



Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities

Jalal M. Fadili, Larbi Boubchir

► To cite this version:

Jalal M. Fadili, Larbi Boubchir. Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities. IEEE Transactions on Image Processing, 2005, 14 (2), pp.231-240. <10.1109/TIP.2004.840704>. <hal-01123882>

HAL Id: hal-01123882

<https://hal.science/hal-01123882v1>

Submitted on 5 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Analytical Form for a Bayesian Wavelet Estimator of Images Using the Bessel K Form Densities

Jalal M. Fadili and Larbi Boubchir

Abstract—A novel Bayesian nonparametric estimator in the wavelet domain is presented. In this approach, a prior model is imposed on the wavelet coefficients designed to capture the sparseness of the wavelet expansion. Seeking probability models for the marginal densities of the wavelet coefficients, the new family of Bessel K forms (BKF) densities are shown to fit very well to the observed histograms. Exploiting this prior, we designed a Bayesian nonlinear denoiser and we derived a closed form for its expression. We then compared it to other priors that have been introduced in the literature, such as the generalized Gaussian density (GGD) or the α -stable models, where no analytical form is available for the corresponding Bayesian denoisers. Specifically, the BKF model turns out to be a good compromise between these two extreme cases (hyperbolic tails for the α -stable and exponential tails for the GGD). Moreover, we demonstrate a high degree of match between observed and estimated prior densities using the BKF model. Finally, a comparative study is carried out to show the effectiveness of our denoiser which clearly outperforms the classical shrinkage or thresholding wavelet-based techniques.

Index Terms—Bayesian denoiser, Bessel K forms (BKF), posterior conditional mean, wavelets.

I. 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 data sets without the imposition of a parametric regression model. During the 1990s, the nonparametric regression literature was arguably dominated by nonlinear wavelet shrinkage and wavelet thresholding estimators [1]–[3]. 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 [4].

Since the seminal papers by Donoho and Johnstone [1], the image processing literature have been inundated by hundreds of papers applying or proposing modifications of the original algorithm in image processing problems. Various data adaptive

wavelet thresholding estimators have been developed (see, for example, the extensive reviews in [5], [6], and the references therein).

Various Bayesian approaches for nonlinear wavelet thresholding and nonlinear wavelet shrinkage estimators have also 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. The prior model is designed to capture the sparseness of wavelet expansions. 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 [7]–[18]. 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 detailed study involving recent classical and Bayesian wavelet methods was carried out in [19] in the development toward high-performance wavelet estimators and their finite sample properties.

A popular prior for each wavelet coefficient is a scale mixture of two normal distributions [8] or one normal distribution and a point mass at zero [9], [20]. In [21], 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 [22] and on nonparametric mixed-effects model [16]. Since the work of Mallat [4], the generalized Gaussian distribution (GGD) has been commonly used as a prior for the wavelet coefficients in the image processing community (see, e.g., [7], [17], and [23]). Simoncelli [24] 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 [18] used α -stable distributions [25], a family of heavy-tailed densities, as a prior to capture the sparseness of the wavelet coefficients at each scale. These authors showed the superiority of the α -stable distributions in fitting the mode and the 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. Furthermore, in both the GGD and the α -stable priors, the derived Bayesian estimator has no closed analytical form in general situation and involves intensive numerical integration.

Using a physical model for image formation (the so-called *transported generator model*), a family of two-parameter probability densities, called Bessel K forms (BKFs), have been proposed in [26] and [27] to model the distribution of arbitrary images that have been filtered by a variety of band-pass filters (e.g., derivative, Gabor, interpolation, steerable filters, etc). It is obvious that wavelet decompositions of an image are members of such class of filters. Therefore, the BKF is a suitable model provided that the resulting wavelet coefficients marginals are: unimodal, symmetric around the mode, and leptokurtic. The first two conditions are widely adopted in the literature and are common to other priors such as the α -stable or the GGD models. The last condition simply means that the prior is a sharply peaked distribution with tails that are heavier as compared to normal density of the same variance. The BKF is then adapted to capture the heavy tail behavior of wavelet coefficients densities. Exploiting this prior in a Bayesian framework, we designed a posterior conditional mean (PCM) estimator. We derived a closed-form for its expression as a main result of our paper.

This paper is organized as follows. In Section II, we define the nonparametric regression problem and introduce some notational aspects. Some necessary preliminaries on the Bessel K forms are provided in Section III. We then apply the BKF to model the prior densities of the wavelet coefficients on a database of test images. A comparison to other priors is also carried out. Section IV establishes the expression of the marginal posterior density of the wavelet coefficients under the BKF prior. We also propose a cumulants-based estimator of the hyperparameters involved in nonparametric regression problems, namely p , c , and σ^2 . Finally, a closed-form expression of the Bayesian denoiser based on an L_2 -loss criterion is stated. Proofs of the main results are deferred to the Appendix. Section V compares the performance of the designed algorithm with classical denoisers on a digitized database of images. Finally, conclusions and directions of future work are drawn.

II. NONPARAMETRIC WAVELET-BASED REGRESSION

Let g_{mn} , $m, n = 0, \dots, N-1$ equally-spaced samples of a real-valued image. N is considered as a power of 2 ($N = 2^J$). Consider the standard nonparametric regression setting

$$y_{mn} = g_{mn} + \epsilon_{mn} \quad (1)$$

where ϵ_{mn} are iid normal random variables with mean zero and variance σ^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 \mathbf{y} , \mathbf{g} , and ϵ denote the matrix representation of the corresponding samples. Let $\mathbf{D} = \mathcal{W}\mathbf{y}$, $\mathbf{S} = \mathcal{W}\mathbf{g}$, and $\mathbf{V} = \mathcal{W}\epsilon$, where \mathcal{W} is the two-dimensional dyadic orthonormal wavelet transform (DWT) operator [28]. In a two-dimensional setting, the subbands HH_j , HL_j , and LH_j , $j = J_c, \dots, J-1$ 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 \mathbf{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 (1) that

$$d_{mn}^{oj} = s_{mn}^{oj} + v_{mn}^{oj}, \quad j = J_c, \dots, J-1; m, n = 0, \dots, 2^j - 1. \quad (2)$$

The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few large detail coefficients in \mathbf{D} contain information about the underlying image \mathbf{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 \mathbf{g} . By thresholding or shrinking the detail coefficients and inverting the DWT, one can obtain an estimate of the underlying image \mathbf{g} . So, the resulting three-step wavelet-based estimation procedure can be summarized by the following diagram:

$$\mathbf{y} \xrightarrow{\text{DWT}} \mathbf{D} \xrightarrow{\text{Nonlinear estimation operator } \eta} \{\hat{\mathbf{S}} = \eta(\mathbf{D})\} \xrightarrow{\text{IDWT}} \hat{\mathbf{g}}$$

where η is a nonlinear (shrinkage or thresholding) operator. Examples of such an operator are the hard and soft thresholding introduced by Donoho *et al.* [1]. Following their work, there are a variety of methods in the literature to choose the threshold level and the thresholding rule (see [19] for a review and small sample performance of these methods).

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 an introduction to Bessel K forms distributions family suitable to characterize the wavelet subband coefficients densities which have been already observed to be sharply peaked and heavily tailed.

III. BESSEL K FORMS MODEL

A. Definition

Let g_F be a filtered version on an image g by the bandpass filter F . Using the *transported generator model*, the density function of g_F has been shown to be [26] for $p > 0$, $c > 0$

$$f(x; c, p) = \frac{1}{\sqrt{\pi}\Gamma(p)} \left(\frac{c}{2}\right)^{-\frac{p}{2}-\frac{1}{4}} \left|\frac{x}{2}\right|^{p-\frac{1}{2}} K_{p-\frac{1}{2}} \left(\sqrt{\frac{2}{c}}|x|\right) \quad (3)$$

where K_ν is the modified Bessel function defined as [29], [30]

$$K_\nu(zx) = \frac{\Gamma(\nu + \frac{1}{2})(2x)^\nu}{\sqrt{\pi}z^\nu} \int_0^{+\infty} \frac{\cos(zt)dt}{(t^2 + x^2)^{\nu+\frac{1}{2}}} \quad (4)$$

$\left(Re(\nu) > -\frac{1}{2}, z > 0, |\arg x| < \frac{\pi}{2}\right)$

p and c are respectively the shape and scale parameters. The associated characteristic function is

$$\varphi(\omega) = \left(1 + \frac{c\omega^2}{2}\right)^{-p}. \quad (5)$$

This probability density function (pdf) is flexible enough to account for a mode not centered at zero. Nonetheless, we restrict ourselves to only two-parameter BKF densities throughout this paper. For $p = 1$, f simply reduces to the double exponential pdf. More generally, f is the p th convolution power of double exponential pdfs. If $p > 1$, we get closer to the Gaussian case (especially when $p \gg 1$, which is intuitively acceptable using a central limit theorem argument). If $p < 1$, the pdf becomes more sharply peaked and the tails are heavier.

B. Properties

The BKF family has many properties, as follows.

- 1) The pdf is unimodal, symmetric around the mode.
- 2) It can be easily shown from the log-characteristic function that the cumulants of a BKF variable are of the form

$$\kappa_{2i} = p \left(\frac{c}{2} \right)^i \frac{(2i)!}{i}, \quad i \geq 1. \quad (6)$$

All odd cumulants are null and even ones are nonzero. Particularly, the variance and the kurtosis of a BKF random variable X are

$$\text{Var}(X) = \kappa_2 = pc, \quad \text{Kurt}(X) = \frac{\kappa_4}{\kappa_2^2} + 3 = \frac{3}{p} + 3. \quad (7)$$

This result is similar to the one established in [26]. Then, the pdf is always leptokurtic with tails heavier than the normal pdf.

- 3) A BKF is a specific kind of normal variance-mean mixture when the mixing variable is Gamma distributed with parameters p and c [26], [27]. Henceforth, to simulate a BKF variable, one can generate normal samples whose variances are randomly generated from a Gamma distribution with parameters p and c . The more general family of distributions, namely the generalized hyperbolic distribution described by [31], could be used if more parameters are needed to describe the pdf (e.g., skewness).
- 4) If the random variable X is BKF distributed with parameters p and c , then X/σ is BKF distributed with parameters p and c/σ^2 .

From (7), the parameters p and c can be easily estimated using the second- and fourth-order moments of X according to

$$\hat{p} = \frac{3}{\text{Kurt}(X) - 3}, \quad \hat{c} = \frac{\text{Var}(X)}{\hat{p}}.$$

There are a few subtleties which should be considered when working with these statistics and we will bear them in mind throughout the paper. First, care must be taken when estimating these quantities from a limited sample. The sample kurtosis and variance can be estimated using the classical formulas. However, it turns out that these estimators are biased at small samples. In the case where one has large number of samples n , this is not a problem. On the other hand, for small samples, one can construct unbiased centered estimators of the cumulants using k statistics [32]. This is an important issue when dealing with samples in the wavelet domain at coarse levels. In addition, the variance of the k statistics estimator is minimum compared to

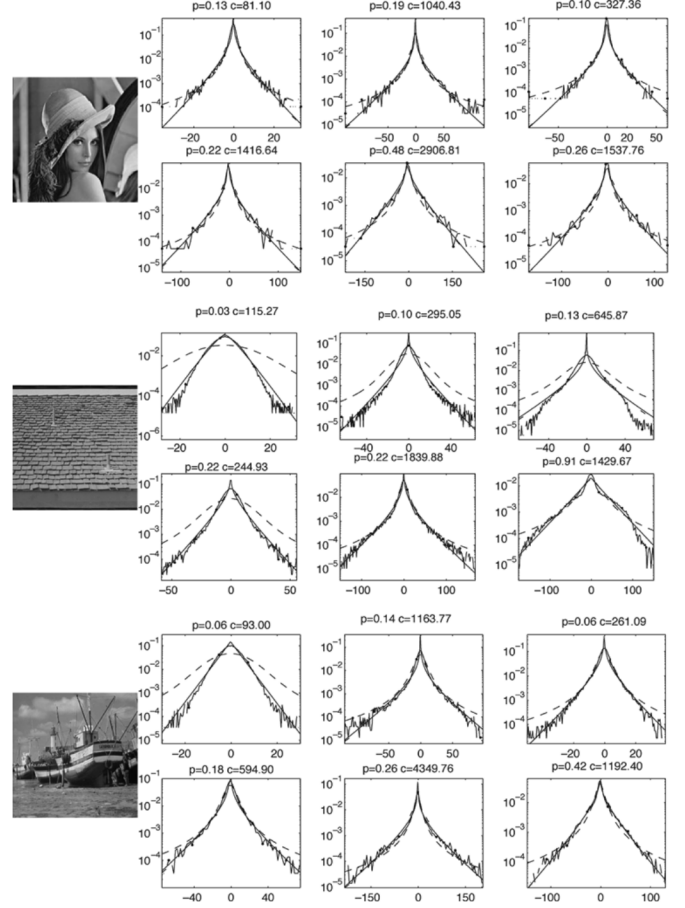


Fig. 1. Estimated and observed densities of the wavelet detail coefficients of three classical images (Lena, Roof textured image, and Boat image). The observed histogram ($- \bullet -$) was fitted using the BKF (solid), α -stable (dashed), and GGD (dotted) models. Only two detail levels are shown for each image (first and second row each time). The three columns correspond to the HH, HL, and LH orientations.

all other unbiased estimators. Therefore, k statistics unbiased cumulants estimators were used

$$\hat{\kappa}_2 = \frac{n}{n-1} \hat{M}_2, \quad \hat{\kappa}_4 = \frac{n^2 \left[(n+1) \hat{M}_4 - 3(n-1) \hat{M}_2^2 \right]}{(n-1)(n-2)(n-3)} \quad (8)$$

where \hat{M}_i is the i th sample central moment.

We now illustrate some prior estimation results for a variety of images taken from a digitized database [33]. Depicted in the first column of Fig. 1 are some images taken from this database (Lena, textured image Roof, and the famous Boat image). Shown are the estimated and the observed densities of the wavelet detail coefficients of each image on log scale. The observed histogram ($- \bullet -$) was fitted using the BKF (solid), α -stable (dashed), and 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 α -stable prior tends to exaggerate the tails of the distribution in some cases (e.g., for textured images), while the GGD prior suffers from a lack of capturing the heavy tails behavior. The BKF model is a good compromise between these two extreme cases (hyperbolic tails for the α -stable and exponential tails for the GGD).

TABLE I
MEAN (AND STANDARD DEVIATION IN PARENTHESES) OF THE KL DIVERGENCE BETWEEN THE OBSERVED AND THE FITTED DENSITIES OF THE WAVELET DETAIL COEFFICIENTS AT THREE DECOMPOSITION SCALES FOR EACH ORIENTATION. THREE PRIORS ARE COMPARED, THE BKF, THE α -STABLE, AND THE GGD MODELS. THE KL DIVERGENCE WAS AVERAGED AT EACH SCALE ORIENTATION OVER 100 IMAGES TAKEN FROM A DIGITIZED DATABASE

HH			
Prior	$j = 1$	$j = 2$	$j = 3$
BKF	-0.0251 (0.1572)	-0.0021 (0.2605)	0.0815 (0.5981)
α -stable	0.1388 (0.1788)	0.1452 (0.2398)	0.3385 (1.8938)
GGD	0.0851 (0.2638)	0.1016 (0.3089)	0.0506 (0.1607)
HL			
BKF	0.0680 (0.6638)	0.0217 (0.2308)	0.0119 (0.1240)
α -stable	0.1270 (0.2063)	0.1431 (0.3026)	0.1001 (0.2460)
GGD	0.1056 (0.2339)	0.0813 (0.1293)	0.0608 (0.0912)
LH			
BKF	-0.0350 (0.3081)	-0.0056 (0.1857)	0.0509 (0.1943)
α -stable	0.1475 (0.1809)	0.2220 (0.5658)	0.1525 (0.5611)
GGD	0.1291 (0.2585)	0.0804 (0.1622)	0.0744 (0.1323)

To quantify the performance of each prior in modeling-observed histograms, many quantities can be used, such as the L_2 norm between pdfs, the Kolmogorov–Smirnov (KS) distance, or the Kullback–Leibler (KL) divergence. It turns out that the BKF pdf is not square integrable for $p < 0.25$. The KS statistic is not easily accessible since it uses the distribution function which is not known in closed form for these prior models. A numerical version could be used instead. Here, we prefer to use the KL divergence between two pdfs discretized at the histogram bins. We have computed this quantity for each fitted prior model over a 100 image database. The KL divergence between the observed and the fitted densities of the wavelet detail coefficients were calculated at three decomposition scales for each orientation. Shown in Table I is the average KL measure for each prior model, scale, and orientation over the 100 images. Standard deviations are also reported in parentheses. This table underscores a high degree of match between the observed and the estimated density using the BKF model. The BKF prior clearly outperforms the α -stable prior. This is mainly due to the fact that the α -stable model exaggerates the tails of the fitted density. While the mean value of the BKF prior KL divergence is usually smaller than that of the GGD, the standard deviation associated with the BKF is sometimes higher than that of the GGD. We can then legitimately claim that the BKF density fits the data at least as well as the GGD. The BKF can prove better than the GGD model in capturing the heavy tails of the observed histogram. Furthermore, a closed-form expression of the Bayesian shrinkage rule associated with the BKF prior can be readily obtained, as we will show in the next section.

IV. BAYESIAN DENOISING

A. Marginal Density of the Wavelet Coefficients

As stated above, in the Bayesian approach, a prior is imposed on the wavelet coefficients designed to describe their distribu-

tion. It is also assumed in the prior model that the wavelet coefficients s_{mn}^{oj} of the true image are mutually independent random variables (and independent of the noise process v_{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. As clearly demonstrated in the previous section, the wavelet detail coefficients at each scale and orientation are BKF distributed

$$s \sim \text{BKF}(p, c). \quad (9)$$

Once the data are observed, the observed wavelet coefficients d are determined and we seek the posterior distribution of s . Under the white Gaussian noise model considered here, the d are independent and Gaussian conditionally on s

$$d|s \sim \mathcal{N}(s, \sigma^2). \quad (10)$$

Therefore, the marginal pdf of d given σ^2 , p , and c

$$f_d(d) = \int_{-\infty}^{+\infty} f_{d|s}(d-s) f_s(s; c, p) ds = \int_{-\infty}^{+\infty} \phi(d; s, \sigma^2) f_s(s; c, p) ds \quad (11)$$

where $\phi(x; s, \sigma^2)$ is the normal pdf with mean s and variance σ^2 and $f_s(x; c, p)$ is the BKF pdf given in (3). Although the proposed BKF prior models the histograms very well, its exact analytic form is not easy to work with. Approximating the modified Bessel function $K_\nu(x)$ using the expansion in [30], also used by authors in [27], we can state the following lemma.

Lemma 1: For $0 < p \leq 1$ and $c > 0$, then the marginal pdf can be well approximated by

$$\tilde{f}_d(d; p, c, \sigma^2) = \left(\frac{\sigma^2}{2c} \right)^{\frac{p}{2}} \phi(d; 0, \sigma^2) (I_+ + I_-) \quad (12)$$

where $I_\pm = e^{((\pm(d/\sigma) + \sigma\sqrt{2/c})/2)^2} D_{-p}(\pm(d/\sigma) + \sigma\sqrt{2/c})$. $D_\nu(x)$ stands for the parabolic cylinder function of fractional order ν [29], [30]. f_d is an even function whose mode is at zero. See the Appendix for the proof.

The parabolic cylinder function involved in (12) is defined in the Appendix. From a practical point of view, there is an extended literature on numerical algorithms for computing the parabolic cylinder function. In [34], authors give Fortran programs for computing $D_\nu(x)$ with real orders ν and real arguments. The methods are based on recursions, MacLaurin series and asymptotic expansions. The numerical computation of $D_\nu(x)$ based on this method is fast and stable. Here, we implemented Matlab and C codes which are translations of the Fortran program given in [34]. A Fortran toolbox with numerous routines for computing special functions, including $D_\nu(x)$, can be retrieved from the web site [35]. The C and Matlab codes for computing $D_\nu(x)$ can be obtained upon request from the author.

Remark: As pointed out by Grenander *et al.* [26], [27], most of filtered natural images have a value of $p \leq 1$, then our condition in Lemma 1 is not restrictive and the approximation is very good.

B. Hyperparameters Estimation

To implement the formula in (12) or (17) in practice, one must elicit the hyperparameters estimation problem, which, in turn, will lead to a data-driven denoising procedure that is adaptive

to each subband. For example, one may apply a maximum likelihood (ML) approach. However, given the parametric form of (12), there is no known closed form expression of the ML estimates and intensive numerical methods must be used. The approach we propose in the present paper is based on higher order statistics, i.e., cumulants. It is obvious to show that adding a Gaussian process to a BKF process will only modify its variance but not the other cumulants. Therefore, we can write the cumulants of d as

$$\kappa_2 = pc + \sigma^2, \text{ and } \kappa_{2i} = p \left(\frac{c}{2}\right)^i \frac{(2i)!}{i} \text{ for } i \geq 2. \quad (13)$$

Again, only even cumulants are nonzero. To estimate the three unknown hyperparameters, namely p , c , and σ^2 , it is enough to estimate the cumulants up to order six. Using the fourth (kurtosis) and sixth-order cumulants, one can easily derive the analytical expression of c . Then, substituting c in the kurtosis formula, an estimate of p can be deduced. The estimate of σ^2 is finally obtained from the second-order cumulant (variance). However, using the sixth-order cumulant (even with the k statistics estimator) can yield to very poor performance since the variance of any cumulant estimate increases dramatically with its order [36, p. 212]. Furthermore, in the presence of noise, the estimate quality is more deteriorated because the cumulant estimator variance is an increasing power function of the noise variance [36].

Based upon these arguments, we here suggest an alternative estimation strategy. If one is able to measure σ , then this estimate could be used. If such is not the case, σ is estimated from the HH orientation at the finer scale using the popular robust estimator [1]

$$\hat{\sigma} = \frac{\text{MAD}(d_{mn}^{\text{HH}_1})}{0.6745} \quad (14)$$

where MAD is the median absolute deviation. Using this estimate, one can easily derive the following estimates of p and c at each subband from the noisy observations using their second- and fourth-order cumulants

$$\hat{p} = \frac{3 \max(\hat{\kappa}_2 - \hat{\sigma}^2, 0)^2}{\hat{\kappa}_4}, \quad \hat{c} = \frac{\max(\hat{\kappa}_2 - \hat{\sigma}^2, 0)}{\hat{p}}. \quad (15)$$

In fact, the zero argument in (15) must be replaced by any positive real very small to keep p or c nonzero. $\hat{\kappa}_2$ and $\hat{\kappa}_4$ are the sample second- and fourth-order cumulants obtained from (8). In practice, if c (or p) is zero, then the prior pdf is simply a point mass at zero and the posterior conditional mean estimator $s_{\text{PCM}}(d)$ presented in the next section is zero for all d .

Shown in Fig. 2 are the estimated (solid) and observed (dashed) marginal densities f_d of the wavelet detail coefficients for the same images as in Fig. 1. These images were corrupted by a zero-mean Gaussian noise [signal-to-noise ratio (SNR)=15 dB]. The observed histogram was fitted using (12). The hyperparameters were estimated for each scale and orientation using (14) and (15). Two detail levels are shown for each image. The estimated parameters p and c are shown at the

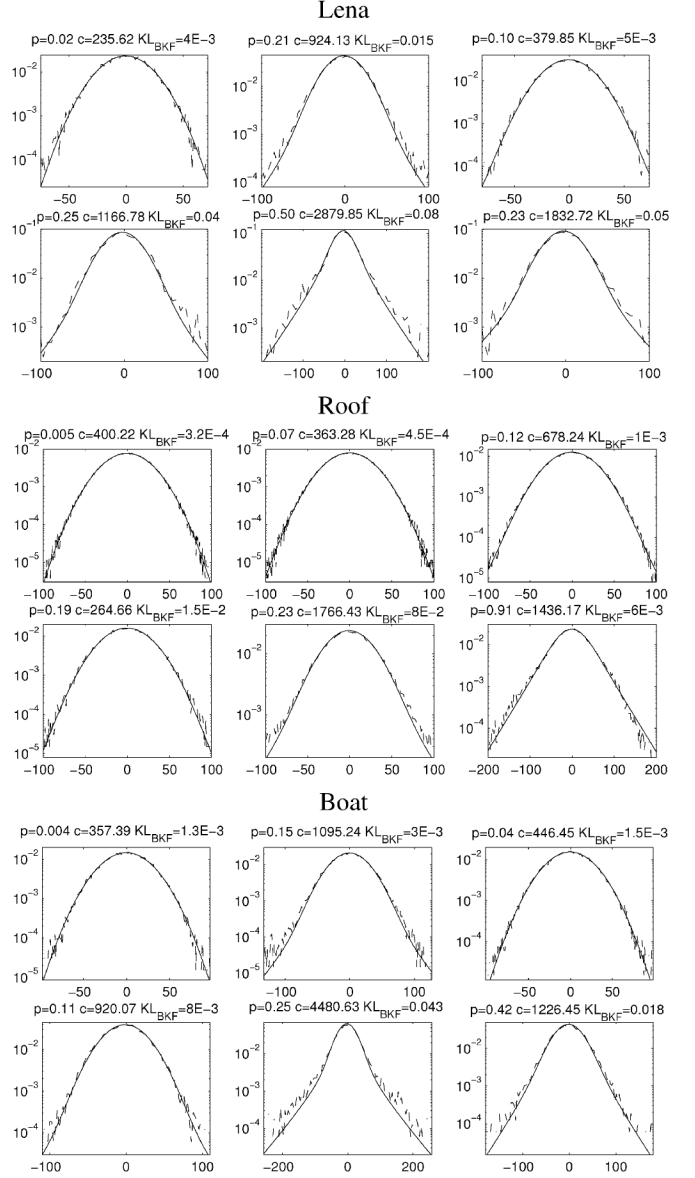


Fig. 2. Estimated (solid) and observed (dashed) marginal densities (in log scale) of the wavelet detail coefficients for the same images as in Fig. 1 corrupted by a zero-mean Gaussian noise (SNR=15 dB). The observed histogram was fitted using (12). Only two detail levels are shown for each image. Again, the three columns correspond to the HH, HL, and LH orientations.

top of each plot. These estimates are in very good agreement with those obtained from the noise free images, depicted at the top of each plot in Fig. 1. We can also observe a high degree of match between the observed histograms of noisy detail wavelet coefficients and the fitted marginal density given by (12).

Remark: As one can notice, the first decomposition level is dominated by noise and the marginal density tends to normal. At finer scales, the signal-to-noise ratio increases and the observed marginal has heavier tails (BKF-like behavior). Furthermore, the estimates of p and c at these fine scales are closer to those of the noise free image.

C. Bayesian Term-by-Term Denoising

1) *Posterior Conditional Mean Estimator:* Let us now turn to the Bayesian framework to obtain wavelet shrinkage

estimates of the unknown image g . Different losses will lead to different Bayesian rules and, therefore, to different wavelet estimates for the unknown image [19]. It is well known that L_2 -based Bayes rules correspond to posterior conditional means (PCM) estimates of wavelet coefficients s (conditionally on the hyperparameters σ^2 , p , and c), given by

$$\begin{aligned} s_{\text{PCM}}(d) &= E_{s|d}(s) = \frac{\int_{-\infty}^{+\infty} s f_{d|s}(d-s) \tilde{f}_s(s; p, c) ds}{\int_{-\infty}^{+\infty} f_{d|s}(d-s) \tilde{f}_s(s; p, c) ds} \\ &= \frac{\int_{-\infty}^{+\infty} s f_{d|s}(d-s) \tilde{f}_s(s; p, c) ds}{\tilde{f}_d(d; p, c, \sigma^2)}. \end{aligned} \quad (16)$$

This expression corresponds to a level-dependent Bayesian coring rule (following terminology in [7]). The denominator in this equation has been established in Lemma 1. The following theorem gives the closed analytical form of the PCM estimator of the wavelet coefficients s under the observation model in (2), given σ^2 , p , and c .

Theorem 1: For $0 < p \leq 1$, and c strictly positive. Under the observation model given in (2), the posterior conditional mean estimator is given by (17), shown at bottom of the next page, which is a nonlinear odd function (see Appendix for the proof).

Depicted in Fig. 3 are the Bayesian rule input-output curves obtained using the result of Theorem 1. The first (respectively, last) two rows show the influence of the shape parameter p (respectively, the ratio \sqrt{c}/σ) on the Bayesian estimator curves for constant ratio \sqrt{c}/σ (respectively, shape parameter p). 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. This asymptotic behavior can be proved using the asymptotic expansion of $D_\nu(x)$ for large $|x|$ [29]. In the context of image denoising, this is accomplished in a data-driven adaptative manner since the hyperparameters are estimated from the data at each orientation and each scale. As far as the influence of p is concerned, the amount of shrinkage decreases as p increases. This can be intuitively understood from the fact that as p increases, the heavier the tails and the higher the probability that smaller values are due to the true function g . Similar observation was made for the α -stable prior in [18]. Another explanation of this behavior is that as p increases, the SNR increases ($\text{SNR} = \sqrt{pc}/\sigma$), yielding less shrinkage amount. The same behavior is also observed in Fig. 3(b) with increasing values of the ratio \sqrt{c}/σ , which can be interpreted in a similar way. The designed denoiser of (17) is only valid for $p \leq 1$ which is not restrictive as most of filtered natural images have a value of $p \leq 1$. Nevertheless, in practice, if p happens to

be large, i.e., the signal pdf tends to a Gaussian, the well-known linear PCM estimator $(pc/(pc + \sigma^2))d$ can be used.

V. EXPERIMENTAL RESULTS

We now assess the performance of the Bayesian denoiser with the BKF prior by comparing it to various denoising methods. We again use the test images of the digitized database [33]. Five other denoising algorithms are considered: Bayesian PCM denoiser using α -stable prior [18], [37], universal threshold hard and soft thresholding [1], and the Stein unbiased risk estimator (SURE) [2]. The best possible uniform thresholding technique with a threshold that minimizes the MSE between the corrupted and the uncontaminated image was also computed. We refer to this as the oracle threshold [23]. With the α -stable prior, no closed form is available for its PCM Bayesian denoiser. Direct numerical integrations are too time consuming and numerically unstable. In the work of [37], we proposed a method using the characteristic functions corresponding to the true image coefficients s and the noise. The Fourier integrals involved in its PCM estimator expression were computed using the fast Fourier transform, which turned out to be numerically efficient and stable [38]. The DWT employs Daubechies compactly-supported wavelet with regularity 4. The coarsest level of decomposition was chosen to be $\log_2 \log N^2 + 1$ from asymptotic considerations [19].

Beside visual quality, we calculated the signal to MSE ratio (SMR) in order to quantify the achieved performance improvement. This is defined in decibels, as follows:

$$\text{SMR}_{\text{dB}} = 10 \log_{10} \frac{\sum_{mn} g_{mn}^2}{\sum_{mn} (\hat{g}_{mn} - g_{mn})^2} \quad (18)$$

where the denominator is the estimated MSE between the original and the denoised images.

Fig. 4 shows the resulting images for each denoising method for the Lena image with an input SNR = 15 dB ($\sigma = 20$). One can see that the visual quality of the Bayesian BKF prior-based denoiser is superior to the other methods. The mean and standard deviation (over ten runs) of SMRs given by the various methods for the Lena image are compared in Fig. 5. The SMRs were computed for each run and each value of input SNR_{in} in the range [5, 20] dB. Again, one can notice that the BKF denoiser outperforms the other methods. It compares favorably with the oracle thresholding but is much better than the SURE, especially at low SNRs. The universal hard thresholding method gives a poorer result although better than soft thresholding which gives an oversmooth estimate. The α -stable prior Bayesian denoiser is underperforming at low-input SNR_{in}. The main reason is

$$s_{\text{PCM}}(d) = p\sigma \frac{e^{\left(\frac{-\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}}{2}\right)^2} D_{-p-1}\left(-\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}\right) - e^{\left(\frac{\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}}{2}\right)^2} D_{-p-1}\left(\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}\right)}{e^{\left(\frac{-\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}}{2}\right)^2} D_{-p}\left(-\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}\right) + e^{\left(\frac{\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}}{2}\right)^2} D_{-p}\left(\frac{d}{\sigma} + \sigma\sqrt{\frac{2}{c}}\right)} \quad (17)$$

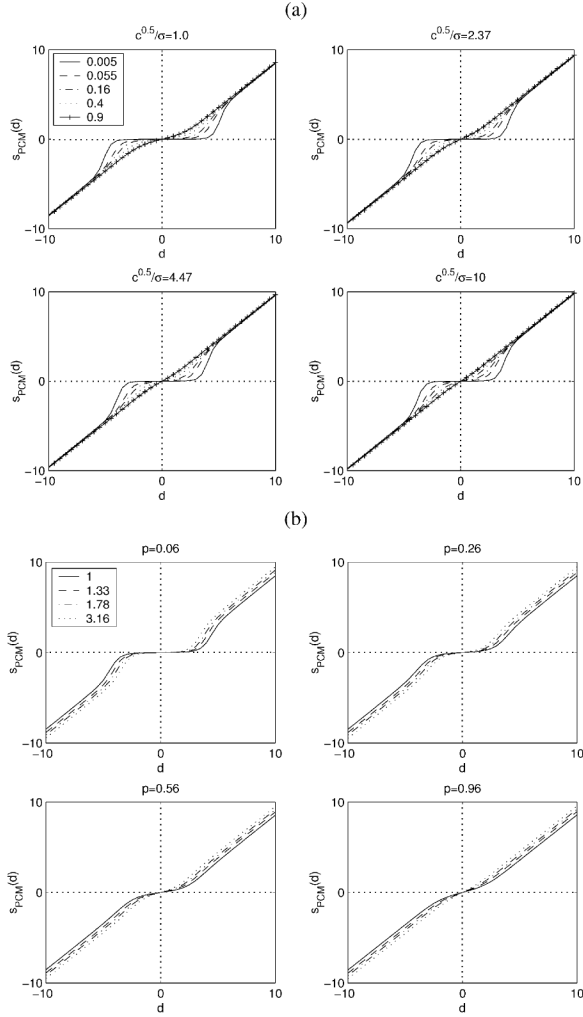


Fig. 3. Bayesian rule input-output curves (s_{PCM} as a function of d). (a) Influence of the BKF prior shape parameter p on the Bayesian estimator curves for constant ratio \sqrt{c}/σ . As shown in the legend, p varies from small (corresponding to strongly heavy-tailed signal, solid line) to large values (slightly heavy-tailed signal, line with + marker). (b) Influence of the ratio \sqrt{c}/σ on the Bayesian estimator curves for constant p : $\sqrt{c}/\sigma = 1$ (solid) to $\sqrt{c}/\sigma = 3.16$ (dotted).

the weakness of the hyperparameters estimator which remains an important issue. Surprisingly, this has not been pointed out by authors in [18]. This general behavior is also observed on the Roof and the Boat images. For the Roof textured image, where the prior distribution is expected to be sharply-peaked and heavily-tailed, the advantages of the BKF Bayesian denoiser are most salient at low SNR_{in} . In Fig. 6, the average SMRs over the ten runs and the whole database (100 images) for each denoising method, as a function of SNR_{in} , are depicted. The general behavior described before is confirmed by this plot. This suggests that the BKF prior is a suitable model adapted to capture the sparseness behavior of the wavelet coefficients for a large class of images.

VI. CONCLUSION

In this paper, a novel wavelet-based Bayesian denoiser has been introduced. Specifically, a new class of prior pdfs, namely the Bessel K forms, has been proposed. We have shown that



Fig. 4. Comparing various denoising methods for Lena image corrupted by noise with an input $\text{SNR}_{in} = 15$ dB ($\sigma = 20$). The Bayesian PCM denoiser with the BKF prior is clearly superior to the other methods confirming the general behavior observed on the test database used in this paper.

the BKF prior is very powerful in capturing the sparseness behavior of the wavelet coefficients over a large class of images. The BKF model also demonstrated a high degree of match between observed and estimated prior densities. Specifically, the BKF model turns out to be a good compromise between two extreme cases: the α -stable prior with exaggerated hyperbolic tails and the GGD prior with exponential tails insufficient to describe correctly the heavy-tailed behavior of the wavelet coefficients distribution. Using this prior, analytical expressions for the posterior marginal distribution as well as the posterior conditional mean estimators have been derived. When compared to other denoising approaches, the experimental results have shown the superiority of the BKF Bayesian denoiser over a large database of images.

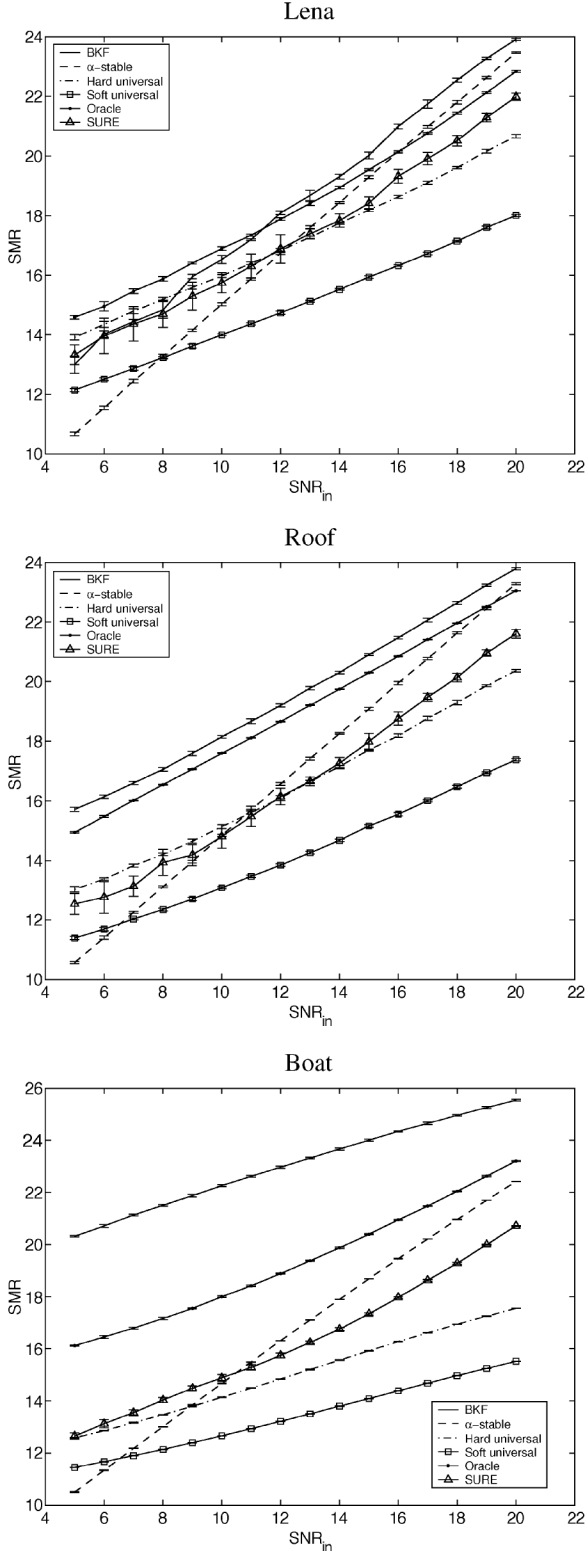


Fig. 5. For each SNR_{in} , mean and standard deviation (over ten runs) of SMRs given by various denoising methods for three test images: Lena, Boat, and Roof.

Despite his good performances, there are still some aspects in this method that have to be investigated. First, we can point out the lack of translation invariance which yields a ringing effect, even if this effect is quite negligible in the BKF denoised images. This problem can be solved either by cycle spinning using Coifman and Donoho algorithm [39] or using translation

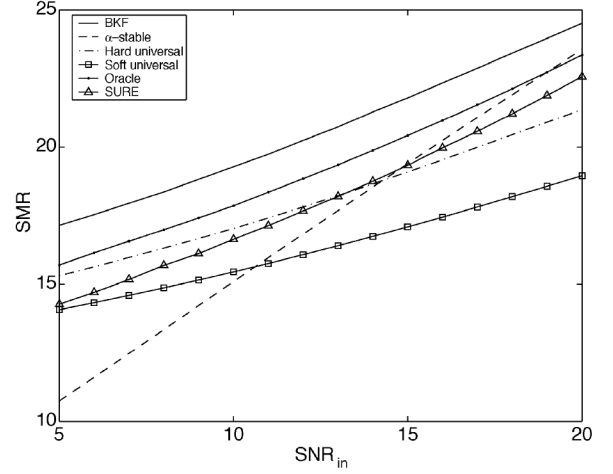


Fig. 6. Average SMRs over the ten runs and the 100 image database for each denoising method, as a function of the SNR_{in} .

invariant transforms, such as the nonorthogonal undecimated (redundant) wavelet transform [40]. The former is simple to implement but is two time consuming. In the latter, the transform is highly redundant and the noise becomes correlated. Furthermore, the interscale 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-based transforms, e.g., curvelets [41] and bandlets [42]. Another way would consist of combining contextual spatial and Bayesian intensity information in an elegant and powerful Markov random field model, in the same vein as in [43] and [44]. Our investigations are now focusing on these aspects.

APPENDIX I PROOF OF LEMMA 1

Proof: First, we will use the following approximation result given in [29], [30]. When $\nu = p - (1/2)$ is fixed, for large $|x|$, the first-order asymptotic expansion of $K_\nu(x)$ is

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x} \quad (19)$$

with a rest given by

$$|R| < \left| \frac{\Gamma(\nu + \frac{3}{2})}{(2x)\Gamma(\nu - \frac{1}{2})} \right|, \text{ for } \nu - \frac{1}{2} \leq 1. \quad (20)$$

This means that for real ν and x real and positive, this rest is less than the absolute value of the first discarded term and is of the same sign, provided that $p = \nu + (1/2) \leq 2$. Applying this result to the BKF pdf, it follows that:

$$\tilde{f}(x; c, p) = \frac{(2c)^{-\frac{p}{2}}}{\Gamma(p)} |x|^{p-1} e^{-\sqrt{\frac{2}{c}}|x|}. \quad (21)$$

Obviously, for $p > 1$, the pdf in (21) is bimodal with modes at $\pm(p-1)\sqrt{(2/c)}$ and not at zero. Therefore, (21) is only valid

for $p \leq 1$. Plugging this expression into (11) and decomposing the integral for positive and negative s , we can write

$$\tilde{f}_d(d; p, c, \sigma^2) = \frac{(2c)^{-\frac{p}{2}}}{\Gamma(p)} \phi(d; 0, \sigma^2) [I_+ + I_-] \quad (22)$$

where $I_{\pm} = \int_0^{+\infty} s^{p-1} e^{-(s^2/2\sigma^2) - s(\pm(d/\sigma^2) + \sqrt{(2/c)})} ds$. To establish the final result, we will need the integral formula in [30, p. 337]

$$\int_0^{+\infty} s^{\mu-1} e^{-\beta s^2 - \gamma s} ds = (2\beta)^{-\frac{\mu}{2}} \Gamma(p) e^{\frac{\gamma^2}{8\beta}} D_{-\mu} \left(\frac{\gamma}{\sqrt{2\beta}} \right) \quad (23)$$

$Re(\mu) > 0, Re(\beta) > 0$

where $D_{\mu}(x)$ is the parabolic cylinder function [30]. It has several integral representations. In our case of interest, the argument μ is always strictly negative; therefore, we choose the following representation [30]:

$$D_{\mu}(x) = \frac{e^{-\frac{x^2}{4}}}{\Gamma(-\mu)} \int_0^{+\infty} e^{-xt - \frac{t^2}{2}} t^{-\mu-1} dt, \text{ for } Re(\mu) < 0 \quad (24)$$

$\mu = p$ and $\beta = (1/2\sigma^2)$ are strictly positive by definition. Then, combining the integral in (23) with (22) and after some algebraic arrangements, the result in (12) follows. ■

APPENDIX II PROOF OF THEOREM 1

Proof: First, let us recall the posterior conditional mean estimator expression

$$s_{\text{PCM}}(d) = E_{s|d}(s) = \frac{\int_{-\infty}^{+\infty} s f_{d|s}(d-s) \tilde{f}_s(s; p, c) ds}{\int_{-\infty}^{+\infty} f_{d|s}(d-s) \tilde{f}_s(s; p, c) ds} \quad (25)$$

The denominator expression has been established in Lemma 1. To derive the numerator expression, we follow the same steps as in Lemma 1. By decomposing the integral in the numerator for positive and negative s , we can show that

$$\int_{-\infty}^{+\infty} s f_{d|s}(d-s) f_s(s) ds = p\sigma \left(\frac{\sigma^2}{2c} \right)^{\frac{p}{2}} \phi(d; 0, \sigma^2) (I'_- - I'_+) \quad (26)$$

where $I'_{\pm} = \int_0^{+\infty} s^p e^{-(s^2/2\sigma^2) - s(\pm(d/\sigma^2) + \sqrt{(2/c)})} ds$. from the integral formula in (23), its is straightforward to show that for $p+1 > 0$

$$I'_{\pm} = e^{\left(\frac{\pm \frac{d}{\sigma} + \sigma \sqrt{\frac{2}{c}}}{2} \right)^2} D_{-p-1} \left(\pm \frac{d}{\sigma} + \sigma \sqrt{\frac{2}{c}} \right) \quad (27)$$

which is valid since $p > 0$. Finally, combining (12), (26), and (27), the result in Theorem 1 follows. This completes the proof. ■

REFERENCES

- [1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [2] —, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [3] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Wavelet shrinkage: asymptopia?," *J. R. Stat. Soc. B*, vol. 57, no. 2, pp. 301–337, 1995.
- [4] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [5] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: Wiley, 1999.
- [6] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge, 2000.
- [7] E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," in *Proc. 3rd Int. Conf. Image Processing*, vol. 1, Lausanne, Switzerland, 1996, pp. 379–382.
- [8] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive bayesian wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 92, pp. 1413–1421, 1997.
- [9] F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a bayesian approach," *J. R. Stat. Soc. B*, vol. 60, pp. 725–749, 1998.
- [10] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
- [11] I. Johnstone and B. Silverman, "Empirical Bayes approaches to mixture problems and wavelet regression," Tech. Rep., Dept. Mathematics, Univ. Bristol, Bristol, U.K., 1998.
- [12] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *J. Amer. Stat. Assoc.*, vol. 93, no. 441, pp. 173–179, 1998.
- [13] M. A. Clyde and E. I. George, "Empirical bayes estimation in wavelet nonparametric regression," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds. New York: Springer-Verlag, 1999, pp. 309–322.
- [14] M. Clyde and E. George, "Flexible empirical bayes estimation for wavelets," *J. R. Stat. Soc. B*, vol. 62, pp. 681–698, 2000.
- [15] M. Vannucci and F. Corradi, "Covariance structure of wavelet coefficients: theory and models in a bayesian perspective," *J. R. Stat. Soc. B*, vol. 61, pp. 971–986, 1999.
- [16] S. Huang and H. Lu, "Bayesian wavelet shrinkage for nonparametric mixed effects models," *Stat. Sinica*, vol. 10, pp. 1021–1040, 2000.
- [17] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.
- [18] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 772–783, May 2001.
- [19] A. Antoniadis, J. Bigot, and T. Sapatinas, "Wavelet estimators in nonparametric regression: a comparative simulation study," *J. Stat. Software*, vol. 6, no. 6, 2001.
- [20] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, no. 2, pp. 391–401, 1998.
- [21] B. Vidakovic and F. Ruggeri, "Bams Method: Theory and simulations," Tech. Rep., Inst. Stat. Decision Sciences, Duke Univ., Durham, NC, 2000.
- [22] S. Huang and N. Cressie, "Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data," *Technometr.*, vol. 42, pp. 262–276, 2000.
- [23] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
- [24] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds. New York: Springer-Verlag, 1999, pp. 291–308.
- [25] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.
- [26] U. Grenander and A. Srivastava, "Probability models for clutter in natural images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 424–429, Apr. 2001.
- [27] A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probabilities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1200–1214, Sep. 2002.
- [28] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.
- [29] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1972.
- [30] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. New York: Academic, 1980.

- [31] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," 1982.
- [32] C. Rose and M. D. Smith, "k-statistics: unbiased estimators of cumulants," in *Mathematical Statistics With Mathematica*. New York: Springer-Verlag, 2002, ch. 7.2, pp. 256–259.
- [33] [Online]. Available: <http://sipi.usc.edu/services/database/database.html>
- [34] S. Zhang and J. Jin, *Computation of Special Functions*. New York: Wiley, 1996.
- [35] —, [Online]. Available: <http://iris-lee3.ece.uiuc.edu/~jjin/routines/routines.html>
- [36] R. A. Fisher, "Moments and product moments of sampling distributions," in *Contributions to Mathematical Statistics*. New York: Wiley, 1950, ch. 20, pp. 199–238.
- [37] J. Mathieu, "Transformée en ondelettes et régression nonparamétrique dans un contexte bayésien," M.S. thesis, Ecole Nationale Supérieure d'Ingénieur, Caen, France, 2002.
- [38] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [39] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," Tech. Rep., Dept. Mathematics, Yale Univ., New Haven, CT, 1995.
- [40] J. Sharcanski, C. R. Jung, and R. T. Clarke, "Adaptive image denoising using scale and space consistency," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1092–1101, Sep. 2002.
- [41] J. L. Starck, E. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
- [42] E. Le Pennec and S. Mallat, "Image compression with geometrical wavelets," in *Proc. Int. Conf. Image Processing*, vol. 1, Thessaloniki, Greece, 2000, pp. 661–664.
- [43] M. Malfait and D. Roose, "Wavelet-based image denoising using a markov random field a priori model," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 549–565, May 1997.
- [44] A. Pizurica, W. Philips, I. Lemahieu, and M. Achenoy, "A joint inter- and intrascale statistical model for bayesian wavelet-based image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, May 2002.



Jalal M. Fadili graduated from the Ecole Nationale Supérieure d'Ingénieurs (ENSI) de Caen, Caen, France, and received the M.Sc. and Ph.D. degrees in signal and image processing from the University of Caen.

He was a Research Associate at the university of Cambridge, Cambridge, U.K., in 2000. He has been an Assistant Professor of signal and image processing since September 2001 at ENSI. His research interests include statistical estimation and detection and multiscale methods in signal and image processing. His areas of application include medical imaging.



Larbi Boubchir received the B.Sc. degree in computer science from the University of Oran, Algeria, in 1998 and the M.Sc. degree in computer science from the University of Tours, Tours, France, in 2002. He is currently pursuing the Ph.D. degree in signal and image processing at the University of Caen, Caen, France, under the supervision of Dr. J. M. Fadili.

His research areas are directed toward Bayesian image restoration using multiscale geometrical transforms.