2\pi}\phi(y)\Big[e^{(y-\lambda)^2/2}(1-\Phi(T-y+\lambda))+e^{(y+\lambda)^2/2}(1-\Phi(T+y+\lambda))\Big]=$$
$$\sqrt{2\pi}\phi(y)r_+(y;\lambda;T),$$

and

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

where

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

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

Thus the conditional mean of β given y under H_1 is

$$E(\beta|y,H_1) = \frac{K_{H_1}(y;\lambda,T)}{m_{H_1}(y;\lambda,T)} = y - \frac{\lambda r_-(y;\lambda,T) - \frac{1}{\sqrt{2\pi}}e^{-(T^2/2+\lambda T)}(e^{Ty}-e^{-Ty})}{r_+(y;\lambda,T)}.$$

Similarly, if we introduce $\Psi(a;T) = \Phi(a+T) - \Phi(a)$, we find

$$m_{H_0}(y;\lambda,T) = I(y;0,T,\lambda) + I(-y;0,T,\lambda) =$$
$$\sqrt{2\pi}\phi(y)\Big[e^{(y-\lambda)^2/2}\Psi(\lambda-y;T) + e^{(y+\lambda)^2/2}\Psi(\lambda+y;T)\Big] =$$
$$\sqrt{2\pi}\phi(y)\rho_+(y;\lambda;T),$$

and

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

where

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

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

Thus the conditional mean of β given y under H_0 is

$$E(\beta|y,H_0) = \frac{K_{H_0}(y;\lambda,T)}{m_{H_0}(y;\lambda,T)} = y - \frac{\lambda\rho_-(y;\lambda,T) + \frac{1}{\sqrt{2\pi}}e^{-(T^2/2+\lambda T)}(e^{Ty}-e^{-Ty})}{\rho_+(y;\lambda,T)}$$

as was given in (19). Finally, the conditional densities of noisy wavelet coefficients are

$$f(y|H_0) = A_0 m_{H_0}(y;\lambda,T) = A_0\sqrt{2\pi}\phi(y)\rho_+(y;\lambda;T), \tag{43}$$

$$f(y|H_1) = A_1 m_{H_1}(y;\lambda,T) = A_1\sqrt{2\pi}\phi(y)r_+(y;\lambda;T), \tag{44}$$

which with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T}-1)$ and $A_1 = (\lambda/2)e^{\lambda T}$ yield

$$\eta = \frac{f(y|H_1)}{f(y|H_1)} = (e^{\lambda T}-1)\frac{r_+(y;\lambda,T)}{\rho_+(y;\lambda,T)}, \tag{45}$$

which completes the proof of the Proposition 2.

VII. Appendix B

Generalization of the Proposition 2 to arbitrary σ

To generalize the procedure from Appendix B to arbitrary values of the noise standard deviation σ, the integrals $I(y;a,b,\lambda)$ and $II(y;a,b,\lambda)$ from (39)-(42) need to be replaced by $\mathcal{I}(y;a,b,\lambda,\sigma)$ and $\mathcal{II}(y;a,b,\lambda,\sigma)$, respectively, where

$$\mathcal{I}(y;a,b,\lambda,\sigma) = \int_a^b e^{-\lambda\beta}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\beta)^2}{2\sigma^2}}d\beta,$$

$$\mathcal{II}(y;a,b,\lambda,\sigma) = \int_a^b \beta e^{-\lambda\beta}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\beta)^2}{2\sigma^2}}d\beta.$$

By introducing a change of variables $\beta' = \frac{\beta}{\sigma}$ it follows that

$$\mathcal{I}(y;a,b,\lambda,\sigma) = \frac{1}{\sigma}\int_a^b e^{-\sigma\lambda\beta'}\phi(y/\sigma-\beta')d\beta' = \frac{1}{\sigma}I(\frac{y}{\sigma};a,b,\sigma\lambda) \tag{46}$$

$$\mathcal{II}(y;a,b,\lambda,\sigma) = \int_a^b \beta' e^{-\sigma\lambda\beta'}\phi(y/\sigma-\beta')d\beta' = II(\frac{y}{\sigma};a,b,\sigma\lambda) \tag{47}$$

This yields

$$f(\beta|y,H_0) = \frac{A_0}{\sigma}m_{H_0}(\frac{y}{\sigma};\sigma\lambda,T) = A_0\sqrt{2\pi}\phi(\frac{y}{\sigma})\rho_+(\frac{y}{\sigma};\sigma\lambda;T) \tag{48}$$

$$f(\beta|y,H_1) = \frac{A_1}{\sigma}m_{H_1}(\frac{y}{\sigma};\sigma\lambda,T) = A_1\sqrt{2\pi}\phi(\frac{y}{\sigma})r_+(\frac{y}{\sigma};\sigma\lambda;T), \tag{49}$$

and

$$E(\beta|y,H_0) = \frac{\sigma K_{H_0}(\frac{y}{\sigma};\sigma\lambda,T)}{m_{H_0}(\frac{y}{\sigma};\sigma\lambda,T)} = y - \frac{\sigma^2\lambda\rho_-(\frac{y}{\sigma};\sigma\lambda,T) + \frac{\sigma}{\sqrt{2\pi}}e^{-(T^2/2+\sigma\lambda T)}(e^{Ty/\sigma}-e^{-Ty/\sigma})}{\rho_+(\frac{y}{\sigma};\sigma\lambda,T)} \tag{50}$$

$$E(\beta|y,H_1) = \frac{\sigma K_{H_1}(\frac{y}{\sigma};\sigma\lambda,T)}{m_{H_1}(\frac{y}{\sigma};\sigma\lambda,T)} = y - \frac{\sigma^2\lambda r_-(\frac{y}{\sigma};\sigma\lambda,T) - \frac{\sigma}{\sqrt{2\pi}}e^{-(T^2/2+\sigma\lambda T)}(e^{Ty/\sigma}-e^{-Ty/\sigma})}{r_+(\frac{y}{\sigma};\sigma\lambda,T)} \tag{51}$$

VIII. Appendix C

Derivations for the shrinkage rule with the generalized Laplacian prior

For the generalized Laplacian prior $f(\beta) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\exp(-|\lambda\beta|^\nu)$, we have

$$\int_{-T}^{T} f(\beta)d\beta = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\int_{-T}^{T}\exp(-|\lambda\beta|^\nu)d\beta = \frac{\lambda\nu}{\Gamma(\frac{1}{\nu})}\int_{0}^{T}\exp(-(\lambda\beta)^\nu)d\beta. \tag{52}$$

By introducing the change of variables $t = (\lambda\beta)^\nu$ it follows that $d\beta = \frac{1}{\lambda\nu}\, t^{\frac{1}{\nu}-1}\, dt$ and thus

$$\int_{-T}^{T} f(\beta)d\beta = \frac{1}{\Gamma\Big(\frac{1}{\nu}\Big)}\int_{0}^{(\lambda T)^\nu} t^{\frac{1}{\nu}-1}e^{-t}dt = \Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big),$$

where $\Gamma_{inc}(x,a) = \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$ is the *incomplete gamma function*. From (13), we have

$$\mu = \frac{P(H_1)}{P(H_0)} = \frac{1-\int_{-T}^{T} f(\beta)d\beta}{\int_{-T}^{T} f(\beta)d\beta} = \frac{1-\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}{\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)}, \tag{53}$$

as it was given in (30). For the conditional densities $f(\beta|H_0)$ and $f(\beta|H_1)$ of noise-free coefficients from (27) and (28), the normalization constants B_0 and B_1 are

$$B_0 = \Big(\int_{-T}^{T}\exp(-|\lambda\beta|^\nu)d\beta\Big)^{-1} = \Big(\frac{2\Gamma(\frac{1}{\nu})}{\lambda\nu}\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)} \tag{54}$$

and

$$B_1 = \Big(2\int_{T}^{\infty}\exp(-|\lambda\beta|^\nu)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(2\int_{T}^{\infty} f(\beta)d\beta\Big)^{-1} = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})}\Big(1-\int_{-T}^{T} f(\beta)d\beta\Big)^{-1} \tag{55}$$

$$= \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})\Big[1-\Gamma_{inc}\Big((\lambda T)^\nu, \frac{1}{\nu}\Big)\Big]}.$$

References

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

[2] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intel.*, vol. 11, no. 7, pp. 674–693, 1989.

[3] A. Cohen and J. Kovačević, "Wavelets: the mathematical background," *Proc. IEEE*, vol. 84, no. 4, pp. 514–522, Apr. 1996.

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

[5] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Information Theory*, vol. 38, no. 8, pp. 617–643, Mar. 1992.

[6] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.

[7] 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.

[8] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Proc.*, vol. 9, no. 9, pp. 1532–1546, Sept. 2000.

[9] B. Vidakovic, "Nonlinear wavelet shrinkage with bayes rules and bayes factors," *J. of the American Statistical Association*, vol. 93, pp. 173–179, 1998.

[10] ——, "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.

[11] 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.

[12] D. Leporini, J. C. Pasquet, and H. Krim, "Best basis representation with prior statistical models," in *Lecture Notes in Statistics*, P. M Ed.

[13] 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.

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

[17] 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.

[18] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet -based denoising exploiting interscale dependency," *IEEE Trans. Signal Proc.*, vol. 50, no. 11, pp. 2744 –2756, Nov. 2002.

[19] E. P. Simoncelli, "Modeling the joint statistics of image in the wavelet domain," in *Proc. SPIE Conf. on Wavelet Applications in Signal and Image Processing VII*, Denver, CO.

[20] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," in *Proc. IEEE Internat. Conf. on Image Proc.*, Chicago, IL, Oct. 1998.

[21] 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.

[22] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree structured image modeling using wavelet-domain hidden markov models," in *Proc. SPIE Technical Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, Denver, CO, July 1999.

[23] G. Fan and X. G. Xia, "Image denoising using local contextual hidden markov model in the wavelet domain," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 125–128, May 2001.

[24] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Proc. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.

[25] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising using a local gaussian scale mixture model in the wavelet domain," in *Proc. SPIE, 45th Annual Meeting*, San Diego, July 2000.

[26] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive wiener denoising using a gaussian scale mixture model in the wavelet domain," in *Proc. IEEE Internat. Conf. on Image Proc.*, Thessaloniki, Greece, Oct. 2001.

[27] ——, "Image denoising using gaussian scale mixtures in the wavelet domain," *IEEE Trans.*

Image Proc. (accepted).

[28] X. Li and M. Orchard, "Spatially adaptive denoising under overcomplete expansion," in *Proc. IEEE Internat. Conf. on Image Proc.*, Vancouver, Canada, Sept. 2000.

[29] 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.

[30] M. Jansen and A. Bultheel, "Geometrical priors for noisefree wavelet coefficients in image denoising," in *Bayesian inference in wavelet based models*, ser. Lecture Notes in Statistics, P. Müller and B. B. Vidakovic, Eds., vol. 141. Springer Verlag, 1999, pp. 223–242.

[31] ——, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 629–639, 2001.

[32] 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.

[33] ——, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Medical Imaging*, vol. 22, no. 3, pp. 323–331, 2003.

[34] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 434–443, May 1968.

[35] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., Springer Verlag, New York, 1995, pp. 125–150.

[36] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.