Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 1370–1382
Pattern Recognition Letters
www.elsevier.com/locate/patrec

A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior

Larbi Boubchir *, Jalal M. Fadili

Image Processing Group, GREYC CNRS UMR 6072 - ENSICAEN, 6, Bd Maréchal Juin, 14050 Caen, France

Received 29 April 2005; received in revised form 3 January 2006
Available online 23 March 2006

Communicated by Prof. L. Younes

Abstract

In this paper, a nonparametric Bayesian estimator in the wavelet domains is presented. In this approach, we propose a prior statistical model based on the α-stable densities adapted to capture the sparseness of the wavelet detail coefficients. An attempt to apply this model in the context of wavelet denoising have been already proposed in (Achim, A., Bezerianos, A., Tsakalides, P., 2001. Novel Bayesian multiscale method for speckle removal in medical ultrasound images. IEEE Trans. Med. Imag. 20, 772–783). However, despite its efficacy in modeling the heavy tail behavior of the empirical wavelet coefficients histograms, their denoiser proves very poor in practice especially at low SNRs. It suffers from many drawbacks such as numerical instability because of the lack of a closed-form expression of the Bayesian shrinkage rule, and the weakness of the estimator of the hyperparameters associated with the α-stable prior. Here, we propose to overcome these limitations using the scale mixture of Gaussians theorem as an analytical approximation for α-stable densities, which is not known in general, in order to obtain a closed-form expression of our Bayesian denoiser.

Keywords: Wavelets; Bayesian denoiser; α-stable; Gaussian mixture model; Posterior conditional mean

1. Introduction

Nonparametric wavelet-based regression has been a fundamental tool in data analysis over the past two decades and is still an expanding area of ongoing research. The goal is to recover an unknown image, say g, based on sampled data that are contaminated with noise. Only very general assumptions about g are made such as that it belongs to a certain class of functions (e.g. Besov space). Nonparametric regression (or denoising) techniques provide a very effective and simple way of finding structure in datasets without imposing of a parametric regression model. During the 1990s, the nonparametric regression literature was arguably dominated by nonlinear wavelet shrinkage and wavelet thresholding estimators (Donoho and Johnstone, 1994, 1995). These estimators are a new subset of an old class of nonparametric regression estimators, namely orthogonal series methods. Moreover, these estimators are easily implemented through fast algorithms so they are very appealing in practical situations (Mallat, 1989).

Since the seminal papers by Donoho and Johnstone (1994, 1995), the image processing literature have been inundated by hundreds of papers applying or proposing modifications of the original algorithm in estimation and/or restoration problems. Various alternatives to wavelet thresholding have been developed in (Vidakovic, 1999). Authors in (Donoho and Johnstone, 1995) proposed the SureShrink estimator. Nason (1996) considered estimators based on the cross-validation principle in order to determine the regularization parameter. In (Abramovich and Benjamini, 1996; Ogden and Parzen, 1996), authors considered thresholding as a multiple hypotheses testing

* Corresponding author. Tel.: +33 231452920; fax: +33 231452698.
E-mail addresses: Larbi.Boubchir@greyc.ensicaen.fr (L. Boubchir), Jalal.Fadili@greyc.ensicaen.fr (J.M. Fadili).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.01.008

procedure. In (Hall et al., 1997; Efromovich, 2000) suggested that the wavelet coefficients could be thresholded in blocks rather than term-by-term. The wavelet block thresholding estimators have excellent mean squared error performances relative to wavelet term-by-term thresholding estimators in finite sample situation.

The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few large detail coefficients contain information about the underlying image. It is then legitimate to impose a prior designated to model the sparsity of the wavelet representation. Various Bayesian approaches for nonlinear wavelet denoising have been recently proposed. These estimators have been shown to be effective and it is argued that they are less ad hoc than the classical proposals discussed above. In the Bayesian approach a prior distribution is imposed on the wavelet coefficients. Then the image is estimated by applying a suitable Bayesian rule to the resulting posterior distribution of the wavelet coefficients. Different choices of loss function lead to different Bayesian rules and hence to different nonlinear wavelet shrinkage and wavelet thresholding rules. Such wavelet estimators have been discussed in several papers, for example (Achim et al., 2001; Simoncelli and Adelson, 1999; Abramovich et al., 1998; Chang et al., 2000a). Moreover, it has been shown that Bayesian wavelet estimators outperform the classical wavelet term-by-term thresholding estimators in terms of mean squared error (MSE) in finite sample situations.

A popular prior for each wavelet coefficient is a scale mixture of two normal distributions (Chipman et al., 1997) or one normal distribution and a point mass at zero (Abramovich et al., 1998). In (Vidakovic and Ruggeri, 2000), authors considered a double exponential prior with a point mass at zero to derive an adaptive multiresolution smoother. Huang et al. proposed two Bayesian approaches based on deterministic/stochastic decomposition (Huang and Cressie, 2000) and on nonparametric mixed-effects model (Huang and Lu, 2000). Since the work of (Mallat, 1989), the Generalized Gaussian Distribution (GGD) has been commonly used as a prior for the wavelet coefficients in the image processing community (see e.g. Chang et al., 2000a,b). Simoncelli (1999) also used a local mixture of Gaussians prior to derive the corresponding Bayesian shrinker. However, the GGD prior suffers from a lack of capturing the heavy tail behavior of the observed wavelet coefficients densities. Based upon this observation, authors in (Achim et al., 2001) used α-stable distributions (Nikias and Shao, 1995), a family of heavy tailed densities, as a prior to capture the sparseness of the wavelet coefficients at each scale. However, in both the GGD and the α-stable priors, the derived Bayesian estimator has no closed analytical form in general situation. This involves intensive numerical integration which is numerically unstable since the integration limits are infinite. Recently, in (Fadili and Boubchir, 2005), the Bessel K forms (BKF) family has been successfully proposed in wavelet-based Bayesian denoising. In that work, a closed-form expression of the L_2-loss Bayesian shrinkage rule associated with the BKF prior was proposed.

In our approach, we propose a prior statistical model based on the α-stable densities adapted to capture the sparseness of the wavelet detail coefficients.[1] An attempt to apply this model in the context of wavelet denoising have been already proposed in (Achim et al., 2001). These authors showed the superiority of the α-stable distributions in fitting the mode and the heavy tail behavior of the wavelet coefficients distributions. However, their hyperparameters estimator is very poor in the presence of contaminating noise and remains an important issue yielding very bad performance of their wavelet denoiser especially at low SNRs. Moreover, explicit forms of the probability density functions (PDF) are not known in general. Therefore, the Bayesian denoiser derived by Achim et al. (2001) suffered from other drawbacks such as numerical instability because of the lack of a closed-form expression of the Bayesian shrinkage rule, and the weakness of the estimator of the hyperparameters associated with the α-stable prior.

In general, with the α-stable prior (which is part of the scale mixture of Gaussians family (Andrews and Mallows, 1974)), the most preoccupying issue is not only that the α-stable distribution does not always possess a closed-form expression, but also that the Bayesian modeling and estimation associated to this prior is complicated. Clearly, even if the prior distribution is analytically known, the integrals involved in the Bayesian analysis and estimation (e.g. posterior distribution, posterior mode, posterior mean), are highly complex to manipulate. One can then circumvent these difficulties using either approximations (e.g. Laplace, finite mixtures), numerical or Monte-Carlo integration methods. Towards this goal, in the present paper, we propose to overcome these difficulties using the finite mixture of Gaussians as a fast and numerically stable analytical approximation for α-stable densities in order to obtain a closed-form expressions for our Bayesian denoiser. We show the effectiveness and the stability of this approximation. Additionally, we propose an approximate maximum likelihood estimator for the hyperparameters of our closed-form expressions. When applied to discrete wavelet transforms of real images, the approximate α-stable model demonstrates a high degree of match between observed and estimated prior densities. Exploiting this prior, we design a Bayesian L_2-loss nonlinear denoiser and we derive a closed-form for its expression.

This paper is organized as follows: In Section 2 we define the wavelet-based nonparametric regression problem. Some necessary preliminaries on the α-stable model are given in Section 3. The approximation by the Gaussian scale-mixture model is proposed in Section 4. Using this approximate α-stable prior, Section 5 is devoted to expose our

[1] Although the presentation here is focusing on the wavelet transform, our methodology is readily applicable to any basis in which the image is sparsely represented.

nonparametric Bayesian estimator. Section 6 compares the performance of the designed algorithm with previously published denoisers on a digitized database of images. Finally, conclusions and directions of future work are drawn.

2. Nonparametric regression with bases

Suppose we have noisy data at regularly sampled pixels:

$$y_{mn} = g_{mn} + \epsilon_{mn} \tag{1}$$

where ϵ_{mn} are iid normal random variables with mean zero and variance σ_ϵ^2 independent of g_{mn}. The goal is to recover the underlying function g from the observed noisy data y_{mn}, without assuming any particular parametric structure for g. Let $\boldsymbol{y}$, $\boldsymbol{g}$ and ϵ denote the matrix representation of the corresponding samples. Let $\boldsymbol{D} = \mathscr{W}\boldsymbol{y}$, $\boldsymbol{S} = \mathscr{W}\boldsymbol{g}$ and $\boldsymbol{V} = \mathscr{W}\epsilon$, where $\mathscr{W}$ is the two-dimensional dyadic orthonormal wavelet transform (DWT) operator (Mallat, 1999). In a two-dimensional setting, the subbands HH_j, HL_j and LH_j, $j = J_c, \ldots, J-1$ correspond to the detail coefficients in diagonal, horizontal and vertical orientations, and the subband LL_{J_c} is the approximation or the smooth component. J_c is the coarsest scale of the decomposition. Let s_{mn}^{oj} be the detail coefficient of the true image $\boldsymbol{g}$ at location (m,n), scale j and orientation o, and similarly for d_{mn}^{oj} and v_{mn}^{oj}. Due to the orthogonality of the basis, v_{mn}^{oj}, the DWT of white noise are also independent normal variables with the same variance. It follows from Eq. (1) that

$$\begin{aligned} c_{mn} &= a_{mn} + \epsilon_{mn} \\ d_{mn}^{oj} &= s_{mn}^{oj} + \epsilon_{mn}, \quad j = J_c, \ldots, J-1; \;\; m,n = 0, \ldots, 2^j - 1 \end{aligned} \tag{2}$$

where a_{mn} (resp. c_{mn}) is the approximation coefficient of the true image g (resp. y) at location (m,n).

The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few large detail coefficients in $\boldsymbol{D}$ contain information about the underlying image $\boldsymbol{g}$, while small values can be attributed to the noise which uniformly contaminates all wavelet coefficients. It is also advisable to keep the approximation coefficients intact because they represent low-frequency terms that usually contain important features about the image $\boldsymbol{g}$. A wide class of nonparametric wavelet estimators can be written as minimizers of the penalized least-squares problem:

$$\chi^2(s_{mn}^{oj}) = \frac{1}{2\sigma^2}\|\boldsymbol{D} - \boldsymbol{S}\|_2^2 + \sum_{o,j,m,n} p_\lambda(|s_{mn}^{oj}|) \tag{3}$$

The regularizing penalty function $p_\lambda(|\cdot|)$ is nondecreasing, nonnegative and not necessarily convex on $\mathbb{R}^+$, and irregular at point zero to produce sparse solutions (Antoniadis and Fan, 1999). It depends on a hyperparameter vector λ. The minimization of this high-dimensional problem reduces to component-wise minimization problem where a unique solution can be easily found under suitable conditions (Antoniadis and Fan, 1999). Antoniadis and Fan (1999) showed that most of classical estimators reviewed in Section 1 have objective criteria which are special cases of Eq. (3) corresponding to specific choices of $p(|\cdot|)$, e.g. soft shrinkage corresponds to the L_1-norm penalty.

In the Bayesian approach a prior distribution is imposed on the wavelet coefficients in order to capture the sparseness of the wavelet expansion. The following section is intended to provide a brief introduction to α-stable distributions family suitable to characterize the wavelet subband coefficients densities which have been already observed to be sharply peaked and heavily tailed.

3. α-Stable distribution

Definition 1 (*Characteristic function*). Let X a random variable (RV). X is said α-stable ($X \sim S_\alpha(\beta,\mu,\gamma)$) if there exist $0 < \alpha \leqslant 2$, $\gamma \geqslant 0$, $-1 \leqslant \beta \leqslant 1$ and $\mu \in \mathbb{R}$ such that the associated characteristic function is given by (Zolotarev formulation):

$$\psi_X(t) = \exp(\mathrm{i}\mu t - \gamma^\alpha |t|^\alpha (1 - \mathrm{i}\beta \operatorname{sign}(t) W(\alpha,t))), \quad t \in \mathbb{R} \tag{4}$$

where

$$W(\alpha,t) = \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1 \\ -\dfrac{2}{\pi}\log|t| & \text{if } \alpha = 1 \end{cases}$$

Therefore, the four parameters α, β, μ and γ uniquely and completely define the α-stable distribution. α is the characteristic exponent. It controls the heaviness of the tails of the PDF. β is the symmetry index. It determines the skewness of the distribution. When $\beta = 0$, X is symmetric α-stable (SαS) RV. μ is the location parameter. If $1 < \alpha \leqslant 2$, this parameter is equal to the mean and the median when $0 < \alpha \leqslant 1$. The scale parameter $\gamma = \sigma^\alpha$. It is a measure of the spread of the samples from a distribution around its mean. If $\alpha = 2$ (Gaussian distribution), the standard deviation of the distribution is equal to $\sqrt{2}\sigma$. It should be also noted that an α-stable RV have infinite variance for $\alpha < 2$ and that $E(|X|^q) < \infty$ if $q \leqslant \alpha$. The latter remark has a fundamental consequence on the integrals involved in the Bayesian shrinkage rules to be properly defined (Mathieu, 2002).

The wavelet detail coefficients densities have been already observed to be symmetric sharply peaked and heavily tailed (Mallat, 1989; Simoncelli and Adelson, 1999; Chang et al., 2000a,b; Achim et al., 2001; Fadili and Boubchir, 2005; Portilla et al., 2003). This is exactly the property which is captured by an SαS distribution where $\beta = 0$ and $\mu = 0$, i.e. $S_{m,n}^{oj} \sim S_\alpha(0,0,\gamma)$.

4. Analytical approximation of SαS PDFs

The PDF of an α-stable RV exists and is continuous. But there is no explicit expression for this PDF except for a few special cases. By taking the inverse Fourier transform of the characteristic function $\psi_X(t)$ given in Eq. (4),

one can obtain an integral representation of the PDF where the integration limits are infinite (Nolan, 1997). This integral can be evaluated analytically only for $\alpha = 2$ in which case the distribution is Gaussian, for $\alpha = 1$ in which case the distribution is Cauchy and finally for $\alpha = \frac{1}{2}$ in which case the distribution is Lévy (Pearson). In (Nolan, 1997), the author proved the existence of an equivalent exact integral representation where the limits of integration are finite. This exact approach is numerically stable but is very time-consuming. Moreover, it cannot offer an explicit expression for the Bayesian denoiser. Nonetheless, we will use this exact method throughout this paper as a reference when assessing the accuracy of our α-stable PDF approximation. Hereafter, we present a fast and numerically stable method, based on the scale mixture of Gaussians to obtain an approximate analytical expression of the α-stable PDF with arbitrary parameters.

4.1. The scale mixture of Gaussians

The concept of mixture is based on a corollary of the mixture theorem of α-stable RVs. This result states that any SαS RV can be represented as the product of a Gaussian RV and a positive α-stable RV (Samorodnitsky and Taqqu, 1994; Andrews and Mallows, 1974):

Proposition 1 (Scale mixture of Gaussians). *Let* $X \sim \mathcal{N}(0, 2\gamma_x)$, *that is, let X be distributed with Gaussian distribution* ($\alpha_X = 2$). *Let Y be a positive stable random variable,* $Y \sim S_{\frac{\alpha_z}{2}}(-1, 0, (\cos(\frac{\pi\alpha_z}{4}))^{\frac{2}{\alpha_z}})$ *and independent form X. Then,*

$$Z = Y^{\frac{1}{2}} X \sim S_{\alpha_z}(0, 0, \gamma_x) \tag{5}$$

If we define $V = Y^{\frac{1}{2}} \geqslant 0$ and $f_V(v) = h(v)$ then the PDF of Z is deduced by using the marginalization property of probabilities:

$$\begin{aligned} f_Z(z) &= \int_0^{+\infty} f_{Z|V}(z|v) f_V(v)\,\mathrm{d}v \\ &= \frac{1}{\sqrt{4\pi\gamma_x}} \int_0^{+\infty} \exp\left(-\frac{z^2}{4\gamma_x v^2}\right) h(v) v^{-1}\,\mathrm{d}v \end{aligned} \tag{6}$$

Sampling $f_Z(z)$ at N points allows to obtain a finite mixture approximation to any SαS PDF:

$$p_{\alpha,0,\mu,\gamma}(z) \approx \frac{\sum_{j=1}^{N} v_j^{-1} \exp\left(-\frac{(z-\mu)^2}{4\gamma_x v_j^2}\right) h(v_j)}{\sqrt{4\pi\gamma_x} \sum_{j=1}^{N} h(v_j)} \tag{7}$$

It should be noted that this analytical expression for the SαS PDF is only an *approximation*, since the continuous integral was approximated by a *finite sum*. Therefore, Eq. (6) should be sampled at a large number of points for the approximation to be excellent. However, to reduce the complexity of the model in Eq. (7) and get fast but good enough approximation, one might prefer to use only a small number of components and to sample Eq. (6) at a few points only. In this case the approximation is coarse and we suggest using the "Expectation–Maximization" (EM) algorithm to fine-tune the components to obtain a better approximation. For instance, as we will show in Section 4.3, only a few mixture components (typically 4–8) are necessary to negotiate a good compromise between the approximation quality, the model complexity and the calculation time.

4.2. The SαS PDF approximation algorithm

Here we describe an algorithm that fits a SαS PDF to observed samples $\{z_m\}_{m=1,\ldots,M}$ using the scale mixture of Gaussians approximation to SαS PDFs. This algorithm follows the next steps:

- *Step 1:* Given the parameters $(\alpha, \beta = 0, \mu = 0, \gamma)$ of the desired SαS, generate the characteristic function (Eq. (4)) of the mixing PDF which is positive stable distributed with parameters: $(\frac{\alpha}{2}, \beta = -1, \mu = 0, \gamma = (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}})$.
- *Step 2:* Evaluate the positive stable PDF f_Y at N equally spaced points taking the inverse FFT of the characteristic function given in Eq. (4), where N is the pre-specified number of components in the mixture.
- *Step 3:* The mixing function is the PDF of the random variable $V = Y^{\frac{1}{2}}$, which is obtained by
$$h(v_i) = 2 v_i f_Y(v_i^2) \tag{8}$$
- *Step 4:* Substitute the mixing function samples calculated in the third step in Eq. (7) allows to obtain the analytical approximation for the SαS PDF:
$$p_{\alpha,0,0,\gamma}(z_m) = \frac{\sum_{j=1}^{N} \phi(z_m; 0, 2\gamma v_j^2) v_j f_Y(v_j^2)}{\sum_{j=1}^{N} v_j f_Y(v_j^2)} \tag{9}$$
where $\phi(z; \mu, \delta^2)$ is the Gaussian PDF with mean μ and variance δ^2.
- *Step 5:* Use the EM algorithm (McLachlan and Peel, 2000) to refine the approximation using the observed samples z_m. In the case of the mixture of Gaussians model we seek to get Maximum Likelihood (ML) estimates such that
$$p_{\alpha,0,0,\gamma}(z_m) = \sum_{j=1}^{N} P(z_m|j) P_j \tag{10}$$
where P_j (mixing proportions) initial values of $\frac{h(v_j)}{\sum_{j=0}^{N} h(v_j)}$ and $P(z_m|j) = \phi(z_m; 0, \sigma_j^2)$ with $\sigma_j^2 = 2\gamma v_j^2$.

4.3. Assessing the approximation quality

Here, we first assess the visual quality of the approximation as the SαS distribution gets far from Gaussian (heavier tails). Only the influence of α is considered because it is the parameter that characterizes the shape of the PDF (heaviness of the tails). It is also obvious that γ (or equivalently σ_x) being the scale parameter, it has no impact on the choice of N and therefore, its influence is not relevant in

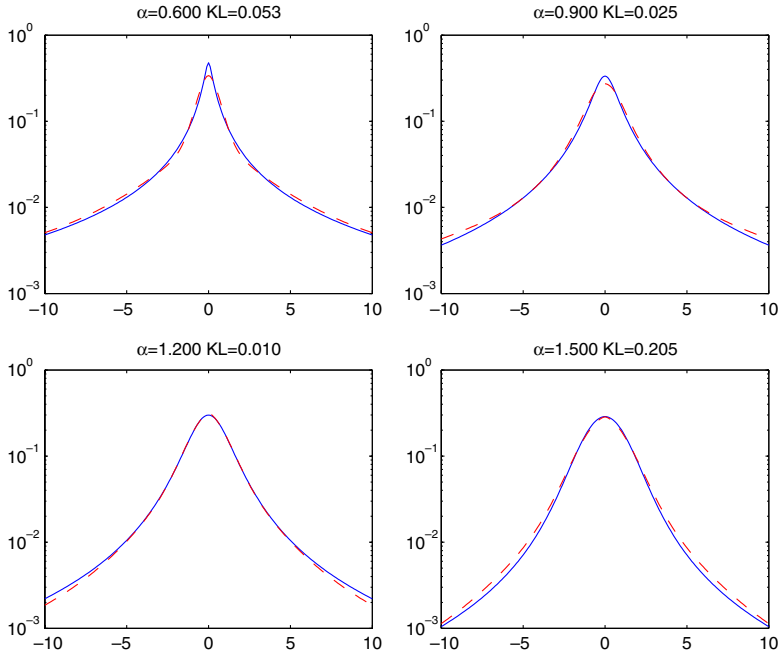

Fig. 1. Comparison on log–log scale between SαS PDFs obtained using the exact integral representation of (Nolan, 1997) (—), and the approximate PDF with a mixture of eight Gaussians (- - -) for various values of the parameter α. The KL divergence between the exact and the approximate PDFs is reported at the top of each plot.

our discussion here. The results depicted in Fig. 1 show that the Gaussian scale-mixture PDF, with a mixture of eight Gaussians, is very close to the exact PDF whatever the value of the tail-heaviness exponent α. The visual quality is also confirmed by small values of the Kullback–Leibler (KL) divergence calculated between the two PDFs.

As we mentioned before, for the approximation to be sufficiently accurate and numerically stable and fast, one has to define properly the number of components in the mixture N. Model complexity measures, e.g. the minimum description length (MDL) or some popular information criteria such as the BIC or the AIC, are commonly adopted to objectively choosing the number of components in a mixture model (Figueiredo et al., 1999; McLachlan and Peel, 2000). These measures attempt to provide a data-driven estimate of N.

The estimation of the optimal number of Gaussians N is defined as the one minimizing the cost function C_{MDL} (Figueiredo et al., 1999). We can show that its expression is given by

$$C_{\mathrm{MDL}}(N) = -\sum_{m=1}^{M} \log \sum_{j=1}^{N} P(z_m|j)P_j + \frac{2N-1}{2}\log(M) \tag{11}$$

where M is the number of available samples.

Fig. 2(a) shows the evolution of the MDL criterion as a function of N for different values of the characteristic exponent α. Except the Gaussian case ($\alpha = 2$) where only one component is necessary, results in Fig. 2(a) clearly demonstrate that the optimal number of Gaussians N is located in the interval $[4, 8]$. This objective choice ensures a trade-off between the approximation quality (in likelihood sense) and the model complexity. This was also confirmed by the KL divergence calculated between the exact and the approximate PDFs. As illustrated in Fig. 2(b), the KL divergence decreases very rapidly and converges to 0 for $N \geqslant 8$ for all values of α except for the Gaussian case where $N = 1$.

From these results, we can legitimately conclude that the value $N = 8$ is sufficient to get a very accurate approximation while keeping the calculation time reasonable.

5. Bayesian denoiser

5.1. Marginal PDF of the observed wavelet coefficients

As stated above, in the Bayesian approach, a prior is imposed on the wavelet coefficients designed to describe their distribution. It is also assumed in the prior model that the wavelet coefficients s_{mn}^{oj} of the true image at each scale and orientation are mutually independent RVs and independent of the noise process ϵ_{mn}^{oj}. Throughout the rest of the paper and for readability, the superscript oj and subscript mn (orientation o at scale j and location (m,n)) will be dropped out. The detail coefficients s at each scale and each orientation are SαS distributed

$$s \sim S_\alpha(0, 0, \gamma = \sigma^\alpha) \tag{12}$$

and the probabilistic model associated with d conditionally on s is Gaussian

$$d|s \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{13}$$

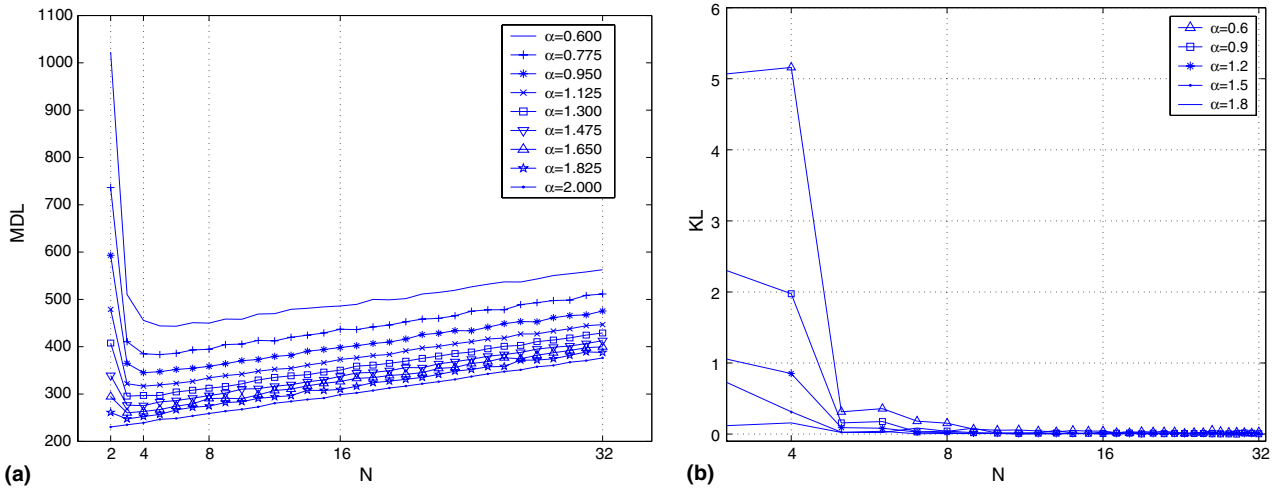

Fig. 2. The MDL criterion (a) and the KL divergence (b) as a function of the number of Gaussians N for different values of the characteristic exponent α.

Using the Bayes rule, the marginal PDF of d can be written

$$p(d|\theta_1,\theta_2) = \int_{-\infty}^{+\infty} p(d|s,\theta_2)p(s|\theta_1)\,\mathrm{d}s = \int_{-\infty}^{+\infty} \phi(d-s;\theta_2)p(s|\theta_1)\,\mathrm{d}s \quad (14)$$

$p(s|\theta_1)$ is the approximate SαS PDF with the hyperparameters set $\theta_1 = \{P(j),\sigma_j\}$, and $\phi(d;\theta_2)$ is the normal noise PDF with variance $\theta_2 = \sigma_\epsilon^2$. The analytical approximation of the marginal PDF of d is given by

$$p(d|\theta_1,\theta_2) = \frac{1}{\sqrt{2\pi}}\sum_j P_j(\sigma_j^2+\sigma_\epsilon^2)^{-\frac{1}{2}}\exp\left(-\frac{d^2}{2(\sigma_j^2+\sigma_\epsilon^2)}\right) \quad (15)$$

5.2. The hyperparameters estimation

In the image denoising context, one must elicit the hyperparameters (θ) estimation problem, which in turn will lead to a data-driven denoising procedure that is adaptive to each subband. To implement the formula in Eq. (15), one has to estimate $\theta = \{P_j,\sigma_j,\sigma_\epsilon\}$, which amounts to estimating $\{\alpha,\gamma=\sigma^\alpha,\sigma_\epsilon\}$. However, the hyperparameters estimation step is a difficult task for SαS RVs especially in the presence of contaminating noise. In addition, this step is very crucial and must be treated very carefully as it will irrevocably influence the performance of the denoising algorithm. As noted by authors in (Fadili and Boubchir, 2005), the weakness of the hyperparameters estimator was a major reason for the poor performance shown by the version of α-stable prior implemented by Achim et al. (2001).

In our scale mixture of Gaussians approximation, the estimation of P_j and σ_j parameter amounts to first estimating the original parameters α and σ, and then fine-tuning using the EM algorithm. However, this task becomes more complicated in the presence of contaminating noise.

In our model, the estimation of α and γ is only useful for initialization, and final estimates are given by the EM step in Algorithm 4.2. Therefore, we have chosen the quantile-based estimator of (McCulloch, 1986) assuming that for reasonable SNRs, the tails of the marginal distribution $p(d|\theta_1,\theta_2)$ are not very sensitive to the presence of noise. The McCulloch method provides an initial and fast estimation for the parameters α and γ under the restriction $0.6 \leqslant \alpha \leqslant 2$. This is not a limitation since our images are supposed at least integrable in L_1 sense, involving that $\alpha \geqslant 1$. As far as the level noise σ_ϵ is concerned, it is estimated from the HH orientation at the finer scale using the popular robust estimator (Donoho and Johnstone, 1994):

$$\hat{\sigma}_\epsilon = \frac{\mathrm{MAD}(d_{mn}^{HH_1})}{0.6745} \quad (16)$$

where MAD is the median absolute deviation. Then, using these initial values of α, σ and σ_ϵ, we apply the algorithm described in Section 4.2 to fine-tune the hyperparameters and get an accurate and stable approximation to the marginal PDF of the observed wavelet coefficients. This can be summarized as follows:

(1) Use the quantile method (McCulloch, 1986) to get $\hat{\alpha}$ and $\hat{\gamma}$, and the MAD estimator to get $\hat{\sigma}_\epsilon$ (Donoho and Johnstone, 1994).
(2) Apply steps 1–4 of the algorithm in Section 4.2 using $\hat{\alpha}$ and $\hat{\gamma}$, and get $\hat{\sigma}_j^2$ and $\widehat{P}_j$.
(3) Define $\hat{\sigma}'_j = \sqrt{\hat{\sigma}_j^2+\hat{\sigma}_\epsilon^2}$.
(4) Use the EM algorithm to fine-tune $\hat{\sigma}'_j$ and $\widehat{P}_j$ according to the mixture model (Eq. (15)).

We now illustrate some prior estimation results for a variety of images taken form a digitized database (http://sipi.usc.edu/services/database/database.html). Depicted in the first column of Fig. 3 are some images (Lena, Barbara and textured image Roof). Shown are the estimated and the

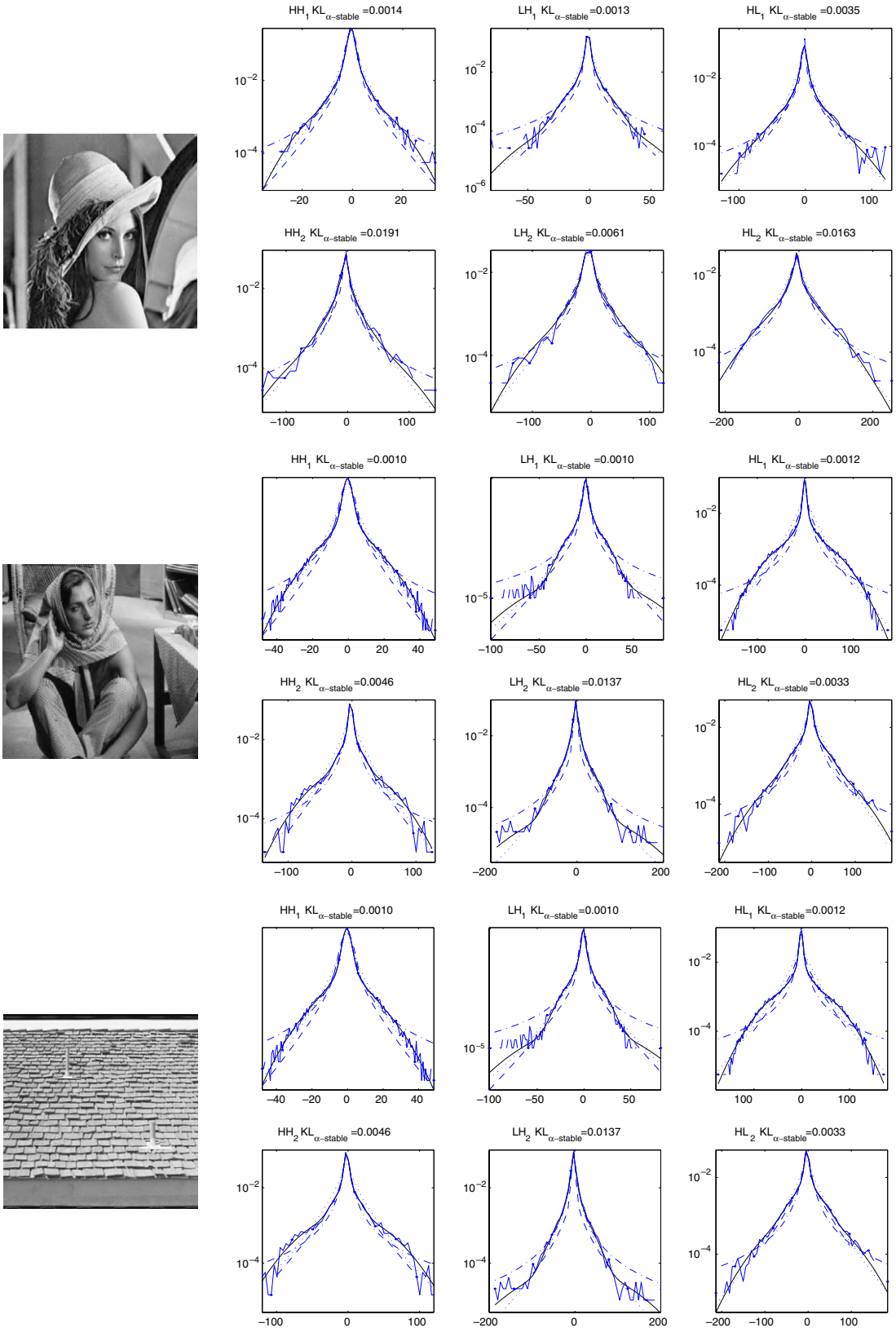

Fig. 3. Estimated and observed marginal densities of the observed wavelet detail coefficients of two classical images (Lena, Barbara and Roof textured image). The observed histogram (-●-) was fitted using the scale-mixture α-stable (—), original α-stable (–·–), BKF (- - -) and GGD (· · ·) models. Only two detail levels are shown for each image (first and second row each time). The three columns correspond to the *HH*, *HL* and *LH* orientations.

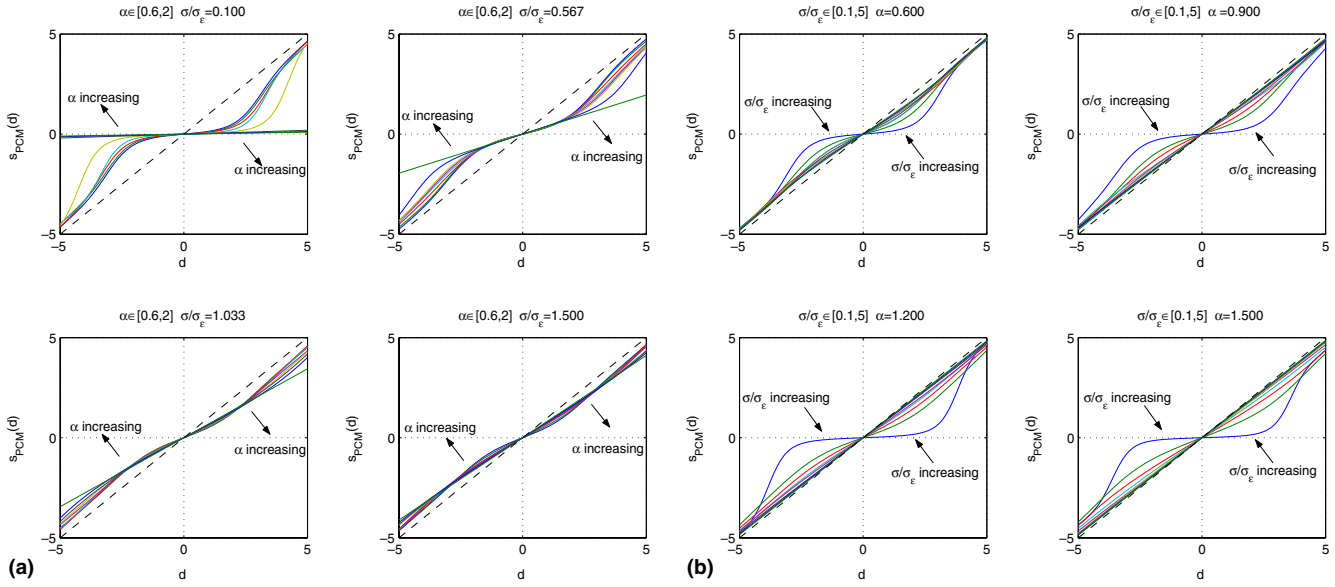

Fig. 4. Bayesian rule input–output curves ($s_{\mathrm{PCM}}(d)$ as a function of d). (a) Influence of the ratio $\frac{\sigma}{\sigma_\epsilon}$ on the Bayesian estimator curves for constant $\alpha \in [0.6, 2]$. (b) Influence of parameter α on the Bayesian estimator curves for constant $\frac{\sigma}{\sigma_\epsilon} \in [0.1, 5]$.

Fig. 5. Visual comparison of various denoising methods on test image Lena. This image is corrupted by Gaussian noise with an input $\mathrm{SNR}_{\mathrm{in}} = 15$ dB.

observed densities of the wavelet detail coefficients of each image on log scale. The observed histogram (-●-) was fitted using the scale-mixture α-stable algorithm with eight Gaussians as we described above (solid). For comparison purposes, we also depict the fit given by the original α-stable as proposed by Achim et al. (2001) (dash-dotted), the BKF (dashed) and the GGD (dotted) models. Only two detail levels are shown for each image (first and second row each time). The three columns correspond to the *HH*, *HL* and *LH* orientations. The original version of the α-stable prior tends to exaggerate the tails of the distribution in some cases (e.g. for textured images). From these results, we can legitimately claim that the Gaussian scale-mixture SαS density fits the observed wavelet detail coefficients very well. It generally outperforms the GGD and the BKF models. Furthermore, a closed-form expression of the L_2-loss Bayesian shrinkage rule associated with our prior can be readily obtained as we will show in the next section.

5.3. Bayesian term-by-term denoising

Different choices of loss function lead to different Bayesian rules and hence to different nonlinear wavelet shrinkage and wavelet thresholding rules. For example, it is well known that the L_1-loss function corresponds to the maximum a posteriori (MAP) estimator. However, except some special cases of SαS distributions (e.g. $\alpha = 2$), it is not easy to derive a general analytical form of the corresponding Bayesian shrinkage rule even with the scale-mixture approximation. Alternatively, we use the L_2-based Bayes rules which correspond to posterior conditional means (PCM) estimates. We here derive a general expression, using the approximate prior PDF, of the PCM estimates of the wavelet coefficients s (conditionally on the hyperparameters):

$$s_{\mathrm{PCM}}(d|\theta) = \frac{\sum_j P_j \frac{d\sigma_j^2}{\sigma_j^2+\sigma_\epsilon^2}\phi(d;\sigma_j^2+\sigma_\epsilon^2)}{\sum_j P_j \phi(d;\sigma_j^2+\sigma_\epsilon^2)} \tag{17}$$

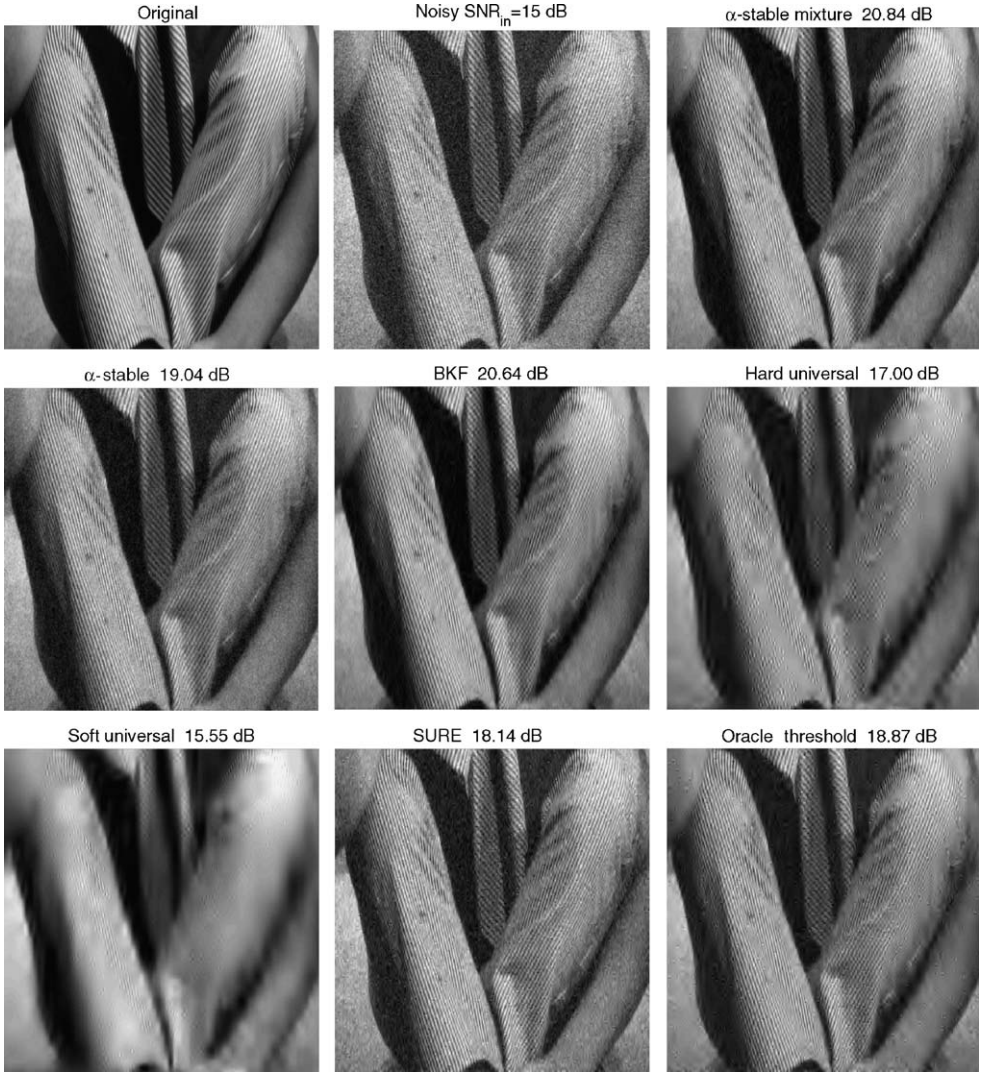

Fig. 6. Visual comparison of various denoising methods on test image Barbara. This image is corrupted by Gaussian noise with an input $\mathrm{SNR}_{\mathrm{in}} = 15$ dB. The Barbara image was zoomed on a textured area of the trousers.

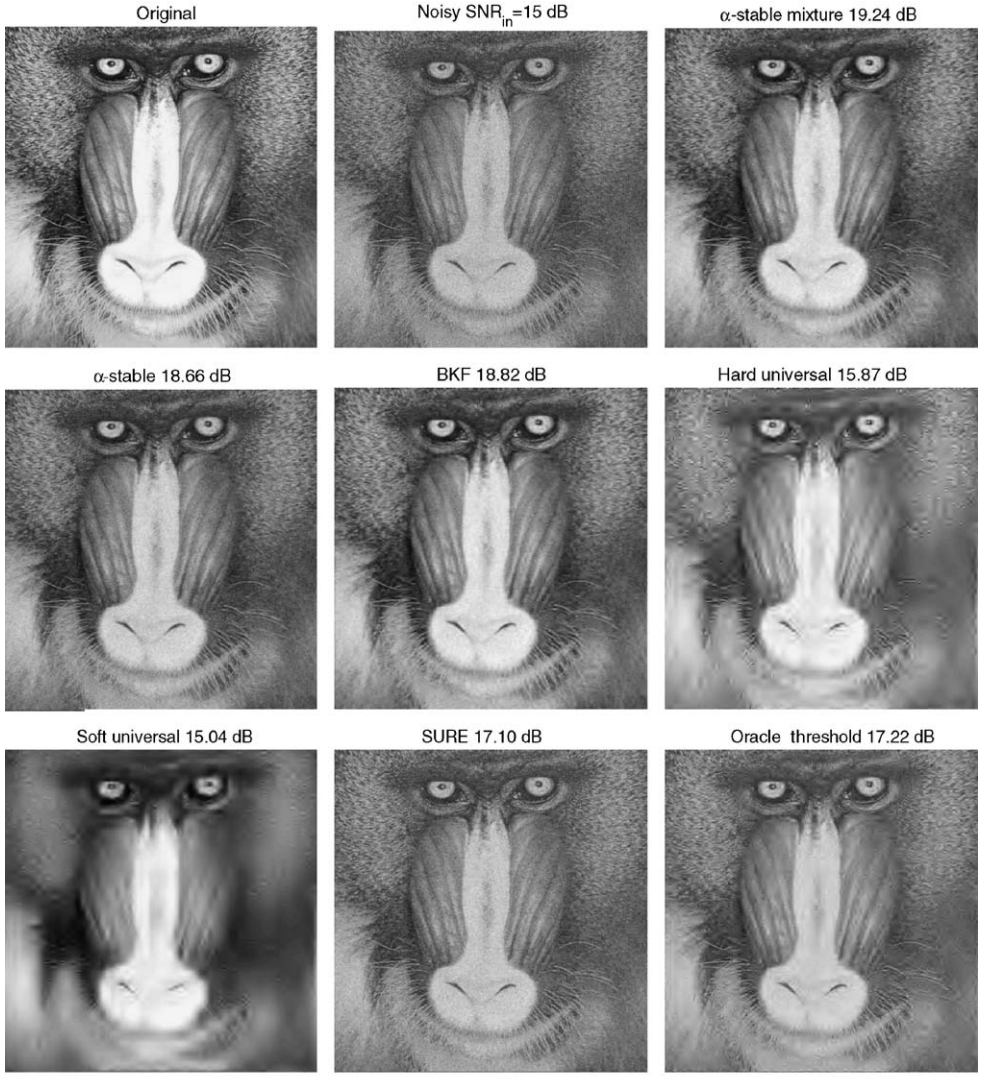

Fig. 7. Visual comparison of various denoising methods on test image Mandrill. This image is corrupted by Gaussian noise with an input $\mathrm{SNR}_{\mathrm{in}} = 15$ dB.

This equation shows that the SαS PCM estimate can be seen as a weighted average of Gaussian PCM estimates, where the weights are given by the mixing proportions and the Gaussian PDFs.

Fig. 4 depicts the Bayesian rule input–output curves obtained using Eq. (17). The plots in (a) (resp. (b)) show the influence of the ratio $\frac{\sigma}{\sigma_\epsilon}$ (resp. characteristic exponent α) on the Bayesian estimator curves for constant exponent $\alpha \in [0.6, 2]$ (resp. ratio $\frac{\sigma}{\sigma_\epsilon}$). The ratio $\frac{\sigma}{\sigma_\epsilon}$ can be seen as a measure of SNR.[2] As demonstrated by these plots, the proposed Bayesian rule shrinks small observed wavelet coefficients heavily and large ones only slightly, approaching the identity line when $|d|$ is very large. As far as the influence of $\frac{\sigma}{\sigma_\epsilon}$ is concerned, the amount of shrinkage decreases as $\frac{\sigma}{\sigma_\epsilon}$ increases. This can be intuitively understood from the fact that the contribution of the true signal becomes salient as the ratio $\frac{\sigma}{\sigma_\epsilon}$ increases, yielding less shrinkage amount. The amount of shrinkage also decreases as α decreases. The explanation of this behavior is that as α decreases, the heavier the tails and the higher the probability that smaller values are due to the true image g. These observations are consistent with others' work (Achim et al., 2001; Fadili and Boubchir, 2005).

6. Experimental results

We now assess the performance of our Bayesian denoiser with the scale-mixture approximation to the α-stable prior, called "α-stable mixture", and we compare it to other previously published denoising methods. For the comparison to be fair, we only chose denoising methods using the same transforms, namely, the DWT. Extension to overcomplete representations which are translation and rotation invariant are the subject of our ongoing research. Six other denoising algorithms are considered:

[2] We here note that this ratio is not rigorously a SNR. The α-stable prior assumes that the qth order moment is finite only for $q \leqslant \alpha < 2$. Thus, strictly speaking, the class of images considered do not have finite variance.

the universal threshold Hard and Soft thresholding (Donoho and Johnstone, 1994), the Stein Unbiased Risk Estimator (SURE) (Donoho and Johnstone, 1995), the Oracle threshold estimator (Oracle), the Bessel K forms (BKF) Bayesian denoiser (Fadili and Boubchir, 2005) and the original version of the α-stable Bayesian denoiser (Achim et al., 2001). In the latter, no closed-form is available for the PCM Bayesian denoiser. We here used an equivalent form involving Fourier integrals as proven in (Mathieu, 2002). The numerically Fourier integrals were implemented using FFT-based methods.

Beside visual quality, we also calculated the signal-to-MSE ratio, commonly called the SNR, in order to quantify the achieved performance improvement. this is defined in decibels as follows:

$$\mathrm{SNR}\ (\mathrm{dB}) = 10\log_{10}\frac{\|g\|_2^2}{\|\hat{g}-g\|_2^2} \tag{18}$$

where the denominator is the estimation risk between the true and the denoised images.

The overall performance was quantified on a digitized database of 100 test images (http://sipi.usc.edu/services/database/database.html). The DWT employs Daubechies compactly supported wavelet with regularity 4. The coarsest level of decomposition was chosen to be $\log_2 \log N + 1$, where N is the size of the image.

Fig. 5 shows the estimated images for each denoising methods for the Lena image with an input $\mathrm{SNR}_{\mathrm{in}} = 15$ dB. One can clearly see that the visual quality of the "α-stable mixture" Bayesian denoiser is superior to the other methods but remains comparable to the BKF Bayesian denoiser. This general behavior is also observed on Barbara and Mandrill test images (Figs. 6 and 7). The zoom on a textured area of Barbara proves that our denoiser ensures a good compromise between the noise rejection and the conservation of fine details in the image (e.g. the stripes of the trousers). Owing to its hyperparameter estimation method, our denoiser overcomes the limitations of the original "exact" α-stable Bayesian denoiser as used in (Mathieu, 2002; Achim et al., 2001).

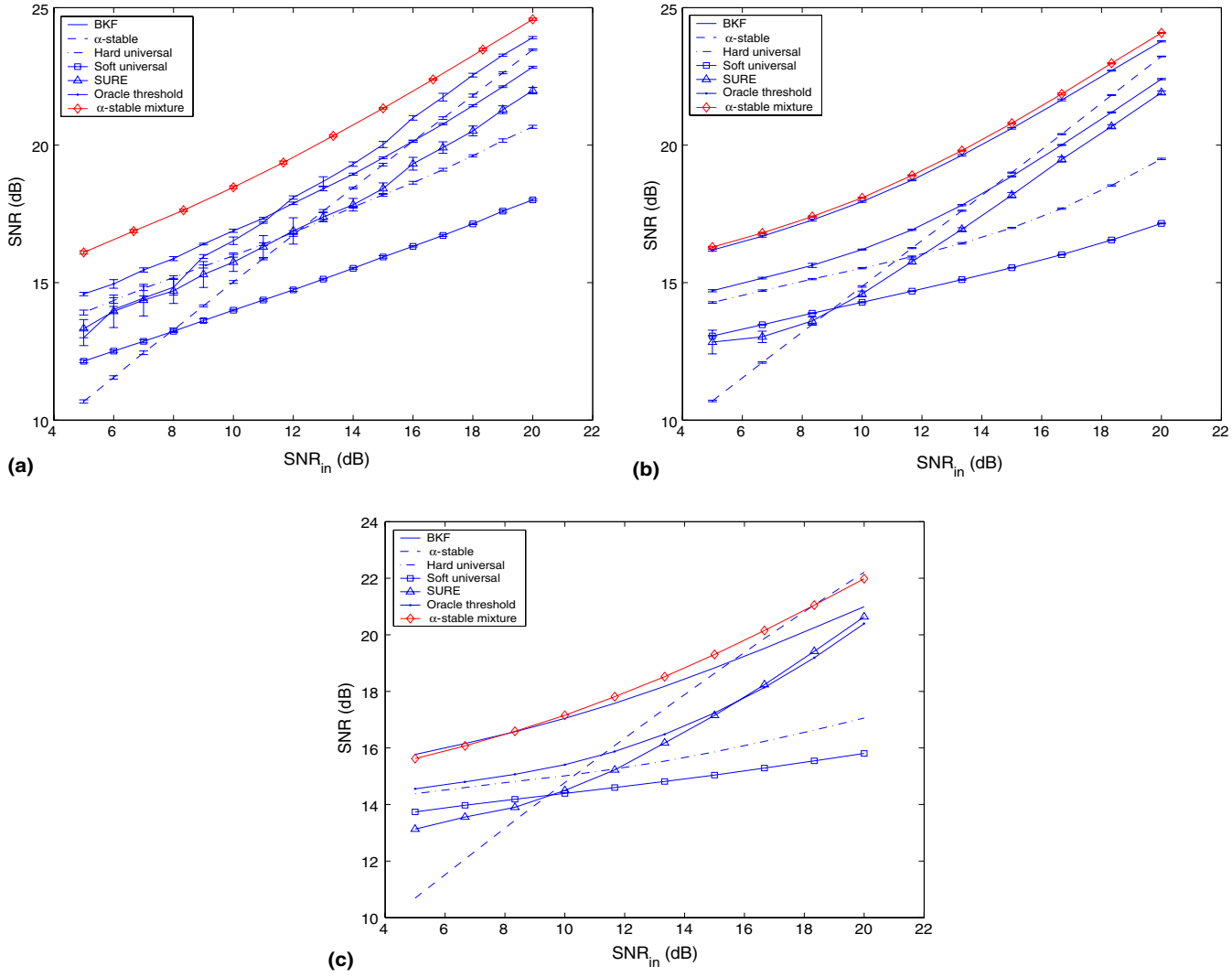

Fig. 8. For each $\mathrm{SNR}_{\mathrm{in}}$, mean and standard deviation (over 50 runs) of SNR given by various methods for (a) Lena, (b) Barbara and (c) Mandrill images.

Table 1
Average SNR over 50 runs and 100 image database for each denoising method, as a function of SNR_{in}

Method	SNR_{in}			
	5	10	15	20
α-stable mixture	17.23	19.54	22.17	25.03
BKF	17.15	19.29	21.8	24.52
α-stable	10.75	15.1	19.4	23.56
Hard universal	15.31	17.03	19.1	21.37
Soft universal	14.08	15.45	17.1	18.96
SURE	14.27	16.64	19.34	22.57
Oracle	15.7	17.87	20.43	23.36

Furthermore, our denoiser is faster and very stable numerically.

To confirm this first experiment, we also carried out a simulation study where we calculated the mean and the standard deviation of the SNR (over 50 runs) for each denoising method on the Lena image. Results are shown in Fig. 8. The SNR_{in} was in the range $[5, 20]$ dB.

In Fig. 8, one can notice that the "α-stable mixture" denoiser outperforms most of the methods, but is still comparable to the BKF approach. It compares favorably with the oracle thresholding but is much better that the SURE especially at low SNRs. The original version of the α-stable denoiser is underperforming at low input SNR_{in}. The main reason is the weakness of the hyperparameters estimator which remains an important issue. In Table 1, we have reported the average SNR over the 50 runs and the whole database (100 images) for each denoising method, as a function of SNR_{in}. The general behavior described before is confirmed by this table. This suggests that the scale-mixture approximation to the "α-stable" prior is an accurate model adapted to capture the sparseness behavior of the wavelet coefficients for a large class of images.

7. Discussion and conclusion

In this paper, a nonlinear nonparametric Bayesian estimator in the orthogonal wavelet domain was presented. An approximation to SαS based on scale mixture of Gaussians was proposed. This approximation has proven accurate and very stable numerically. The EM algorithm was used to refine a first estimation step which serves as a good starting point for the EM algorithm. The number of mixture components problem was solved objectively using the MDL criterion, and we concluded that Gaussians are enough to provide a fast and accurate approximation.

Using this approximate analytical expression for the prior, we also derived the expressions of the posterior marginal distribution as well as the PCM estimator. Experimental results on a large database have shown the superiority of our Bayesian denoiser compared to other denoising approaches.

Despite his good performance, there are still some aspects in this method that have to be investigated and improved. First, we can point out the lack of translation and rotation invariance which yields a ringing effect, although this effect is somewhat negligible in the Bayesian denoised images. This problem can be solved either by cycle spinning or using translation invariant transforms such as the nonorthogonal undecimated (redundant) wavelet transform. The former is simple to implement but is too time-consuming. In the latter, the transform is highly redundant and the noise becomes correlated. Furthermore, the intra- and inter-scale independence property used in our context is no longer valid. Another important issue toward the goal of refining our method is to take into account the geometrical information in images. This can be done using specific geometry-adapted transforms such as Curvelets (Starck et al., 2002). Our investigations are now focusing on these aspects.

References

Abramovich, F., Benjamini, Y., 1996. Adaptive thresholding of wavelet coefficients. Comput. Statist. Data Anal. 22, 351–361.

Abramovich, F., Sapatinas, T., Silverman, B., 1998. Wavelet thresholding via a Bayesian approach. J. Roy. Statist. Soc. B 60, 725–749.

Achim, A., Bezerianos, A., Tsakalides, P., 2001. Novel Bayesian multiscale method for speckle removal in medical ultrasound images. IEEE Trans. Med. Imag. 20, 772–783.

Andrews, D.F., Mallows, C.L., 1974. Scale mixture of normal distributions. J. Roy. Statist. Soc. Ser. B 36 (1), 99–102.

Antoniadis, A., Fan, J., 1999. Regularization of wavelets approximations. J. Amer. Statist. Assoc. 96 (455), 939–963.

Chang, S., Yu, B., Vetterli, M., 2000a. Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. Image Process. 9 (9), 1522–1531.

Chang, S., Yu, B., Vetterli, M., 2000b. Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. Image Process. 9 (9), 1532–1546.

Chipman, H., Kolaczyk, E., McCulloch, R., 1997. Adaptive Bayesian wavelet shrinkage. J. Amer. Statist. Assoc. 92, 1413–1421.

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81 (3), 425–455.

Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. 90 (432), 1200–1224.

Efromovich, S., 2000. Sharp linear and block shrinkage wavelet estimation. Statist. Probab. Lett. 49, 323–329.

Fadili, J., Boubchir, L., 2005. Analytical form for a Bayesian wavelet estimator of images using the Bassel k forms densities. IEEE Trans. Image Process. 14 (2), 231–240.

Figueiredo, M.A.T., Leitao, J.M.N., Jain, A.K., 1999. On fitting mixture models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 54–69.

Hall, P., Penev, S., Kerkyacharian, G., Picard, D., 1997. Numerical performance of block thresholded wavelet estimators. Statist. Comput. 7, 115–124.

Huang, S., Cressie, N., 2000. Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. Technometrics 42, 262–276.

Huang, S., Lu, H., 2000. Bayesian wavelet shrinkage for nonparametric mixed effects models. Statist. Sin. 10, 1021–1040.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. PAMI 11 (7), 674–693.

Mallat, S.G., 1999. A Wavelet Tour of Signal Processing, second ed. Academic Press, San Diego.

Mathieu, J., 2002. Transformée en ondelettes et régression non-paramétrique dans un contexte bayesien. Master's thesis. Ecole Nationale Supérieure d'Ingénieurs, Caen, France.

McCulloch, J.H., 1986. Simple consistent estimators of stable distribution parameters. Commun. Statist.-Simul. 15 (4), 1109–1136.

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. Wiley, New York.

Nason, G.P., 1996. Wavelet shrinkage by cross-validation. J. Roy. Statist. Soc. B 58, 463–479.

Nikias, C.L., Shao, M., 1995. Signal Processing with Alpha-Stable Distributions and Applications. Wiley-Interscience.

Nolan, J.P., 1997. Numerical calculation of stable densities and distribution functions. Statist.-Stochast. Models 13, 759–774.

Ogden, R.T., Parzen, E., 1996. Data dependent wavelet thresholding in nonparametric regression with change-point applications. Comput. Statist. Data Anal. 22, 53–70.

Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P., 2003. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process. 12 (11), 1338–1351.

Samorodnitsky, G., Taqqu, M.S., 1994. Stable Non-Gausian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall, New York.

Simoncelli, E.P., 1999. Bayesian denoising of visual images in the wavelet domain. In: Muller, P., Vidakovic, B. (Eds.), Bayesian Inference in Wavelet Based Models. Springer-Verlag, New York, pp. 291–308.

Simoncelli, E.P., Adelson, E.H., 1999. Noise removal via Bayesian wavelet coring In: Third Internat. Conf. on Image Processing, vol. 1. IEEE Signal Processing Society, Lausanne, pp. 379–382.

Starck, J.L., Candès, E.J., Donoho, D.L., 2002. The curvelet transform for image denoising. IEEE Trans. Image Process. 11 (6), 670–684.

Vidakovic, B., 1999. Statistical Modeling by Wavelets. John Wiley & Sons, New York.

Vidakovic, B., Ruggeri, F., 2000. Bams method: Theory and simulations, Tech. Rep. Institute of Statistics and Decision Sciences, Duke University.