



**HAL**  
open science

## Smooth sigmoid wavelet shrinkage for non-parametric estimation

Abdourrahmane M. Atto, Dominique Pastor, Grégoire Mercier

### ► To cite this version:

Abdourrahmane M. Atto, Dominique Pastor, Grégoire Mercier. Smooth sigmoid wavelet shrinkage for non-parametric estimation. ICASSP 2008: IEEE international conference on acoustics, speech and signal processing, March 30 - April 4, Las Vegas, USA, Mar 2008, Las Vegas, United States. pp.3265 - 3268, <10.1109/ICASSP.2008.4518347>. <hal-02136546>

**HAL Id: hal-02136546**

**<https://hal.science/hal-02136546v1>**

Submitted on 22 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# SMOOTH SIGMOID WAVELET SHRINKAGE FOR NON-PARAMETRIC ESTIMATION

*Abdourrahmane M. ATTO, Dominique PASTOR and Gregoire MERCIER*

Institut TELECOM / CNRS UMR 2872 TAMCIC  
TELECOM Bretagne - Technopôle Brest-Iroise  
CS 83818 - 29238 Brest Cedex 3, FRANCE

## ABSTRACT

This paper presents a new sigmoid-based WaveShrink function. The shrinkage obtained via this function is particularly suitable to reduce noise without impacting significantly the statistical properties of the signal to be recovered. The proposed WaveShrink function depends on a parameter that makes it possible to control the attenuation degree imposed to the data, and thus, allows for a flexible shrinkage.

**Index Terms**— Estimation, Wavelet transforms, Gaussian noise, Image enhancement, Image restoration.

## 1. INTRODUCTION

This paper addresses the choice of the shrinkage function in non-parametric estimation of a signal according to the WaveShrink procedure described in [1]. The method 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 shrunked wavelet coefficients. In Additive, White and Gaussian Noise (AWGN), such estimation is very efficient for smooth signals because the wavelet transform of such signals is sparse: these signals are characterized by few large wavelet coefficients [1].

The contribution of this work is a whole family of smooth and non zero-forcing WaveShrink functions. These functions depend on a parameter that allows for a very flexible adjustment of the shrinkage. The new functions presented in this paper derive from the sigmoid function and, thus, are called Smooth Sigmoid-Based Shrinkage (SSBS) functions. SSBS is an adjustable shrinkage thanks to a parameter that controls the attenuation degree imposed to the wavelet coefficients.

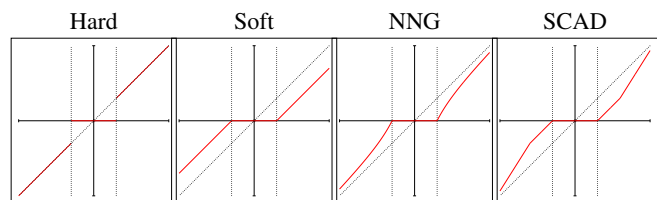
Due to the smoothness of the SSBS functions, the shrinkage is performed with less variability so that the statistical properties of the signal are better preserved than when using standard thresholding functions. Since the SSBS functions introduce little variability among coefficients with closed amplitudes, they achieved an almost artifact-free denoising, which is relevant in many applications. Moreover, the performance of the proposed shrinkage is comparable to that obtained with the BLS-GSM method, introduced in [2] and considered as the best parametric method.

## 2. SMOOTH SIGMOID-BASED SHRINKAGE

### 2.1. SSBS functions

This section introduces a new family of shrinkage functions for wavelet coefficient estimation of a signal corrupted by AWGN. Before that, let us describe very briefly, the limitations of some popular thresholding and shrinkage functions met in the literature on the topic. The standard WaveShrink functions considered are the hard

and soft thresholding functions (proposed in [1]); the Non-Negative Garrote (NNG) function (proposed for wavelet shrinkage in [3]) and the Smoothly Clipped Absolute Deviation (SCAD) function of [4]. The graphs of these functions are given in figure 1. On the first hand,



**Fig. 1.** Some standard WaveShrink functions.

the hard thresholding function is not everywhere continuous and its discontinuities generate a high variance in the estimated signal; on the other hand, the soft thresholding function is continuous, but creates an attenuation on large coefficients, which results in an over-smoothing and an important bias for the estimated signal [3]. The NNG and SCAD functions achieve a certain compromise between the hard and the soft thresholding functions as can be seen in figure 1. However, all the standard WaveShrink functions presented above include zero-forcing. This zero-forcing induces singularities of the thresholding function. As a consequence, this results in a significant variance of the estimation due to the sensitivity of the inverse wavelet transform.

In addition, thresholding rules assume that the wavelet representation is sparse. Note that smooth signals yield sparse wavelet representations in the sense given by [1]: for such signals, large coefficients are very few in number. In contrast, wavelet representations of natural images, which tend to be piecewise regular rather than smooth, fail to be sparse enough since large coefficients are not very few. This justifies the introduction of more flexible WaveShrink methods for correcting the drawbacks of thresholding rules. As an example, a WaveShrink parametrisation such as that proposed in [5] results in non zero-forcing and is shown to be more efficient than soft thresholding estimation.

Therefore, a suitable shrinkage function should satisfy the following two properties. **(P1) Smoothness:** a smooth shrinkage function induces small variability among coefficients with close values. **(P2) Penalized shrinkage:** a strong (resp. a weak) attenuation for small (resp. large) coefficients is required because small (resp. large) coefficients contain less (resp. more) information about the signal.

Now, consider the family of real-valued functions:

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

for  $x \in \mathbb{R}$  and where  $(t, \tau, \lambda) \in \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+$ .

Basically, these functions are suitable shrinkage functions in the sense given above since they satisfy properties **(P1)** and **(P2)**. Each  $\delta_{t,\tau,\lambda}$  is the product of the soft thresholding function with a sigmoid-like function. As such, the functions  $\delta_{t,\tau,\lambda}$  will hereafter be called Smooth Sigmoid-Based Shrinkage (SSBS) functions.

The hard and soft thresholding functions can be regarded as functions  $\delta_{t,\tau,\lambda}$  for degenerated values of parameter  $\tau$ . In fact, for fixed  $t = \lambda > 0$ , the function  $\delta_{t,\tau,t}(x)$  tends to the soft thresholding function  $\text{sgn}(x)(|x| - t)_+$  when  $\tau$  tends to  $+\infty$ . Now, put  $t = 0$  and let  $\tau$  tends to  $+\infty$ . Then,  $\delta_{0,\tau,\lambda}(x)$  tends to  $\delta_{0,\infty,\lambda}(x)$ , which is a hard thresholding function defined by:

$$\delta_{0,\infty,\lambda}(x) = \begin{cases} x \mathbb{1}_{\{|x| > \lambda\}} & \text{if } x \in \mathbb{R} \setminus \{-\lambda, \lambda\}, \\ \pm \lambda / 2 & \text{if } x = \pm \lambda. \end{cases} \quad (2)$$

For fixed values of  $t$ ,  $\tau$  and  $\lambda$ ,  $\delta_{t,\tau,\lambda}(x)$  behaves as the soft thresholding function  $\text{sgn}(x)(|x| - t)_+$  when  $x \rightarrow \infty$ . It follows that parameter  $t$  imposes an attenuation of  $t$ . Such an attenuation may be undesirable when  $t$  is large since, in this case, soft thresholding is known to over-smooth the estimate. For instance, this will occur if  $t$  is either the universal or the minimax threshold, which both tend to  $\infty$  when the sample size tends to  $\infty$ . Therefore, we now restrict our attention to the particular case where  $t = 0$ , the following property being satisfied for this case. **(P3) Vanishing attenuation at infinity**: to avoid over-smoothing of the estimate, the attenuation decreases to zero when the amplitude of the coefficient tends to infinity. In the sequel, we put  $\delta_{\tau,\lambda} = \delta_{0,\tau,\lambda}$ . Therefore, we have

$$\delta_{\tau,\lambda}(x) = \frac{x}{1 + e^{-\tau(|x| - \lambda)}}. \quad (3)$$

The functions  $\delta_{\tau,\lambda}$  not only satisfy properties **(P1)** and **(P2)** but also guarantee vanishing attenuation (property **(P3)**). From now on, we only consider the SSBS functions  $\delta_{\tau,\lambda}$ .

## 2.2. The role played by the SSBS parameters

This section provides an interpretation of the SSBS parameters. First, note that  $\lambda$  plays the role of a threshold since  $\delta_{\infty,\lambda}$  is a hard thresholding function with threshold height  $\lambda$  (see section 2.1 above).

For every given  $\lambda$ , it is easy to see that, in Cartesian-coordinates,  $A = (\lambda, \lambda/2)$ ,  $O = (0, 0)$  and  $A' = (-\lambda, -\lambda/2)$  belong to the curve of every function  $(\delta_{\tau,\lambda})_{\tau > 0}$ . According to Eq. (3), we have  $\delta_{\tau,\lambda}(\pm \lambda) = \pm \lambda/2$  and  $\delta_{\tau,\lambda}(0) = 0$  for any  $\tau > 0$ . As an illustration, some SSBS graphs are plotted in figure 2, for different values of  $\tau$ .

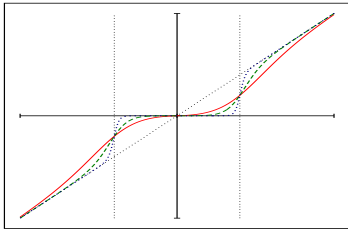


Fig. 2. The SSBS functions  $\delta_{\tau,\lambda}$  for different values of  $\tau$ .

It follows that  $\tau$  parameterizes the curvature of the arc  $\widehat{A'OA}$ , that is, the arc of the SSBS function in the interval  $]-\lambda, \lambda[$ . This curvature directly relates to the attenuation degree we want to apply to the wavelet coefficients. Consider the graph of figure 3, where a SSBS function is plotted in the positive half plan. Due to the anti-symmetry of the SSBS function, we only focus on the curvature of arc  $\widehat{OA}$ . Let  $C$  be the intersection between the abscissa axis and the

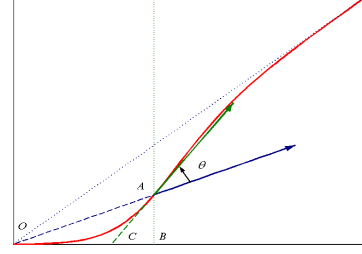


Fig. 3. Graph of  $\delta_{\tau,\lambda}$  in the positive half plan.

tangent at point  $A$  to the curve of the SSBS function, the equation of this tangent being  $y = 0.25(2 + \tau\lambda)(x - \lambda) + 0.5\lambda$ . The coordinates of point  $C$  are  $C = (\tau\lambda^2 / (2 + \tau\lambda), 0)$ . We can easily control the arc  $\widehat{OA}$  curvature via the angle, denoted by  $\theta$ , between vector  $\overrightarrow{OA}$ , which is fixed, and vector  $\overrightarrow{CA}$ , which is carried by the tangent to the curve of  $\delta_{\tau,\lambda}$  at point  $A$ . The larger  $\theta$ , the stronger the attenuation of the coefficients with amplitudes less than or equal to  $\lambda$ . For a fixed  $\lambda$ , the relation between angle  $\theta$  and parameter  $\tau$  is

$$\cos \theta = \frac{\overrightarrow{OA} \cdot \overrightarrow{CA}}{\|\overrightarrow{OA}\| \cdot \|\overrightarrow{CA}\|} = \frac{10 + \tau\lambda}{\sqrt{5(20 + 4\tau\lambda + \tau^2\lambda^2)}}. \quad (4)$$

It easily follows from Eq. (4) that  $\theta \in ]0, \arccos \sqrt{5}/5[$ ; when  $\theta = \arccos \sqrt{5}/5$ , then  $\tau = +\infty$  and  $\delta_{\tau,\lambda}$  is the hard thresholding function of Eq. (2). From Eq. (4), we derive that  $\tau$  can be written as a function of  $\theta$  and  $\lambda$  as follows:

$$\tau(\theta, \lambda) = \frac{10}{\lambda} \frac{\sin^2 \theta + 2 \sin \theta \cos \theta}{5 \cos^2 \theta - 1}. \quad (5)$$

In practice, when  $\lambda$  is fixed, the foregoing makes it possible to control the attenuation degree we want to impose to the data in  $]0, \lambda[$  by choosing  $\theta$ , which is rather natural, and calculating  $\tau$  according to Eq. (5). Since we can control the shrinkage by choosing  $\theta$ ,  $\delta_{\theta,\lambda} = \delta_{\tau(\theta,\lambda),\lambda}$  henceforth denotes the SSBS function where  $\tau(\theta, \lambda)$  is given by Eq. (5).

## 3. EXPERIMENTAL RESULTS

In this section,  $\text{Hard}_\lambda$ ,  $\text{Soft}_\lambda$ ,  $\text{NNG}_\lambda$  and  $\text{SCAD}_\lambda$  denote the Hard, Soft, NNG, and SCAD thresholding functions with threshold height  $\lambda$ , respectively. Similarly,  $\text{SSBS}_{\theta,\lambda}$  will stand for SSBS with parameters  $\theta$  and  $\lambda$ . The Stationary (also call shift-invariant or redundant) Wavelet Transform (SWT) is used for the wavelet decomposition. The analyzing wavelet is the Haar wavelet, and the shrinkage concerns the detail coefficients of the first four decomposition levels.

Many experimental tests show that SSBS achieves better Peak Signal-to-Noise Ratios (PSNRs) than standard thresholding shrinkages and PSNRs close to those attained with BLS-GSM. The PSNR, often used to assess the quality of a compressed image, is given by  $\text{PSNR} = 10 \log_{10}(d^2 / \text{MSE})$ , where  $d$  gives the dynamics of the signal ( $d = 255$  in the standard case of 8 bit-coded images).

More specifically, we consider the  $512 \times 512$  'Lena' image corrupted by AWGN. Table 1 presents the average PSNRs obtained by SSBS as well as hard, soft, NNG and SCAD thresholdings. Each PSNR is computed over 10 noise realizations. The threshold used is either the minimax or the universal threshold. In table 1, the universal (resp. minimax) threshold will be denoted by  $\lambda_u$  (resp.  $\lambda_m$ ). The different values considered for the AWGN standard deviation  $\sigma$  are 5, 15, 25, and 35. Parameter  $\tau$  is obtained from Eq.

(5) by setting  $\theta = \pi/6$  and  $\theta = \pi/4$  respectively. This table shows that SSBS outperforms standard shrinkages. In addition, the performance measurements presented in table 1 suggest comparing the non-parametric SSBS to the parametric BLS-GSM presented in [2]. The latter models neighborhoods of 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 computationally expensive but the results are very good. As mentioned above, BLS-GSM is regarded as the best parametric denoising method. For example, the results obtained by BLS-GSM for the same input PSNRs as those used in the preceding tests are given in table 2 for a comparison to SSBS.

**Table 1.** Average PSNRs computed over 10 AWGN realizations. PSNR<sub>0</sub> denotes the input PSNR.

Method	$\sigma = 5$		$\sigma = 15$		$\sigma = 25$	
	PSNR <sub>0</sub> = 34.15	PSNR <sub>0</sub> = 24.61	PSNR <sub>0</sub> = 24.61	PSNR <sub>0</sub> = 20.17	PSNR <sub>0</sub> = 20.17	PSNR <sub>0</sub> = 20.17
	$\lambda = \lambda_u, \lambda = \lambda_m$	$\lambda = \lambda_u, \lambda = \lambda_m$	$\lambda = \lambda_u, \lambda = \lambda_m$	$\lambda = \lambda_u, \lambda = \lambda_m$	$\lambda = \lambda_u, \lambda = \lambda_m$	$\lambda = \lambda_u, \lambda = \lambda_m$
Hard <sub><math>\lambda</math></sub>	35.37	36.31	30.41	31.50	28.18	29.31
Soft <sub><math>\lambda</math></sub>	31.57	32.97	27.03	28.20	25.33	26.34
NNG <sub><math>\lambda</math></sub>	33.76	35.16	28.56	29.92	26.50	27.72
SCAD <sub><math>\lambda</math></sub>	33.03	34.50	27.87	29.21	25.87	27.05
SSBS <sub><math>\frac{\pi}{6}, \lambda</math></sub>	36.34	37.38	31.09	31.83	28.75	29.12
SSBS <sub><math>\frac{\pi}{4}, \lambda</math></sub>	35.77	36.90	30.68	31.94	28.40	29.63

**Table 2.** PSNRs computed when denoising the ‘Lena’ image corrupted by AWGN. The SSBS <sub>$\theta, \lambda$</sub>  PSNRs given are the best PSNRs achieved when  $\theta = \pi/6, \pi/4$  and  $\lambda = \lambda_u, \lambda_m$ .

Method	$\sigma = 5$		$\sigma = 15$		$\sigma = 25$	
	PSNR <sub>0</sub> = 34.15	PSNR <sub>0</sub> = 24.61	PSNR <sub>0</sub> = 24.61	PSNR <sub>0</sub> = 20.17	PSNR <sub>0</sub> = 20.17	PSNR <sub>0</sub> = 20.17
SSBS	37.38	31.94	31.94	29.63	29.63	29.63
LD-SSBS	37.72	32.57	32.57	30.26	30.26	30.26
BLS-GSM	38.22	33.52	33.52	31.28	31.28	31.28

According to these results, SSBS is about 1.5 dB less efficient than BLS-GSM in terms of PSNR. In fact, it is very easy to improve the PSNRs of SSBS by introducing some level dependency, interscale dependency or block-thresholding procedures. For instance, if we proceed by adjusting the SSBS method at every decomposition level  $j$  with the threshold value  $\lambda/2^{(j-1)/2}$ , the PSNRs obtained with this Level Dependent SSBS (LD-SSBS) are those given in table 2.

At this stage, it is now relevant to consider and assess the visual quality of denoised images obtained via the different shrinkage methods considered in this paper. Denoised images obtained via BLS-GSM, SSBS and LD-SSBS are given in figure 4. The 512 × 512 ‘Lena’ image is corrupted by AWGN with standard deviation  $\sigma = 35$ . For the SSBS method, the thresholds are again the universal and the minimax thresholds computed over the whole original image. As far as LD-SSBS is concerned, these thresholds are adjusted at each decomposition level as described above. Parameter  $\tau$  is obtained from Eq. (5) by setting  $\theta = \pi/6$ .

We observe that SSBS denoising achieves a good trade-off between noise reduction, texture/edge enhancement and computational load and is almost free of artifacts. In contrast, LD-SSBS and BLS-GSM better restore textures and edges but introduce artifacts that are easily visible in figure 4.

In addition, the following experimental results suggest that SSBS better preserves the statistical properties of the signal of interest than standard thresholding functions, because SSBS yields coefficients that are approximately distributed as the signal coefficients.

Consider the Generalized Gaussian Distribution (GGD) with parameters  $\alpha$  and  $\beta$ , defined for every real value  $x$  by:  $p_{\alpha, \beta}(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta}$ , where  $\Gamma$  is given for  $z > 0$  by  $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ . The GGD models are sometimes suitable for describing distributions of wavelet coefficients of real-world signals. For instance, the horizontal detail coefficients located at the first decomposition level of the SWT of the ‘Lena’ image can be approximately modeled by the GGD with  $\alpha = 1.1$  and  $\beta = 0.56$  (see figure 5 (a)).

Now, consider at the same decomposition level, the horizontal detail SWT coefficients of a noisy ‘Lena’ image; noise being AWGN with standard deviation  $\sigma = 35$ . The histogram of these coefficients is plotted in figure 5 (b). The flexibility of the SSBS function makes it possible to obtain distributions of the shrunk coefficients close to the distribution of the original wavelet coefficients, as can be seen in figure 5 (c-d). In contrast, this could not be achieved with standard shrinkage functions. In fact, the distributions obtained for the hard, soft, NNG and SCAD shrinkages deviate significantly from the distribution of the original image wavelet coefficients since these shrinkages force to zero too many small coefficients. Figure 5 (e-h) gives the histograms of the coefficients resulting from the hard, soft, NNG and SCAD shrinkages. The threshold is set to  $\sigma$  so as to emphasize the effect of the zero-forcing on the distribution of the shrunk coefficients. The minimax and the universal thresholds being larger than  $\sigma$ , the results are even worse with these standard thresholds: a very high peak at zero and very few large coefficients scattered on both sides of this peak.

## 4. CONCLUSION

In this work, we have introduced a whole family of shrinkage functions, the SSBS functions. The SSBS functions are smooth and perform penalized and very regular wavelet shrinkage. This shrinkage is non-parametric because no prior information about the signal is taken into account. The SSBS method is simple and flexible in the sense that SSBS functions allow for a fine tuning of the shrinkage.

The experimental results of section 3 show that SSBS achieves almost artifact-free denoising and good PSNRs, even in comparison to BLS-GSM. We can reasonably expect to improve SSBS performance by taking the intrascale and interscale dependencies into account.

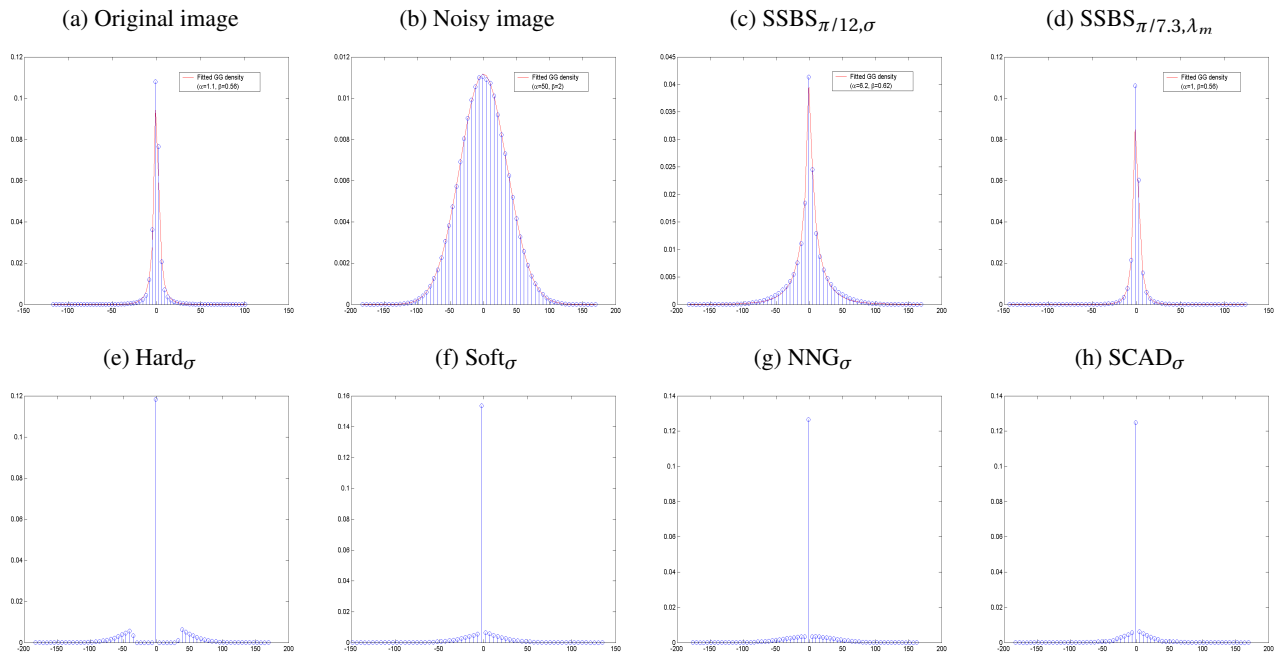
Extension of this work involves addressing the influence and the optimization of the three parameters  $t$ ,  $\tau$  and  $\lambda$ . In particular, Stein’s unbiased risk estimate or its variant stated in [5, Theorem 1] could be used as a criterion to optimize the choice of these parameters.

## 5. REFERENCES

- [1] D.L. Donoho and I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.
- [2] 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.
- [3] H.Y. Gao, “Waveshrink shrinkage denoising using the non-negative garrote,” *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, 1998.
- [4] A. Antoniadis and J. Fan, “Regularization of wavelet approximations,” *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 939–955, Sep. 2001.



**Fig. 4.** BLS-GSM and SSBS denoising of ‘Lena’ image corrupted by AWGN with standard deviation  $\sigma = 35$ .



**Fig. 5.** (a-b) Histograms of the detail coefficients located at the first decomposition level of the ‘Lena’ image and a noisy ‘Lena’ image; noise is AWGN with  $\sigma = 35$ . The histograms are fitted with Generalized Gaussian densities. (c-h) Histograms of the shrunken noisy detail SWT coefficients located at the first decomposition level of the SWT of ‘Lena’ image.