16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SPARSITY FROM BINARY HYPOTHESIS TESTING AND APPLICATION TO NON-PARAMETRIC ESTIMATION

Dominique Pastor and Abdourrahmane M. Atto

Institut TELECOM, TELECOM Bretagne
LAB-STICC, CNRS UMR 3192
Technopôle Brest-Iroise, CS 83818
29238 Brest Cedex 3, FRANCE
email: {dominique.pastor, am.atto}@telecom-bretagne.eu
phone: + (33) 2 29 00 14 87

ABSTRACT

This paper presents and discusses an alternative notion of sparsity. This notion derives from a theoretical result in binary hypothesis testing and slightly differs from the standard notion of sparsity introduced by Donoho and Johnstone. As an application of this alternative notion of sparsity and as an extension of the detection threshold recently proposed, level-dependent detection thresholds are introduced. The performance of level-dependent detection thresholds is illustrated in the context of non-parametric estimation by soft thresholding in the wavelet domain. Experimental results show that the resulting approach performs well in comparison with one of the best up-to-date parametric method. In connection with some results concerning the statistical properties of wavelet coefficients associated with strictly stationary random processes, prospects are suggested for estimating unknown signals in non-necessarily white or Gaussian noise.

1. INTRODUCTION

Sparsity has gained much interest in the signal processing community since Donoho and Johnstone's seminal work in 1994 [1]. The reason is the existence of transforms that are sparse in the following sense: a sparse transform makes it possible to represent many signals by coefficients that have small or null amplitudes, except a few ones whose amplitudes are large. In this sense, the orthonormal discrete wavelet transform (DWT) is sparse for smooth and piecewise regular signals. Following [1], sparsity is often introduced from an estimation point of view and used to estimate or reconstruct a signal on the basis of its noisy samples. In continuation of [2], the purpose of this paper is to define and discuss another notion of sparsity. This notion of sparsity is based on [3, Theorem VII.1], a result concerning binary hypothesis testing and, more specifically, the detection in additive white Gaussian noise (AWGN) of a random signal with unknown distribution and prior. This alternative notion of sparsity is presented in section 2. In section 3, two applications of this notion are presented. Both concern the non-parametric estimation of an unknown deterministic signal, when the non-parametric estimation is performed by soft thresholding of the wavelet coefficients associated with a noisy observation of the signal. Soft thresholding requires to choose a threshold for selecting the coefficients that contain information about the signal to recover. In section 3.2, the detection threshold, originally introduced in [2] as an alternative to the standard universal and minimax thresholds, is presented as a consequence of the new notion of sparsity. The second application, treated in section 3.3, is an extension of the detection threshold. In fact, level-dependent detection thresholds are proposed in section 3.3 for the estimation of signals by soft thresholding. The resulting approach has performance measurements close to those obtained with BLS-GSM [4], a reference amongst the best up-to-date parametric techniques. Section 4 relates the notion of sparsity proposed in this paper to estimators of the noise standard deviation, especially for situations where too many large coefficients for the signal are present among the detail wavelet coefficients. Section 5 concludes this paper by propounding several extensions for the estimation of signals in possibly coloured and/or non-Gaussian noise.

2. SPARSITY FOR DETECTING SIGNALS WITH UNKNOWN DISTRIBUTIONS AND PRIORS IN AWGN

We begin by recalling a result, namely proposition 1, established in [2]. The discussion that follows this statement concerns the notion of sparsity that proposition 1 suggests. As mentioned in [2], if the random variables considered in proposition 1 below are replaced by n-dimensional real random vectors and the absolute values by the standard Euclidean norm in $\mathbb{R}^n$, the statement thus obtained still holds true and is an easy extension of [3, Theorem VII.1]. Henceforth, $V(\rho, p)$ stands for the function defined for every nonnegative real number ρ and every $0 \leqslant p \leqslant 1$ by

$$V(\rho,p) = p\Big[\mathrm{F}\left(\rho+\xi(\rho,p)\right) - \mathrm{F}\left(\rho-\xi(\rho,p)\right)\Big] + 2(1-p)\Big[1-\mathrm{F}(\xi(\rho,p))\Big], \quad (1)$$

where F is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$ and

$$\xi(\rho,p) = \frac{\rho}{2} + \frac{1}{\rho}\left[\ln\frac{1-p}{p} + \ln\left(1+\sqrt{1-\frac{p^2}{(1-p)^2}e^{-\rho^2}}\right)\right]. \quad (2)$$

As usual, if a property $\mathcal{P}$ holds true almost surely, we write $\mathcal{P}$ (a-s). The thresholding test $\mathcal{T}_h$ with threshold height $h \geqslant 0$ is the measurable map of $\mathbb{R}$ into $\mathbb{R}$ defined for every real number y by

$$\mathcal{T}_h(y) = \begin{cases} 1 & \text{if } |y| \geqslant h \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 1 *Consider the following binary hypothesis testing problem*

$$\begin{cases} \mathcal{H}_0: & Y \sim \mathcal{N}(0,\sigma^2) \\ \mathcal{H}_1: & Y = \Theta + X,\ \Theta \neq 0 \text{ (a-s)}, \begin{cases} |\Theta| \geqslant \mathsf{a} \geqslant 0 \text{ (a-s)}, \\ X \sim \mathcal{N}(0,\sigma^2), \end{cases} \end{cases}$$

where Y, Θ, X are real random variables such that Θ and X are independent and σ is some positive real value.

If the a priori *probability of occurrence of hypothesis* $\mathcal{H}_1$ *is less than or equal to some value* $\mathrm{p}^* \leqslant 1/2$*, then* $V(\mathsf{a}/\sigma, \mathrm{p}^*)$ *is a sharp upper bound for the probabilities of error of the Bayes test* $\mathcal{L}$ *with the least probability of error among all possible tests and the thresholding test* $\mathcal{T}_{\sigma\xi(\mathsf{a}/\sigma,\mathrm{p}^*)}$ *with threshold height* $\sigma\xi(\mathsf{a}/\sigma, \mathrm{p}^*)$*. The bound* $V(\mathsf{a}/\sigma, \mathrm{p}^*)$ *is sharp because attained by both* $\mathcal{L}$ *and* $\mathcal{T}_{\sigma\xi(\mathsf{a}/\sigma,\mathrm{p}^*)}$ *if* $|\Theta| = \mathsf{a}$ (a-s)*, with* $\mathrm{P}[\Theta = \mathsf{a}] = \mathrm{P}[\Theta = -\mathsf{a}] = 1/2$ *and the probability of occurrence of hypothesis* $\mathcal{H}_1$ *is* p^**.*

Basically, the assumptions made about the signal Θ in the statement above are aimed at bounding our lack of *prior* knowledge: we assume that the signal is less present than absent since the probability of occurrence of the alternative hypothesis $\mathcal{H}_1$ is assumed to be less than or equal to one half; the signal is assumed to be relatively big in the sense that its amplitude is above or equal to a.

We now consider the following model. Denoting the set of natural numbers by $\mathbb{N}$, suppose that $\{Y_i\}_{i\in\mathbb{N}}$ is a sequence of real random variables such that, for each $i \in \mathbb{N}$, Y_i obeys the following binary hypothesis model:

$$\begin{cases} \mathcal{H}_{0,i}: & Y_i = X_i, \\ \mathcal{H}_{1,i}: & Y_i = \Theta_i + X_i, \end{cases} \tag{3}$$

where $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of independent, centred and identically Gaussian distributed random variables with standard deviation $\sigma > 0$ and $\{\Theta_i\}_{i\in\mathbb{N}}$ is a sequence of random variables that are not necessarily independent or identically distributed. The sequence $\{\Theta_i\}_{i\in\mathbb{N}}$ can be regarded as the coefficients associated with some signal of interest, the sequence $\{X_i\}_{i\in\mathbb{N}}$ represents some AWGN and $\{Y_i\}_{i\in\mathbb{N}}$ is a noisy observation of the signal coefficients. In addition, we make the following assumptions. These assumptions are of the same type as those of proposition 1. In association with the binary hypothesis model of Eq. (3), they define the notion of sparsity considered in the sequel.

(**A**) [Amplitude] There exists $\mathsf{a} \geqslant 0$ such that $|\Theta_i| \geqslant \mathsf{a}$ (a-s) for every $i \in \mathbb{N}$.

(**O**) [Occurrence] For each $i \in \mathbb{N}$, Θ_i and X_i are independent and the probability of occurrence of the alternative hypothesis $\mathcal{H}_{1,i}$ is less than or equal to $\mathrm{p}^* \leqslant 1/2$.

Assumption **(A)** specifies the existence of a minimum amplitude a for the signal coefficients whereas assumption **(O)** sets a maximum value p^* for the probabilities of occurrence of the alternative hypotheses $\mathcal{H}_{1,i}$. Therefore, when a is large and p^* is small, it follows from assumptions **(A)** and **(O)** that large coefficients are few in number, which corresponds to the standard notion of sparsity. In contrast to the standard notion of sparsity, the new one, defined by assumptions **(A)** and **(O)**, does not require a to be large and the number of large signal coefficients to be small. This notion of sparsity involves the case where the signal coefficients are random variables. Above all, proposition 1 provides us with the thresholding test $\mathcal{T}_{\sigma\xi(\mathsf{a}/\sigma,\mathrm{p}^*)}$ with threshold $\sigma\xi(\mathsf{a}/\sigma, \mathrm{p}^*)$ to distinguish the coefficients containing significant information about the signal from those due to noise alone. In addition, under assumptions **(A)** and **(O)**, this test has the same upper-bound $V(\mathsf{a}/\sigma, \mathrm{p}^*)$ for the probability of error than the Bayes test. The next section describes how the notion of sparsity defined by assumptions **(A)** and **(O)** and the threshold introduced just above can be used in non-parametric estimation by soft thresholding.

3. DETECTION THRESHOLDS FOR NON-PARAMETRIC ESTIMATION BY SOFT THRESHOLDING

The notion of sparsity specified by assumptions **(A)** and **(O)** can be applied in two ways to non-parametric estimation by soft thresholding. First, as described in [2], it can be used to tune the soft thresholding function with a different threshold from those usually proposed; this will be recalled in section 3.2. Second, and this is the topic of section 3.3, it makes it possible to introduce a level-dependent approach where soft thresholding is adjusted at each decomposition level with a dedicated detection threshold. As a preamble, some basics about non-parametric estimation by sparse transform and soft thresholding are recalled.

3.1 Background

Let us consider the samples of some signal corrupted by independent AWGN with standard deviation $\sigma > 0$. The non-parametric estimation of a signal proposed in [1] is performed along the following lines. First, a linear orthonormal transform $\mathcal{W}$ is applied to the data. This transform is sparse in the sense that it represents the signal by a relatively small number of coefficients whose amplitudes are large in comparison to those resulting from noise. The second step is a non-linear filtering of the coefficients returned by $\mathcal{W}$. The purpose of this filtering is to eliminate the noise components by forcing them to zero and, possibly, to denoise the signal components. This filtering is performed by a thresholding function $\delta_\lambda(\cdot)$ that depends on a threshold λ whose main role is to distinguish the noisy signal components from those due to noise alone. Basically, a coefficient whose absolute value exceeds λ is regarded as a component of the noisy signal; a coefficient with absolute value below λ is considered as noise. The last step reconstructs the estimate of the signal on the basis of the filtered coefficients. The performance of this method is evaluated through a cost or risk function, which is generally the Mean Square Error (MSE) of the estimate.

The computation of the estimate thus requires to choose $\mathcal{W}$, the thresholding function $\delta_\lambda(\cdot)$ and the threshold value λ. For the sparse transform, it is customary to use the DWT. Regarding the thresholding function, the soft thresholding function is a good choice for its properties of smoothness and adaptation (see [5]). This function is defined for every real value x by

$$\delta_\lambda(x) = \begin{cases} x - \operatorname{sgn}(x)\lambda & \text{if } |x| \geqslant \lambda, \\ 0 & \text{elsewhere,} \end{cases} \tag{4}$$

where $\operatorname{sgn}(x) = 1$ (resp. -1) if $x \geqslant 0$ (resp. $x < 0$). Let Y_i and Θ_i, $i = 1, 2, \ldots, N$, be the noisy wavelet coefficients and the signal wavelet coefficients, respectively. Then, the MSE associated with $\delta_\lambda(\cdot)$ is $\mathrm{MSE} = (1/N)\sum_{i=1}^{N} \mathbb{E}\left(\Theta_i - \delta_\lambda(Y_i)\right)^2$. As far as the threshold is concerned, the literature on the topic usually distinguishes between the universal and the minimax thresholds introduced in [1]. The universal threshold $\lambda_u(N)$ is an estimate of the maximum of the amplitude that can be attained by the noise components. We have $\lambda_u(N) = \sigma\sqrt{2\ln N}$ where N is the sample size. The expression of the minimax threshold is useless in the sequel. Several authors have suggested that the universal and minimax thresholds are actually too large for many practical applications (see [6, 7] among others). By considering the notion of sparsity of section 2, another threshold is proposed in [2] and this threshold, called the *detection threshold*, yields better performance measurements than standard ones. This is briefly recalled in the next section before presenting a level-dependent extension.

3.2 Soft thresholding by detection threshold

The following summary of [2] is useful for a better understanding of the next section and is a first application of the new notion of sparsity introduced in section 2. Let us assume that the coefficients $\{Y_i\}_{1\leqslant i\leqslant N}$ returned by the transform $\mathcal{W}$

satisfy Eq. (3) with assumptions **(A)** and **(O)**. Since the universal threshold $\lambda_u(N)$ can be regarded as the maximum amplitude of the noise coefficients when N is large enough, the minimum amplitude in Eq. (3) is assumed to be $\mathsf{a} = \lambda_u(N)$ in this section. As far as the maximum probability of occurrence p^* is concerned, we consider the leat favourable case $\mathrm{p}^* = 1/2$. This model is acceptable to describe the statistical behaviour of the wavelet coefficients for smooth or piecewise regular signals ([1, 8]). Since the threshold of the thresholding function is aimed at distinguishing small from large coefficients, section 2 leads to choose this threshold equal to the so-called detection threshold

$$\begin{aligned}\lambda_d(N) &= \sigma\xi(\lambda_u(N)/\sigma)\\ &= \sigma\sqrt{\ln N/2} + \sigma \ln\left(1+\sqrt{1-1/N^2}\right)/\sqrt{2\ln N}.\end{aligned}$$

This threshold accepts or rejects the null hypothesis with a probability of error less than or equal to $V(\sqrt{2\ln N}, 1/2)$, which is a decreasing function of N. For small values of N, the threshold $\lambda_d(N)$ is close to the minimax threshold; for large values of N (above or equal to 2048), the value of $\lambda_d(N)$ is about $\lambda_u(N)/2$ and smaller than the minimax threshold. Note the following. In the model above, the coefficients returned for the signal are not necessarily deterministic but can be random variables. However, the experimental results given in [2] and below are obtained by considering the case where the coefficients Θ_i are deterministic because they derive from the application of $\mathcal{W}$ to the samples $f(t_i)$, $i = 1, \ldots, N$, of some deterministic function. In [2], an upper bound for the risk of the soft thresholding estimation is computed when the detection threshold is used. For a wide class of signals, and when the number of observations is large, this upper bound is proved to be from about twice to four times smaller than the standard upper bounds given for the universal and the minimax thresholds. This result is experimentally verified: for a large class of synthetic signals and standard images, the detection threshold actually achieves smaller risks for the estimation by soft thresholding than the universal and the minimax thresholds.

Summarizing the theoretical and experimental results of [2], the minimax and universal thresholds are suitable for recovering smooth signals; the detection threshold is suitable for estimating less smooth signals, including piecewise regular signals, which are known to be over-smoothed when using the minimax or the universal threshold. In fact, smooth signals yield very sparse wavelet representations in the sense given by [1]: for such signals, large coefficients are indeed very few in number. In contrast, wavelet representations of natural images, which are piecewise regular rather than smooth, fail to be sparse enough since large coefficients are not very few. Assumption **(O)** then makes it possible to derive thresholds adapted to less smooth signals. Some other results are given below since the soft thresholding with level-dependent detection thresholds presented hereafter is compared to soft thresholding with detection threshold.

3.3 Soft Thresholding with level-dependent detection thresholds

The idea developed in this section is another application of the notion of sparsity presented in section 2. We still assume that the coefficients returned by the sparse transform obey the model of Eq. (3) with assumptions **(A)** and **(O)**. However, we now modify assumption **(A)** as follows. In fact, it is known that the amplitudes of the signal detail coefficients tend to decrease when the decomposition level increases (see [9, Theorem 9.7, p. 389], for instance). Therefore, we introduce the following model for the minimum amplitudes of the wavelet coefficients. Since $\lambda_u(N)$ is considered as the maximum amplitude of the noise coefficients, we still consider that the minimum amplitude of the detail coefficients that carry significant information about the signals is $\mathsf{a}_1 = \lambda_u(N)$ at the first decomposition level $j = 1$. On the other hand, [9, Theorem 9.7, p. 389] suggests assuming that the minimum amplitude of the coefficients associated with the signal at decomposition level j is $\mathsf{a}_j = \mathsf{a}_{j-1}/\sqrt{2}$ for $j = 2, \ldots, J$. Therefore, we have $\mathsf{a}_j = \sigma\sqrt{2\ln N}/\sqrt{2^{j-1}}$. As far as **(O)** is concerned, we consider the least favourable case again by setting $\mathrm{p}^* = 1/2$, even though, for smooth or piecewise regular signals, the proportion of significant coefficients is known to increase with the decomposition level [7, Section 10.2.4, p. 460]. The possibility to relate the probability of occurrence of $\mathcal{H}_{1,i}$ to the decomposition level is postponed to a forthcoming work. The foregoing and the results of section 2 lead us to adjust the soft thresholding function at each decomposition level $j = 1, 2, \ldots, J$, $J \leqslant \log_2(N) - 1$, with the level-dependent detection threshold

$$\lambda_d(j, N) = \sigma\sqrt{\ln N/2^j} + \sigma\frac{\ln\left(1+\sqrt{1-1/N^2}\right)}{\sqrt{2^j \ln N}}. \quad (5)$$

This threshold value straightforwardly derives from Eq. (2) by setting $\rho = \mathsf{a}_j$ and $p = 1/2$. By so proceeding, the soft thresholding function becomes level-dependent.

We conclude this section by experimental results dedicated to image denoising. We consider the Stationary Wavelet Transform (SWT), particularly suitable for denoising because it is translation invariant and redundant [10, 7]. The SWT is performed with the Symlet wavelet of order 8 ('sym8' in the Matlab Wavelet toolbox). The decomposition levels are $j = 1, 2, \ldots, J$ with $J = 4$. We denoise the 512×512 'Lena' image additively corrupted by independent WGN with standard deviation $\sigma = 9, 18, 27, 36$. We measure the performance of the denoising by calculating the PSNR before and after denoising. We recall that the PSNR is defined by $\mathrm{PSNR} = 10\log_{10}\left(255^2/\mathrm{MSE}\right)$. For instance, for compression applications, the quality of a compressed image is considered to be good if the PSNR exceeds 30 dB.

Table 1 presents the PSNR values obtained over 10 trials for each standard deviation tested. In this table, PSNR[BLS-GSM] stands for the PSNR obtained with the BLS-GSM method described in [4] and which is regarded as a reference amongst parametric methods; PSNR[$\lambda(N)$] is the PSNR obtained by soft thresholding with threshold height λ, λ being either the detection threshold $\lambda_d(N)$, the universal threshold $\lambda_u(N)$ or the minimax threshold $\lambda_m(N)$.

As mentioned in section 3.2, the detection threshold performs better than the universal and minimax thresholds. But, above all, the PSNRs obtained by soft thresholding with level-dependent detection thresholds are larger than those yielded by the other non-parametric methods and significantly approach those achieved by BLS-GSM. Figure 1 displays some examples of denoised images yielded by the several methods considered in this section.

4. SPARSITY AND ESTIMATION OF THE NOISE STANDARD DEVIATION.

In section 2 and its applications above, the value of the noise standard deviation is assumed to be known. When this value is unknown, it is generally estimated via the MAD estimator [1]. The MAD estimator is rather natural when the transform is sparse in the sense of [1] for the following reason. If the signal is smooth, the signal wavelet coefficients are very few among the detail coefficients at the first decomposition level. In this case, the MAD estimator applied to the detail coefficients of the first decomposition level is robust and yields a good estimate of the noise standard deviation ([1]). However, as already highlighted, wavelet representations of natural images, which are piecewise regular rather

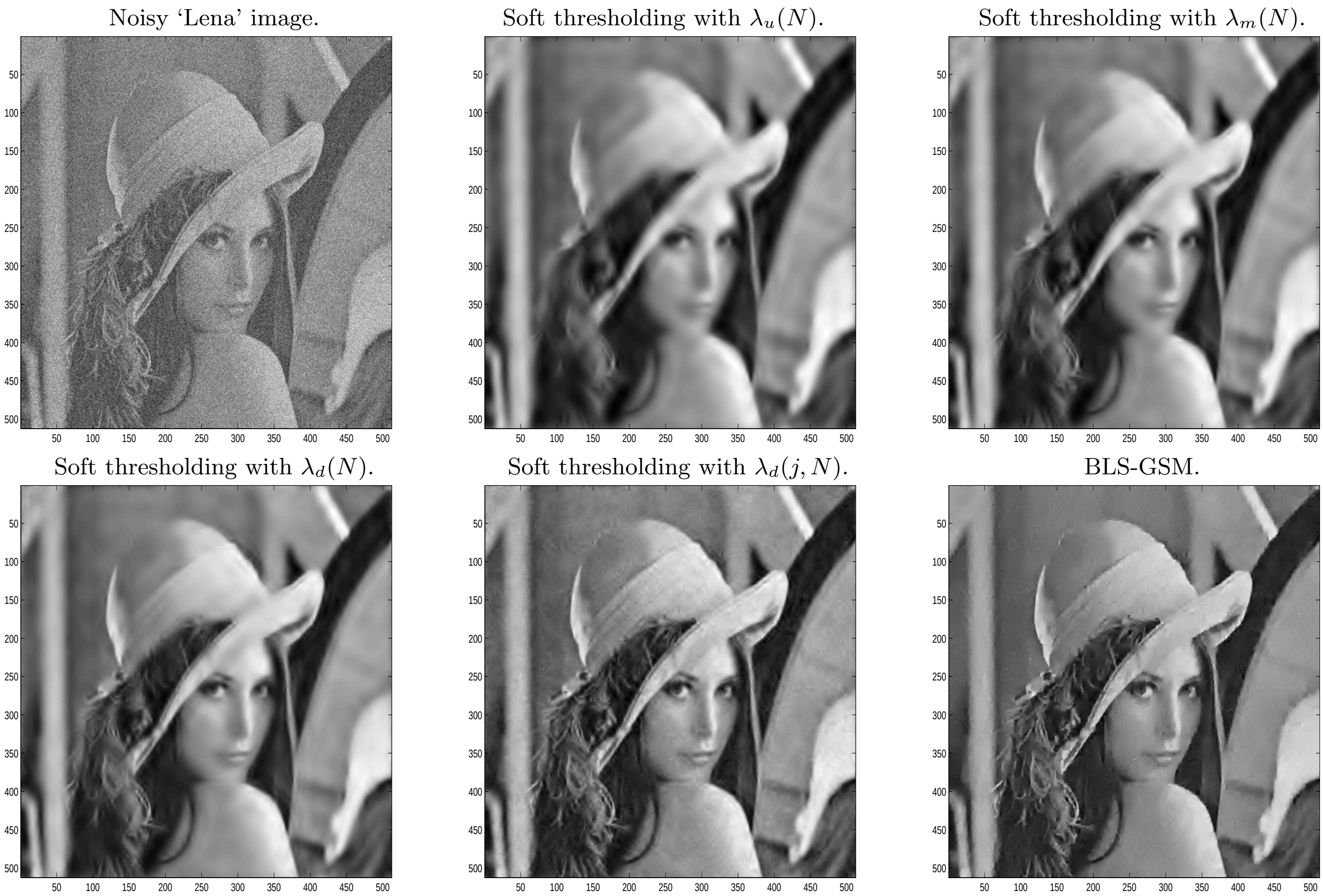

Figure 1: Noisy 'Lena' image and denoised versions. Noise is white and Gaussian with standard deviation $\sigma = 36$.

Table 1: PSNRs for different values of σ when the standard 'Lena' image is corrupted by independent AWGN with standard deviation σ. The Initial PSNR is the PSNR of the noisy image.

σ	9	18	27	36
Initial PSNR	29.0	23.0	19.5	17.0
PSNR[$\lambda_u(N)$]	29.3	26.6	25.3	24.5
PSNR[$\lambda_m(N)$]	30.6	27.8	26.3	25.4
PSNR[$\lambda_d(N)$]	32.1	29.1	27.5	26.5
PSNR[$\lambda_d(j, N)$]	33.6	31.0	29.5	28.4
PSNR[BLS-GSM]	35.7	32.7	30.9	30.1

than smooth, fail to be sparse enough. This is the reason why assumption **(O)** is not very constraining by merely assuming that the alternative hypotheses occur with probabilities less than or equal to $p^* \leqslant 1/2$. When the proportion of signal coefficients is large among the wavelet detail coefficients (for instance, when decomposing piecewise regular signals), the MAD estimator is no longer robust. An alternative to the MAD estimator could be derived from sparsity assumptions such as **(A)** and **(O)**. Indeed, with sparsity assumptions that embrace **(A)** and **(O)**, the noise standard deviation is proved in [11] to be the only positive real number satisfying a specific convergence criterion when the sample size and the minimum amplitude of the signals tend to infinity and the observations are independent. This convergence involves neither the probability distributions nor the probabilities of occurrence of the alternative hypotheses. Estimators of the noise standard deviation, derived from this theoretical result, are proposed in [11] and [12]. They have been tested in applications different from those considered in the present paper. Therefore, the performance of these estimators should be studied and compared to that of the MAD estimator for non-parametric estimation. According to the theoretical and experimental results presented in [11] and [12], these estimators can be expected to perform well even when many signal coefficients are present among the detail wavelet coefficients.

5. CONCLUSION

We have introduced a notion of sparsity that derives from a result established for binary hypothesis testing. This notion of sparsity is similar to that propsoed by Donoho and Johnstone in [1]. The differences have been discussed and these differences are exploited in two applications of this notion of sparsity to non-parametric estimation. The use of this notion of sparsity for the estimation of the noise standard deviation in comparison with the standard MAD estimator has also been suggested as an introduction to forthcoming work on the topic. In future work, a theoretical upper-bound on the MSE should be derived for comparison with the upper-bound established for the detection threshold ([2]) and the upper-bounds given for the universal and minimax thresholds ([1]), in the context of soft thresholding estimation in the wavelet domain.

We also would like to highlight how the contents of this paper connect to results concerning the statistical properties of the wavelet coefficients associated with strictly stationary random processes. More specifically, consider some signal in additive and independent strictly stationary noise. Do not assume that noise is either Gaussian or white. According to results such as those given in [13], [14] and [15], the sequences of coefficients returned by the wavelet, wavelet packet and M-band wavelet packet transforms of the input noise tend to be white and Gaussian in a distributional sense specified in [14] and [15]. The tendency to whiteness and Gaussianity is obtained when the decomposition level and the regularity of the decomposition filters are large enough. It is worth emphasizing that, for most paths of the decomposition tree, the regularity of the filters must be chosen large enough with respect to an already large enough decomposition level. However, experimental results given in the aforementioned papers suggest that the tendency to Gaussianity and whiteness is obtained for a large class of random processes of practical interest and for reasonable values for the regularity and the decomposition level. Therefore, it is reasonable to consider that the sequence of the signal wavelet coefficients is additively corrupted by independent white Gaussian noise. Under sparsity assumptions, the noise standard deviation, if unknown, can be estimated. Then, if we assume that **(A)** and **(O)** are satisfied, the non-parametric estimation of the signal can be carried out as proposed in this paper. We plan to study such an approach in connection to results - such as those stated in [16, 17], among others - about sparsity and Besov spaces. To complete this concluding prospects, note also that the soft thresholding function can successfully be replaced by the smooth shrinkage function proposed in [18]. This shrinkage function is not only continuous but also avoids the over-smoothing and important estimation bias incurred by using the soft thresholding function [19]. In a forthcoming paper, we will propose an approach that combines this shrinkage function with level-dependent detection thresholds such as those introduced in section 3.3.

REFERENCES

[1] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.

[2] A. M. Atto, D. Pastor, and G. Mercier, "Detection thresholds for non-parametric estimation," *Signal, Image and Video processing*, 2008, http://dx.doi.org/10.1007/s11760-008-0051-x.

[3] D. Pastor, R. Gay, and A. Gronenboom, "A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 228–238, Jan. 2002.

[4] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[5] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[6] A. Bruce and H. Gao, "Understanding waveshrink: Variance and bias estimation," StatSci, Research Report 36, 1996.

[7] S. Mallat, *A wavelet tour of signal processing, second edition.* Academic Press, 1999.

[8] I. Johnstone, "Wavelets and the theory of non-parametric function estimation," *Journal of the Royal Statistical Society*, vol. A, no. 357, pp. 2475–2493, 1999.

[9] S. Mallat, *A wavelet tour of signal processing.* Academic Press, 1998.

[10] R. Coifman and D. Donoho, *Translation invariant de-noising.* Lecture Notes in Statistics, 1995, no. 103, pp. 125–150.

[11] D. Pastor, "A theoretical result for processing signals that have unknown distributions and priors in white gaussian noise," *Computational Statistics & Data Analysis, CSDA*, vol. 52, no. 6, pp. 3167–3186, 2008, http://dx.doi.org/10.1016/j.csda.2007.10.011.

[12] D. Pastor and A. Amehraye, "Algorithms and applications for estimating the standard deviation of awgn when observations are not signal-free," *Journal of Computers (JCP)*, vol. 2, no. 7, pp. 1–10, September 2007.

[13] D. Leporini and J. Pesquet, "High-order wavelet packets and cumulant field analysis," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 863–877, Apr. 1999.

[14] A. Atto and D. Pastor, "Central limit theorems for wavelet packet decompositions of stationary random processes," submitted to IEEE Transactions on Information Theory, http://fr.arxiv.org/abs/0802.0797.

[15] ——, "Limit distributions for wavelet packet coefficients of band-limited stationary random processes," submitted to EUSIPCO 2008.

[16] D. L. Donoho and I. M. Johnstone, "Neo-classical minimax problems, thresholding and adaptive function estimation," *Bernoulli 2*, no. 1, pp. 39–62, 1996.

[17] ——, "Asymptotic minimaxity of wavelet estimators with sampled data," *Statistica Sinica*, vol. 9, no. 1, pp. 1–32, 1999.

[18] A. M. Atto, D. Pastor, and G. Mercier, "Smooth sigmoid wavelet shrinkage for non-parametric estimation," in *ICASSP 2008 : IEEE international conference on acoustics, speech and signal processing, March 30 - April 4, Las Vegas, USA*, 2008.

[19] H. Gao, "Waveshrink shrinkage denoising using the non-negative garrote," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, 1998.