Bayesian Denoising in the Wavelet-domain Using an Analytical Approximate α-stable prior

Larbi Boubchir, Jalal M. Fadili and Daniel Bloyet
GREYC UMR CNRS 6072
ENSICAEN, 6 Bd du Maréchal Juin 14050 Caen, France
E-mail: {boubchir,jfadili,dbloyet}@greyc.ismra.fr

Abstract

A nonparametric Bayesian estimator in the wavelet domain is presented. In this approach, we propose a prior model based on the α-stable densities to capture the sparseness of the wavelet coefficients. An attempt to apply this model image wavelet-denoising have been already proposed in [2]. However, despite its efficacy in modeling the heavy-tail behaviour of the empirical detail coefficients densities, their denoiser proves very poor in practice and suffers from many drawbacks such as the weakness of the hyperparameters estimator associated with the α-stable prior. Here, we propose to overcome these limitations using the scale-mixture of Gaussians as an analytical approximation for α-stable densities. Exploiting this prior, we design a Bayesian L_2-loss nonlinear denoiser.

1. Introduction

Since the seminal papers by Donoho & Johnstone [4], 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. 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. Then the image is estimated by applying a suitable Bayesian rule to the resulting posterior distribution of the wavelet coefficients.

Various prior choices can be found in the statistical literature, see [3] for a detailed review. Popular priors in the image processing community are the GGD and the α-stable priors [6, 2]. However, in these cases, the derived Bayesian estimator has no closed analytical form in general situation. This involves intensive numerical integration which is numerically unstable time-consuming. Recently, in [5], the Bessel K forms (BKF) family has been proposed with its closed-form Bayesian shrinker.

Here, we propose a prior statistical model based on the α-stable densities. The authors in [2] have already shown 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. The Bayesian denoiser derived by [2] 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. We here propose to overcome these limitations using the scale-mixture of Gaussians as a fast and numerically stable analytical approximation for α-stable densities. Additionally, we propose an approximate maximum likelihood estimator for the hyperparameters. When applied to DWT 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.

2. Nonparametric regression

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}. 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}. The DWT of white noise are also independent normal variables with the same variance. It follows from Eq.1 that:

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

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

where a_{mn} (resp. c_{mn}) is the approximation coefficient of the true image g (resp. y) at location (m, n). It is advisable to keep the approximation coefficients intact because they represent low-frequency terms that usually contain important features about the image g.

3. α-stable distributions family

Let X be an α-stable random variable (RV), $X \sim S_\alpha(\beta, \mu, \gamma)$, $0 < \alpha \leq 2$, $\gamma \geq 0$, $-1 \leq \beta \leq 1$ and $\mu \in \mathbb{R}$. X can be uniquely defined by its characteristic function [8]. α controls the heaviness of the tails of the PDF. β is the symmetry index ($\beta = 0$ then X is a symmetric α-stable ($S\alpha S$) RV). μ is the location parameter. $\gamma = \sigma^\alpha$ is the scale parameter. The wavelet detail coefficients densities have been already observed to be sharply peaked and heavily tailed [6]. This is exactly the property which is captured by an $S\alpha S$ distribution where $\beta = 0$ and $\mu = 0$, i.e. $s^{oj}_{m,n} \sim S_\alpha(0, 0, \gamma)$.

4. Analytical approximation of $S\alpha S$ 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.

4.1. The scale-mixture of Gaussians

The concept of mixture is based on a corollary of the mixture theorem of α-stable RVs:

Corollary 1 (Scale-Mixture of Gaussians)
Let $X \sim \mathcal{N}(0, 2\gamma_x)$. Let Y be a positive stable random variable, $Y = V^2 \sim S_{\frac{\alpha_z}{2}}(-1, 0, (\cos(\frac{\pi\alpha_z}{4}))^{\frac{2}{\alpha_z}})$ and independent form X. Then:

$$Z = Y^{\frac{1}{2}} X \sim S_{\alpha_z}(0, 0, \gamma_x)$$
$$f_Z(z) = \frac{1}{\sqrt{4\pi\gamma_x}} \int_{-\infty}^{+\infty} \exp\left(-\frac{z^2}{4\gamma_x v^2}\right) h(v) v^{-1} dv \quad (3)$$

Sampling $f_Z(z)$ at N points allows to obtain a mixture approximation to any $S\alpha S$ PDF:

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

This analytical expression for the $S\alpha S$ PDF is only an approximation, since the the continuous integral was approximated by a finite sum. However, to reduce the the complexity of the model in Eq.4 and get fast but good enough approximation, one might prefer to use only a small number of components and to sample Eq.3 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. However, using an Minimum Description Length criterion, only a few mixture components (typically 4 to 8) are necessary to negotiate a good compromise between the approximation quality, the model complexity and the calculation time.

4.2. The $S\alpha S$ PDF approximation algorithm

We describe an algorithm that fits a $S\alpha S$ PDF to observed samples $\{z_m\}_{m=1,\dots,M}$ using the scale-mixture of Gaussians approximation to $S\alpha S$ PDFs. This algorithm follows the next steps:

1. Generate the characteristic function of the mixing PDF which is positive stable distributed with parameters: $(\frac{\alpha}{2}, \beta = -1, \mu = 0, \gamma = (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}})$.
2. Evaluate the positive stable PDF f_Z at N (number of components in the mixture) equally spaced points taking the inverse FFT of the characteristic function.
3. Substitute the mixing function samples $h(v_i) = 2v_i f_Y(v_i^2)$ allows to obtain the analytical approximation for the $S\alpha S$ PDF:
$$p_{\alpha,\gamma}(z_m) = \frac{\sum_{j=1}^N \phi(z_m; 0, 2\gamma v_j^2) v_j f_Z(v_j^2)}{\sum_{j=1}^N v_j f_Z(v_j^2)} \quad (5)$$
where $\phi(z; \mu, \delta^2)$ is the Gaussian PDF with mean μ and variance δ^2.
4. Use the EM algorithm 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 estimates such that:
$$p_{\alpha,\gamma}(z_m) = \sum_{j=1}^N P(z_m|j) P_j \quad (6)$$
where P_j are the mixing proportions and $P(z_m|j) = \phi(z_m; 0, \sigma_j^2)$.

The algorithm is presented in the following iterative form:

- Initialization:
$$P_{old,j} = \frac{h(v_j)}{\sum_{j=0}^N h(v_j)} \quad (7)$$
and
$$P_{old}(z_m|j) = \phi(z_m; 0, 2\gamma v_j^2) \quad (8)$$
- Repeat until convergence:
E step:
$$P_{new}(j|z_m) = \frac{P_{old}(z_m|j) P_{old,j}}{\sum_m P_{old}(z_m|j) P_{old,j}} \quad (9)$$
M step:
$$\sigma^2_{new,j} = \frac{\sum_m P_{old}(j|z_m) z_m^2}{\sum_m P_{old}(j|z_m)} \quad (10)$$

$$P_{new,j} = \frac{1}{M} \sum_m P_{old}(j|z_m) \tag{11}$$

5. Bayesian denoiser

5.1. Marginal PDF of the observed wavelet coefficients

In the Bayesian approach, a prior is imposed on the wavelet coefficients designed to describe their distribution. The detail coefficients s at each scale and each orientation are $S\alpha S$ distributed:

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

and the probabilistic model associated to d conditionally on s is Gaussian. Using the Bayes rule, the analytical approximation of the marginal PDF of d is given by:

$$p(d) = \frac{1}{\sqrt{2\pi}} \sum_j P_j (\sigma_j^2 + \sigma_\epsilon^2)^{-\frac{1}{2}} \exp\left(-\frac{d^2}{2(\sigma_j^2 + \sigma_\epsilon^2)}\right) \tag{13}$$

where σ_ϵ is the level noise.

5.2. The hyperparameters estimation

In the image denoising context, one must elicit the hyperparameters (θ) estimation problem in each subband. To implement the formula in Eq.13, 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 $S\alpha S$ RVs especially in the presence of contaminating noise.

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. The estimation of these parameters is only useful for initialization, and final estimates are given by the EM algorithm. Therefore, we have chosen the quantile-based estimator of McCulloch [7] assuming that for reasonable SNRs, the tails of the marginal distribution $p(d)$ are not very sensitive to the presence of noise. Again, the level noise σ_ϵ is estimated from the HH orientation at the finer scale:

$$\hat{\sigma}_\epsilon = \frac{\mathrm{MAD}(d_{mn}^{HH_1})}{0.6745} \tag{14}$$

We now illustrate some prior estimation results. Fig.1 show the estimated and the observed densities of the wavelet detail coefficients for the Barbara image on log scale. The observed histogram (-•-) was fitted using the scale-mixture α-stable algorithm with 8 Gaussians as we described above (solid). For comparison purposes, we also depict the fit given by the original α-stable as proposed by [2] (dash-dotted), the BKF (dashed) and the GGD (dotted) models. From these results, we can legitimately claim that the Gaussian scale-mixture $S\alpha S$ density fits the observed wavelet detail coefficients very well. It generally outperforms the GGD and the BKF models.

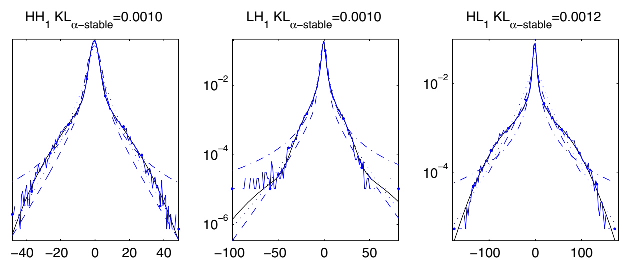

Figure 1. Estimated and observed marginal densities of the observed wavelet detail coefficients for the Barbara image.

5.3. Bayesian term-by-term denoising

Using the approximate prior PDF, it is easy to show that the L_2-based Bayes rules, which correspond to posterior conditional mean (PCM) estimate is (conditionally on the hyperparameters):

$$s_{\mathrm{PCM}}(d|\theta) = \frac{\sum_j P_j \frac{d\,\sigma_j^2}{\sigma_j^2+\sigma_\epsilon^2} \phi(d; 0, \sigma_j^2 + \sigma_\epsilon^2)}{\sum_j P_j \phi(d; 0, \sigma_j^2 + \sigma_\epsilon^2)} \tag{15}$$

This equation shows that the $S\alpha S$ 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.

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 to other previously published denoising methods. Six other denoising algorithms are considered: the universal threshold Hard and Soft thresholding, the Stein Unbiased Risk Estimator (SURE), the Oracle threshold estimator (Oracle), the Bessel K forms (BKF) Bayesian denoiser [5], original version of the α-stable Bayesian denoiser [2]. In the latter, no closed-form is available for the PCM Bayesian denoiser. We here used an equivalent form involving Fourier integrals. The Fourier integrals were numerically implemented using FFT-based methods. The overall performance was quantified on a digitized database of 100 test images [1]. 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 [3] where N is the size of the image.

Fig.2 shows a zoom on a textured area of the estimated images for each denoising methods for the Barbara image with an input $\mathrm{SNR_{in}} = 15\mathrm{dB}$. 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. Our denoiser ensures an excellent 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 [2]. Furthermore, our denoiser is faster and very stable numerically.

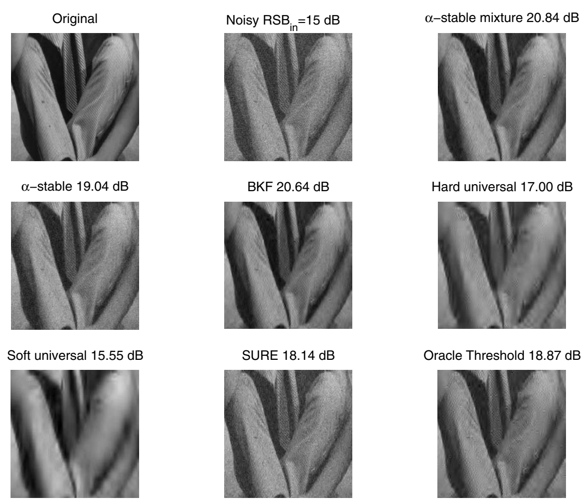

Figure 2. Visual comparison of various denoising methods on Barbara image.

In Fig.3, we have depicted the average PSNR over the 50 runs and the whole database (100 images) for each denoising methods, as a function of $\mathrm{SNR_{in}}$. One can notice that the "α-stable mixture" denoiser outperforms most of the methods, but is 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 $\mathrm{SNR_{in}}$. The main reason is the weakness of the hyperparameters estimator which remains an important issue. These findings suggest 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. Conclusion

In this paper, a nonlinear nonparametric Bayesian estimator in the Wavelet domain was presented. An approximation to $S\alpha S$ 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. 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. Our efforts are now directed towards extension of these Bayesian models to directional transforms such as curvelets or bandelets.

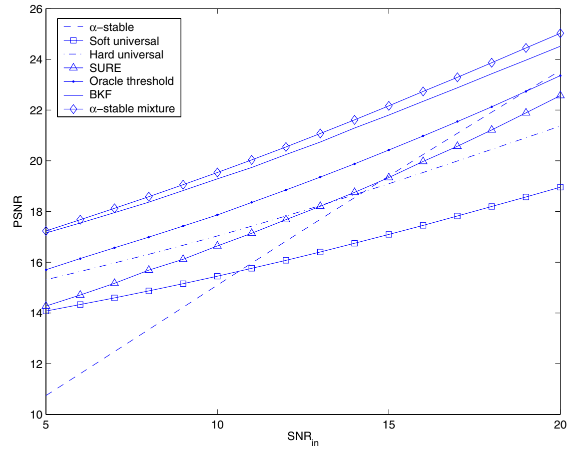

Figure 3. Average PSNR over the 50 runs and the 100 images database.

References

[1] http://sipi.usc.edu/services/database/database.html.

[2] A. Achim, A. Bezerianos, and P. Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imag.*, 20:772–783, 2001.

[3] A. Antoniadis, J. Bigot, and T. Sapatinas. Wavelet estimators in nonparametric regression: A comparative simulation study. *Journal of Statistical Software*, 6(6), 2001.

[4] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[5] J. M. Fadili and L. Boubchir. Analytical form for a bayesian wavelet estimator of images using the bassel k forms densities. *IEEE Trans. Image Processing*, 2003 (in press).

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

[7] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communication on Statistics-Simulation*, 15(4):1109–1136, 1986.

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