

BAYESIAN DENOISING BASED ON THE MAP ESTIMATION IN WAVELET-DOMAIN USING BESSEL K FORM PRIOR

Larbi Boubchir and Jalal M. Fadili

Image processing group, GREYC UMR CNRS 6072
ENSICAEN, 6 Bd du Maréchal Juin 14050 Caen, France

ABSTRACT

In this paper, a nonparametric Bayesian estimator in the wavelet domain using the Bessel K Form (BKF) distribution will be presented. Our first contribution is to show how the BKF prior is suited to characterize images belonging to Besov spaces. Exploiting this prior, our second contribution is to design a Bayesian L_1 -loss maximum a posteriori estimator nonlinear denoiser, for which we formally establish the mathematical properties. Finally, a comparative study is carried to show the effectiveness of our Bayesian denoiser compared to other denoising approaches.

1. INTRODUCTION

In the last decade, the nonparametric wavelet-based regression has been a fundamental tool in data analysis. Nonparametric regression (or denoising) estimators provide a powerful tool for recovering an unknown image, say g , based on sampled data that are contaminated with noise using multi-scale decompositions. Many of these estimators have been developed based on Donoho and Johnstone's work [1]. Since then, various Bayesian approaches for nonlinear wavelet shrinkage estimators have been developed recently, and various priors have been proposed to model the statistical behavior of the noiseless wavelet coefficients. These estimators impose a prior distribution on wavelet coefficients designed to capture the sparseness of the wavelet expansions. Then the image is estimated by applying a Bayesian rule to the resulting posterior distribution of the wavelet coefficients. Various prior choices can be found in the statistical literature (see [2] for a detailed review). For example, a popular prior is the Generalized Gaussian Distribution (GGD) proposed by Mallat [3], and used by many others, or the α -stable prior used by Achim [4]. However, the GGD prior suffers from a lack of capturing the heavy tail behavior of the observed wavelet coefficients densities. The α -stable prior show their superiority in fitting the mode and the tail behavior of the wavelet coefficients distributions. But 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.

In our approach, we propose a prior statistical model based on BKF. 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. Then, the BKF is adapted to capture the heavy tail behavior of wavelet coefficients densities. In this paper, our first contribution is to show how the BKF prior is suited to characterize images belonging to Besov spaces. More specifically, we shed the light on the relationship between the parameters of the BKF prior and those of the Besov space within which realizations of such a prior are likely (almost surely) to fall. Exploiting this prior, our second contribution is to design a Bayesian L_1 -loss maximum a posteriori estimator nonlinear denoiser, for which we formally establish the mathematical properties.

2. NONPARAMETRIC WAVELET-BASED REGRESSION

Let $g_{m,n}$, $m, n = 0, \dots, N - 1$ equally-spaced sampled of a real-valued image g . N is considered as a power of 2 ($N = 2^J$). The goal is to recover the underlying function g from the observed noisy data y_{mn} corrupted with Gaussian white noise, without assuming any particular parametric structure for g . Let x_{mn}^{oj} (resp. d_{mn}^{oj}) be the detail coefficient of the true image (resp. observed noisy image) at location (m, n) , scale j and orientation o . Due to the orthogonality of the basis, the DWT of white noise are also independent normal variables with the same variance. It follows from Eq.1 that:

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

where J_c is the coarsest scale of the decomposition. The approximation coefficients are kept intact because they represent low-frequency terms that contain important features about g .

3. THE BESSEL K FORMS DISTRIBUTIONS

Let g_F be a filtered version of an image g by the bandpass filter F . Using the *transported generator model*, the density function of g_F has been shown to be [5] for $p > 0, c > 0$

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

where K_ν , p and c are respectively the modified Bessel function (defined as [6]), the shape and scale parameters. We can also reparametrize this PDF by defining $\sigma_x = \sqrt{pc}$ which will be useful in the following:

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

If $p > 1$, we get closer to the Gaussian case. If $p < 1$, the PDF becomes more sharply peaked and the tails are heavier. The wavelet detail coefficients densities have been already observed to be sharply peaked and heavily tailed [3]. This is exactly the property which is captured by a BKF distribution where $0 < p \leq 1$ and c strictly positive. The latter property has been extensively illustrated in [7].

4. BKF PRIOR AND BESOV SPACES

The Besov spaces form an important spaces class in the image processing domain. The advantage of Besov spaces is that they are much more general tool in describing the smoothness and the regularity properties of functions, e.g. piece-wise smooth or with isolated singularities, etc. Here, instead of the original definition of Besov spaces through the modulus of continuity, we will focus on a definition of the Besov space norm using a practical characterization with the wavelet coefficients.

Definition 1 Let $g = \sum_{j,k} x_{j,k} \psi_{j,k}$ where $x_{j,k}$ are the wavelet coefficients and ψ is a wavelet with sufficient number of vanishing moments [8]. The Besov norm for the function $g \in B_{r,q}^s$ is related to a sequence space norm on its wavelet coefficients [8] and is given by:

$$\|x\|_{B_{r,q}^s} = \begin{cases} |x_{0,0}| + \left\{ \sum_{j=0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |x_{j,k}|^r \right)^{\frac{q}{r}} \right\}^{\frac{1}{q}} & \text{if } 1 \leq q < \infty \\ |x_{0,0}| + \sup_{j \geq 0} \left\{ 2^{js'} \left(\sum_{k=0}^{2^j-1} |x_{j,k}|^r \right)^{\frac{1}{r}} \right\} & \text{if } q = \infty \end{cases} \quad (4)$$

for $1 \leq r < \infty$ where $s' = (s + \frac{1}{2} - \frac{1}{r})$. s can be viewed as a regularity parameter of the image g .

This norm equivalence can also be related to the prior distribution of wavelet coefficients at each detail scale. The following theorem gives an explicit relationship between the parameters of the BKF prior model and the Besov space within which realizations of the BKF prior will fall (almost surely).

Theorem 1 Let $x_{j,k} \sim BKF(p, \sigma_j)$ with parameterization (3), where $0 < p \leq 1$ and $\sigma_j = \sigma_0 2^{-j\beta}$ (the scale invariance property of images), with $(\sigma_0 > 0, \beta \geq 0)$. Then, for a fixed $x_{0,0}$, $g \in B_{r,q}^s$ almost surely if and only if $\beta > (s + \frac{1}{2})$, for $1 \leq r < \infty$ and $1 \leq q \leq \infty$.

The proof parallels that of [9] that was also followed by [10] to prove a theorem similar to ours with the GGD prior, although our bound is tighter than theirs and is valid even for the case $q = \infty$. Actually, we can generalize this result to the more general family of scale-mixture of Gaussian priors under appropriate conditions. This will be the subject of a forthcoming paper.

5. MAP BAYESIAN DENOISER

5.1. Marginal PDF of the wavelet coefficients

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 x_{mn}^o of the true image are mutually independent random variables and independent of the noise process ϵ_{mn} . The detail coefficients x at each scale and each orientation are BKF distributed (using parameterization 2):

$$x \sim BKF(p, c) \quad (5)$$

and the probabilistic model associated with d conditionally on x is Gaussian. As far as the marginal PDF of the noisy coefficients is concerned, it was formally shown in [7] that a very accurate analytical approximation of this PDF is given by:

$$\tilde{f}_d(d; p, c, \sigma^2) = \left(\frac{\sigma^2}{2c}\right)^{\frac{p}{2}} \phi(d; 0, \sigma^2) (I_+ + I_-) \quad (6)$$

$$\text{where } I_{\pm} = e^{\left(\frac{\pm d}{\sigma} + \sigma \sqrt{\frac{2}{c}}\right)^2} D_{-p} \left(\pm \frac{d}{\sigma} + \sigma \sqrt{\frac{2}{c}}\right)$$

$\phi(d; 0, \sigma^2)$ is the Gaussian PDF with variance σ^2 . $D_\nu(x)$ stands for the Parabolic Cylinder function of fractional order ν , which is valid for $0 < p \leq 1$ and $c > 0$.

5.2. Hyperparameters estimation

In the image denoising context, one must elicit the hyperparameters estimation problem in each sub-band to implement the denoiser. 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. Thus, 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 MAD estimator [1]. Using this estimate one can easily derive the following estimates of p and c at each sub-band

from the noisy observations using their 2nd and 4th order cumulants:

$$\hat{p} = 3 \max(\hat{\kappa}_2 - \hat{\sigma}^2, 0)^2 / \hat{\kappa}_4, \hat{c} = \max(\hat{\kappa}_2 - \hat{\sigma}^2, 0) / \hat{p} \quad (7)$$

5.3. MAP term-by-term denoising

In the Bayesian framework, to obtain wavelet shrinkage estimates of the unknown image g different losses will lead to different Bayesian rules [2]. Here, We use the L_1 -based Bayes rules correspond to the MAP estimator and we derive an analytical expression for its expression. The MAP estimator of x given the noisy observation d can be easily derived as being:

$$\hat{x}_{MAP}(d) = \arg \max_x f_{x|d}(x|d) = \arg \max_x f_{d|x}(d-x) f_x(x; p, c) \quad (8)$$

The following proposition gives a general analytical expression of the MAP estimates of the wavelet coefficients x under the observation model in Eq.1, conditionally on the hyperparameters $\{\sigma^2, p, c\}$.

Proposition 1 For $0 < p \leq 1$ and c strictly positive.

- Under the observation of Eq.1, the BKF MAP estimator is given by:

$$x_{MAP}(d) = \begin{cases} 0 & |d| \leq \lambda \\ \frac{\text{sgn}(d)}{2} \left(A + \sqrt{A^2 + 4(p-1)\sigma^2} \right) & |d| > \lambda \end{cases} \quad (9)$$

$$\text{where } A = |d| - \sqrt{\frac{2}{c}} \sigma^2$$

$$\text{and } \lambda = \sqrt{2} \sigma \left(\sqrt{2(1-p)} + \frac{\sigma}{\sqrt{c}} \right)$$

- As $d \rightarrow \infty$,

$$x_{MAP}(d) = d \left(1 - \sqrt{\frac{2}{c}} \frac{\sigma^2}{|d|} + \frac{(p-1)\sigma^2}{d^2} + O(|d|^{-3}) \right) \quad (10)$$

- The BKF MAP estimator is equivalent to universal hard thresholding for $\frac{\sigma^2}{c} = \log N$ as $p \rightarrow 1$ (laplacian prior) or large N .

Fig.1 shows the Bayesian rule input-output curves obtained using the result of proposition 1. The MAP estimator is of shrinkage-type, even-symmetric and is continuous in both d and λ . It is always below the identity line and approaches it when $|d| \rightarrow \infty$ at the rate $O(|d|^{-1})$. The last result stated is that the universal threshold seems to be a quite pessimistic bound attained when the signal-to-noise ratio is low. The same conclusion was also drawn by [11] when dealing with the GGD prior.

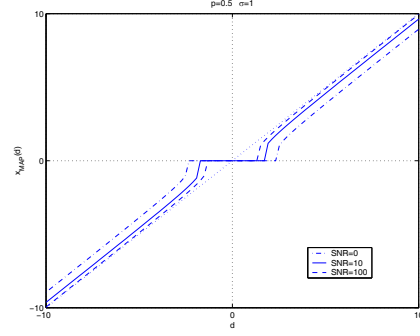


Fig. 1. The MAP Bayesian rule input-output curves ($x_{MAP}(d)$) as a function of d with different values of SNR (varying c).

6. EXPERIMENTAL RESULTS

We now assess the performance of our BKF Bayesian denoiser based on the MAP estimation by comparing it to various denoising methods. Six other denoising algorithms are considered: the universal threshold Hard and Soft thresholding, the Stein Unbiased Risk Estimator (SURE), the GGD denoiser (based on the MAP estimation) [11] and the original version of the α -stable Bayesian denoiser [4]. The Oracle threshold estimator, with the best threshold obtained from the original image, was also used as a reference.

Beside visual quality, we also calculated the PSNR (signal-to-MSE ratio) in order to quantify the achieved performance improvement. The overall performance was quantified on a digitized database of 100 test images [12]. The DWT employs Daubechies compactly-supported wavelet with regularity 4. The coarsest level of decomposition was chosen to be $(\log_2 \log N + 1)$ from asymptotic consideration [2].

Fig.2 shows the resulting images for each denoising methods for the Lena image with an input $\text{SNR}_{\text{in}} = 15\text{dB}$ ($\sigma=20$). One can see clearly that the visual quality of our BKF Bayesian denoiser is superior to the other methods but remains comparable to the GGD in this situation of good SNR.

In Fig.3, we have depicted the average PSNR over the 20 runs and the whole database (100 test images) for each denoising methods, as a function of SNR_{in} . The behavior described before is confirmed by this plot. That is, the BKF prior MAP estimator outperforms the other estimators. The difference in performance between the BKF and both the GGD and the Oracle threshold is less salient at low SNRs. The opposite is observed for the α -stable prior as the difference increases favorably for the BKF at low SNRs.

7. CONCLUSION

In this paper, a wavelet-based Bayesian denoiser based on the MAP estimation using the BKF prior was presented. We

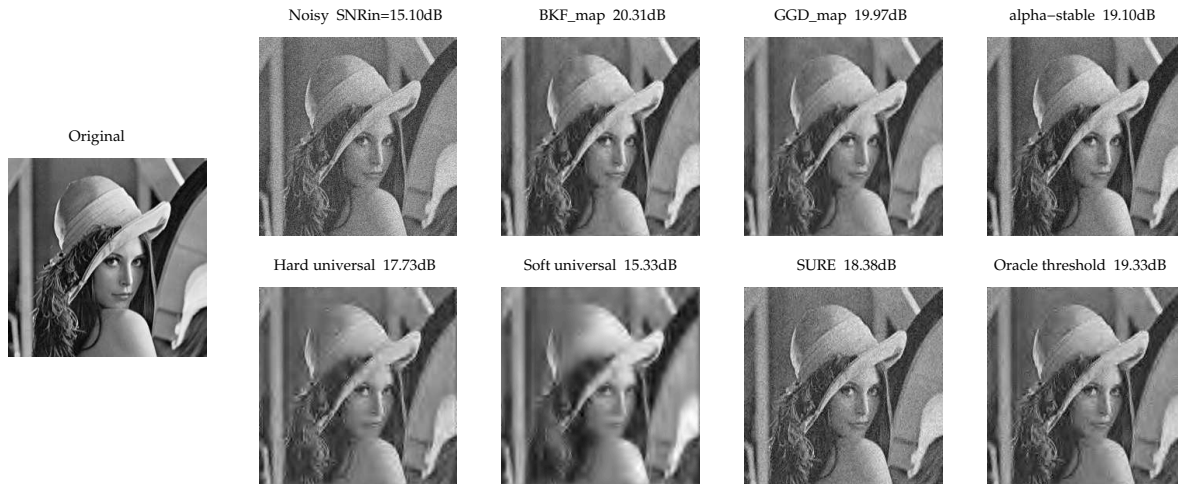


Fig. 2. Visual comparison of various denoising methods on Lena image. This image is corrupted by Gaussian noise with an input $\text{SNR}_{\text{in}} = 15\text{dB}$ ($\sigma=20$). The Bayesian MAP denoiser with the BKF prior is clearly superior to the other methods.

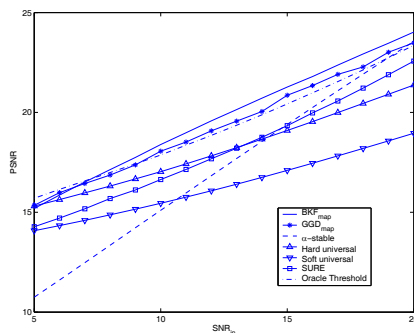


Fig. 3. Average PSNR over the 20 runs and the 100 image database for each denoising methods, as a function of the SNR_{in} .

have shown that, under suitable conditions, the BKF prior is well adapted to characterize functions in the Besov spaces. Experimental results on a large database have shown the superiority of our Bayesian denoiser compared to the other denoising approaches. This suggests that the BKF prior is an accurate model adapted to capture the sparseness behavior of the wavelet coefficients for a large class of images, which confers to the corresponding MAP estimator good denoising properties. Our effort are now directed towards extension of these Bayesian models to translation-invariant and directional transforms such as curvelets.

8. REFERENCES

- [1] David L. Donoho and Iain M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.
- [2] A. Antoniadis and J. Bigot and T. Sapatinas, “Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study”, *Journal of Statistical Software*, vol. 6, no. 6, 2001.
- [3] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, *IEEE Trans. PAMI*, vol. 11, no. 7, pp. 674-693, 1989.
- [4] A. Achim, A. Bezerianos and P. Tsakalides, “Novel Bayesian multiscale method for speckle removal in medical ultrasound images”, *IEEE Trans. Med. Imag.*, vol. 20, pp. 772-783, 2001.
- [5] U. Grenander and A. Srivastava, “Probability models for clutter in natural images”, *IEEE PAMI*, vol. 23, no. 4, pp. 424-429, 2001.
- [6] I.S. Gradshteyn and I.M. Ryzhik, “Table of integrals, series and products”, *Academic Press, A. Jeffrey Edition*, 1980.
- [7] J. Fadili and L. Boubchir, “Analytical Form for a Bayesian Wavelet Estimator of Images Using The Bessel K forms Densities”, *IEEE Trans. Image Processing*, vol. 14, no. 2, pp. 231-240, February 2005.
- [8] Yves Meyer, “Wavelets and operators”, *Cambridge University Press, English Edition*, 1992.
- [9] F. Abramovich, T. Sapatinas and Bernard Silverman, “Wavelet thresholding via a Bayesian approach”, *J. R. Statist. Soc. B*, vol. 60, pp. 725-749, 1998.
- [10] H. Choi and R. Baraniuk, “Wavelet statistical models and Besov spaces”, *SPIE Technical Conference on Wavelet Applications in Signal Processing*, Denver, CO, July 1999.
- [11] P. Moulin and J. Liu, “Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors”, *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 909-919, 1999.
- [12] “<http://sipi.usc.edu/services/database/Database.html>”.