WAVELET THRESHOLDING FOR A WIDE CLASS OF NOISE DISTRIBUTIONS

D. Leporini and J.-C. Pesquet
Laboratoire des Signaux et Systèmes,
CNRS/UPS and GDR ISIS, ESE, Gif-sur-Yvette Cedex, France
e-mail: `leporini@lss.supelec.fr`

ABSTRACT

Wavelet thresholding techniques are becoming popular in the signal processing community for denoising applications. Near-minimax properties were in particular established for simple threshold estimates over wide classes of regular functions. In this paper, we establish close connections between wavelet thresholding techniques and MAP estimation using exponential power prior distributions for a wide class of noise distributions, including heavy-tailed noises. We subsequently prove that a great variety of estimators are derived from a MAP criterion. A simulation example is presented to substantiate the proposed approach.

1 Introduction

There has recently been a great research interest in wavelet thresholding techniques for signal and image denoising applications [2, 7, 10]. In [3], near-minimax properties of threshold estimates (namely hard and soft thresholding) were established over classes of smooth function spaces (Besov and Triebel bodies). The model generally adopted for the observed process $\boldsymbol{y} \in \mathbb{R}^K$ is the following:

$$y(k) = x(k) + n(k), \quad k \in \{1, \ldots, K\},$$

where $\boldsymbol{n}$ is often assumed to be i.i.d. Gaussian with zero mean and finite variance σ^2. Estimation of the underlying unknown signal $\boldsymbol{x}$ is of interest. Let $\boldsymbol{W}_x$, $\boldsymbol{W}_y$ and $\boldsymbol{W}_n$ denote respectively the vector of wavelet coefficients of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{n}$. The components of these vectors are such that:

$$W_y(k) = W_x(k) + W_n(k), \quad k \in \{1, \ldots, K\}.$$

Thresholding techniques successfully utilize the unitary transform property of the wavelet decomposition to statistically distinguish the signal components $\boldsymbol{W}_x$ from those of the noise $\boldsymbol{W}_n$. We recall that the hard and soft threshold estimates of the wavelet coefficients $\boldsymbol{W}_x$ are respectively obtained according to:

$$\begin{aligned}\widehat{W}_x^{\text{hard}}(k) &= W_y(k)\mathbb{I}_{|W_y(k)|>\chi}, \\ \widehat{W}_x^{\text{soft}}(k) &= \operatorname{sign}(W_y(k))\max(0, |W_y(k)| - \chi),\end{aligned}$$

where $\chi > 0$ denotes the threshold value and $\mathbb{I}_A$ is the usual indicator function on a set A. The problem typically encountered with such approaches first concerns the choice of a thresholding policy and, subsequently, of the threshold value [5]. In the sequel, we exhibit close connections between wavelet thresholding and Maximum *A Posteriori* (MAP) estimation using exponential power prior distributions. One of the main advantage of this approach is to naturally provide a thresholding rule, and consequently, a thresholding value adapted to the signal/noise under study.

2 Connections between MAP estimation and thresholding rules

We choose to model the wavelet coefficients $\boldsymbol{W}_x$ by level-dependent distributions from the family of Exponential Power Distribution $\mathcal{EPD}(\alpha, \beta)$:

$$f_{\alpha,\beta}(\cdot) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|\cdot|/\alpha)^\beta}, \quad (\alpha, \beta) \in \mathbb{R}_+^{*2}.$$

Note that the $\mathcal{EPD}(\alpha, \beta)$ model was first proposed by Mallat [4] for wavelet coefficient representation of signals and images, and subsequently applied to image coding [1], and image denoising using *a posteriori* mean estimates [10]. Such estimates are however unable to provide thresholding rules. For the sake of simplicity, we further assume in the sequel that the wavelet coefficients $\boldsymbol{W}_x$ are independent (which does not imply that the signal itself is i.i.d.). We proceed to determine the MAP estimate of $W_x(k)$:

$$\widehat{W}_x(k) = \arg\min_{\omega} \mathcal{F}_{\alpha,\beta}(\omega), \tag{1}$$

with

$$\mathcal{F}_{\alpha,\beta}(\omega) = \frac{(W_y(k) - \omega)^2}{2} + \frac{\sigma^2}{\alpha^\beta}|\omega|^\beta. \tag{2}$$

As expected, we prove hereafter that the parameter β determines the nature of MAP estimates, and establish close connections with wavelet thresholding techniques. This result actually appears as a particular case of nonsmooth regularization for local strong homogeneity recovery recently introduced in [6]. In the case of wavelet decompositions, this regularization approach may be interpreted as the possibility to obtain large zones with zero coefficients. We now present one of our main results.

Proposition 1 *The MAP estimation with $\beta \in (0,1]$ leads to thresholding rules in the sense that:*

$$\widehat{W}_x(k) = 0 \quad \textit{iff} \quad |W_y(k)| < \chi_{\alpha,\beta},$$

where the threshold value $\chi_{\alpha,\beta}$ may be analytically derived according to:

$$\chi_{\alpha,\beta} = \frac{2-\beta}{2(1-\beta)} \left(\frac{2\sigma^2(1-\beta)}{\alpha^\beta} \right)^{1/(2-\beta)}. \qquad (3)$$

Moreover, we also obtain:

$$\lim_{|W_y(k)|\to\infty} \widehat{W}_x(k) = W_y(k),$$

$$\lim_{\substack{W_y(k)\to\pm\chi_{\alpha,\beta} \\ |W_y(k)|>\chi_{\alpha,\beta}}} \widehat{W}_x(k) = \pm \left(\frac{2\sigma^2(1-\beta)}{\alpha^\beta} \right)^{1/(2-\beta)}.$$

Proof. The first derivative of $\mathcal{F}_{\alpha,\beta}$ reads:

$$\mathcal{F}'_{\alpha,\beta}(\omega) = \omega - W_y(k) + \frac{\beta\sigma^2}{\alpha^\beta}\text{sign}(\omega)|\omega|^{\beta-1}.$$

By checking the sign of $\mathcal{F}'_{\alpha,\beta}$, we straightforwardly see that any minimizer of (2) belongs to the interval $[\min(0, W_y(k)), \max(0, W_y(k))]$, so that MAP estimation indeed corresponds to a shrinkage method. For simplicity, we only consider the case where $W_y(k) \geq 0$ (so that ω_0 satisfying $\mathcal{F}'_{\alpha,\beta}(\omega_0) = 0$, if defined, is also positive). We first derive, after some algebra, that:

$$\exists\omega_0 \mid \mathcal{F}'_{\alpha,\beta}(\omega_0) = 0 \iff W_y(k) \geq \tilde{\omega}_0 + \frac{\beta\sigma^2}{\alpha^\beta}\tilde{\omega}_0^{\beta-1},$$

with

$$\tilde{\omega}_0 = \left(\frac{\beta(1-\beta)\sigma^2}{\alpha^\beta} \right)^{1/(2-\beta)}.$$

Note that this result only provides a necessary condition (*i.e.* a lower bound on the threshold value) since the function $\mathcal{F}_{\alpha,\beta}$ is not differentiable at 0. We consequently must find the value $\tilde{\omega}_1 > \tilde{\omega}_0$ satisfying:

$$\mathcal{F}'_{\alpha,\beta}(\tilde{\omega}_1) = 0 \iff W_y(k) = \tilde{\omega}_1 + \frac{\beta\sigma^2}{\alpha^\beta}\tilde{\omega}_1^{\beta-1},$$

and

$$\mathcal{F}_{\alpha,\beta}(\tilde{\omega}_1) = \mathcal{F}_{\alpha,\beta}(0).$$

It may be proved that $\tilde{\omega}_1$ may be analytically derived according to:

$$\tilde{\omega}_1 = \left(\frac{2\sigma^2(1-\beta)}{\alpha^\beta} \right)^{1/(2-\beta)},$$

which finally provides the expected threshold value:

$$\chi_{\alpha,\beta} = \tilde{\omega}_1 + \frac{\beta\sigma^2}{\alpha^\beta}\tilde{\omega}_1^{\beta-1}.$$

□

This property indicates that Exponential Power Distributions provides a great variety of thresholding policies based on MAP estimates for $\beta \in (0,1]$. In particular, the MAP estimate obtained with $\beta = 1$ corresponds to a soft thresholding of the wavelet coefficients where the threshold value is given by σ^2/α (obtained when $\beta \to 1$ in (3)). Note that this result is not surprising as the distribution $f_{\alpha,1}(\cdot)$ is a Double Exponential $\mathcal{DE}(1/\alpha)$ law. For concreteness we also present the case where $\beta = 1/2$.

Proposition 2 *The MAP estimate corresponding to $\beta = 1/2$ is given by:*

$$\widehat{W}_x(k) = \begin{cases} 0 & \textit{if } |W_y(k)| < \frac{3}{2}\left(\frac{\sigma^2}{\sqrt{\alpha}}\right)^{2/3}, \\ F_{1/2}(W_y(k)) & \textit{otherwise,} \end{cases}$$

where the function $F_{1/2}(\cdot)$ is defined by:

$$F_{1/2}(W_y(k)) = \frac{4}{3}\text{sign}(W_y(k))|W_y(k)| \cos\left(\frac{\pi - \arccos(\sigma^2/4\sqrt{\alpha}(|W_y(k)|/3)^{3/2})}{3} \right).$$

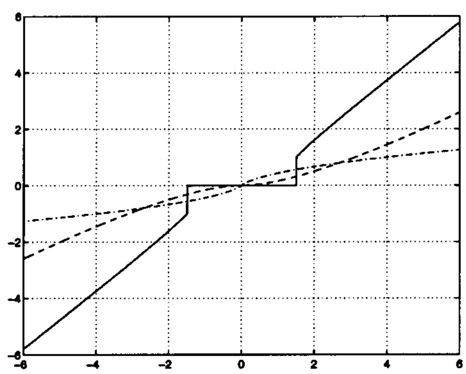

Figure 1: MAP estimators corresponding to $\beta = 1/2$ (solid line), $\beta = 3/2$ (dashed line), and $\beta = 3$ (dashdotted line).

The function $F_{1/2}(\cdot)$ is presented in Fig. 1 for $\alpha = \sigma = 1$. Note that, in the general case, the analytic expression of MAP estimates cannot be derived for $|W_y(k)| > \chi_{\alpha,\beta}$, even if we previously showed that it is asymptotically linear. We now consider the case where $\beta > 1$.

Proposition 3 *The MAP estimates with $\beta > 1$ correspond to differentiable increasing functions. Moreover, when $|W_y(k)|$ tends to infinity, we find that:*

$$\widehat{W}_x(k) \sim W_y(k) \textit{ for } \beta \in (1,2),$$

$$\widehat{W}_x(k) \sim \frac{\alpha^\beta}{\beta\sigma^2} W_y(k)^{1/(\beta-1)} \textit{ for } \beta > 2.$$

Actually, it may be proved that the previously obtained results are a particular case of a more general property involving non-Gaussian noises:

Proposition 4 *Under the following assumptions*

$$\begin{aligned} W_x(k) &\sim \mathcal{EPD}(\alpha_1, \beta_1) \\ W_n(k) &\sim \mathcal{EPD}(\alpha_2, \beta_2), \end{aligned}$$

the MAP estimation with $\beta_1 \in (0,1]$ and $\beta_2 > \beta_1$ corresponds to thresholding rules. In particular, hard thresholding rules are obtained when $\beta_2 \leq 1$.

Proof. We first examine the case $\beta_2 \leq 1$, and again assume that $W_y(k) \geq 0$. It may easily be proved that the function to be minimized is concave in the considered interval $[0, W_y(k)]$, so that the minimum is reached either at 0 or at $W_y(k)$, and therefore leads to a hard thresholding. The threshold value is obtained by using the inequality $\mathcal{F}_{\alpha_1,\beta_1,\alpha_2,\beta_2}(0) \leq \mathcal{F}_{\alpha_1,\beta_1,\alpha_2,\beta_2}(W_y(k))$, which leads to:

$$\chi_{\alpha_1,\beta_1,\alpha_2,\beta_2} = \left(\frac{\alpha_2^{\beta_2}}{\alpha_1^{\beta_1}}\right)^{1/(\beta_2-\beta_1)}.$$

When $\beta_2 > 1$, we proceed to calculate $\mathcal{F}'_{\alpha_1,\beta_1,\alpha_2,\beta_2}$ and derive, similarly to the proof of Proposition 1:

$$\exists \omega_0 \mid \mathcal{F}'_{\alpha_1,\beta_1,\alpha_2,\beta_2}(\omega_0) = 0 \Longleftrightarrow W_y(k) \geq \tilde{\theta}_0,$$

with

$$\tilde{\theta}_0 = \left(\frac{\beta_1 \alpha_2^{\beta_2}}{\beta_2 \alpha_1^{\beta_1}}\right)^{1/(\beta_2-\beta_1)} \left\{ \left(\frac{1-\beta_1}{\beta_2-1}\right)^{(\beta_2-1)/(\beta_2-\beta_1)} + \left(\frac{1-\beta_1}{\beta_2-1}\right)^{(\beta_1-1)/(\beta_2-\beta_1)} \right\}.$$

As previously stated, this value only provides a lower bound on the threshold value, which however cannot be analytically derived in the general case. □

Note that this property may be of interest in dealing with heavy-tailed noise distributions. We extend this idea by considering another class of noise distributions:

Proposition 5 *Under the following assumptions*

$$\begin{aligned} W_x(k) &\sim \mathcal{EPD}(\alpha_1, \beta) \\ W_n(k) &\sim \mathcal{C}(0, 1/\alpha_2), \end{aligned}$$

where $\mathcal{C}(0, 1/\alpha_2)$ denotes the centred Cauchy distribution, the MAP estimation with $\beta \in (0,1)$ or $\beta = 1$ and $\alpha_1\alpha_2 > 1$ leads to double thresholding rules in the sense that

$$\exists (\chi^1_{\alpha_1,\alpha_2,\beta}, \chi^2_{\alpha_1,\alpha_2,\beta}) \in \mathbb{R}_+^2 \mid \widehat{W}_x(k) \neq 0 \Rightarrow \chi^1_{\alpha_1,\alpha_2,\beta} < |W_y(k)| < \chi^2_{\alpha 1,\alpha 2,\beta}.$$

Interestingly, these estimates are closely related to the constrained minimax thresholding introduced in [8] where the boundedness of the signal coefficients is however assumed. For illustration, we focus on the case where $\beta = 1$. In this case, the nonconvex function to be minimized may be written:

$$\omega \longmapsto \log\left(1 + \alpha_2^2 (W_y(k) - \omega)^2\right) + \frac{|\omega|}{\alpha_1}.$$

Proposition 6 *Provided that $\alpha_1\alpha_2 \geq 1$, the minimization of the previous expression leads to the double (soft) thresholding rule defined by:*

$$\chi^1_{\alpha_1,\alpha_2,1} = \alpha_1 - \frac{\sqrt{\alpha_1^2\alpha_2^2 - 1}}{\alpha_2},$$

while $\chi^2_{\alpha_1,\alpha_2,1} > \chi^1_{\alpha_1,\alpha_2,1}$ satisfies:

$$\log\left(1 + \alpha_2^2 (\chi^2_{\alpha_1,\alpha_2,1})^2\right) = \log\left(1 + \alpha_2^2 (\chi^1_{\alpha_1,\alpha_2,1})^2\right) + \frac{\chi^2_{\alpha_1,\alpha_2,1} - \chi^1_{\alpha_1,\alpha_2,1}}{\alpha_1}$$

Proof. Defining

$$\mathcal{F}_{\alpha_1,\alpha_2}(\omega) = \log\left(1 + \alpha_2^2 (W_y(k) - \omega)^2\right) + \frac{|\omega|}{\alpha_1},$$

we again consider the case where $W_y(k) \geq 0$. The derivative $\mathcal{F}'_{\alpha 1,\alpha_2}$ is then given by:

$$\mathcal{F}'_{\alpha_1,\alpha_2}(\omega) = \frac{1}{\alpha_1} + \frac{2\alpha_2^2(\omega - W_y(k))}{1 + \alpha_2^2(\omega - W_y(k))^2}.$$

It is straightforward to show that the existence of a minimum ω_0 satisfying $\mathcal{F}'_{\alpha_1,\alpha_2}(\omega_0) = 0$ is equivalent to:

$$W_y(k) \geq \tilde{\omega}_0 = \alpha_1 - \frac{1}{\alpha_2}\sqrt{\alpha_1^2\alpha_2^2 - 1}.$$

Moreover, this minimum is simply given by:

$$\omega_0 = W_y(k) - \tilde{\omega}_0.$$

We must then study the sign of $\mathcal{F}_{\alpha_1,\alpha_2}(0) - \mathcal{F}_{\alpha_1,\alpha_2}(\omega_0)$. It may be proved after some algebra that:

$$\begin{cases} \mathcal{F}_{\alpha_1,\alpha_2}(0) > \mathcal{F}_{\alpha_1,\alpha_2}(\omega_0) & \text{if } \tilde{\omega}_0 < W_y(k) < \chi^2_{\alpha_1,\alpha_2,1}, \\ \mathcal{F}_{\alpha_1,\alpha_2}(0) < \mathcal{F}_{\alpha_1,\alpha_2}(\omega_0) & \text{if } W_y(k) > \chi^2_{\alpha_1,\alpha_2,1}, \end{cases}$$

where the threshold value $\chi^2_{\alpha_1,\alpha_2,1}$ is as previously defined by an implicit equation. Finally, the double (soft) thresholding MAP estimate then reads:

$$\widehat{W}_x(k) = \begin{cases} \operatorname{sign}(W_y(k)) \max(0, |W_y(k)| - \tilde{\omega}_0) & \text{if } |W_y(k)| < \chi^2_{\alpha_1,\alpha_2,1}, \\ 0 & \text{otherwise.} \end{cases}$$

Note finally that the case $\alpha_1\alpha_2 < 1$ directly leads to $\widehat{W}_x(k) = 0$. □

In the sequel, we apply this original soft thresholding rule to a denoising application based on this latter statistical model.

3 Parameter estimation

Let $\boldsymbol{\theta} = [(\alpha_1^j)_{0 \leq j \leq J}, \alpha_2]$, where α_1^j stands for the dispersion parameter of the signal components at resolution level j, denote the parameter vector of the model. Note that the noise decomposition coefficients $\boldsymbol{W}_n$ are assumed to be independently distributed according to $\mathcal{C}(0, 1/\alpha_2)$. In order to estimate the model parameters, we propose to use an underlying signal estimate based on median filtering. Parameter estimates are subsequently obtained from the wavelet expansion of this signal. More precisely, we first derive the level-dependent dispersion parameters $(\alpha_1^j)_{0 \leq j \leq J}$ using a maximum likelihood approach, while the parameter α_2 is estimated using the method of sample characteristic function [9], by using the fact that:

$$1/\alpha_2 = -\log|\varphi(1)|,$$

where $\varphi(t)$ corresponds to the characteristic function of the noise.

4 A denoising example

To illustrate the previous results, we consider a denoising application involving Cauchy noise distributions, which corresponds to a particular case of α-stable processes (with $\alpha = 1$). We recall that, in this case, the wavelet coefficients $\boldsymbol{W}_n$ also correspond to a Cauchy process.

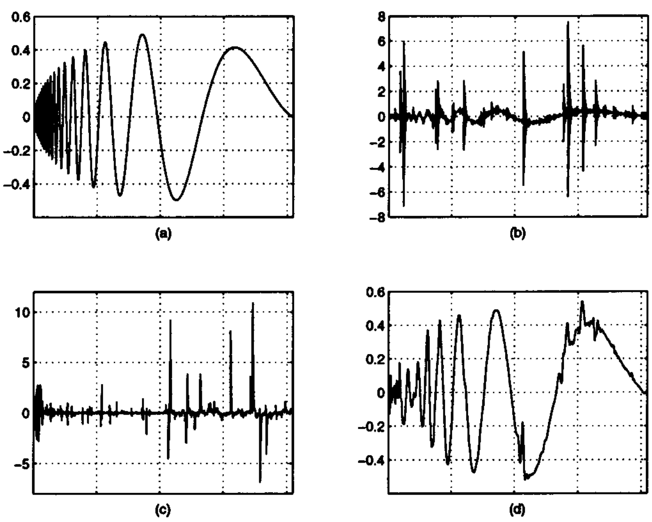

Figure 2: (a) original signal, (b) and (c) noisy signal and its wavelet decomposition respectively, and (d) reconstructed process using MAP estimation.

The original process is presented in Fig. 1 and corresponds to the classical Doppler signal from the Stanford database. The noisy process is obtained with a correlated $\mathcal{C}(0, 10^{-2})$ noise, and is also displayed in Fig. 1, along with its wavelet expansion. As expected, the estimated process using the MAP criterion suppresses the severe outliers generated by the heavy-tailed distribution.

5 Conclusion

In this paper, we have established close connections between MAP estimation (or equivalently wavelet regularization), and wavelet thresholding techniques using exponential power prior distributions. In particular, the crucial role of the exponent parameter β was exhibited, and original hard and soft thresholding rules were obtained in the case of non-Gaussian noise distributions. An application to heavy-tailed noises was finally presented to substantiate the proposed approach.

References

[1] R. W. Buccigrossi and E. P. Simoncelli. Progressive wavelet image coding based on a conditional probability model. In *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, Munich, Germany, April 21-24 1997.

[2] D. L. Donoho. Denoising by Soft-Thresholding. *IEEE Trans. Informat. Theory*, 41:613–627, May 1995.

[3] D. L. Donoho and I. M. Johnstone. Minimax Estimation via Wavelet Shrinkage. Technical Report, Stanford University, 1994.

[4] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, PAMI-11:674–693, Jul. 1989.

[5] G. P. Nason. Choice of the Threshold Parameter in Wavelet Function Estimation. In A. Antoniadis, editor, *Wavelets and statistics*. Lecture Notes in Statistics, Springer Verlag, 1995.

[6] M. Nikolova. Estimées localement fortement homogènes. *C. R. Acad. Sci. Paris*, Série 1, 325:665–670, 1997.

[7] J.-C. Pesquet, H. Krim, D. Leporini, and E. Hamman. Bayesian Approach to Best Basis Selection. In *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, pages 2634–2638, Atlanta, USA, May 7-9 1996.

[8] I. C. Schick and H. Krim. Robust Wavelet Thresholding for Noise Suppression. In *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, pages 3421–3424, Munich, Germany, April 21-24 1997.

[9] M. Shao and C. L. Nikias. Signal Processing with Fractional Lower-Order Moments: Stable Processes and Their Applications. *Proc. IEEE*, 81:984–1009, July 1993.

[10] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Proc. IEEE Conf. on Image Processing*, Lausanne, Switzerland, Sept. 1996.