OPTIMAL SURE PARAMETERS FOR SIGMOIDAL WAVELET SHRINKAGE

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

Institut TELECOM, TELECOM Bretagne
LAB-STICC, CNRS UMR 3192
Technopôle Brest-Iroise, CS 83818
29238 Brest Cedex 3, FRANCE

ABSTRACT

This paper provides a SURE optimization for the parameters of a sub-class of smooth sigmoid based shrinkage functions. The optimization is performed on an unbiased estimation risk obtained by using sigmoid shrinkage functions. The SURE sigmoid shrinkage performance measurements are compared to those of the SURELET (SURE linear expansion of thresholds) parameterization. It is shown that the SURE sigmoid shrinkage performs better than the SURELET parameterization on most standard images. The relevance of SigShrink is reinforced by the flexible spatial adaptation it provides thanks to its parameters: a threshold for discriminating large and small data and an attenuation degree to control the shrinkage imposed to data with small amplitudes.

1. INTRODUCTION

The WaveShrink (Wavelet Shrinkage) estimation of a signal involves projecting the observed noisy signal on a wavelet basis, estimating the signal coefficients with a thresholding or shrinkage function and reconstructing an estimate of the signal by means of the inverse wavelet transform of the shrunken wavelet coefficients. The Smooth Sigmoid-Based Shrinkage (SSBS) functions introduced in [1] constitute a wide class of WaveShrink functions. The SSBS functions derive from the sigmoid function and perform an adjustable wavelet shrinkage thanks to parameters that control the attenuation degree imposed to the wavelet coefficients.

The present work addresses (Section 3) the optimization of a sub-class of the SSBS functions, the non-zero-forcing SSBS functions, hereafter called the SigShrink (Sigmoid Shrinkage) functions. The optimization of the SigShrink parameters is performed in the sense of the new SURE (Stein Unbiased Risk of Estimation, [2]) proposed in [3]. The SURE SigShrink estimation performance is then compared to that of the SURELET (SURE linear expansion of thresholds) parameterization (proposed in [3]). The SURELET parameterization consists in a sum of Derivatives Of Gaussian (DOGs) distribution. It is shown that the SURE SigShrink is more relevant than the sum of DOGs both in terms of estimation risk and spatial adaptation.

Before addressing the main topics of the paper, the following Section 2 briefly describes the non-parametric estimation by wavelet shrinkage and the class of SigShrink functions.

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

2. SIGMOID SHRINKAGE IN THE WAVELET DOMAIN

2.1 Shrinkage in the wavelet domain

Let $\{c_i\}_{1\leqslant i\leqslant N}$ be a sequence of wavelet coefficients,

$$c_i = d_i + \epsilon_i, \qquad i = 1,2,\dots,N, \tag{1}$$

where $\boldsymbol{d} = \{d_i\}_{1\leqslant i\leqslant N}$ represents the wavelet coefficients of the signal of interest and the random variables $\{\epsilon_i\}_{1\leqslant i\leqslant N}$ are independent and identically distributed (iid), Gaussian with null mean and variance σ^2, in short, $\epsilon_i \sim \mathcal{N}(0,\sigma^2)$ for every $i = 1,2,\dots,N$.

The non-parametric estimation by shrinkage in the wavelet domain (WaveShrink estimation in the sense of [4]) involves estimating the wavelet coefficients $\boldsymbol{d} = \{d_i\}_{1\leqslant i\leqslant N}$ of the signal under consideration by $\widehat{\boldsymbol{d}} = \{\delta(c_i)\}_{1\leqslant i\leqslant N}$, where $\delta(\cdot)$ is a thresholding or shrinkage function. This processing is proved to be a relevant strategy when the wavelet transform achieves a sparse representation of the signal in the sense that, among the coefficients d_i, $i = 1,2,\dots,N$, only a few of them have large amplitudes and, as such, characterize the signal. In this respect, thresholding estimators such as "keep or kill" and "shrink or kill" rules are nearly optimal, in the Mean Square Error (MSE) sense, in comparison with oracles (see [4] for further details).

However, wavelet representations of many signals encountered in practical applications such as speech and image processing fail to sparse enough (see examples given in [5]). In order to obtain a shrinkage better adapted for estimating less sparse signals, [1] proposes to construct shrinkage functions satisfying the following properties.

- **(P1) Smoothness**: of the shrinkage function so as to induce small variability among data with close values.
- **(P2) Penalized shrinkage**: a strong (resp. a weak) attenuation is imposed for small (resp. large) data.
- **(P3) Vanishing attenuation at infinity**: the attenuation decreases to zero when the amplitude of the coefficient tends to infinity.

For a signal whose wavelet representation fails to be sparse enough, it is more convenient to impose the penalized shrinkage property **(P2)** instead of zero-forcing since small coefficients may contain significant information about the signal. Property **(P1)** guarantees the regularity of the shrinkage process and the role of property **(P3)** is to avoid over-smoothing of the estimate (noise mainly affect small wavelet coefficients).

2.2 Smooth sigmoid-based shrinkage

The family of real-valued functions defined by

$$\delta_{\tau,\lambda}(x) = \frac{x}{1+e^{-\tau(|x|-\lambda)}}, \tag{2}$$

for $(\tau,\lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+$, are suitable functions for the estimation of less sparse signals since they satisfy **(P1)**, **(P2)** and **(P3)** properties. Each $\delta_{\tau,\lambda}$ is the product of the identity function with a sigmoid-like function. A function $\delta_{\tau,\lambda}$ will hereafter be called a SigShrink (Sigmoid Shrinkage) function. The class of SigShrink functions pertains to a wide class of smooth sigmoid-based shrinkage (SSBS) functions given in [1].

The parameter λ in Eq. (2) is the SigShrink "threshold". This follows from that $\delta_{\tau,\lambda}$ tends to a hard-thresholding function with threshold height λ when τ tends to ∞. For fixed λ, the SigShrink parameter τ controls the attenuation degree we want to impose to the wavelet coefficients in the interval $]0,\lambda[$. Indeed, the points $O = (0,0)$ and $A = (\lambda, \lambda/2)$ are fixed points for the class of SigShrink functions, and the parameter τ relates to the curvature of the SigShrink arc $\widehat{OA}$ in the following sense: if θ is the angle between the fixed vector $\overrightarrow{OA}$ and the vector carried by the tangent to the curve of $\delta_{\tau,\lambda}$ at point A, then $\cos\theta = \overrightarrow{OA}.\overrightarrow{CA}/||\overrightarrow{OA}||.||\overrightarrow{CA}|| = (10+\tau\lambda)\big/\sqrt{5(20+4\tau\lambda+\tau^2\lambda^2)}$, and thus, we derive that $\tau = \tau(\theta,\lambda)$ can be written as a function of θ and λ as follows: $\tau(\theta,\lambda) = 10\lambda^{-1}\left([\sin^2\theta + 2\sin\theta\cos\theta]\big/[5\cos^2\theta - 1]\right)$. The larger θ, the stronger the attenuation of the coefficients with amplitudes less than or equal to λ (see [1]).

In contrast to thresholding functions, SigShrink allows for artifact-free denoising: the smoothness of the SigShrink functions allows for reducing noise without impacting significantly the signal. This is due to the fact that the shrinkage is performed with less variability among coefficients with close values. In contrast to the SURELET parameterization ([3]), SigShrink allows for adjustable denoising: the above interpretation of the SigShrink parameters allows choosing the denoising level. From hard denoising to smooth denoising, there exists a wide class of regularities that can be attained for the denoised signal by adjusting the attenuation degree and threshold. The following Section 2.3 details these adjustable and artifact-free SigShrink properties (the risk comparison is reported in Section 3).

2.3 Smooth adaptation (adjustable and artifact-free denoising)

The shrinkage performed by the SigShrink method is adjustable *via* the attenuation degree θ and the threshold λ.

Figure 2 gives denoising examples for different values of θ and λ. The denoising concerns the 'Lena' image corrupted by independent additive, white and Gaussian noise (AWGN) with standard deviation $\sigma = 35$ (figure 1). The wavelet transform used is the stationary wavelet transform [6] with the Haar wavelet. The thresholds used are the standard minimax and universal thresholds [4]. In these figures, $\text{SigShrink}_{\theta,\lambda}$ stands for the SigShrink function which parameters are θ and λ. The PSNR (Peak Signal to Noise Ratio, in deciBel unit, dB),

$$\text{PSNR} = 10\log_{10}\left(\frac{255^2}{\text{MSE}}\right), \tag{3}$$

is also given to assess the quality of the denoised images.

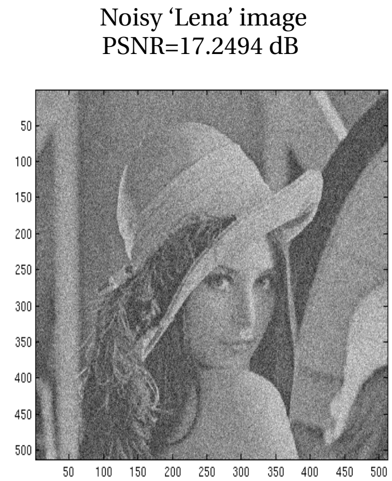

Figure 1: 'Lena' image corrupted by AWGN with standard deviation $\sigma = 35$.

For a fixed attenuation degree (angle θ), we observe that the smoother denoising is obtained with the larger threshold (universal threshold). A small value for the threshold (minimax threshold) leads to better preservation of the textural information contained in the image (compare in figure 2, image (a) versus image (d); image (b) versus image (e); image (c) versus image (f)).

Now, for a fixed threshold λ, the SigShrink shape is controllable *via* angle θ. The attenuation degree θ reflects the regularity of the shrinkage and the attenuation imposed to data with small amplitudes (mainly noise coefficients). The larger θ, the better the noise reduction. However, SigShrink functions are more regular for small values of θ, and thus, small values for θ lead to less artifacts (in figure 2, compare images (d), (e) and (f)).

It follows that SigShrink denoising is flexible thanks to parameters θ and λ, preserves the image features and leads to artifact-free denoising. Note that the SigShrink parameter selection discussed in this section is heuristic in the sense that the threshold is chosen on the basis of the well-known properties of the standard minimax and universal thresholds for discriminating signal and noise wavelet coefficients, whereas the attenuation degree is chosen so as to guarantee a regular (artifact-free) denoising. It is now interesting to compare the risk performance of the SigShrink denoising to that of the most performant shrinkage function, the SURELET approach (SURE optimization for a shrinkage function parameterized as a Linear Expansion of Thresholds, LET) [3]. The following addresses the SURE-based optimization of the SigShrink parameters and provides this comparison.

3. SURE-BASED OPTIMIZATION OF SIGSHRINK PARAMETERS

3.1 SURE-based optimization of SigShrink parameters

Consider the WaveShrink estimation described in section 2.1. The risk function used to measure the accuracy of a

WaveShrink estimator is the standard MSE. This risk is

$$r_\delta(\boldsymbol{d},\widehat{\boldsymbol{d}}) = \frac{1}{N}\mathbb{E}\|\boldsymbol{d}-\widehat{\boldsymbol{d}}\|^2 = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big(d_i-\delta(c_i)\Big)^2 \qquad (4)$$

for a shrinkage function δ. The SURE approach [2] involves estimating unbiasedly the risk $r_\delta(\boldsymbol{d},\widehat{\boldsymbol{d}})$. The SURE optimization then consists in finding the set of parameters that minimizes this unbiased estimate.

From [3, Theorem 1], we have that

$$r_\delta(\boldsymbol{d},\widehat{\boldsymbol{d}}) = \frac{1}{N}\left(||\boldsymbol{d}||^2_{\ell_2}+\sum_{i=1}^{N}\mathbb{E}\left(\delta^2(c_i)-2c_i\delta(c_i)+2\sigma^2\delta'(c_i)\right)\right), \qquad (5)$$

where δ can be any differentiable shrinkage function that does not explode at infinity (see [3] for details). A SigShrink function is such a shrinkage function. The derivate of the SigShrink function $\delta_{\tau,\lambda}$ is

$$\delta'_{\tau,\lambda}(x) = \frac{1+(1+\tau|x|)e^{-\tau(|x|-\lambda)}}{(1+e^{-\tau(|x|-\lambda)})^2}. \qquad (6)$$

Proposition 1 *The quantity $\vartheta + ||\boldsymbol{d}||^2_{\ell_2}/N$, where $||\cdot||_{\ell_2}$ denotes ℓ_2-norm and*

$$\vartheta(\tau,\lambda) = \frac{1}{N}\sum_{i=1}^{N}\frac{2\sigma^2-c_i^2+2(\sigma^2+\sigma^2\tau|c_i|-c_i^2)e^{-\tau(|c_i|-\lambda)}}{(1+e^{-\tau(|c_i|-\lambda)})^2}, \qquad (7)$$

is an unbiased estimator of the risk $r_{\delta_{\tau,\lambda}}(\boldsymbol{d},\widehat{\boldsymbol{d}})$, where $\delta_{\tau,\lambda}$ is a SigShrink function.

As a consequence of proposition 1, we get that minimizing $r_{\delta_{\tau,\lambda}}(\boldsymbol{d},\widehat{\boldsymbol{d}})$ of Eq. (4) amounts to minimizing the unbiased (SURE) estimator ϑ given by Eq. (7). The next section presents experimental tests for illustrating the SURE SigShrink denoising of some natural images corrupted by AWGN. For every tested image and every noise standard deviation considered, the optimal SURE SigShrink parameters are those minimizing ϑ, the values $\{c_i\}_{1\leqslant i\leqslant N}$ representing the wavelet coefficients of the noisy image.

3.2 Experimental results

The SURE optimization approach for SigShrink is now given for some standard test images corrupted by AWGN.

The SigShrink estimation is compared with that of the SURELET “sum of DOGs” (Derivatives Of Gaussian) parameterization. SURELET is a SURE-based method that moreover includes an inter-scale predictor with *a priori* information about the position of significant wavelet coefficients. For the comparison with SigShrink, we only use the “sum of DOGs” parameterization, that is the SURELET method without inter-scale predictor and Gaussian smoothing. By so proceeding, we thus compare two shrinkage functions: SigShrink versus “sum of DOGs”. The sum of DOGs initially proposed in [3] is given by

$$\delta_{a_1,a_2,\ldots,a_K}(x) = \sum_{k=1}^{K} a_k x e^{-\frac{k-1}{2T^2}x^2}, \qquad (8)$$

where T is some pre-specified parameter. After a discussion on the choice of the number K of derivatives in (8), [3] concludes, on the basis of experimental tests, that results are very similar for $K \geqslant 2$ and thus, they consider the following case (where $K = 2$):

$$\delta_{a_1,a_2}(x) = \left(a_1 + a_2 e^{-\frac{1}{12\sigma^2}x^2}\right)x, \qquad (9)$$

where σ is the noise standard deviation.

From Eq. (2), it follows that the difference between a SigShrink function and the sum of DOGs of Eq. (9) are the weights $1/1+e^{-\tau(|x|-\lambda)}$ and $a_1 + a_2 e^{-\frac{1}{12\sigma^2}x^2}$ imposed to the wavelet coefficient represented by the variable x. Since these weights depend on two parameters, we can expect performance of the same order for both the SigShrink and the sum of DOGs parameterizations. Note also that the SigShrink weight is based on the ℓ_1-norm (absolute value of x) whereas the sum of DOGs takes into account the ℓ_2-norm (quadratic) of x. Since the signal representation is assumed to be sparse, the SigShrink might probably lead to slightly better results than the sum of DOGs. These observations are confirmed by the following experimental tests.

In the sequel, the SURE SigShrink parameters (attenuation degree and threshold) are those obtained by performing the SURE optimization at every detail (horizontal, vertical, diagonal) sub-image located at the different resolution levels concerned (4 resolution levels in our tests).

We consider the standard 2-dimensional Discrete Wavelet Transform (DWT) for the PSNR comparison between the two methods since the authors of [3] provides their free MatLab code[1] based on the DWT for the SURELET method. The wavelet used is the Symlet wavelet of order 8 (‘sym8’ in the Matlab Wavelet toolbox) The tests are carried out with the following values for the noise standard deviation: $\sigma = 5, 15, 25, 35$. For every value σ, 25 tests have been performed based on different noise realizations. Every test involves: performing a DWT for the tested image corrupted by AWGN, computing the optimal SURE SigShrink parameters, applying the SigShrink function with these parameters to denoise the wavelet coefficients and building an estimate of the corresponding image by applying the inverse DWT to the shrunken coefficients. For every test, the PSNR is calculated for the original image and the denoised image.

Table 1 gives the average values for the 25 PSNRs obtained by the SURE SigShrink and the SURELET “sum of DOGs” methods.

We use the Matlab routine `fmincon` to compute the optimal SURE SigShrink parameters. This function computes the minimum of a constrained multivariable function by using nonlinear programming methods (see Matlab help for the details).

From table 1, it follows that the two methods yield PSNRs of the same order, with a slight advantage for the SigShrink method.

To conclude this section, we now illustrate, in figure 3, the SigShrink denoising for an ultrasonic image of breast cancer. The contrasts of the denoised SigShrink images are slightly enhanced in order to highlight that SigShrink denoises the image and preserves feature information without introducing artifacts (because of the regularity of the SigShrink function).

[1] avalaible at `http://bigwww.epfl.ch/demo/suredenoising/`

Figure 2: SigShrink denoising of 'Lena' image corrupted by AWGN with standard deviation $\sigma = 35$ (see figure 1). The universal threshold λ_u and the minimax threshold λ_m are used.

Figure 3: SigShrink denoising for an ultrasonic image of breast cancer. The SWT with four resolution levels and the biorthogonal spline wavelet with order 3 for decomposition and with order 1 for reconstruction ('bior1.3' in Matlab Wavelet toolbox) are used. The noise standard deviation is estimated by the median absolute deviation normalised by the constant 0.6745 (see [4]).

Table 1: Average values of the PSNRs computed over 25 noise realizations, when denoising test images by the SURE SigShrink and SURELET parameterizations (without additional processing such as interscale predictors). The tested images are corrupted by AWGN with standard deviation σ. The DWT is computed by using the 'sym8' wavelet.

Images	'House'	'Peppers'	'Barbara'	'Lena'	'Flin'	'Finger'	'Boat'	'Barco'
$\boldsymbol{\sigma} = \mathbf{5}$ ($\Longrightarrow$ Input PSNR = 34.1514).								
SigShrink	37.4880	36.6827	36.3980	37.5518	35.3128	35.8805	36.3608	36.9928
SURELET	37.3752	36.6708	36.3767	37.5023	35.3102	35.9472	36.3489	35.9698
$\boldsymbol{\sigma} = \mathbf{15}$ ($\Longrightarrow$ Input PSNR = 24.6090).								
SigShrink	31.6472	30.0930	29.3972	32.0571	28.3815	29.4191	30.2895	30.4545
SURELET	31.2834	29.9621	29.2817	31.9059	28.3502	29.4365	30.2706	27.4525
$\boldsymbol{\sigma} = \mathbf{25}$ ($\Longrightarrow$ Input PSNR = 20.1720).								
SigShrink	29.2948	27.3111	26.5146	29.7435	25.6407	26.6262	27.8216	27.9599
SURELET	28.8085	26.9941	26.4404	29.5937	25.5953	26.7659	27.8227	23.6221
$\boldsymbol{\sigma} = \mathbf{35}$ ($\Longrightarrow$ Input PSNR = 17.2494).								
SigShrink	27.7840	25.5818	24.8910	28.2782	23.9326	24.9625	26.3764	26.5068
SURELET	27.2768	25.1307	24.8383	28.1462	23.8954	25.0756	26.3880	21.3570

4. CONCLUSION

The paper proposes a SURE approach for estimating the SigShrink (Sigmoid Shrinkage) parameters. It also provides a comparison of the two novel WaveShrink functions, namely the SigShrink function of [1] and the sum of DOGs (also called LET, Linear Expansion of Thresholds) of [3]. The experimental tests show that the SURE SigShrink performs slightly better than the SURE sum of DOGs, with no additional computational load.

In practical applications, the advantage of SigShrink is the spatial adaptation that results from its parameter interpretation: as highlighted in Section 2 above, the function has been constructed in order to perform without necessarily using the SURE optimization. The SigShrink parameters are meaningful: a threshold for discriminating large and small coefficients and an attenuation degree to control the shrinkage imposed to small data. This results in a very flexible shrinkage in the sense that it allows for a fine tuning of the denoising by varying two parameters (see Section 2.3). As a consequence, if the SURE results are not satisfactory, or if we wish to reduce noise without impacting significantly the signal, then we can choose more adapted (smaller) attenuation degree and threshold than those obtained from the SURE approach.

To conclude this work, we can reasonably expect to improve SigShrink denoising performance by introducing inter-scale or/and intra-scale predictor, which could provide information about the position of significant wavelet coefficients. Such predictors are used by the more popular WaveShrink toolboxes, namely the SURELET [3] and the BLS-GSM [7] (free MatLab code [2]).

REFERENCES

[1] 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.

[2] C. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.

[3] 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.

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

[5] 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.

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

[7] 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.

[2] http://decsai.ugr.es/~javier/denoise/software/index.htm