Sparsity Measure and the Detection of Significant Data

Abdourrahmane Atto, Dominique Pastor, Grégoire Mercier

Sparsity Measure and the Detection of Significant Data

Abdourrahmane M. Atto [1], Dominique Pastor [2], Grégoire Mercier [3]

***Abstract*—The paper provides a formal description of the sparsity of a representation via the detection thresholds. The formalism proposed derives from theoretical results about the detection of significant coefficients when data are observed in presence of additive white Gaussian noise. The detection thresholds depend on two parameters describing the sparsity degree for the representation of a signal. The standard universal and minimax thresholds correspond to detection thresholds associated with different sparsity degrees.**

***Index Terms*—Sparsity measure, Wavelets, Detection thresholds.**

I. Introduction

The detection thresholds are synthesized by considering a risk function which is the probability of erroneously deciding that a coefficient is significant when it is not the case. They depend on two parameters that can be used to bound the sparsity degree of the wavelet representation [1]. These thresholds are optimal in the sense that they lead to the same upper bound for the probability of error than the Bayes test with minimal probability of error among all possible tests [2], for a certain class of signals, including sparse signals. It is shown in this paper that the standard minimax and universal thresholds are detection thresholds corresponding to different degrees of sparsity. The selection of appropriate detection thresholds with respect to the wavelet decomposition properties of some signals such as smooth and piecewise regular signals is also discussed.

II. Detection thresholds and Sparsity degree

Consider the following decision problem with binary hypothesis model $(\mathscr{H}_0, \mathscr{H}_1)$, where $\mathscr{H}_0 : c_i \sim \mathscr{N}(0, \sigma^2)$ versus $\mathscr{H}_1 : c_i = d_i + \epsilon_i, |d_i| \geqslant \mathsf{a} > 0, \epsilon_i \sim \mathscr{N}(0, \sigma^2)$.

Let ξ be the function defined for $\rho > 0$ and $0 \leqslant p \leqslant 1/2$ by

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

Assume that the *a priori* probability of occurrence of hypothesis $\mathscr{H}_1$ is less than or equal to some value $p^* \leqslant 1/2$. Then, for deciding $\mathscr{H}_0$ versus $\mathscr{H}_1$, the thresholding test with threshold height $\lambda_D(\mathsf{a}, p^*)$,

$$\lambda_D(\mathsf{a}, p^*) = \sigma\xi(\mathsf{a}/\sigma, p^*), \quad (2)$$

has the same sharp upper bound for its probability of error than the Bayes test with the least probability of error (see [2] for details).

Parameter p^* reflects the presence (quantity) of significant coefficients of the signal amongst the noisy coefficients. Assuming that p^* is less than or equal to $1/2$ ensures that the representation of the signal is at least sparse (in the weak sense).

Parameter a can be seen as the minimum amplitude considered to be significant for a signal coefficient. Parameters p^* and a thus allow to formalize more precisely the sparsity degree of the signal representation (see [1]).

In what follows, we consider $\sigma = 1$ for the sake of simplicity. The following proposition unifies the minimax, universal, and detection thresholds.

[1] TELECOM Bretagne, *am.atto@telecom-bretagne.eu*
[2] TELECOM Bretagne, *dominique.pastor@telecom-bretagne.eu*
[3] TELECOM Bretagne, *gregoire.mercier@telecom-bretagne.eu*

Proposition 1: For any positive real value η, there exist $\mathsf{a}_0 > 0$ and p_0^*, with $0 \leqslant p_0^* \leqslant 1/2$, such that

$$\lambda_D(\mathsf{a}_0, p_0^*) = \eta. \quad (3)$$

Proof: The result simply follows by noting that ξ is continuous, positive, $\lim_{p\to 0}\xi(\rho, p) = 0$ for fixed $\rho > 0$, and $\lim_{\rho\to+\infty}\xi(\rho, p) = +\infty$ for fixed p, $0 < p \leqslant 1/2$. ■

Let $\lambda_u(N) = \sigma\sqrt{2\ln N}$, the so-called universal threshold. This threshold reflects the maximum amplitude of the white Gaussian noise coefficients. Indeed, if $\epsilon_i \overset{\text{iid}}{\sim} \mathscr{N}(0, \sigma^2)$, it follows from [3, p. 187]; [4, p. 454], that

$$\lim_{N\to+\infty} \mathrm{P}\left[\lambda_u(N) - \alpha_N \leqslant \max\left\{|\epsilon_i|\right\}_{1\leqslant i\leqslant N} \leqslant \lambda_u(N)\right] = 1, \quad (4)$$

with $\alpha_N = \sigma\ln\ln N/\ln N$. Thus, the maximum amplitude of $\{\epsilon_i\}_{1\leqslant i\leqslant N}$ has a strong probability of being close to the universal threshold when N is large.

From proposition 1, it follows that for any $N \geqslant 2$, there exist some values a, p^* such that $\lambda_D(\mathsf{a}, p^*) = \lambda_u(N)$. Figure 1 shows the level curves $\lambda_D(\mathsf{a}, p^*) = \lambda_u(N)$ for different values of N. It appears that large values of a are associated with small values of p^* (strong sparsity) and *vice versa* (weak sparsity).

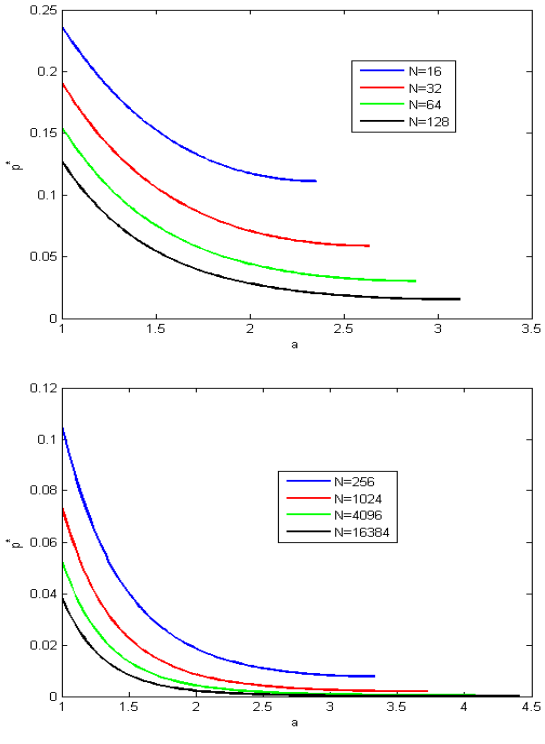


Fig. 1. Level curves $\lambda_D(\mathsf{a}, p^*) = \sqrt{2\ln N}$ for different values of N.

The same remark (as for the universal threshold) holds true for the *minimax threshold*. The minimax threshold $\lambda_m(N)$ is defined as the largest value λ among the values attaining a minimax risk bound given in [5].

III. Detection of significant wavelet coefficients

Detection thresholds are well-adapted to estimate wavelet coefficients corrupted by AWGN because of the sparsity of the wavelet transform [2]. Moreover, these thresholds are adaptable to the wavelet transform decomposition schemes: sparsity ensures that for reasonable resolution levels, signal coefficients are less present than noise coefficients among the detail wavelet coefficients and that signal coefficients have large amplitudes (in comparison to noise coefficients).

More precisely, it is known that for smooth or piecewise regular signals, the proportion of significant coefficients, which plays a role similar to that of p^*, increases with the resolution level [4, Section 10.2.4, p. 460]. Therefore, if we can give, first, upper-bounds $(p_j^*)_{j=1,2,\cdots,J}; p_j^* \leqslant 1/2$ for every $j = 1,2,\cdots,J$, and second , lower-bounds $(\mathsf{a}_j)_{j=1,2,\cdots,J}$ for the amplitudes of the significant wavelet coefficients, then we can derive level-dependent detection thresholds that can select significant wavelet coefficients at every resolution level. Since significant information tends to be absent among the first resolution level detail wavelet coefficients, it is reasonable to set $\mathsf{a}_1 = \sigma\sqrt{2\ln N}$, that is the universal threshold. Now, when the resolution level increases, it follows from [4, Theorem 6.4] that a convenient choice for $\mathsf{a}_j, j > 1$ is $\mathsf{a}_j = \mathsf{a}_1/\sqrt{2^{j-1}}$ when the signal of interest is smooth or piecewise regular.

In addition, since noise tends to be less present when the resolution level increases, p_j^* must be an increasing function of j. Note that detection thresholds are defined for $p_j^* \leqslant 1/2$. It is thus essential to stop the shrinkage at a resolution level J for which p_J^* is less than or equal to $1/2$. We propose the use of exponentially or geometrically increasing sequences for the values $(p_j^*)_{j=1,2,\cdots,J}$ since p_1^* must be a very small value (significant information tends to be absent among the first resolution level detail wavelet coefficients) and the presence of significant information increases significantly as the resolution level increases. In the following, we consider a sequence $(p_j^*)_{j=1,2,\cdots,J}$ such that $p_{j+1}^* = (p_j^*)^{1/\mu}$ with $\mu > 1$.

Summarizing, we consider the thresholds $\lambda_D(\mathsf{a}_j, p_j^*)$, where λ_D is defined by Eq. (2) and (a_j, p_j^*), for $j = 1,2,\cdots,J$ are given by

$$\mathsf{a}_j = \sigma\sqrt{\ln N}/2^{j/2-1}, \tag{5}$$

and

$$p_j^* = 1/2^{\mu^{J-j}}. \tag{6}$$

IV. Detection thresholds in practice

Experimental tests are carried by using the Stationary Wavelet Transform (SWT) and the biorthogonal spline wavelet with order 3 for decomposition and with order 1 for reconstruction ('bior1.3' in Matlab Wavelet toolbox). The maximum decomposition level is fixed to $J = 4$. The SWT [6] has appreciable properties in denoising. Its redundancy makes it possible to reduce residual noise and some possible artifacts incurred by the translation sensitivity of the orthonormal wavelet transform.

The detection thresholds are used to calibrate the *Smooth Sigmoid Based Shrinkage* (SSBS) functions of [7]. The SSBS functions are smooth functions and they allow for a flexible control of the shrinkage through parameters which model the attenuation imposed to small, median and large data. This allows correcting the main drawbacks of the soft and hard shrinkage functions. The SSBS functions are defined by:

$$\delta_{t,\tau,\lambda}(x) = \frac{\mathrm{sgn}(x)(|x| - t)_+}{1 + e^{-\tau(|x|-\lambda)}}, \tag{7}$$

for $x \in \mathbb{R}$, $(t,\tau,\lambda) \in \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+$, where $\mathrm{sgn}(x) = 1$ (resp. -1) if $x \geqslant 0$ (resp. $x < 0$), and $(x)_+ = x$ (resp. 0) if $x \geqslant 0$ (resp. $x < 0$).

The results below were obtained with the following values for the SSBS parameters. We consider the values

- $t_0 = 0$ and $\theta_0 = \pi/10$ when the noise standard deviation σ is less than or equal to 5,
- $t_0 = \sigma/5$ and $\theta_0 = \pi/6$ when $5 < \sigma \leqslant 15$,
- $t_0 = \sigma/3$ and $\theta_0 = \pi/5$ when σ is larger than 15.

In addition, we use as threshold heights $\lambda_0(j) = \lambda_D(\mathsf{a}_j, p_j^*)$, the detection thresholds defined by Eq. (2), where (a_j, p_j^*), for $j = 1,2,\cdots,J$ are given by Eqs. (5) and (6). We use the value $\mu = 2.35$ in Eq. (6), which tend to be a good compromise for the different test images used, and we assume that $p_J^* = 1/2$.

The PSNR (in deciBel unit, dB) is used to assess the quality of a denoised image,

$$\mathrm{PSNR} = 10\log_{10}\left(255^2/\mathrm{MSE}\right). \tag{8}$$

PSNRs achieved by SSBS are given for several values of the noise standard deviation in table I, in comparison to the BLS-GSM method of [8]. The latter (free MatLab software [1]) is a parametric method using redundant wavelet transform and models neighbourhoods of wavelet coefficients with Gaussian vectors multiplied by random positive scalars. BLS-GSM also takes into account the orientation and the interscale dependencies of the wavelet coefficients. It is actually the best parametric method using redundant wavelet transform and it is computationally expensive (see [9] for an appreciation of the BLS-GSM computing time).

According to table I, the performance of SSBS is comparable to that obtained with BLS-GSM in terms of PSNR. Indeed, the difference in PSNR between SSBS and BLS-GSM is about 1 dB, which can be regarded as a good result since SSBS is a simple non-parametric method (in the sense that neither interscale nor intra-scale predictors are included in the shrinkage process).

In addition, we now address the sensitivity of the SSBS calibrated by the detection thresholds according to the wavelet filters. Table II presents the PSNRs obtained by using different filters. It follows from this table that, for a given wavelet family, there is no significant variability of the results with respect to the length of the wavelet filter used: short impulse response wavelet filters perform better with the 'House' image while long impulse response filters perform better with the 'Fingerprint' image. Thus, the best PSNR still depends on the input image. In addition, there is no significant difference between the results obtained with respect to the wavelet family used, when the length of the filters are sensibly of the same order.

V. Conclusion

This paper highlights some remarkable properties of the detection thresholds. Detection thresholds depend on two parameters that describe the sparsity of the wavelet representation in terms of "minimum significant amplitude" for the signal and "probability of occurrence" of the significant signal coefficients in the sequence of wavelet coefficients. It is shown that the universal and minimax thresholds are particular detection thresholds corresponding to different degrees of sparsity.

On the other hand, this paper analyzes the combination between detection thresholds and the SSBS functions. The SSBS functions are a family of smooth sigmoid based shrinkage functions which perform a penalized shrinkage. The experimental

[1] avalaible at http://decsai.ugr.es/~javier/denoise/software/index.htm

TABLE I

MEANS AND VARIANCES OF THE PSNRS COMPUTED OVER 25 NOISE REALIZATIONS, WHEN DENOISING TEST IMAGES BY THE SSBS AND BLS-GSM METHODS. THE TESTED IMAGES ARE CORRUPTED BY AWGN WITH STANDARD DEVIATION σ. THE SWT IS COMPUTED BY USING THE SPLINE 'BIOR1.3' WAVELET. THE SSBS PARAMETERS $(t_0, \theta_0, \lambda_0)$ GIVEN IN SECTION IV ARE USED.

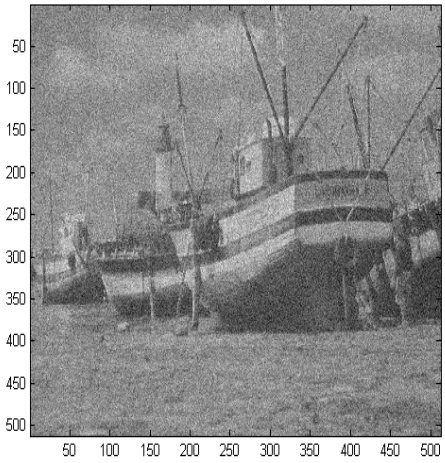
Image		'House'	'Peppers'	'Barbara'	'Lena'	'Fingerprint'	'Boat'
$\boldsymbol{\sigma = 5}$ ($\Longrightarrow$ Input PSNR = 34.1514).							
Mean(PSNR)	SSBS	37.9508	37.4567	35.8365	37.8641	35.2408	36.2997
	BLS-GSM	38.2248	37.5750	37.1966	38.1847	36.3801	36.7190
Var(PSNR) $\times 10^{03}$	SSBS	0.6045	0.6371	0.2058	0.1528	0.0783	0.1244
	BLS-GSM	0.0080	0.0028	0.0004	0.0004	0.0012	0.0008
$\boldsymbol{\sigma = 15}$ ($\Longrightarrow$ Input PSNR = 24.6090).							
Mean(PSNR)	SSBS	32.9988	31.5997	28.4616	32.8278	29.1183	30.8712
	BLS-GSM	33.8043	32.0354	30.7868	33.4822	29.9361	31.6285
Var(PSNR)	SSBS	0.0023	0.0016	0.0003	0.0007	0.0001	0.0004
	BLS-GSM	0.0011	0.0027	0.0001	0.0002	0.0002	0.0002
$\boldsymbol{\sigma = 25}$ ($\Longrightarrow$ Input PSNR = 20.1720).							
Mean(PSNR)	SSBS	30.6291	28.6862	25.6226	30.5796	26.3489	28.4944
	BLS-GSM	31.6253	29.4782	27.8179	31.2575	27.1177	29.2595
Var(PSNR)	SSBS	0.0020	0.0024	0.0001	0.0005	0.0004	0.0006
	BLS-GSM	0.0034	0.0070	0.0008	0.0004	0.0002	0.0005
$\boldsymbol{\sigma = 35}$ ($\Longrightarrow$ Input PSNR = 17.2494).							
Mean(PSNR)	SSBS	29.0606	26.6922	24.4933	29.1528	24.6512	27.1420
	BLS-GSM	30.0584	27.8208	25.9688	29.7596	25.2411	27.7355
Var(PSNR)	SSBS	0.0043	0.0023	0.0002	0.0006	0.0004	0.0006
	BLS-GSM	0.0037	0.0032	0.0006	0.0007	0.0004	0.0012

TABLE II

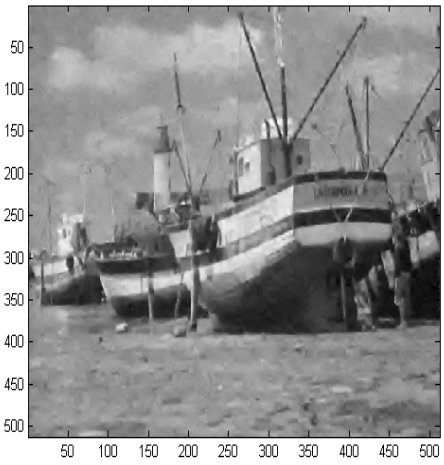
AVERAGE PSNRS COMPUTED OVER 25 NOISE REALIZATIONS, WHEN DENOISING TEST IMAGES BY THE SSBS METHOD. THE TESTED IMAGES ARE CORRUPTED BY AWGN WITH STANDARD DEVIATION $\sigma = 35$. THE SWT IS COMPUTED BY USING DIFFERENT DAUBECHIES, SPLINE, AND SYMLET WAVELETS REFERENCED AS IN THE MATLAB WAVELET TOOLBOX. THE SSBS PARAMETERS $(t_0, \theta_0, \lambda_0)$ GIVEN IN SECTION IV ARE USED.

Image	'House'	'Peppers'	'Barbara'	'Lena'	'Fingerprint'	'Boat'
Daubechies filters						
'db1'	28.6500	26.2633	24.1965	28.6561	23.1275	26.6957
'db2'	28.2556	26.0992	24.3287	28.8418	24.3481	26.7808
'db4'	28.0620	25.6922	24.4194	28.7922	24.8339	26.6889
'db8'	27.5814	25.1000	24.3617	28.5160	24.8711	26.4323
Symlet filters						
'sym1'	28.6889	26.2879	24.2026	28.6630	23.1335	26.6878
'sym2'	28.2613	26.0705	24.3386	28.8387	24.3446	26.7780
'sym4'	28.2320	25.9667	24.4539	28.9027	24.8237	26.7770
'sym8'	28.0018	25.7022	24.4534	28.8234	24.9062	26.6888
Spline biorthogonal filters						
'bior1.1'	28.6650	26.2695	24.1935	28.6614	23.1377	26.6907
'bior2.2'	27.6983	25.7494	24.4346	28.2951	24.8211	26.6846
'bior4.4'	28.0672	25.8622	24.4830	28.7680	24.6440	26.6894
'bior6.8'	28.0400	25.7525	24.5354	28.8420	24.9629	26.7517

(a) Noisy image
PSNR=17.25 dB



(b) SSBS
PSNR=27.1220 dB



(c) BLS-GSM
PSNR=27.7766 dB

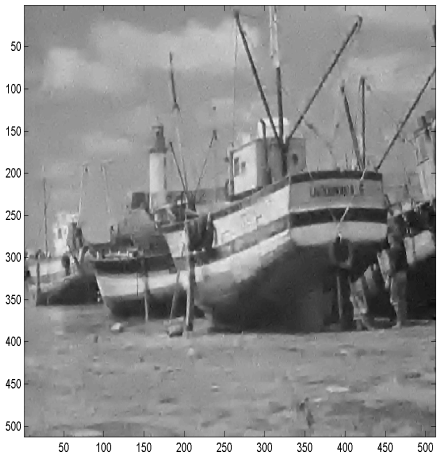


Fig. 2. SSBS and BLS-GSM denoising of noisy 'Boat' image corrupted by AWGN with standard deviation $\sigma = 35$.

results show that SSBS functions adjusted with these detection thresholds achieve denoising PSNR comparable to that of the best parametric and computationally expensive method, the BLS-GSM of [8]. This performance is remarkable for a non-parametric method where no interscale or intra-scale predictors are used to provide information about significant wavelet coefficients.

REFERENCES

[1] D. Pastor and A. M. Atto, "Sparsity from binary hypothesis testing and application to non-parametric estimation," *European Signal Processing Conference, EUSIPCO,* Lausanne, Switzerland, August 25-29, 2008.

[2] A. M. Atto, D. Pastor, and G. Mercier, "Detection threshold for non-parametric estimation," *Signal, Image and Video Processing, Springer*, vol. 2, no. 3, Sept. 2008.

[3] S. M. Berman, *Sojourns and extremes of stochastic processes.* Wadsworth and Brooks/Cole, 1992.

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

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

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

[7] A. M. Atto, D. Pastor, and G. Mercier, "Smooth sigmoid wavelet shrinkage for non-parametric estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP,* Las Vegas, Nevada, USA, 30 march - 4 april, 2008.

[8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. 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.

[9] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, Mar. 2007.