

Adaptive parameter selection for block wavelet-thresholding deconvolution

F. Navarro ^{*,**} M.J. Fadili ^{**} C. Chesneau ^{***}

^{*} LMNO CNRS-Université de Caen, 14032 Caen Cedex, France
(e-mail: Fabien.Navarro@math.unicaen.fr).

^{**} GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen Cedex, France (e-mail: Jalal.Fadili@greyc.ensicaen.fr).

^{***} LMNO CNRS-Université de Caen, 14032 Caen Cedex, France
(e-mail: Christophe.Chesneau@math.unicaen.fr).

Abstract: In this paper, we propose a data-driven block thresholding procedure for wavelet-based non-blind deconvolution. The approach consists in appropriately writing the problem in the wavelet domain and then selecting both the block size and threshold parameter at each resolution level by minimizing Stein's unbiased risk estimate. The resulting algorithm is simple to implement and fast. Numerical illustrations are provided to assess the performance of the estimator.

Keywords: Inverse problem; Estimation parameters; Deconvolution; Signal reconstruction; Coloured noise.

1. INTRODUCTION

1.1 Problem statement

Let $A : H \rightarrow K$ be a known linear operator between the Hilbert spaces H and K endowed with the corresponding inner products and associated norms. Assume that we observe the data Y according to

$$Y = Af + \epsilon W, \quad (1)$$

where $f \in H$ is the unknown function we wish to recover, $\epsilon > 0$ is the noise level and W is a zero-mean Gaussian white noise. We aim to estimate a function f from the observations Y .

1.2 Motivating example

In this paper, we consider the problem of deconvolution, but other examples can also be considered as well, e.g., density with error in variable, tomography, etc., if the operator possesses K the so-called Wavelet-Vaguelette Decomposition (WVD) of Donoho [1995]. The degradation process can be modelled as a blurring operator A which is characterised by the kernel g (i.e. also often referred to as a point spread function (PSF) or impulse response)

$$Af : \Omega \rightarrow \mathbb{R}$$

$$t \mapsto Af(t) = g \otimes f(t) = \int_{\Omega} g(s)f(t-s)ds, \quad t \in \Omega$$

where $\Omega = [0, 1]$ and both f and g belong to $H = \mathbb{L}_{\text{per}}^2(\Omega) = \{h; h \text{ is 1-periodic on } \Omega \text{ and } \int_{\Omega} |h(t)|^2 dt < \infty\}$. Deconvolution is a challenging ill-posed inverse problem.

1.3 Overview of previous work and contributions

There is already an extensive literature on deconvolution for the case where A is known (see, e.g., Donoho [1995],

Cavalier [2008], Cavalier et al. [2004], Cavalier [2002], Chesneau [2007], Johnstone [1999]).

In this work, we consider adaptive block estimators based on the Wavelet-Vaguelette Decomposition (WVD) of Donoho [1995]. Minimality of James-Stein block thresholding over Besov balls with fixed block size has been established for deconvolution in Chesneau et al. [2010]. The performance of Block thresholding estimators strongly depends on the block size L and threshold level λ . In the nonparametric regression setting, in order to choose these two key parameters in an optimal way, Cai and Zhou [2009] proposed an adaptive James-Stein block thresholding estimator whose parameters minimize the Stein's unbiased risk estimate (SURE) and established its minimax rates of convergence under the mean squared error over Besov balls. Johnstone [1999] described a SURE-based level-dependent term-by-term thresholding estimator for a class of linear inverse problems possessing a WVD. Recently, the SURE has proved to be a powerful tool for signal/image restoration Luisier et al. [2007], Pesquet et al. [2009], Vaiter et al. [2012]. For image denoising problems, a Stein risk estimator have been proposed in Peyré et al. [2011] to both adapt the block-sparsity structure and the threshold.

In this work, we extend several of the previous works to solve linear inverse problems, mainly deconvolution here, with wavelet block thresholding estimators. Our estimator is constructed using a periodized Meyer wavelet basis and two popular block thresholding estimators; namely Block-Soft and Block James-Stein (JS) (defined in the sequel). The corresponding optimal block size and threshold parameter are determined by minimizing the SURE after rewriting the model appropriately in the wavelet domain.

The paper is organized as follows. Section 2 briefly reviews wavelets and Besov balls and describes the block thresholding-based deconvolution estimator. Section 3 dis-

cusses numerical results, before conclusions are drawn in Section 4.

2. NONLINEAR ESTIMATION VIA BLOCK THRESHOLDING

2.1 Wavelets and Besov balls

For the purpose of this paper, we use the periodized Meyer wavelet bases on $\mathbb{L}_{\text{per}}^2(\Omega)$. Any function $f \in \mathbb{L}_{\text{per}}^2(\Omega)$ can be expanded into a wavelet series

$$f(t) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(t), \quad t \in \Omega,$$

where

$$\alpha_{j_0,k} = \langle f, \phi_{j_0,k} \rangle, \quad \beta_{j,k} = \langle f, \psi_{j,k} \rangle \quad (2)$$

with $\langle f, g \rangle = \int_{\Omega} f \bar{g}$, and $\phi_{j_0,k}$ and $\psi_{j,k}$ are scaled and translated versions of the father and mother wavelets ϕ and ψ . The family $\{\phi_{j_0,k}, k = 0, \dots, 2^{j_0} - 1; \psi_{j,k}, j \geq j_0, k = 0, \dots, 2^j - 1\}$ forms an orthonormal basis of $\mathbb{L}_{\text{per}}^2(\Omega)$. Meyer wavelets are compactly supported in the Fourier domain with

$$\begin{cases} \mathcal{D}_j = \text{supp}((\phi_{j,k})_{\ell}) \subset [-2^{j+1}c, 2^{j+1}c], \\ \mathcal{C}_j = \text{supp}((\psi_{j,k})_{\ell}) \subset [-2^{j+2}c, -2^j c] \cup [2^j c, 2^{j+2}c], \end{cases}$$

where $c = 2\pi/3$, $g_{\ell} = \langle g, e_{\ell} \rangle$ with $e_{\ell} = e^{2\pi i \ell t}$, and supp denotes the support. See Meyer [1992] for a detailed account.

We say that a function f belongs to the Besov ball $\mathbf{B}_{p,r}^s(M)$ if and only if $\int_{\Omega} |f(t)|^p dt \leq M$, and there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (2) satisfy

$$\begin{aligned} & 2^{j_0(1/2-1/p)} \left(\sum_{k=0}^{2^{j_0}-1} |\alpha_{j_0,k}|^p \right)^{1/p} + \\ & \left(\sum_{j=j_0}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*, \end{aligned}$$

with a smoothness parameter $0 < s < q$, and the norm parameters p and r . Besov spaces capture a variety of smoothness features in a function including spatially inhomogeneous behavior, see Meyer [1992].

2.2 Smoothness of the kernel g

In this study, we focus on the standard context of ordinary smoothness of g . That is, that there exist three constants, $c_g > 0$, $C_g > 0$ and $\delta > 1$ such that, for any $\ell \in \mathbb{Z}$,

$$c_g \frac{1}{(1 + \ell^2)^{\delta/2}} \leq |g_{\ell}| \leq C_g \frac{1}{(1 + \ell^2)^{\delta/2}}. \quad (3)$$

This assumption controls the decay of the Fourier coefficients of g . It is a standard hypothesis usually adopted in the field of nonparametric estimation for deconvolution problems, see e.g. Cai [2002], Donoho [1995], Pensky and Vidakovic [1999].

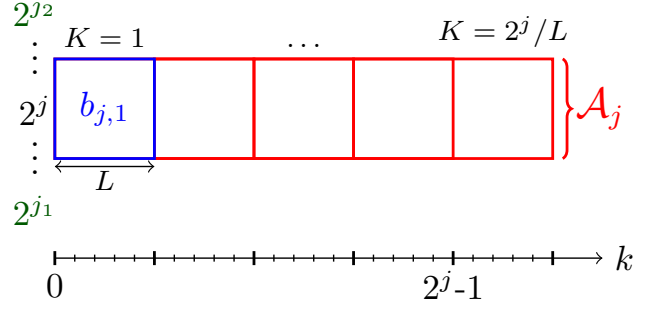


Fig. 1. Illustration of the blocks partition at a given scale.

2.3 Wavelet deconvolution in the Fourier domain

We set, $f_{\ell} = \langle f, e_{\ell} \rangle$, $W_{\ell} = \langle W, e_{\ell} \rangle$, where W_{ℓ} are i.i.d. standard (complex-valued) Gaussian random variables. Then, re-writing the model (1) in the Fourier domain

$$y_{\ell} = f_{\ell} g_{\ell} + \epsilon W_{\ell}, \quad \ell \in \mathbb{Z}. \quad (4)$$

Moreover, we can compute the wavelet transform via the Parseval formula

$$\beta_{j,k} = \sum_{\ell \in \mathbb{Z}} f_{\ell} (\psi_{j,k})_{\ell},$$

where $(\cdot)_{\ell} = \langle \cdot, e_{\ell} \rangle$. Then, combining the latter with (4), we get

$$\sum_{\ell \in \mathcal{C}_j} \frac{y_{\ell}}{g_{\ell}} (\psi_{j,k})_{\ell} = \sum_{\ell \in \mathcal{C}_j} f_{\ell} (\psi_{j,k})_{\ell} + \epsilon \sum_{\ell \in \mathcal{C}_j} \frac{e_{\ell}}{g_{\ell}} (\psi_{j,k})_{\ell}.$$

Thus, we can perform wavelet transform and deconvolution simultaneously and (4) reduces to a general sequence model with correlated noise

$$z_{j,k} = \beta_{j,k} + \epsilon W_{j,k}, \quad (5)$$

where $W_{j,k} = \sum_{\ell \in \mathcal{C}_j} \frac{e_{\ell}}{g_{\ell}} (\psi_{j,k})_{\ell} \sim \mathcal{N}\left(0, \epsilon^2 \sum_{\ell \in \mathcal{C}_j} \frac{|(\psi_{j,k})_{\ell}|^2}{|g_{\ell}|^2}\right)$.

Let us denote by σ_j the standard deviations of the noise process $W_{j,k}$ and let $\lambda_j = \lambda \sigma_j$ be a sequence of thresholds to be applied to the wavelet coefficients at level j . In view of (3), we have at level j that $\sigma_j \sim \epsilon 2^{j\delta}$. We define the estimator of the wavelet coefficients of f as

$$\widehat{\beta}_{j,k}^* = \eta^*(z_{j,k}, \lambda_j), \quad (6)$$

where η^* is either Block-Soft or Block-JS thresholding defined in (8).

2.4 Block deconvolution estimator

Let the observed signal Y be the vector of n observations Y_1, \dots, Y_n . Let $j_1 = \lfloor \ln_2(\ln n) \rfloor$ be the coarsest resolution level, and $j_2 = \lfloor (1/(2\delta + 1)) \ln_2(n/\ln n) \rfloor$, where, for any $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the integer part of a . For any $j \in \{j_1, \dots, j_2\}$, let $L \geq 1$ be the block size.

Let $A_j = \{1, \dots, \lfloor 2^j L^{-1} \rfloor\}$ be the set indexing the blocks at resolution j . For each j , let $\{b_{j,K}\}_{K \in A_j}$ be a uniform and disjoint open covering of $\{0, \dots, 2^j - 1\}$, i.e. $\bigcup_{K \in A_j} b_{j,K} = \{0, \dots, 2^j - 1\}$ and for any $(K, K') \in A_j^2$ with $K \neq K'$, $b_{j,K} \cap b_{j,K'} = \emptyset$ and $\text{Card}(b_{j,K}) = L$, where $b_{j,K} = \{k \in \{0, \dots, 2^j - 1\}; (K-1)L \leq k \leq KL - 1\}$ is the K th block (see Fig.1).

We define our block estimator \widehat{f}_λ of f by

$$\widehat{f}_\lambda(t) = \sum_{k=0}^{2^{j_1}-1} \widehat{\alpha}_{j_1,k} \phi_{j_1,k}(t) + \sum_{j=j_1}^{j_2} \sum_{K \in A_j} \sum_{k \in b_{j,K}} \widehat{\beta}_{j,k}^* \psi_{j,k}(t), \quad (7)$$

$t \in [0, 1]$, where for any resolution level j and position $k \in b_{j,K}$ within the K th block, the wavelet coefficients of f are estimated via (6) with one of the rules

$$\begin{aligned} \eta^{\text{Soft}}(z_{j,k}, \lambda) &= \left(1 - \frac{\lambda}{\|z_{b_{j,K}}\|_2}\right)_+ z_{j,k}, \\ \eta^{\text{JS}}(z_{j,k}, \lambda) &= \left(1 - \frac{\lambda^2}{\|z_{b_{j,K}}\|_2^2}\right)_+ z_{j,k}, \end{aligned} \quad (8)$$

with, $\|z_{b_{j,K}}\|_2^2 = \sum_{k \in b_{j,K}} |z_{j,k}|^2$, and $(a)_+ = \max(a, 0)$. $\widehat{\alpha}_{j_1,k} = \sum_{\ell \in \mathcal{D}_{j_1}} \frac{y_\ell}{g_\ell} (\phi_{j_1,k})_\ell$ is the empirical approximation coefficient.

Note that (7) is simple to implement and fast (thanks to the Fast Fourier Transform algorithm) which allows us to perform the selection procedure in a reasonable time.

Block thresholding estimators depend on both the block size L and threshold parameter λ which mainly determine the performance of the resulting estimator. So it is important to select these parameters in an optimal way.

2.5 Unbiased risk estimation for automatic parameter selection

Since the discrete wavelet transform is orthogonal, the risk of an estimator \widehat{f} of f will be the same as that of its wavelets coefficients. The risk $R(\widehat{f}, f) = \mathbb{E}(\|\widehat{f} - f\|^2)$ will measure the estimation error in the Mean Square Error sense. Since the blocks are also non-overlapping, an estimator of the risk is derived by summing the risk estimates over blocks and resolution levels j .

More precisely, since each of the thresholding rules in (8) is weakly differentiable with an essentially bounded gradient, Stein lemma allows to get an estimator of the risk in each block, which solely depends on the observation. Denoting $z_{b_{j,K}} = (z_{j,k})_{k \in b_{j,K}}$, the SURE on each block $b_{j,K}$ is given by

$$\begin{aligned} J^{\text{Soft}}(z_{b_{j,K}}, \lambda_j, \sigma_j) &= L\sigma_j^2 + (L\|z_{b_{j,K}}\|^2 - 2L\sigma_j^2) \mathbb{1}_{\{\|z_{b_{j,K}}\| < \lambda_j\}} \\ &\quad + (L\lambda_j^2 - 2\sigma_j^2(L-1)) \frac{\lambda_j}{\|z_{b_{j,K}}\|} \mathbb{1}_{\{\|z_{b_{j,K}}\| \geq \lambda_j\}} \\ J^{\text{JS}}(z_{b_{j,K}}, \lambda_j, \sigma_j) &= L\sigma_j^2 + (L\|z_{b_{j,K}}\|^2 - 2L\sigma_j^2) \mathbb{1}_{\{\|z_{b_{j,K}}\| < \lambda_j\}} \\ &\quad + \frac{L\lambda_j^2 - 2\sigma_j^2(L-2)}{\|z_{b_{j,K}}\|^2/\lambda_j^2} \mathbb{1}_{\{\|z_{b_{j,K}}\| \geq \lambda_j\}}, \end{aligned}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function of the event in its argument. Thus, the overall SURE of $\widehat{\beta}_{j,k}^*$ is computed by simple summation

$$J^*(\{z_{j,k}\}_{j,k}, \lambda) = \sum_{j=j_1}^{j_2} \sum_{K \in A_j} \sum_{k \in b_{j,K}} J^*(z_{b_{j,K}}, \lambda_j, \sigma_j). \quad (9)$$

This is indeed an unbiased estimator of the risk. Consequently, our approach can be used as a principled way to objectively choose the optimal parameters λ and L that minimize (9).

2.6 Deconvolution algorithm

The deconvolution algorithm can be summarized as follows:

Algorithm 1. deconvolution algorithm

Parameters: The observed blurred and noisy signal Y , the PSF g .

Initialization:

- Block size: $L = \lfloor \ln n \rfloor$.
- Coarsest decomposition scale $j_1 = \lfloor \ln_2(\ln n) \rfloor$.

Step one: Fourier and wavelet domain.

- Apply an inverse filtering on the data using FFT (i.e. $1/g_\ell$).
- Perform one Forward wavelet transform: $z_{j,k} = \sum_{\ell \in \mathcal{C}_j} \frac{y_\ell}{g_\ell} (\psi_{j,k})_\ell$.
- Compute and store the variance σ_j^2 .

Step two: Compute the risk.

- For each possible dyadic block $b_{j,K}$ compute $J^*(z_{b_{j,K}}, \lambda_j, \sigma_j)$ and sum over the scale to get $J^*(\{z_{b_{j,K}}\}, \lambda_j, \sigma_j)$.

Step three: Optimal parameters.

- Get the optimal parameters λ and L by a grid search.

Output: Get deconvolved signal \widehat{f}_λ at the optimal λ and L .

3. SIMULATION EXPERIMENTS

Three test functions, representing different degrees of irregularity, were used. The signals have been blurred with a Laplacian PSF and were corrupted by a zero-mean white Gaussian noise. Fig. 2 depicts the SURE and true risk curves as a function of the block size and the threshold λ , for each thresholding rule, plotted at each scale with the optimal block size (1st row) and summed over scales for each block size used (2nd row). It can be observed that the SURE gives a very reliable estimate for the risk, and in turn, also a high-quality estimate of the optimal λ . The restoration quality can be assessed both visually and quantitatively in Fig. 3, which shows the restored signals at the optimally chosen λ and L , with an input BSNR = 30 dB.

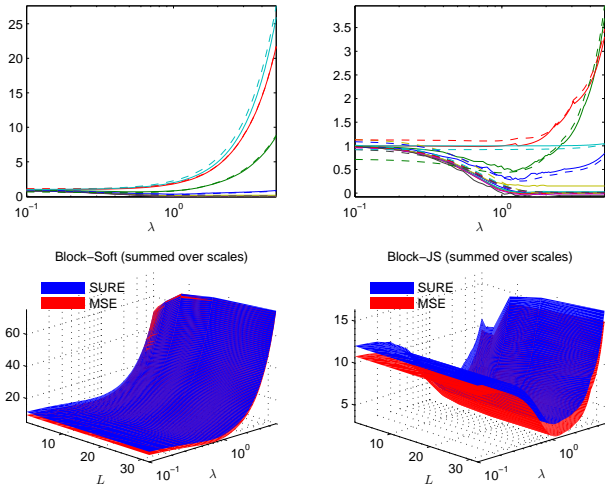
4. CONCLUSION

In this paper, a data-driven block wavelet-based deconvolution estimator was presented. Its usefulness has been illustrated on automatic selection of the both the block size L and threshold level λ for signal deconvolution. Although we focused on one-dimensional convolution, the approach can handle other operators A and easily extended to a multi-dimensional setting (e.g. images). We reported some simulation experiments to support our findings. Our efforts are now directed towards establishing the theoretical minimax performance of this estimator.

REFERENCES

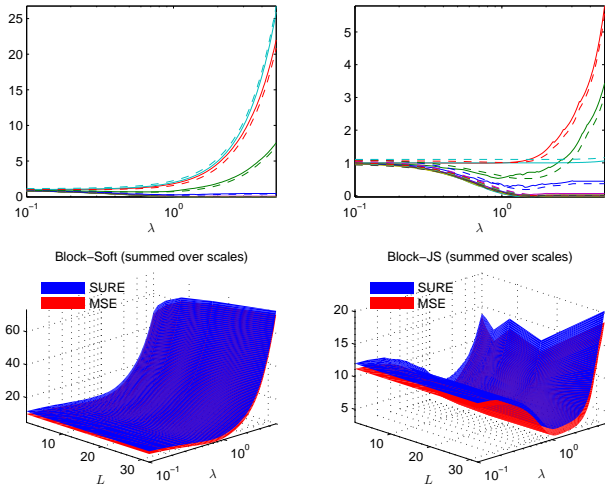
- T. Cai. On adaptive wavelet estimation of a derivative and other related linear inverse problems. *J. Statist. Plann. Inference*, volume 108, pages 329–349, 2002.

Block-Soft Solid: SURE Dashed: MSE (shown per scale) Block-JS Solid: SURE Dashed: MSE (shown per scale)



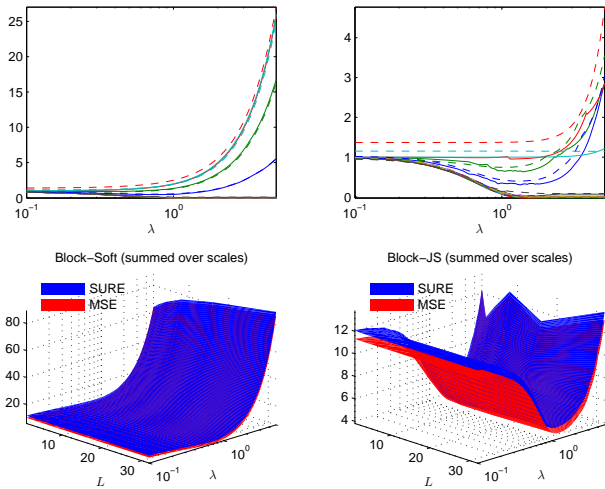
(a) Piece-Polynomial

Block-Soft Solid: SURE Dashed: MSE (shown per scale) Block-JS Solid: SURE Dashed: MSE (shown per scale)



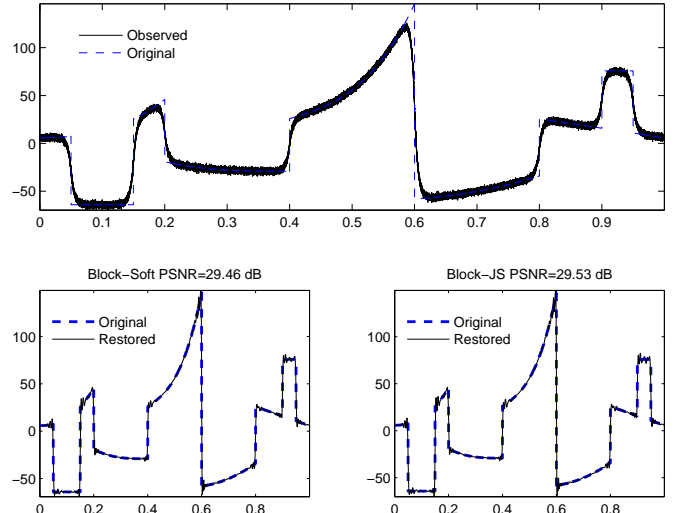
(b) Piece-Regular

Block-Soft Solid: SURE Dashed: MSE (shown per scale) Block-JS Solid: SURE Dashed: MSE (shown per scale)

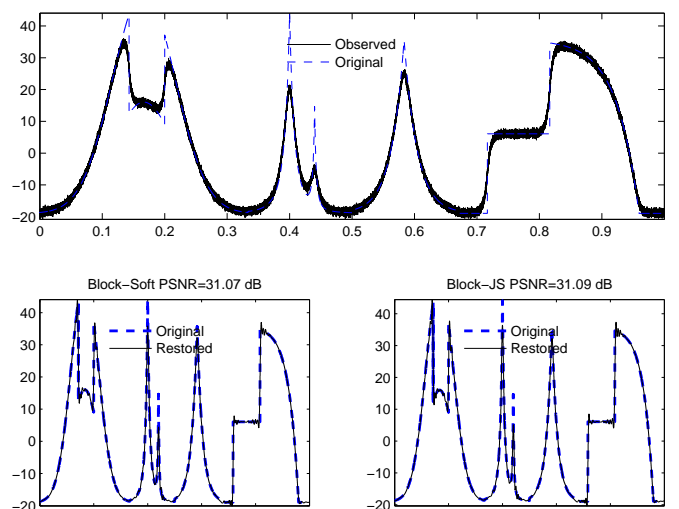


(c) Bumps

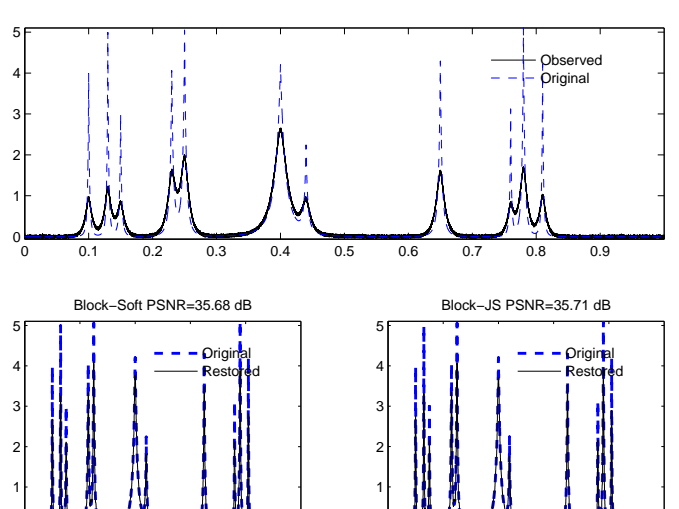
Fig. 2. SURE and risk as a function of the threshold for Soft (left) and JS (right) block thresholding both at each scale with the optimal block size used (top) and summed over scales for each block size used (bottom).



(a) Piece-Polynomial



(b) Piece-Regular



(c) Bumps

Fig. 3. Top panel: Original and observed signal (top); Bottom panels: Block-Soft (left) and Block-JS (right) estimates at the optimal λ and L .

- T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. *Ann. Stat.*, volume 37, pages 569–595, 2009.
- L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, volume 24 (2008).
- L. Cavalier and A.B. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory and Related Fields*, volume 123, pages 323–354, 2002.
- L. Cavalier, Y. Golubev, O. Lepski and A.B. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Probab. Theory and Related Fields*, volume 48, pages 426–446, 2004.
- C. Chesneau. Wavelet estimation via block thresholding: A minimax study under the L^p risk. *Stat. Sinica*, volume 18, pages 1007–1024, 2008.
- C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for wavelet-based image deconvolution. *Elec. J. Statistics*, volume 4, pages 415–435, 2010.
- D.L. Donoho. Nonlinear Solution of Inverse Problems by Wavelet-Vaguelette Decomposition. *Appl. Comput. Harmon. Anal.*, volume 2, pages 101–126, 1995.
- J. Fan and J.Y. Koo. Wavelet deconvolution. *IEEE Trans. Inf. Theory*, volume 48, pages 734–747, 2002.
- J. Johannes. Deconvolution with unknown error distribution. *Ann. Stat.*, volume 38, pages 2301–2323, 2009.
- I. Johnstone. Wavelet Shrinkage For Correlated Data And Inverse Problems: Adaptivity Results. *Statistica Sinica*, volume 9, pages 51–83, 1999.
- F. Luisier, T. Blu, and M. Unser, A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding. *IEEE Trans. Image Process.*, volume 16, pages 593–606, 2007.
- Y. Meyer. Wavelets and Operators. *Cambridge University Press*, Cambridge, 1992.
- M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat.*, volume 27, pages 2033–2053, 1999.
- J.C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, volume 57, pages 4616–4632, 2009.
- G. Peyré, J. Fadili and C. Chesneau. Adaptive Structured Block Sparsity Via Dyadic Partitioning. *Proc. EUSIPCO 2011*, pages 303–307, 2011.
- S. Vaïter, C. Deledalle, G. Peyré, J. Fadili, and C. Dossal. Local behavior of sparse analysis regularization: Applications to risk estimation. *to appear in Appl. Comput. Harmon. Anal.*, 2012.