Stein Block Thresholding For Image Denoising

C. Chesneau [a],* J. Fadili [b] J.-L. Starck [c]

[a]*Laboratoire de Mathématiques Nicolas Oresme, CNRS-Université de Caen, Campus II, Science 3, 14032, Caen Cedex, France.*

[b]*GREYC CNRS-ENSICAEN-Université de Caen, Image Processing Group, 14050, Caen Cedex, France.*

[c]*Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, 91191 Gif-Sur-Yvette cedex, France.*

Abstract

In this paper, we investigate the minimax properties of Stein block thresholding in any dimension d with a particular emphasis on $d = 2$. Towards this goal, we consider a frame coefficient space over which minimaxity is proved. The choice of this space is inspired by the characterization provided in [4] of family of smoothness spaces on $\mathbb{R}^d$, a subclass of so-called decomposition spaces [23]. These smoothness spaces cover the classical case of Besov spaces, as well as smoothness spaces corresponding to curvelet-type constructions. Our main theoretical result investigates the minimax rates over these decomposition spaces, and shows that our block estimator can achieve the optimal minimax rate, or is at least nearly-minimax (up to a log factor) in the least favorable situation. Another contribution is that the minimax rates given here are stated for a general noise sequence model in the transform coefficient domain beyond the usual i.i.d. Gaussian case. The choice of the threshold parameter is theoretically discussed and its optimal value is stated for some noise models such as the (non-necessarily i.i.d.) Gaussian case. We provide a simple, fast and a practical procedure. We also report a comprehensive simulation study to support our theoretical findings. The practical performance of our Stein block denoising compares very favorably to the BLS-GSM state-of-the art denoising algorithm on a large set of test images. A toolbox is made available for download on the Internet to reproduce the results discussed in this paper.

Key words: block denoising, Stein block, wavelet transform, curvelet transform, fast algorithm

---

* Corresponding author: chesneau@math.unicaen.fr

arXiv:0809.3486v1 [math.ST] 20 Sep 2008

1 Introduction

Consider the nonparametric regression model:

$$Y_{\mathbf{i}} = f(\mathbf{i}/n) + \epsilon_{\mathbf{i}}, \qquad \mathbf{i} \in \{1, ..., n\}^d, \tag{1.1}$$

where $d \in \mathbb{N}^*$ is the dimension of the data, $(Y_{\mathbf{i}})_{\mathbf{i} \in \{1,...,n\}^d}$ are the observations regularly sampled on a d-dimensional Cartesian grid, $(\epsilon_{\mathbf{i}})_{\mathbf{i} \in \{1,...,n\}^d}$ are independent and identically distributed (i.i.d.) $\mathcal{N}(0,1)$, and $f : [0,1]^d \to \mathbb{R}$ is an unknown function. The goal is to estimate f from the observations. We want to build an adaptive estimator $\widehat{f}$ (i.e. its construction depends on the observations only) such that the mean integrated squared error (MISE) defined by $R(\widehat{f}, f) = \mathbb{E}\left(\int_{[0,1]^d} \left(\widehat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x}\right)$ is as small as possible for a wide class of f. A now classical approach to the study of nonparametric problems of the form (1.1) is to, first, transform the data to obtain a sequence of coefficients, second, analyze and process the coefficients (e.g. shrinkage, thresholding), and finally, reconstruct the estimate from the processed coefficients. This approach has already proven to be very successful by several authors and a good survey may be found in [28, 29, 30]. In particular, it is now well established that the quality of the estimation is closely linked to the sparsity of the sequence of coefficients representing f in the transform domain. Therefore, in this paper, we focus our attention on transform-domain shrinkage methods, such as those operating in the wavelet domain.

1.1 The one-dimensional case

First of all, let's consider the one-dimensional case $d = 1$. The most standard of wavelet shrinkage methods is VisuShrink of [22]. It is constructed through individual (or term-by-term) thresholding of the empirical wavelet coefficients. It enjoys good theoretical (and practical) properties. In particular, it achieves the optimal rate of convergence up to a logarithmic term over the Hölder class under the MISE. In other words, if $\widehat{f}^V$ denotes VisuShrink, and $\Lambda^s(M)$ the Hölder smoothness class, then there exists a constant $C > 0$ such that

$$\sup_{f \in \Lambda^s(M)} R(\widehat{f}^V, f) \leq C n^{-2s/(1+2s)} (\log n)^{2s/(1+2s)}. \tag{1.2}$$

Other term-by-term shrinkage rules have been developed. See, for instance, the firm shrinkage [25] or the non-negative garrote shrinkage [24]. In particular, they satisfy (1.2) but improve the value of the constant C. An exhaustive account of other shrinkage methods is provided in [3] that the interested reader may refer to.

The individual approach achieves a degree of trade-off between variance and bias contribution to the MISE. However, this trade-off is not optimal; it removes too many terms from the observed wavelet expansion, with the consequence the estimator is too biased and has a sub-optimal MISE convergence rate (and also in other L_p metrics $1 \leq p \leq \infty$). One way to increase estimation precision is by exploiting information about neighboring coefficients. In other words, empirical wavelet coefficients tend to form clusters that could be thresholded in blocks (or groups) rather than individually. This would allow threshold decisions to be made more accurately and permit convergence rates to be improved. Such a procedure has been introduced in [26, 27] who studied wavelet shrinkage methods based on block thresholding. The procedure first divides the wavelet coefficients at each resolution level into non-overlapping blocks and then keeps all the coefficients within a block if, and only if, the magnitude of the sum of the squared empirical coefficients within that block is greater than a fixed threshold. The original procedure developed by [26, 27] is defined with the block size $(\log n)^2$. BlockShrink of [6, 8] is the optimal version of this procedure. It uses a different block size, $\log n$, and enjoys a number of advantages over the conventional individual thresholding. In particular, it achieves the optimal rate of convergence over the Hölder class under the MISE. In other words, if $\widehat{f}^B$ denotes the BlockShrink estimate, then there exists a constant $C > 0$ such that

$$\sup_{f \in \Lambda^s(M)} R(\widehat{f}^B, f) \leq C n^{-2s/(1+2s)}. \tag{1.3}$$

Clearly, in comparison to VisuShrink, BlockShrink removes the extra logarithmic term. The minimax properties of BlockShrink under the L_p risk have been studied in [20]. Other local block thresholding rules have been developed. Among them, there is BlockJS of [7, 8] which combines James-Stein rule (see [40]) with the wavelet methodology. In particular, it satisfies (1.3) but improves the value of the constant C. From a practical point view, it is better than BlockShrink. Further details about the theoretical performances of BlockJS can be found in [17]. We refer to [3] and [9] for a comprehensive simulation study. Variations of BlockJS are BlockSure of [21] and SureBlock of [10]. The distinctive aspect of these block thresholding procedures is to provide data-driven algorithms to chose the threshold parameter. Let's also mention the work of [1] who considered wavelet block denoising in a Bayesian framework to obtain level-dependent block shrinkage and thresholding estimates.

1.2   The multi-dimensional case

Denoising is a long-standing problem in image processing. Since the seminal papers by Donoho & Johnstone [22], the image processing literature has been

inundated by hundreds of papers applying or proposing modifications of the original algorithm in image denoising. Owing to recent advances in computational harmonic analysis, many multi-scale geometrical transforms, such as ridgelets [16], curvelets [14, 13] or bandlets [36], were shown to be very effective in sparsely representing the geometrical content in images. Thanks to the sparsity (or more precisely compressibility) property of these expansions, it is reasonable to assume that essentially only a few large coefficients will contain information about the underlying image, while small values can be attributed to the noise which uniformly contaminates all transform coefficients. Thus, the wavelet thresholding/shrinkage procedure can be mimicked for these transforms, even though some care should be taken when the transform is redundant (corresponding to a frame or a tight frame). The modus operandi is again the same, first apply the transform, then perform a non-linear operator on the coefficients (each coefficient individually or in group of coefficients), and finally apply the inverse transform to get an image estimate. Among the many transform-domain image denoising algorithms to date, we would like to cite [38, 39, 37, 33] which are amongst the most efficient in the literature. Except [33], all cited approaches use orthodox Bayesian machinery and assume different forms of multivariate priors over blocks of neighboring coefficients and even interscale dependency. Nonetheless, none of those papers provide a study of the theoretical performance of the estimators.

From a theoretical point of view, Candès [12] has shown that the ridgelet-based individual coefficient thresholding estimator is nearly minimax for recovering piecewise smooth images away from discontinuities along lines. Individual thresholding of curvelet tight frame coefficients yields an estimator that achieves a nearly-optimal minimax rate $O(n^{-4/3})$[1] (up to logarithmic factor) uniformly over the class of piecewise C^2 images away from singularities along C^2 curves— so-called C^2-C^2 images [15][2] . Similarly, Le Pennec et al. [35] have recently proved that individual thresholding in an adaptively selected best bandlet orthobasis is nearly-minimax for C^α functions away from C^α edges.

In the image processing community, block thresholding/shrinkage in a non-Bayesian framework has been used very little. In [18, 19] the authors propose a multi-channel block denoising algorithm in the wavelet domain. The hyperparameters associated to their method (e.g. threshold), are derived using Stein's risk estimator. Yu et al. [41] advocated the use of BlockJS [7] to denoise audio signal in the time-frequency domain with anisotropic block size. To the best of our knowledge, no theoretical study of the minimax properties of block thresholding/shrinkage for images, and more generally for multi-dimensional data, has been reported in the literature.

---

[1] It is supposed that the image has size $n \times n$.
[2] Known as the cartoon model.

In this paper, we propose a generalization of Stein block thresholding to any dimension d. We investigate its minimax properties with a particular emphasis on $d = 2$. Towards this goal, we consider a frame coefficient space over which minimaxity is proved; see (3.2). The choice of this space is inspired by the characterization provided in [4] of family of smoothness spaces on $\mathbb{R}^d$, a subclass of so-called decomposition spaces [4, 23]. We will elaborate more on these (sparsity) smoothness spaces later in subsection 3.2. From this characterization, it turns out that our frame coefficient spaces are closely related to smoothness spaces that cover the classical case of Besov spaces, as well as smoothness spaces corresponding to curvelet-type constructions in $\mathbb{R}^d$, $d \geq 2$. Therefore, for $d = 2$ our denoiser will apply to both images with smoothness in Besov spaces for which wavelets are known to provide a sparse representation, and also to images that are compressible in the curvelet domain.

Our main theoretical result investigates the minimax rates over these decomposition spaces, and shows that our block estimator can achieve the optimal minimax rate, or is at least nearly-minimax (up to a log factor) in the least favorable situation. Another novelty is that the minimax rates given here are stated for a general noise sequence model in the transform coefficient domain beyond the usual i.i.d. Gaussian case. Thus, our result is particularly useful when the transform used corresponds to a frame, where a bounded zero-mean white Gaussian noise in the original domain is transformed into a bounded zero-mean correlated Gaussian process with a covariance matrix given by the Gram matrix of the frame.

The choice of the threshold parameter is theoretically discussed and its optimal value is stated for some noise models such as the (non-necessarily i.i.d.) Gaussian case. We provide a simple, fast and a practical procedure. We report a comprehensive simulation study to support our theoretical findings. It turns out that the only two parameters of our Stein block denoiser—the block size and the threshold— dictated by the theory work well for a large set of test images and various transforms. Moreover, the practical performance of our Stein block denoising compares very favorably to state-of-the art methods such as the BLS-GSM of [38]. Our procedure is however much simpler to implement and has a much lower computational cost than orthodox Bayesian methods such as BLS-GSM, since it does not involve any computationally consuming integration nor optimization steps. A toolbox is made available for download on the Internet to reproduce the results discussed in this paper.

1.4 Organization of the paper

The paper is organized as follows. Section 2 is devoted to the one-dimensional BlockJS procedure introduced in [7]. In Section 3, we extend BlockJS to the multi-dimensional case and a fairly general noise model beyond the i.i.d. Gaussian case. This section also contains our main theoretical results. In Section 4, a comprehensive experimental study is reported and discussed. We finally conclude in Section 5 and point to some perspectives. The proofs of the results are deferred to the appendix awaiting inspection by the interested reader.

2 The one-dimensional BlockJS

In this section, we present the construction and the theoretical performance of the one-dimensional BlockJS procedure developed by [7].

Consider the one-dimensional nonparametric regression model:

$$Y_i = f(i/n) + \epsilon_i, \qquad i = 1, ..., n, \tag{2.1}$$

where $(Y_i)_{i=1,...,n}$ are the observations, $(\epsilon_i)_{i=1,...,n}$ are i.i.d. $\mathcal{N}(0,1)$, and $f : [0,1] \to \mathbb{R}$ is an unknown function. The goal is to estimate f from the observations. In the orthogonal wavelet framework, (2.1) amounts to the sequence model

$$y_{j,k} = \theta_{j,k} + n^{-1/2} z_{j,k}, \quad j = 0, ..., J, \qquad k = 0, ..., 2^j - 1, \tag{2.2}$$

where $J = \lfloor \log_2 n \rfloor$, $(y_{j,k})_{j,k}$ are the observations, for each j, $(z_{j,k})_k$ are i.i.d. $\mathcal{N}(0,1)$, and $(\theta_{j,k})_{j,k}$ are approximately the true wavelet coefficients of f. Since they determine completely f, the goal is to estimate these coefficients as accurately as possible. To assess the performance of an estimator $\widehat{\theta} = (\widehat{\theta}_{j,k})_{j,k}$ of $\theta = (\theta_{j,k})_{j,k}$, we adopt the minimax approach under the expected squared error over a given Besov body. The expected squared error is defined by $R(\widehat{\theta}, \theta) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{\theta}_{j,k} - \theta_{j,k})^2\right)$, and the Besov body by

$$\Theta_{p,q}^s(M) = \left\{ \theta = (\theta_{j,k})_{j,k}; \quad \left(\sum_{j=0}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\theta_{j,k}|^p \right)^{1/p} \right)^q \right)^{1/q} \leq M \right\}.$$

In this notation, $s > 0$ is a smoothness parameter, $0 < p \leq +\infty$ and $0 < q \leq +\infty$ are norm parameters [3], and $M \in (0, \infty)$ denotes the radius of the ball. The Besov body contains a wide class of $\theta = (\theta_{j,k})_{j,k}$. It includes the Hölder body $\Theta_{\infty,\infty}^s(M)$ and the Sobolev body $\Theta_{2,2}^s(M)$.

---

[3] This is a slight abuse of terminology as for $0 < p, q < 1$, Besov spaces are rather complete quasinormed linear spaces.

The goal of the minimax approach is to construct an adaptive estimator $\widehat{\theta} = (\widehat{\theta}_{j,k})_{j,k}$ such that $\sup_{\theta \in \Theta_{p,q}^s(M)} R(\widehat{\theta}, \theta)$ is as small as possible. A candidate is the BlockJS procedure whose paradigm is described below.

Let $L = \lfloor \log n \rfloor$ be the block size, $j_0 = \lfloor \log_2 L \rfloor$ the coarsest decomposition scale and, for any j, $\Lambda_j = \{0, ..., 2^j - 1\}$ is the set of locations at scale j. For any $j \in \{j_0, ..., J\}$, let $A_j = \{1, ..., \lfloor 2^j L^{-1} \rfloor\}$ be the set of block indices at scale j, and for any $K \in A_j$, $U_{j,K} = \{k \in \Lambda_j; (K-1)L \leq k \leq KL - 1\}$ is the set indexing the locations of coefficients within the Kth block. Let λ_* be a threshold parameter chosen as the root of $x - \log x = 3$ (i.e. $\lambda_* = 4.50524...$). Now estimate $\theta = (\theta_{j,k})_{j,k}$ by $\widehat{\theta}^* = (\widehat{\theta}_{j,k}^*)_{j,k}$ where, for any $k \in U_{j,K}$ and $K \in A_j$,

$$
\widehat{\theta}_{j,k}^* = \begin{cases} y_{j,k}, & \text{if } j \in \{0, ..., j_0 - 1\}, \\ y_{j,k} \left(1 - \dfrac{\lambda_* n^{-1}}{\frac{1}{L} \sum_{k \in U_{j,K}} y_{j,k}^2}\right)_+, & \text{if } j \in \{j_0, ..., J\}, \\ 0, & \text{if } j \in \mathbb{N} - \{0, ..., J\}. \end{cases}
\tag{2.3}
$$

where $(x)_+ = \max(x, 0)$. Thus, at the coarsest scales $j \in \{0, ..., j_0\}$, the observed coefficients $(y_{j,k})_{k \in \Lambda_j}$ are left intact as usual. For $k \in \Lambda_j$ and $j \in \mathbb{N} - \{0, ..., J\}$, $\theta_{j,k}$ is estimated by zero. For $k \in U_{j,K}$, $K \in A_j$ and $j \in \{j_0, ..., J\}$, if the mean energy within the Kth block $\sum_{k \in U_{j,K}} y_{j,k}^2 / L$ is larger than $\lambda_* n^{-1}$ then $y_{j,k}$ is shrunk by the amount $y_{j,k} \frac{\lambda_* n^{-1}}{\frac{1}{L} \sum_{k \in U_{j,K}} y_{j,k}^2}$; otherwise, $\theta_{j,k}$ is estimated by zero. Note that $\dfrac{\frac{1}{L} \sum_{k \in U_{j,K}} y_{j,k}^2}{n^{-1}}$ can be interpreted as a local measure of signal-to-noise ratio in the block $U_{j,K}$. Such a block thresholding originates from the James-Stein rule introduced in [40].

The block length $L = \lfloor \log n \rfloor$ and the value $\lambda_* = 4.50524$ are chosen based on theoretical considerations; under this calibration, the BlockJS is (near) optimal in terms of minimax rate and adaptivity. This is summarized in the following theorem.

Theorem 2.1 ([7]) *Consider the model* (2.2) *for n large enough. Let $\widehat{\theta}^*$ be given as* (2.3). *Then there exists a constant $C > 0$ such that*

$$
\sup_{\theta \in \Theta_{p,q}^s(M)} R(\widehat{\theta}^*, \theta) \leq C \begin{cases} n^{-2s/(2s+1)}, & \text{for } p \geq 2, \\ n^{-2s/(2s+1)} (\log n)^{(2-p)/(p(2s+1))}, & \text{for } p < 2, \ sp \geq 1. \end{cases}
\tag{2.4}
$$

The rates of convergence (2.4) are optimal, except in the case $p < 2$ where there is an extra logarithmic term. They are better than those achieved by

standard individual thresholding (hard, soft, non-negative garotte, etc); we gain a logarithmic factor for $p \geq 2$. See [22].

3   The multi-dimensional BlockJS

This section is the core of our proposal where we introduce a BlockJS-type procedure for multi-dimensional data. The goal is to adapt its construction in such a way that it preserves its optimal properties over a wide class of functions.

3.1   The sequence model

Our approach begins by projecting the model (1.1) onto a collection of atoms $(\varphi_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ that forms a (tight) frame. This gives rise to a sequence space model obtained by calculating the noisy coefficient $y_{j,\ell,\mathbf{k}} = \langle Y, \varphi_{j,\ell,\mathbf{k}} \rangle$ for any element of the frame $\varphi_{j,\ell,\mathbf{k}}$. We then have a multi-dimensional sequence of coefficients $(y_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ defined by

$$y_{j,\ell,\mathbf{k}} = \theta_{j,\ell,\mathbf{k}} + n^{-r/2} z_{j,\ell,\mathbf{k}}, \quad j = 0, ..., J, \quad \ell \in B_j, \quad \mathbf{k} \in D_j, \qquad (3.1)$$

where $J = \lfloor \log_2 n \rfloor$, $r \in [1, d]$, $d \in \mathbb{N}^*$, $B_j = \{1, ..., \lfloor c_* 2^{vj} \rfloor\}$, $c_* \geq 1$, $v \in [0, 1]$, $\mathbf{k} = (k_1, ..., k_d)$, $D_j = \prod_{i=1}^{d} \{0, ..., \lfloor 2^{\mu_i j} \rfloor - 1\}$, $(\mu_i)_{i=1,...,d}$ is a sequence of positive real numbers, $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are random variables and $(\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are unknown coefficients. Let $d_* = \sum_{i=1}^{d} \mu_i$.

The indices j and $\mathbf{k}$ are respectively the scale and position parameters. ℓ is a generic integer indexing for example the orientation (subband) which may be scale-dependent. The parameters $(\mu_i)_{i=1,...,d}$ allow to handle anisotropic subbands. To illustrate the meaning of these parameters, let's see how they specialize in some popular transforms. For example, with the separable two-dimensional wavelet transform, we have $v = 0$, $c^* = 3$, and $\mu_1 = \mu_2 = 1$. Thus, as expected, we get three isotropic subbands at each scale. For the second generation curvelet transform [13], we have $v = 1/2$, $\mu_1 = 1$ and $\mu_2 = 1/2$ which corresponds to the parabolic scaling of curvelets.

3.1.1   Assumptions on the noise sequence

Let $L = \lfloor (r \log n)^{1/d} \rfloor$ be the block length, $j_0 = \lfloor (1/\min_{i=1,...,d} \mu_i) \log_2 L \rfloor$ is the coarsest decomposition scale, and $J_* = \lfloor (r/(d_* + \delta + v)) \log_2 n \rfloor$. For any $j \in \{j_0, ..., J_*\}$, let

8

- $\mathcal{A}_j = \prod_{i=1}^d \{1, ..., \lfloor 2^{\mu_i j} L^{-1} \rfloor\}$ be the set indexing the blocks at scale j.
- For each block index $\mathbf{K} = (K_1, ..., K_d) \in \mathcal{A}_j$, $U_{j,\mathbf{K}} = \{\mathbf{k} \in D_j; \ (K_1 - 1)L \le k_1 \le K_1 L - 1, ..., (K_d - 1)L \le k_d \le K_d L - 1\}$ is the set indexing the positions of coefficients within the $\mathbf{K}$th block $U_{j,\mathbf{K}}$.

Our assumptions on the noise model are as follows. Suppose that there exist $\delta \ge 0$, $\lambda_* > 0$, $Q_1 > 0$ and $Q_2 > 0$ independent of n such that

(A1) $\sup_{j \in \{0,...,J\}} \sup_{\ell \in B_j} 2^{-j(d_* + \delta)} \sum_{\mathbf{k} \in D_j} \mathbb{E} \left(z_{j,\ell,\mathbf{k}}^2 \right) \le Q_1$.

(A2)
$$\sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \mathbb{E} \left(z_{j,\ell,\mathbf{k}}^2 \mathbf{1}_{\left\{ \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 > \lambda_* 2^{\delta j} L^d / 4 \right\}} \right) \le Q_2.$$

Assumptions (A1) and (A2) are satisfied for a wide class of noise models on the sequence $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ (not necessarily independent or identically distributed). Several such noise models are characterized in Propositions 3.1 and 3.2 below.

Remark 3.1 (*Comments on δ) The parameter δ is connected to the nature of the model. For standard models, and in particular, the d-dimensional non-parametric regression corresponding to the problem of denoising (see Section 4), δ is set to zero. The presence of δ in our assumptions, definitions and results is motivated by potential applicability of the multi-dimensional BlockJS (to be defined in Subsection 3.3) to other inverse problems such as deconvolution. The role of δ becomes of interest when addressing such inverse problems. This will be the focus of a future work. To illustrate the importance of δ in one-dimensional deconvolution, see [31].*

3.2   The smoothness space

We wish to estimate $(\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ from $(y_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ defined by (3.1). To measure the performance of an estimator $\widehat{\theta} = (\widehat{\theta}_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ of $\theta = (\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$, we consider the minimax approach under the expected multi-dimensional squared error over a multi-dimensional frame coefficient space. The expected multi-dimensional squared error is defined by

$$R\left(\widehat{\theta}, \theta\right) = \sum_{j=0}^{\infty} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \mathbb{E} \left((\widehat{\theta}_{j,\ell,\mathbf{k}} - \theta_{j,\ell,\mathbf{k}})^2 \right)$$

and the multi-dimensional frame coefficient smoothness/sparseness space by

$$\Theta_{p,q}^s(M) = \left\{ \theta = (\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}; \ \left(\sum_{j=0}^{\infty} \sum_{\ell \in B_j} \left(2^{j(s + d_*/2 - d_*/p)} \left(\sum_{\mathbf{k} \in D_j} |\theta_{j,\ell,\mathbf{k}}|^p \right)^{1/p} \right)^q \right)^{1/q} \le M \right\},$$

(3.2)

with a smoothness parameter s, $0 < p \leq +\infty$ and $0 < q \leq +\infty$. We recall that $d_* = \sum_{i=1}^{d} \mu_i$.

The definition of these smoothness spaces is motivated by the work of [4]. These authors studied decomposition spaces associated to appropriate structured uniform partition of the unity in the frequency space $\mathbb{R}^d$. They considered construction of tight frames adapted to form atomic decomposition of the associated decomposition spaces, and established norm equivalence between these smoothness/sparseness spaces and the sequence norm defined in (3.2). That is, the decomposition space norm can be completely characterized by the sparsity or decay behavior of the associated frame coefficients.

For example, in the case of a "uniform" dyadic partition of the unity, the smoothness/sparseness space is a Besov space $B_{p,q}^s$, for which suitable wavelet expansion [4] is known to provide a sparse representation [34]. In this case, from subsection 3.1 we have $d^* = d$, and $\Theta_{p,q}^s(M)$ is a d-dimensional Besov ball.

Curvelets in arbitrary dimensions correspond to partitioning the frequency plane into dyadic coronae, which are then angularly localized near regions of side length 2^j in the radial direction and $2^{j/2}$ in all the other directions [11]. For $d = 2$, the angular wedges obey the parabolic scaling law [14]. This partition of the frequency plane is significantly different from dyadic decompositions, and as a consequence, sparseness for curvelet expansions cannot be described in terms of classical smoothness spaces. For $d = 2$, Borup and Nielsen [4, Lemma 10] showed that the smoothness/sparseness space (3.2) and the smoothness/sparseness of the second-generation curvelets [13] are the same, in which case $d^* = 3/2$. Embedding results for curvelet-type decomposition spaces relative to Besov spaces were also provided in [4]. Furthermore, it was shown that piecewise C^2 images away from piecewise-C^2 singularities, which are sparsely represented in the curvelet tight frame [14], are contained in $\Theta_{2/3,2/3}^{3/2+\beta}$, $\forall \beta > 0$. Even though the role and the range of β has not been clarified by the authors in [4].

3.3 Multi-dimensional block estimator

As for the one-dimensional case, we wish to construct an adaptive estimator $\widehat{\theta} = (\widehat{\theta}_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ such that $\sup_{\theta \in \Theta_{p,q}^s(M)} R\left(\widehat{\theta}, \theta\right)$ is as small as possible. To reach this goal, we propose a multi-dimensional version of the BlockJS procedure introduced in [7].

From subsection 3.1.1, recall the definitions of L, j_0, J_*, $\mathcal{A}_j$ and $U_{j,\mathbf{K}}$. We

---

[4] With a wavelet having sufficient regularity and number of vanishing moments [34].

estimate $\theta = (\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ by $\widehat{\theta}^* = (\widehat{\theta}^*_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ where, for any $\mathbf{k} \in U_{j,\mathbf{K}}$, $\mathbf{K} \in \mathcal{A}_j$ and $\ell \in B_j$,

$$
\widehat{\theta}^*_{j,\ell,\mathbf{k}} = \begin{cases} y_{j,\ell,\mathbf{k}}, & \text{if } j \in \{0, ..., j_0 - 1\}, \\ y_{j,\ell,\mathbf{k}} \left(1 - \dfrac{\lambda_* n^{-r} 2^{\delta j}}{\frac{1}{L^d} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} y^2_{j,\ell,\mathbf{k}}} \right)_+, & \text{if } j \in \{j_0, ..., J_*\}, & (3.3) \\ 0, & \text{if } j \in \mathbb{N} - \{0, ..., J_*\}. \end{cases}
$$

In this definition, δ and λ_* denote the constants involved in (A1) and (A2). Again, the coarsest scale coefficients are left unaltered, while the other coefficients are either thresholded or shrunk depending whether the local measure of signal-to-noise ratio $\dfrac{\frac{1}{L^d} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} y^2_{j,\ell,\mathbf{k}}}{n^{-r}}$ within the block $U_{j,\mathbf{K}}$ is larger that the threshold $\lambda_* 2^{\delta j}$. Notice that the dimension d of the model appears in the definition of L, the length of each block $U_{j,\mathbf{K}}$. This point is crucial; L optimizes the theoretical and practical performance of the considered multi-dimensional BlockJS procedure. As far as the choice of the threshold parameter λ_* is concerned, it will be discussed in Subsection 3.5 below.

3.4 Minimax theorem

Theorem 3.1 below investigates the minimax rate of (3.3) over $\mathbf{\Theta}^s_{p,q}$.

Theorem 3.1 *Consider the model* (3.1) *for n large enough. Suppose that* (A1) *and* (A2) *are satisfied. Let $\widehat{\theta}^*$ be given as in* (3.3).

- *There exists a constant $C > 0$ such that*

$$
\sup_{\theta \in \mathbf{\Theta}^s_{p,q}(M)} R\left(\widehat{\theta}^*, \theta\right) \leq C \rho_n,
$$

  where

$$
\rho_n = \begin{cases} n^{-2sr/(2s+\delta+d_*+\upsilon)}, & \text{for } q \leq 2 \leq p, \\ (\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}, & \text{for } q \leq p < 2, \ sp > d_* \vee (1 - p/2)(\delta + d_* + \upsilon). \end{cases} \quad (3.4)
$$

- *If $\upsilon = 0$, the minimax rates* (3.4) *hold without the restriction $q \leq p \wedge 2$.*

The rates of convergence (3.4) are optimal for a wide class of variables $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$. If we take $d_* = d = \mu_1 = 1$, $r = 1$, $c_* = 1$ and $\upsilon = \delta = 0$, then we recover the rates exhibited in the one-dimensional wavelet case expressed in Theorem 2.1. There is only a minor difference on the power of the logarithmic term for $p < 2$. Thus, Theorem 3.1 can be viewed as a generalization of Theorem 2.1.

11

In the case of d-dimensional isotropic Besov spaces, where wavelets (corresponding to $\upsilon = 0$, $\mu_1 = \mu_2 = 1$ and then $d_* = d$) provide optimally sparse representations, Theorem 3.1 can be applied without the restriction $q \leq p \wedge 2$. Therefore, for $p \geq 2$, Theorem 3.1 states that Stein block thresholding gets rid of the logarithmic factor, hence achieving the optimal minimax rate over those Besov spaces. For $p < 2$, the block estimator is nearly-minimax.

As far as curvelet-type decomposition spaces are concerned, from section 3.1 we have $\mu_1 = 1$, $\mu_2 = \frac{1}{2}$, $d_* = \mu_1 + \mu_2 = \frac{3}{2}$, $r = d = 2$, $\upsilon = \frac{1}{2}$, $\delta = 0$. This gives the rates

$$
\rho_n = \begin{cases} n^{-2s/(s+1)}, & \text{for } q \leq 2 \leq p, \\ (\log n/n)^{2s/(s+1)}, & \text{for } q \leq p < 2,\ sp > \dfrac{3}{2} \vee (2 - p). \end{cases}
$$

where the logarithmic factor disappears only for $q \leq 2 \leq p$. Following the discussion of section 3.2, C^2-C^2 images correspond to a smoothness space $\Theta^s_{p,q}$ with $p = q = 2/3$. Moreover, $\exists \kappa > 0$ such that taking $s = 2 + \kappa$ satisfies the condition of Theorem 3.1, and C^2-C^2 images are contained in $\Theta^s_{2/3,2/3}$ with such a choice. We then arrive at the rate $O(n^{-4/3})$ (ignoring the logarithmic factor). This is consistent with the results of [32], which established that no estimator can achieve a better rate than the optimal minimax rate $O(n^{-4/3})$ uniformly over the C^2-C^2 class. On the other hand, individual thresholding in the curvelet tight frame has also the nearly-minimax rate $O(n^{-4/3})$ [15] uniformly over the class of C^2-C^2 images. Nonetheless, the experimental results reported in this paper indicate that block curvelet thresholding outperforms in practice term-by-term thresholding on a wide variety of images, although the improvement can be of a limited extent.

3.5 On the (theoretical) choice of the threshold

To apply Theorem 3.1, it is enough to determine δ and λ_* such that (A1) and (A2) are satisfied. The parameter δ is imposed by the nature of the model; it can be easily fixed as in our denoising experiments where it was set to $\delta = 0$. The choice of the threshold λ_* is more involved. This choice is crucial towards good performance of the estimator $\widehat{\theta}^*$. From a theoretical point of view, since the constant C of the bound (3.4) increases with growing λ, the optimal threshold is the smallest real number λ_* such that (A2) is fulfilled. In the following, we first provide the explicit expression of λ_* in the situation of a non-necessarily i.i.d. Gaussian noise sequence $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$. This result is then refined in the case of a white Gaussian noise.

Proposition 3.1 below determines a suitable threshold λ_* satisfying (A1) and

(A2) when $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are Gaussian random variables (not necessarily i.i.d.).

Proposition 3.1 *Consider the model* (3.1) *for n large enough. Suppose that, for any $j \in \{0, ..., J\}$ and any $\ell \in B_j$, $(z_{j,\ell,\mathbf{k}})_{\mathbf{k}}$ is a centered Gaussian process. Assume that there exists two constants $Q_3 > 0$ and $Q_4 > 0$ (independent of n) such that*

- (A3): $\sup_{j \in \{0,...,J\}} \sup_{\ell \in B_j} \sup_{\mathbf{k} \in D_j} 2^{-2\delta j} \mathbb{E}\left(z_{j,\ell,\mathbf{k}}^4\right) \leq Q_3$.
- (A4): *for any $a = (a_{\mathbf{k}})_{\mathbf{k} \in D_j}$ such that $\sup_{j \in \{0,...,J\}} \sup_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} a_{\mathbf{k}}^2 \leq 1$, we have*

$$\sup_{j \in \{0,...,J\}} \sup_{\ell \in B_j} \sup_{\mathbf{K} \in \mathcal{A}_j} 2^{-\delta j} \mathbb{E}\left(\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} a_{\mathbf{k}} z_{j,\ell,\mathbf{k}}\right)^2\right) \leq Q_4.$$

Then (A1) *and* (A2) *are satisfied with $\lambda_* = 4\left((2Q_4)^{1/2} + Q_3^{1/4}\right)^2$. Therefore Theorem 3.1 can be applied to $\widehat{\theta}^*$ defined by* (3.3) *with such a λ_*.*

This result is useful as it establishes that the block denoising procedure and the minimax rates of Theorem 2.1 apply to the case of frames where a bounded zero-mean white Gaussian noise in the original domain is transformed into a bounded zero-mean correlated Gaussian process.

If additional information is considered on $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$, the threshold constant λ_* defined in Proposition 3.1 can be improved. This is the case when $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are i.i.d. $\mathcal{N}(0,1)$ as is the case if the transform were orthogonal (e.g. orthogonal wavelet transform). The statement is made formal in the following proposition.

Proposition 3.2 *Consider the model* (3.1) *for n large enough. Suppose that, for any $j \in \{0, ..., J\}$ and any $\ell \in B_j$, $(z_{j,\ell,\mathbf{k}})_{\mathbf{k}}$ are i.i.d. $\mathcal{N}(0,1)$ as is the case when the transform used corresponds to an orthobasis. Theorem 3.1 can be applied with the estimator $\widehat{\theta}^*$ defined by* (3.3) *with $\delta = 0$ and λ_* the root of $x - \log x = 3$, i.e. $\lambda_* = 4.50524...$.*

The optimal threshold constant λ_* described in Proposition 3.2 corresponds to the one isolated by [7].
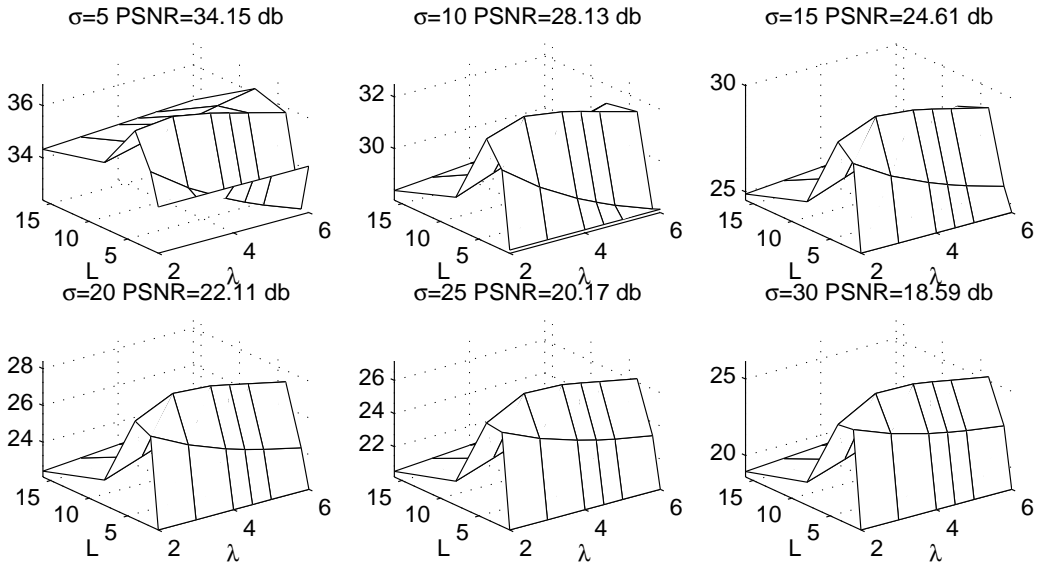
4   Application to image block denoising

4.1   Impact of threshold and block size

In this first experiment, the goal is twofold: first assess the impact of the threshold and the block size on the performance of block denoising, and second investigate the validity of their choice as prescribed by the theory. For a $n \times n$ image f and its estimate $\hat{f}$, the denoising performance is measured in terms of peak signal-to-noise ratio (PSNR) in decibels (dB)

$$\mathrm{PSNR} = 20 \log_{10} \frac{n \|f\|_\infty}{\|\hat{f} - f\|_2} \ \mathrm{dB} \ .$$

In this experiment, as well as in the rest of paper, three popular transforms are used: the orthogonal wavelet transform (DWT), its translation invariant version (UDWT) and the second generation fast discrete curvelet transform (FDCT) with the wrapping implementation [13]. The Symmlet wavelet with 6 vanishing moments was used throughout all experiments. For each transform, two images were tested Barbara (512×512) and Peppers (256×256), and each image, was contaminated with zero-mean white Gaussian noise with increasing standard deviation $\sigma \in \{5, 10, 15, 20, 25, 30\}$, corresponding to input PSNR values $\{34.15, 28.13, 24.61, 22.11, 20.17, 18.59, 14.15\}$ dB. At each combination of test image and noise level, ten noisy versions were generated. Then, block denoising was ten applied to each of the ten noisy images for each block size $L \in \{1, 2, 4, 8, 16\}$ and threshold $\lambda \in \{2, 3, 4, 4.5, 5, 6\}$, and the average output PSNR over the ten realizations was computed. This yields one plot of average output PSNR as a function of λ and L at each combination (image-noise level-transform). The results are depicted in Fig.1, Fig.2 and Fig.3 for respectively the DWT, UDWT and FDCT. One can see that the maximum of PSNR occurs at $L = 4$ (for $\lambda \geq 3$) whatever the transform and image, and this value turns to be the choice dictated by the theoretical procedure. As far as the influence of λ is concerned, the PSNR attains its exact highest peak at different values of λ depending on the image, transform and noise level. For the DWT, this maximum PSNR takes place near the theoretical threshold $\lambda_* \approx 4.5$ as expected from Proposition 3.2. Even with the other redundant transforms, that correspond to tight frames for which Proposition 3.2 is not rigorously valid, a sort of plateau is reached near $\lambda = 4.5$. Only a minor improvement can be gained by taking a higher threshold λ; see e.g. Fig.2 or 3 with Peppers for $\sigma \geq 20$. Note that this improvement by taking a higher λ for redundant transforms (i.e. non i.i.d. Gaussian noise) is formally predicted by Proposition 3.1. Even though the estimate of Proposition 3.1 was expected to be rather crude. To summarize, the value 4.50524... intended to work for orthobases seems to yield good results also with redundant transforms.

14

Barbara 512×512

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db

Peppers 256×256

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db

Fig. 1. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the DWT domain.

4.2   Comparative study

Block vs term-by-term It is instructive to quantify the improvement brought by block denoising compared to term-by-term thresholding. For reliable comparison, we applied the denoising algorithms to six standard grayscale images with different contents of size 512×512 (Barbara, Lena, Boat and Fingerprint) and 256×256 (House and Peppers). All images were normalized

15

Barbara 512×512

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db

Peppers 256×256

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

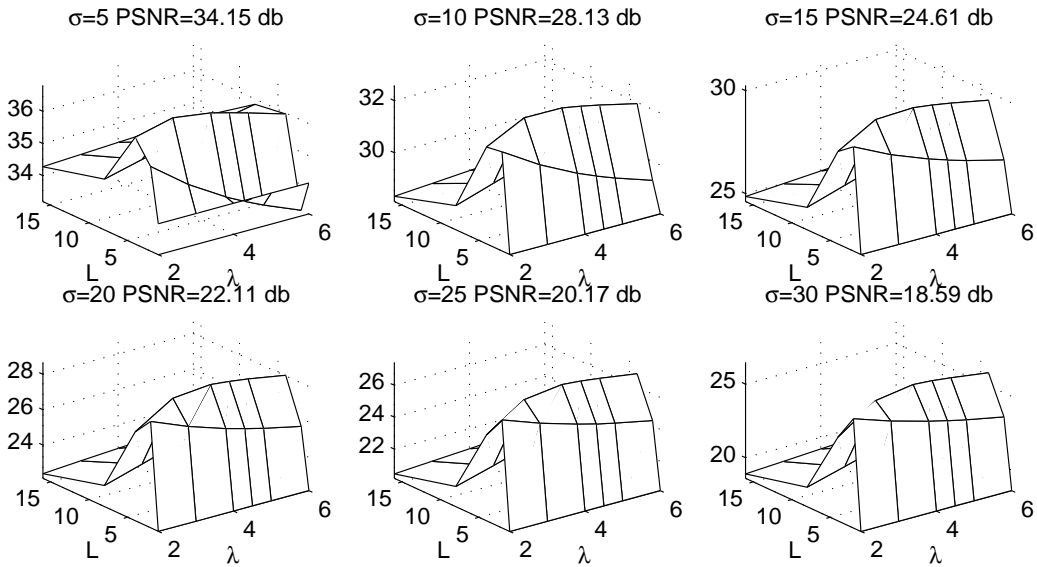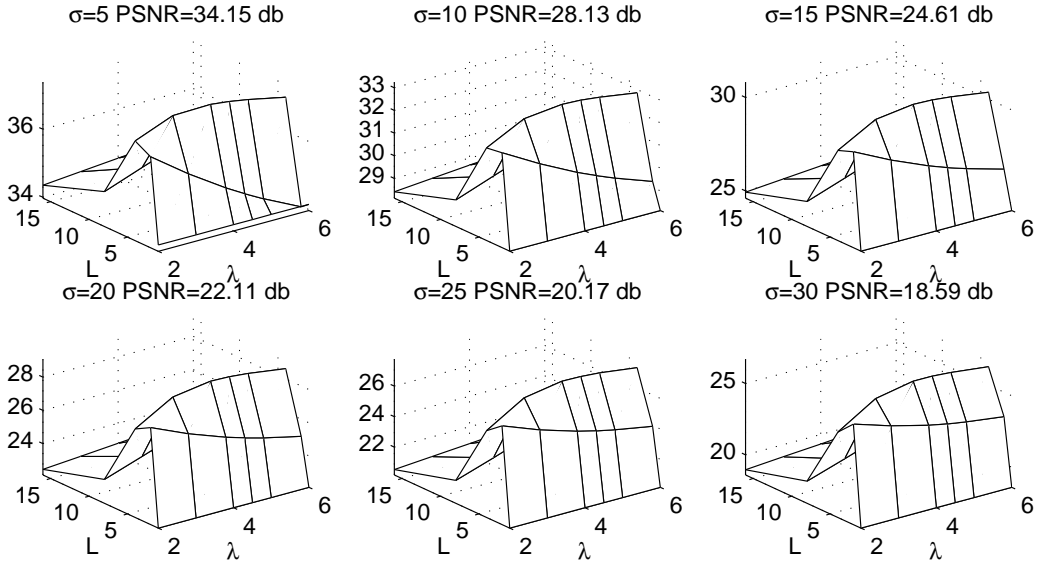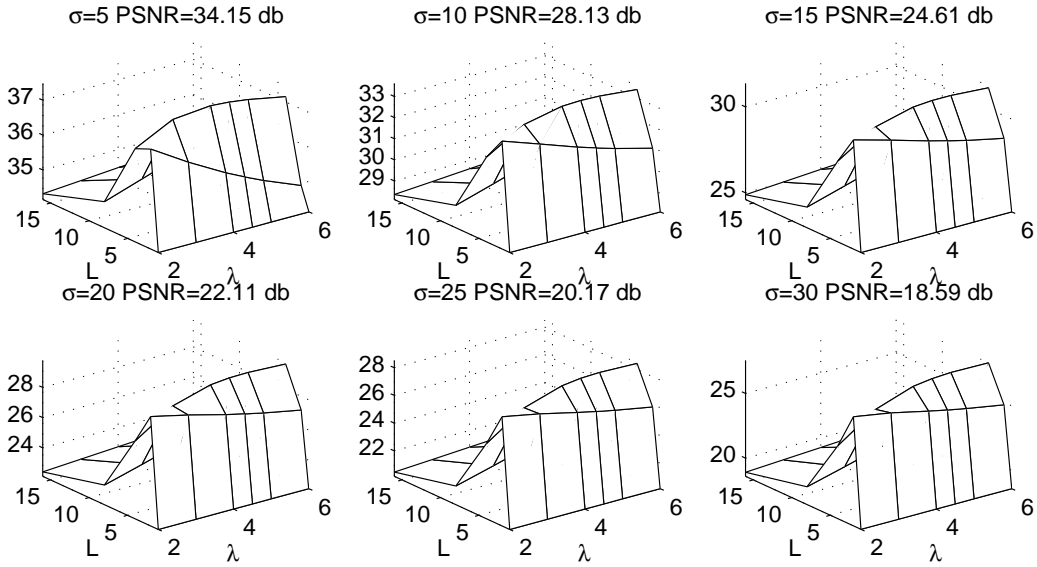σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db

Fig. 2. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the UDWT domain.

to a maximum grayscale value 255. The images were corrupted by a zero-mean white Gaussian noise with standard deviation $\sigma \in \{5, 10, 15, 20, 25, 30\}$. The output PSNR was averaged over ten realizations, and all algorithms were applied to the same noisy versions. The threshold used with individual thresholding was set to the classical value 3σ for the (orthogonal) DWT, and 3σ for all scales and 4σ at the finest scale for the (redundant) UDWT and FDCT. The results are displayed in Fig.4. Each plot corresponds to PSNR improvement over DWT term-by-term thresholding as a function of σ. To summarize,

16

Barbara 512×512

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db

Peppers 256×256

σ=5 PSNR=34.15 db   σ=10 PSNR=28.13 db   σ=15 PSNR=24.61 db

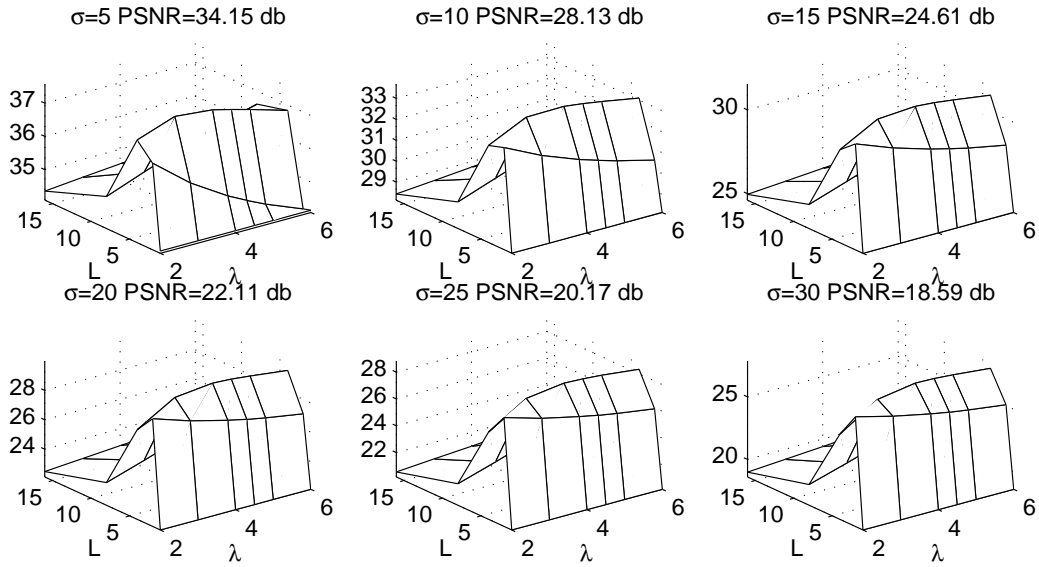σ=20 PSNR=22.11 db   σ=25 PSNR=20.17 db   σ=30 PSNR=18.59 db
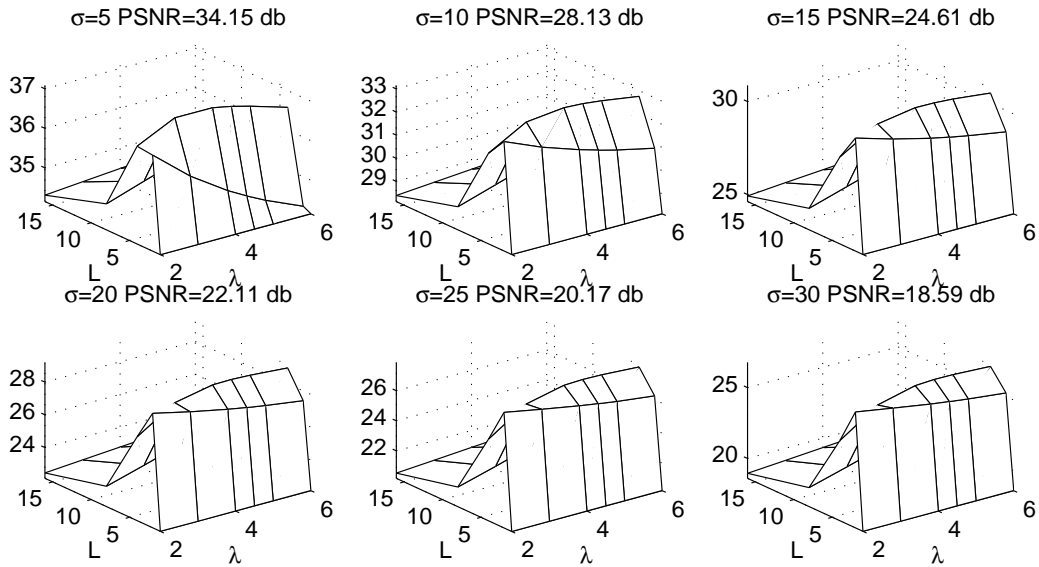
Fig. 3. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the FDCT domain.

- Block shrinkage improves the denoising results in general compared to individual thresholding. Even though the improvement extent decreases with increasing σ. The PSNR increase brought by block denoising with a given transform compared to individual thresholding with the same transform can be up to 2.55 dB.
- Owing to block shrinkage, even the orthogonal DWT becomes competitive with redundant transforms. For Barbara, block denoising with DWT is even better than individual thresholding in the translation-invariant UDWT.

17

- For some images (e.g. Peppers or House), block denoising with curvelets can be slightly outperformed by its term-by-term thresholding counterpart for $\sigma = 50$.
- As expected, no transform is the best for all images. Block denoising with curvelets is more beneficial to images with high frequency content (e.g. anisotropic oscillating patterns in Barbara). For the other images, and except Peppers, block denoising with UDWT or curvelets are comparable (~ 0.2 dB difference).

Note that the additional computational burden of block shrinkage compared to individual thresholding is limited: respectively 0.1s, 1s and 0.7s for the DWT, UDWT and FDCT with 512×512 images, and less than 0.03s, 0.2s and 0.1 for 256×256 images. The algorithms were run under Matlab with an Intel Xeon 3GHz CPU, 8Gb RAM.

Block vs BLS-GSM   The described block denoising procedure has been compared to one of state-of-the-art denoising methods in the literature BLS-GSM [38]. BLS-GSM is a widely used reference in image denoising experiments reported in the literature. BLS-GSM uses a sophisticated prior model of the joint distribution within each block of coefficients, and then computes the Bayesian posterior conditional mean estimator by numerical integration. For fair comparison, BLS-GSM was also adapted and implemented with the curvelet transform. The two algorithms were applied to the same ten realizations of additive white Gaussian noise with σ in the same range as before. The output PSNR values averaged over the ten realizations for each of the six tested image are tabulated in Table 2. By inspection of this table, the performance of block denoising and BLS-GSM remain comparable whatever the transform and image. None of them outperforms the other for all transforms and all images. When comparing both algorithms for the DWT transform, the maximum difference between the corresponding PSNR values is 0.5 dB in favor of block shrinkage. For the UDWT and FDCT, the maximum difference is ~ 0.6 dB in BLS advantage. Visual inspection of Fig.5 and 6 is in agreement with the quantitative study we have just discussed. For each transform, differences between the two denoisers are hardly visible. Our procedure is however much simpler to implement and has a much lower computational cost than BLS-GSM as can be seen from Table 1. Our algorithm can be up to 10 times faster than BLS-GSM while reaching comparable denoising performance. As stated in the previous paragraph, the bulk of computation in our algorithm is essentially invested in computing the forward and inverse transforms.

Barbara 512×512

Lena 512×512

House 256×256

Boat 512×512

Fingerprint 512×512

Peppers 256×256

19
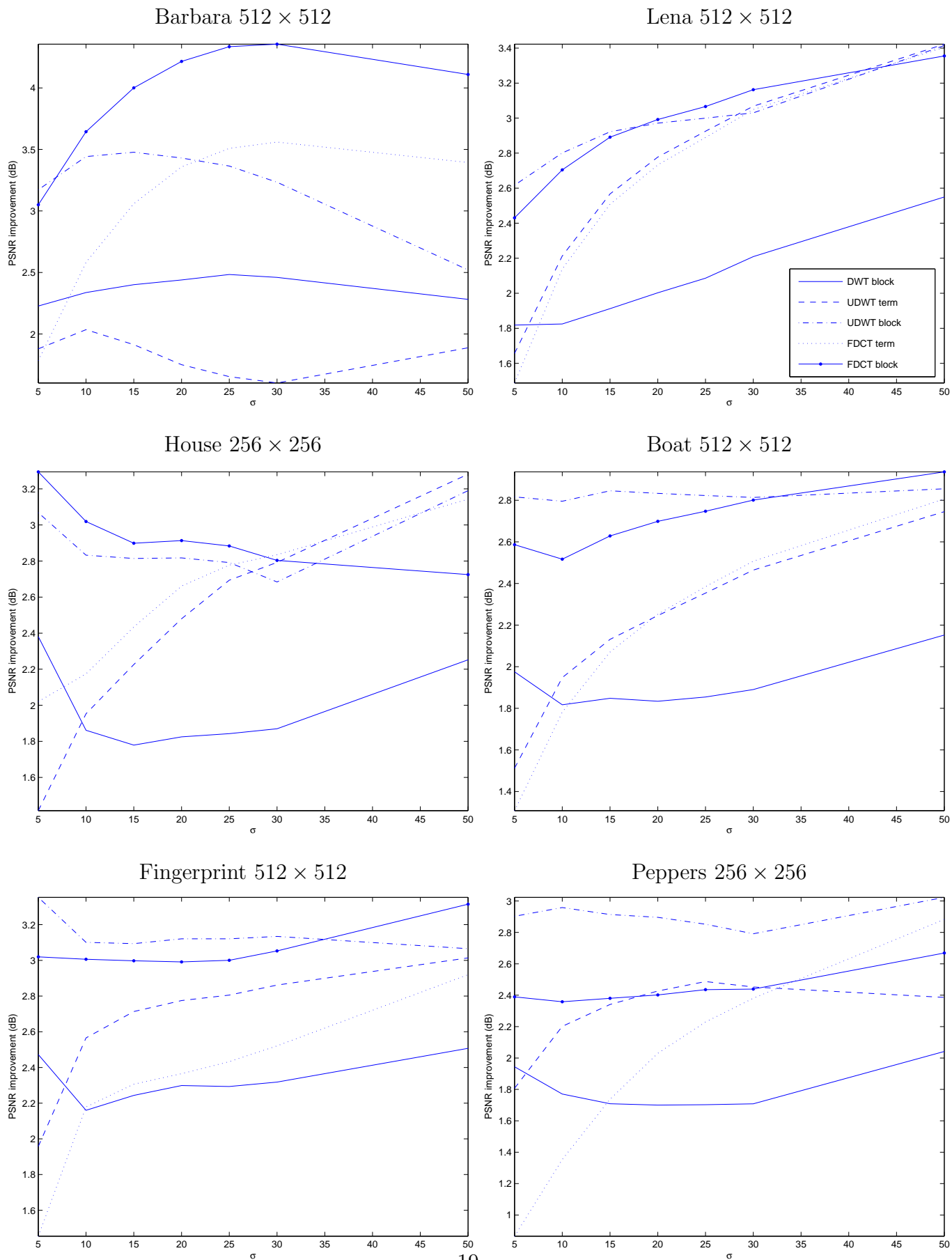
Fig. 4. Block vs term-by-term thresholding. Each plot corresponds to PSNR improvement over DWT term-by-term thresholding as a function of σ.
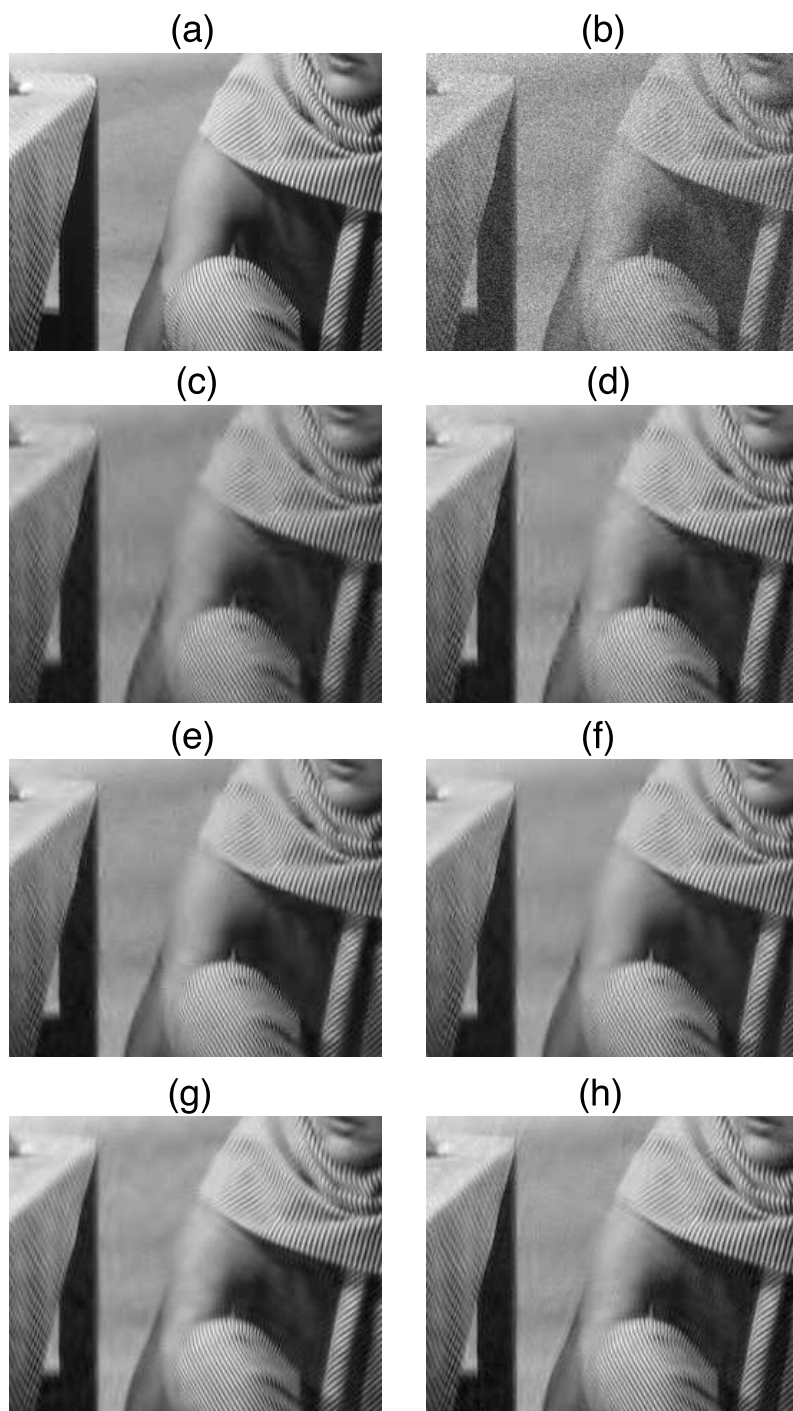
Fig. 5. Visual comparison of our block denoising to BLS-GSM on Barbara 512×512. (a) original. (b) noisy $\sigma = 20$. (c), (e) and (g) block denoising with respectively DWT (28.04 dB), UDWT (29.01 dB) and FDCT (30 dB). (d), (f) and (h) BLS-GSM with respectively DWT (28.6 dB), UDWT (29.3 dB) and FDCT (30.07 dB).

Fig. 6. Visual comparison of our block denoising to BLS-GSM on Lena 512×512. (a) original. (b) noisy $\sigma = 20$. (c), (e) and (g) block denoising with respectively DWT (30.51 dB), UDWT (31.47 dB) and FDCT (31.48 dB). (d), (f) and (h) BLS-GSM with respectively DWT (30.62 dB), UDWT (32 dB) and FDCT (31.6 dB).

512 × 512 image	DWT	UDWT	FDCT
Block	0.22	2.6	5.8
BLS-GSM	3	26	30

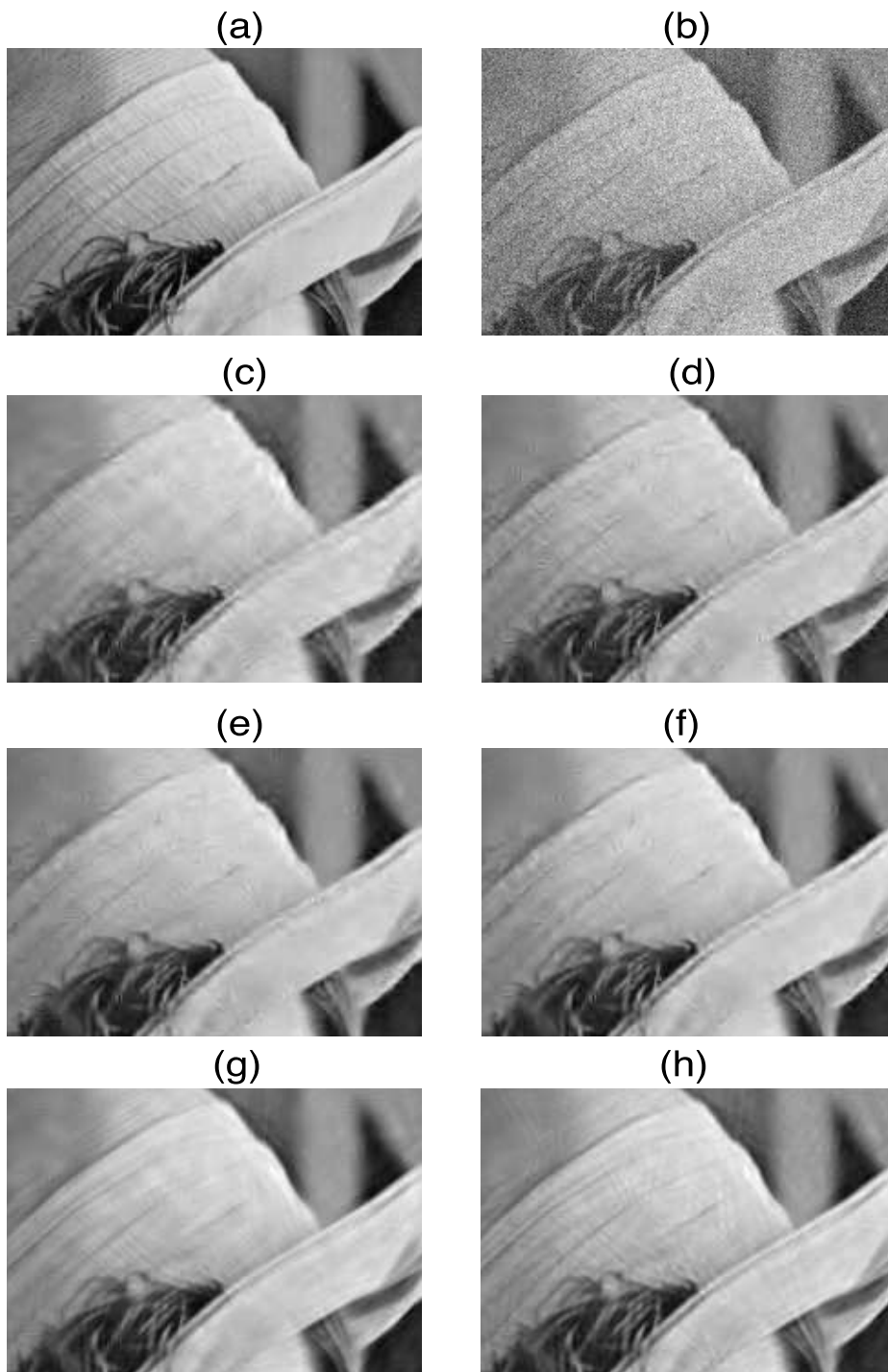| 256 × 256 image	DWT	UDWT	FDCT
Block	0.045	0.45	1.2
BLS-GSM	1	5.5	6.6

Table 1

Execution times in seconds for 512×512 images and 256×256 images. The algorithms were run under Matlab with an Intel Xeon 3GHz CPU, 8Gb RAM.

Barbara 512 × 512

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	36.81	32.50	30.07	28.41	27.16	26.16	23.74
BLS-GSM DWT	36.87	32.65	30.26	28.61	27.40	26.40	23.90
Block UDWT	37.37	33.24	30.80	29.09	27.77	26.70	24.01
BLS-GSM UDWT	37.44	33.43	31.06	29.40	28.16	27.13	24.49
Block FDCT	37.57	33.68	31.52	30.00	28.83	27.86	25.38
BLS-GSM FDCT	37.63	33.82	31.64	30.08	28.93	28.01	25.36

Lena 512 × 512

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	37.61	34.05	31.99	30.62	29.58	28.71	26.36
BLS-GSM DWT	37.41	33.97	31.68	30.62	29.62	28.70	26.36
Block UDWT	38.02	34.75	32.85	31.48	30.41	29.53	27.16
BLS-GSM UDWT	38.16	35.15	33.34	32.02	30.97	30.13	27.78
Block FDCT	38.09	34.78	32.86	31.45	30.43	29.55	27.12
BLS-GSM FDCT	38.10	34.93	33.03	31.60	30.53	29.65	27.02

House 256 × 256

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	37.63	33.47	31.33	29.86	28.76	27.79	25.41
BLS-GSM DWT	37.43	33.97	31.77	29.88	29.17	28.43	26.12
Block UDWT	38.10	34.31	32.31	30.86	29.75	28.80	26.35
BLS-GSM UDWT	38.17	34.79	32.95	31.52	30.41	29.49	27.00
Block FDCT	38.35	34.36	32.04	30.32	29.70	28.71	25.90
BLS-GSM FDCT	38.47	34.69	32.47	30.92	29.71	28.72	25.93

Boat 512 × 512

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	36.41	32.52	30.41	28.93	27.81	26.97	24.83
BLS-GSM DWT	36.06	32.36	30.36	29.04	27.35	26.76	24.86
Block UDWT	36.89	33.15	31.11	29.67	28.59	27.71	25.45
BLS-GSM UDWT	36.85	33.46	31.52	30.14	29.09	28.22	26.00
Block FDCT	36.89	33.07	31.03	29.65	28.59	27.70	25.49
BLS-GSM FDCT	36.74	33.17	31.20	29.80	28.77	27.88	25.52

Fingerprint 512 × 512

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	35.74	31.37	29.10	27.53	26.33	25.34	22.84
BLS-GSM DWT	35.53	31.08	28.82	27.08	26.01	25.11	22.72
Block UDWT	36.22	31.89	29.62	28.06	26.87	25.90	23.37
BLS-GSM UDWT	36.54	32.23	29.91	28.36	27.20	26.30	23.85
Block FDCT	36.13	31.98	29.66	28.03	26.84	25.92	23.51
BLS-GSM FDCT	36.34	32.14	29.82	28.21	27.05	26.14	23.70

Peppers 256 × 256

σ	5	10	15	20	25	30	50
PSNR$_{in}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15
Block DWT	36.81	32.56	30.28	28.64	27.42	26.42	23.77
BLS-GSM DWT	36.69	32.50	30.38	28.90	27.65	26.70	23.55
Block UDWT	37.48	33.60	31.37	29.74	28.52	27.52	24.71
BLS-GSM UDWT	37.59	33.96	31.78	30.17	28.99	27.97	25.16
Block FDCT	37.09	33.14	30.86	29.17	28.01	27.09	24.38
BLS-GSM FDCT	37.15	33.32	31.10	29.44	28.19	26.85	24.27

Table 2

Comparison of average PSNR over ten realizations of block denoising and BLS-GSM, with three transforms.

Following the philosophy of reproducible research, a toolbox is made available freely for download at the address

<div align="center">

`http://www.greyc.ensicaen.fr/~jfadili/software.html`

</div>

This toolbox is a collection of Matlab functions, scripts and datasets for image block denoising. It requires at least WaveLab 8.02 [5] to run properly. The toolbox implements the proposed block denoising procedure with several transforms and contains all scripts to reproduce the figures and tables reported in this paper.

5   Conclusion

In this paper, an Stein block thresholding algorithm for denoising d-dimensional data is proposed with a particular focus on 2D image. Our block denoising is a generalization of one-dimensional BlockJS to d dimensions, with other transforms that orthogonal wavelets, and handles noise in the coefficient domain beyond the i.i.d. Gaussian case. Its minimax properties are investigated, and a fast and appealing algorithm is described. The practical performance of the designed denoiser were shown to be very promising with several transforms and a variety of test images. It turns out that the proposed block denoiser is much faster than state-of-the art competitors in the literature while reaching comparable denoising performance.

We believe however that there is still room for improvement of our procedure. For instance, for $d = 2$, it would be interesting to investigate both theoretically and in practice how our results can be adapted to anisotropic blocks with possibly varying sizes. The rationale behind such a modification is to adapt the blocks to the geometry of the neighborhood. We expect that the analysis in this case, if possible, would be much more involved. Another interesting line of research would be to try to improve our convergence rates by relaxing the condition $q \leq p \wedge 2$. At this moment, given our definition of the smoothness space and our derivations in the proof (see appendix), we have not found a way around it yet. As remarked in subsection 3.1.1, a parameter δ was introduced, whose role becomes of interest when addressing linear inverse problems such as deconvolution. Extension of BlockJS to linear inverse problems remains also an open question. All these aspects need further investigation that we leave for a future work.

Appendix: Proofs

In this section, C represents a positive constant which may differ from one term to another. We suppose that n is large enough.

A Proof of Theorem 3.1

We have the decomposition:

$$R(\widehat{\theta}^*, \theta) = R_1 + R_2 + R_3, \tag{A.1}$$

where

$$R_1 = \sum_{j=0}^{j_0-1} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \mathbb{E}\left((\widehat{\theta}^*_{j,\ell,\mathbf{k}} - \theta_{j,\ell,\mathbf{k}})^2 \right), \quad R_2 = \sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \mathbb{E}\left((\widehat{\theta}^*_{j,\ell,\mathbf{k}} - \theta_{j,\ell,\mathbf{k}})^2 \right),$$

$$R_3 = \sum_{j=J_*+1}^{\infty} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta_{j,\ell,\mathbf{k}}^2.$$

Let us bound the terms R_1, R_3 and R_2 (by order of difficulty).

The upper bound for R_1. It follows from (A1) that

$$R_1 = n^{-r} \sum_{j=0}^{j_0-1} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \mathbb{E}\left(z_{j,\ell,\mathbf{k}}^2 \right) \leq Q_1 n^{-r} \sum_{j=0}^{j_0-1} 2^{j(d_*+\delta)} \mathrm{Card}(B_j)$$

$$= c_* Q_1 n^{-r} \sum_{j=0}^{j_0-1} 2^{j(d_*+\delta+\upsilon)} \leq C 2^{j_0(d_*+\delta+\upsilon)} n^{-r}$$

$$\leq C L^{(1/\min_{i=1,\dots,d} \mu_i))(d_*+\delta+\upsilon)} n^{-r} \leq C (\log n)^{(1/(d \min_{i=1,\dots,d} \mu_i))(d_*+\delta+\upsilon)} n^{-r}$$

$$\leq C n^{-2sr/(2s+\delta+d_*+\upsilon)}. \tag{A.2}$$

We used the inequality $2s/(2s+\delta+d_*+\upsilon) < 1$ which implies that, for a large enough n, $(\log n)^{(1/(d \min_{i=1,\dots,d} \mu_i))(d_*+\delta+\upsilon)} n^{(2s/(2s+\delta+d_*+\upsilon)-1)r} \leq 1$.

The upper bound for R_3. We distinguish the case $q \leq 2 \leq p$ and the case $q \leq p < 2$.

For $q \leq 2 \leq p$, we have $\mathbf{\Theta}^s_{p,q}(M) \subseteq \mathbf{\Theta}^s_{2,q}(M) \subseteq \mathbf{\Theta}^s_{2,2}(M)$. Hence

24

$$R_3 \le M^2 \sum_{j=J_*+1}^{\infty} 2^{-2js} \le C2^{-2J_*s} \le Cn^{-2sr/(d_*+\delta+\upsilon)} \le Cn^{-2sr/(2s+\delta+d_*+\upsilon)}.$$

$$(A.3)$$

For $q \le p < 2$, we have $\boldsymbol{\Theta}_{p,q}^s(M) \subseteq \boldsymbol{\Theta}_{2,q}^{s-d_*/p+d_*/2}(M) \subseteq \boldsymbol{\Theta}_{2,2}^{s-d_*/p+d_*/2}(M)$. We have

$$
\begin{aligned}
& s/(2s+\delta+d_*+\upsilon) \le (s-d_*/p+d_*/2)/(d_*+\delta+\upsilon) \\
\Leftrightarrow \quad & s(d_*+\delta+\upsilon) \le (s-d_*/p+d_*/2)(2s+\delta+d_*+\upsilon) \\
\Leftrightarrow \quad & 0 \le 2s^2 - (d_*/p - d_*/2)(2s+\delta+d_*+\upsilon) \\
\Leftrightarrow \quad & 0 \le 2s(s-d_*/p) + sd_* - (d_*/p - d_*/2)(\delta+d_*+\upsilon) \ .
\end{aligned}
$$

This implies that, if $sp > d_*$ and $s > (1/p - 1/2)(\delta + d_* + \upsilon)$, we have $s/(2s+\delta+d_*+\upsilon) \le (s-d_*/p+d_*/2)/(d_*+\delta+\upsilon)$. Therefore,

$$
\begin{aligned}
R_3 &\le M^2 \sum_{j=J_*+1}^{\infty} 2^{-2j(s-d_*/p+d_*/2)} \le C2^{-2J_*(s-d_*/p+d_*/2)} \\
&\le Cn^{-2r(s-d_*/p+d_*/2)/(d_*+\delta+\upsilon)} \le Cn^{-2sr/(2s+\delta+d_*+\upsilon)}.
\end{aligned}
\qquad (A.4)
$$

Putting (A.3) and (A.4) together, we obtain the desired upper bound.

The upper bound for R_2. We need the following result which will be proved later.

Lemma A.1 *Let $(v_i)_{i\in\mathbb{N}^*}$ be a sequence of real numbers and $(w_i)_{i\in\mathbb{N}^*}$ be a sequence of random variables. Set, for any $i \in \mathbb{N}^*$,*

$$u_i = v_i + w_i.$$

Then, for any $m \in \mathbb{N}^*$ and any $\lambda > 0$, the sequence of estimates $(\tilde{u}_i)_{i=1,\dots,m}$ defined by $\tilde{u}_i = u_i \left(1 - \lambda^2 \left(\sum_{i=1}^{m} u_i^2\right)^{-1}\right)_+$ satisfies

$$\sum_{i=1}^{m}(\tilde{u}_i - v_i)^2 \le 10 \sum_{i=1}^{m} w_i^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^{m} w_i^2\right)^{1/2} > \lambda/2\right\}} + 10 \min\left(\sum_{i=1}^{m} v_i^2, \lambda^2/4\right).$$

Lemma A.1 yields

$$R_2 = \sum_{j=j_0}^{J_*} \sum_{\ell\in B_j} \sum_{\mathbf{K}\in\mathcal{A}_j} \sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \mathbb{E}\left((\widehat{\theta}_{j,\ell,\mathbf{k}}^* - \theta_{j,\ell,\mathbf{k}})^2\right) \le 10(B_1 + B_2), \qquad (A.5)$$

where

$$B_1 = n^{-r} \sum_{j=j_0}^{J_*} \sum_{\ell\in B_j} \sum_{\mathbf{K}\in\mathcal{A}_j} \sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \mathbb{E}\left(z_{j,\ell,\mathbf{k}}^2 \mathbf{1}_{\left\{\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 > \lambda_* 2^{\delta j} L^d/4\right\}}\right)$$

and

$$B_2 = \sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \min \left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2, \lambda_* 2^{\delta j} L^d n^{-r}/4 \right).$$

Using (A2), we bound B_1 by

$$B_1 \leq Q_2 n^{-r} \leq Q_2 n^{-2sr/(2s+\delta+d_*+\upsilon)}. \tag{A.6}$$

To bound B_2, we again distinguish the case $q \leq 2 \leq p$ and the case $q \leq p < 2$.

For $q \leq 2 \leq p$, we have $\theta \in \mathbf{\Theta}_{p,q}^s(M) \subseteq \mathbf{\Theta}_{2,q}^s(M) \subseteq \mathbf{\Theta}_{2,2}^s(M)$. Let j_s be the integer $j_s = \lfloor (r/(2s+\delta+d_*+\upsilon)) \log_2 n \rfloor$. We then obtain the bound

$$B_2 \leq 4^{-1} \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j\delta} \text{Card}(\mathcal{A}_j) \text{Card}(B_j) + \sum_{j=j_s+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta_{j,\ell,\mathbf{k}}^2$$

$$\leq 4^{-1} c_* \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j(d_*+\delta+\upsilon)} L^{-d} + \sum_{j=j_s+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta_{j,\ell,\mathbf{k}}^2$$

$$\leq C n^{-r} 2^{j_s(d_*+\delta+\upsilon)} + M^2 \sum_{j=j_s+1}^{J_*} 2^{-2js}$$

$$\leq C n^{-r} 2^{j_s(d_*+\delta+\upsilon)} + C 2^{-2j_s s} \leq C n^{-2sr/(2s+\delta+d_*+\upsilon)}. \tag{A.7}$$

Putting (A.5), (A.6) and (A.7) together, it follows immediately that

$$R_2 \leq C n^{-2sr/(2s+\delta+d_*+\upsilon)}. \tag{A.8}$$

Let's now turn to the case $q \leq p < 2$. Let j_s^* be the integer $j_s^* = \lfloor (r/(2s+\delta+d_*+\upsilon)) \log_2(n/\log n) \rfloor$. We have

$$B_2 \leq D_1 + D_2 + D_3, \tag{A.9}$$

where

$$D_1 = 4^{-1} \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s^*} 2^{j\delta} \text{Card}(\mathcal{A}_j) \text{Card}(B_j),$$

$$D_2 = 4^{-1} \lambda_* L^d n^{-r} \sum_{j=j_s^*+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} 2^{\delta j} \mathbf{1}_{\left\{ \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 > \lambda_* 2^{\delta j} L^d n^{-r}/4 \right\}}$$

and

$$D_3 = \sum_{j=j_s^*+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \mathbf{1}_{\left\{ \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \leq \lambda_* 2^{\delta j} L^d n^{-r}/4 \right\}}.$$

We have

$$D_1 \leq 4^{-1} c_* \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s^*} 2^{j(d_*+\delta+\upsilon)} L^{-d} \leq C n^{-r} 2^{j_s^*(d_*+\delta+\upsilon)}$$

$$\leq C (\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}. \tag{A.10}$$

26

Moreover, using the classical inequality $\|\theta\|_2^p \le \|\theta\|_p^p, p < 2$, we obtain

$$
\begin{aligned}
D_2 &\le C L^d n^{-r} (L^d n^{-r})^{-p/2} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \right)^{p/2} \\
&\le C(\log n/n)^{r(1-p/2)} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} |\theta_{j,\ell,\mathbf{k}}|^p. \qquad (\text{A.11})
\end{aligned}
$$

Since $q \le p$, we have $\boldsymbol{\Theta}_{p,q}^s(M) \subseteq \boldsymbol{\Theta}_{p,p}^s(M)$. Combining this with $sp > d_*$ and $s > (1/p - 1/2)(\delta + d_* + \upsilon)$, we obtain

$$
\begin{aligned}
D_2 &\le C(\log n/n)^{r(1-p/2)} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} 2^{-j(s+d_*/2-d_*/p)p} \\
&\le C(\log n/n)^{r(1-p/2)} 2^{-j_s^*(s+d_*/2-d_*/p-\delta/p+\delta/2)p} \\
&\le C(\log n/n)^{(2s+\upsilon(1-p/2))r/(2s+\delta+d_*+\upsilon)} \\
&\le C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}. \qquad (\text{A.12})
\end{aligned}
$$

We have, for any $\mathbf{k} \in U_{j,\mathbf{K}}$, the inclusion $\left\{ \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \le \lambda_* 2^{\delta j} L^d n^{-r}/4 \right\} \subseteq \left\{ |\theta_{j,\ell,\mathbf{k}}| \le (\lambda_* 2^{\delta j} L^d n^{-r})^{1/2}/2 \right\}$. Therefore,

$$
\begin{aligned}
D_3 &\le \sum_{j=j_s^*+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \mathbf{1}_{\left\{ |\theta_{j,\ell,\mathbf{k}}| \le (\lambda_* 2^{\delta j} L^d n^{-r})^{1/2}/2 \right\}} \\
&\le C(\lambda_* L^d n^{-r})^{1-p/2} \sum_{j=j_s^*+1}^{J_*} 2^{j\delta(1-p/2)} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} |\theta_{j,\ell,\mathbf{k}}|^p,
\end{aligned}
$$

which is the same bound as for D_2 in (A.11). Then using similar arguments as those used for in (A.12), we arrive at

$$
D_3 \le C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}. \qquad (\text{A.13})
$$

Inserting (A.10), (A.12) and (A.13) into (A.9), it follows that

$$
R_2 \le C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}. \qquad (\text{A.14})
$$

Finally, bringing (A.1), (A.2), (A.3),(A.4), (A.8) and (A.14) together we obtain

$$
\sup_{\theta \in \boldsymbol{\Theta}_{p,q}^s(M)} R(\widehat{\theta}^*, \theta) \le R_1 + R_2 + R_3 \le C\rho_n,
$$

where ρ_n is defined by (3.4). This ends the proof of Theorem 3.1.

27

B   Proof of Lemma A.1

We have

$$\sum_{i=1}^{m}(\tilde{u}_i - v_i)^2 = \max(A, B),\tag{B.1}$$

where

$$A = \sum_{i=1}^{m}\left(w_i - \lambda^2 u_i \left(\sum_{i=1}^{m}u_i^2\right)^{-1}\right)^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}>\lambda\right\}}, \quad B = \sum_{i=1}^{m}v_i^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}\le\lambda\right\}}.$$

Let's bound A and B, in turn.

The upper bound for A**.** Using the elementary inequality $(a-b)^2 \le 2(a^2 + b^2)$, we have

$$A \le 2\sum_{i=1}^{m}\left(w_i^2 + \lambda^4 u_i^2 \left(\sum_{i=1}^{m}u_i^2\right)^{-2}\right)\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}>\lambda\right\}}$$

$$= 2\left(\sum_{i=1}^{m}w_i^2 + \lambda^4 \left(\sum_{i=1}^{m}u_i^2\right)^{-1}\right)\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}>\lambda\right\}}$$

$$\le 2\left(\sum_{i=1}^{m}w_i^2 + \lambda^2\right)\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}>\lambda\right\}}.\tag{B.2}$$

Set

$$D = 2\left(\sum_{i=1}^{m}w_i^2 + \lambda^2\right)\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}u_i^2\right)^{1/2}>\lambda\right\}}.$$

We have the decomposition

$$D = D_1 + D_2,\tag{B.3}$$

where

$$D_1 = D\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}w_i^2\right)^{1/2}>\lambda/2\right\}}, \qquad D_2 = D\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}w_i^2\right)^{1/2}\le\lambda/2\right\}}.$$

We clearly have

$$D_1 \le 2\left(\sum_{i=1}^{m}w_i^2 + \lambda^2\right)\mathbf{1}_{\left\{\left(\sum_{i=1}^{m}w_i^2\right)^{1/2}>\lambda/2\right\}} \le 10\sum_{i=1}^{m}w_i^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^{m}w_i^2\right)^{1/2}>\lambda/2\right\}}.$$
$$\tag{B.4}$$

28

Using the Minkowski inequality, we have the inclusion $\left\{ \left(\sum_{i=1}^m u_i^2 \right)^{1/2} > \lambda \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\} \subseteq \left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} > \lambda/2 \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\}$. Therefore

$$
\begin{aligned}
D_2 &\le 2 \left(\sum_{i=1}^m w_i^2 + \lambda^2 \right) \mathbf{1}_{\left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} > \lambda/2 \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\}} \\
&\le 10 \min \left(\sum_{i=1}^m v_i^2, \lambda^2/4 \right).
\end{aligned}
\tag{B.5}
$$

If we combine (B.2), (B.3), (B.4) and (B.5), we obtain

$$
A \le D \le 10 \sum_{i=1}^m w_i^2 \mathbf{1}_{\left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}} + 10 \min \left(\sum_{i=1}^m v_i^2, \lambda^2/4 \right).
\tag{B.6}
$$

The upper bound for B. We have the decomposition

$$
B = G_1 + G_2
\tag{B.7}
$$

$$
G_1 = B \mathbf{1}_{\left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}}, \qquad G_2 = B \mathbf{1}_{\left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\}}.
$$

Using the Minkowski inequality, we have again the inclusion $\left\{ \left(\sum_{i=1}^m u_i^2 \right)^{1/2} \le \lambda \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\} \subseteq \left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} \le 3 \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}$. It follows that

$$
\begin{aligned}
G_1 &\le \sum_{i=1}^m v_i^2 \mathbf{1}_{\left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} \le 3 \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}} \\
&\le 9 \sum_{i=1}^m w_i^2 \mathbf{1}_{\left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}}.
\end{aligned}
\tag{B.8}
$$

Another application of the Minkowski inequality leads to the inclusion $\left\{ \left(\sum_{i=1}^m u_i^2 \right)^{1/2} \le \lambda \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\} \subseteq \left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} \le 3\lambda/2 \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\}$. It follows that

$$
\begin{aligned}
G_2 &\le \sum_{i=1}^m v_i^2 \mathbf{1}_{\left\{ \left(\sum_{i=1}^m v_i^2 \right)^{1/2} \le 3\lambda/2 \right\} \cap \left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} \le \lambda/2 \right\}} \\
&\le \min \left(\sum_{i=1}^m v_i^2, 9\lambda^2/4 \right).
\end{aligned}
\tag{B.9}
$$

Therefore, if we combine (B.7), (B.8) and (B.9), we obtain

$$
B \le 9 \sum_{i=1}^m w_i^2 \mathbf{1}_{\left\{ \left(\sum_{i=1}^m w_i^2 \right)^{1/2} > \lambda/2 \right\}} + \min \left(\sum_{i=1}^m v_i^2, 9\lambda^2/4 \right).
\tag{B.10}
$$

Putting (B.1), (B.6) and (B.10) together, we have

$$\sum_{i=1}^{m}(\tilde{u}_i - v_i)^2 = \max\left(A, B\right)$$

$$\leq 10 \sum_{i=1}^{m} w_i^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^{m} w_i^2\right)^{1/2} > \lambda/2\right\}} + 10 \min\left(\sum_{i=1}^{m} v_i^2, \lambda^2/4\right).$$

Lemma A.1 is proved.

C   Proof of Proposition 3.1

First of all, notice that the Jensen inequality, (A3) and the fact that $\mathrm{Card}(D_j) \leq 2^{jd_*}$ imply

$$\sup_{j \in \{0,\dots,J\}} \sup_{\ell \in B_j} 2^{-j(d_*+\delta)} \sum_{\mathbf{k} \in D_j} \mathbb{E}\left(z_{j,\ell,\mathbf{k}}^2\right) \leq \sup_{j \in \{0,\dots,J\}} 2^{-j(d_*+\delta)} \sup_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \left(\mathbb{E}\left(z_{j,\ell,\mathbf{k}}^4\right)\right)^{1/2}$$

$$\leq Q_3^{1/2} \sup_{j \in \{0,\dots,J\}} 2^{-jd_*} \mathrm{Card}(D_j)$$

$$\leq Q_3^{1/2}.$$

Therefore (A1) is satisfied.

Let's now turn to (A4). Again, the Jensen inequality yields

$$\sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \mathbb{E}\left(z_{j,\ell,\mathbf{k}}^2 \mathbf{1}_{\left\{\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} > (\lambda_* 2^{\delta j} L^d)^{1/2}/2\right\}}\right)$$

$$\leq \sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \left(\mathbb{E}\left(z_{j,\ell,\mathbf{k}}^4\right)\right)^{1/2} \left(\mathbb{P}\left(\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} > (\lambda_* 2^{\delta j} L^d)^{1/2}/2\right)\right)^{1/2}.$$

Using (A3), it comes that

30

$$\sum_{j=j_0}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{K} \in \mathcal{A}_j} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} \left(\mathbb{E}\left(z_{j,\ell,\mathbf{k}}^4 \right) \right)^{1/2} \left(\mathbb{P}\left(\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 \right)^{1/2} > (\lambda_* 2^{\delta j} L^d)^{1/2}/2 \right) \right)^{1/2}$$

$$\leq C 2^{J_*(d_*+\delta+\upsilon)} Q_3^{1/2} \sup_{j \in \{j_0,\dots,J_*\}} \sup_{\ell \in B_j} \sup_{\mathbf{K} \in \mathcal{A}_j} \left(\mathbb{P}\left(\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 \right)^{1/2} > (\lambda_* 2^{\delta j} L^d)^{1/2}/2 \right) \right)^{1/2}$$

$$\leq C n^r Q_3^{1/2} \sup_{j \in \{j_0,\dots,J_*\}} \sup_{\ell \in B_j} \sup_{\mathbf{K} \in \mathcal{A}_j} \left(\mathbb{P}\left(\left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 \right)^{1/2} > (\lambda_* 2^{\delta j} L^d)^{1/2}/2 \right) \right)^{1/2}.$$

$$(C.1)$$

To bound the probability term, we introduce the Borell inequality. For further details about this inequality, see, for instance, [2].

Lemma C.1 (The Borell inequality) *Let $\mathcal{D}$ be a subset of $\mathbb{R}$ and $(\eta_t)_{t \in \mathcal{D}}$ be a centered Gaussian process. Suppose that*

$$\mathbb{E}\left(\sup_{t \in \mathcal{D}} \eta_t \right) \leq N \quad \text{and} \quad \sup_{t \in \mathcal{D}} \mathbb{V}(\eta_t) \leq Z.$$

Then, for any $x > 0$, we have

$$\mathbb{P}\left(\sup_{t \in \mathcal{D}} \eta_t \geq x + N \right) \leq \exp(-x^2/(2Z)).$$

Let us consider the set $\mathcal{S}_2$ defined by $\mathcal{S}_2 = \{a = (a_{\mathbf{k}}) \in \mathbb{R}^*; \ \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} a_{\mathbf{k}}^2 \leq 1\}$, and the centered Gaussian process $\mathcal{Z}(a)$ defined by

$$\mathcal{Z}(a) = \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} a_{\mathbf{k}} z_{j,\ell,\mathbf{k}}.$$

We have by the Cauchy-Schwartz inequality

$$\sup_{a \in \mathcal{S}_2} \mathcal{Z}(a) = \sup_{a \in \mathcal{S}_2} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} a_{\mathbf{k}} z_{j,\ell,\mathbf{k}} = \left(\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2 \right)^{1/2}.$$

In order to use Lemma C.1, we have to investigate the upper bounds for $\mathbb{E}(\sup_{a \in \mathcal{S}_2} \mathcal{Z}(a))$ and $\sup_{a \in \mathcal{S}_2} \mathbb{V}(\mathcal{Z}(a))$.

The upper bound for $\mathbb{E}(\sup_{a \in \mathcal{S}_2} \mathcal{Z}(a))$. The Jensen inequality and (A3) imply that

$$\mathbb{E}\left(\sup_{a\in\mathcal{S}_2}\mathcal{Z}(a)\right) = \mathbb{E}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2}\right) \leq \left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \mathbb{E}(z_{j,\ell,\mathbf{k}}^2)\right)^{1/2}$$

$$\leq \left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \left(\mathbb{E}(z_{j,\ell,\mathbf{k}}^4)\right)^{1/2}\right)^{1/2} \leq Q_3^{1/4} 2^{\delta j/2} L^{d/2}$$

$$\leq Q_3^{1/4} 2^{\delta j/2} (\log n)^{1/2}.$$

So, $N = Q_3^{1/4} 2^{\delta j/2} (\log n)^{1/2}$.

The upper bound for $\sup_{a\in\mathcal{S}_2} \mathbb{V}(\mathcal{Z}(a))$. By assumption, for any $j\in\mathbb{N}$ and $\mathbf{k}\in D_j$, we have $\mathbb{E}(z_{j,\ell,\mathbf{k}}) = 0$. The assumption (A4) yields

$$\sup_{a\in\mathcal{S}_2} \mathbb{V}\left(\mathcal{Z}(a)\right) = \sup_{a\in\mathcal{S}_2} \mathbb{E}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} a_{\mathbf{k}} z_{j,\ell,\mathbf{k}}\right)^2\right) \leq Q_4 2^{\delta j}.$$

It is then sufficient to take $Z = Q_4 2^{\delta j}$.

Combining the obtained expressions of N and Z with Lemma C.1, for any $j\in\{j_0,...,J_*\}$, $\mathbf{K}\in\mathcal{A}_j$ and $\mathbf{k}\in U_{j,\mathbf{K}}$, we have

$$\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} \geq (\lambda_* 2^{\delta j} L^d)^{1/2}/2\right)$$

$$= \mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} \geq (\lambda_*^{1/2}/2 - Q_3^{1/4})(2^{\delta j} L^d)^{1/2} + Q_3^{1/4}(2^{\delta j} L^d)^{1/2}\right)$$

$$= \mathbb{P}\left(\sup_{a\in\mathcal{S}_2} \mathcal{Z}(a) \geq (\lambda_*^{1/2}/2 - Q_3^{1/4})(2^{\delta j} L^d)^{1/2} + N\right)$$

$$\leq \exp\left(-(\lambda_*^{1/2}/2 - Q_3^{1/4})^2 2^{\delta j} L^d/(2Z)\right) \leq n^{-r(\lambda_*^{1/2}/2 - Q_3^{1/4})^2/(2Q_4)}.$$

Since $\lambda_* = 4\left((2Q_4)^{1/2} + Q_3^{1/4}\right)^2$, it follows that

$$\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} \geq (\lambda_* 2^{\delta j} L^d)^{1/2}/2\right) \leq n^{-r}. \tag{C.2}$$

Putting (C.1) and (C.2) together, we have proved (A2). This ends the proof of Proposition 3.1.

D Proof of Proposition 3.2

The proof of this proposition is similar to the one of Theorem 3.1. The only difference is that, instead of using Lemma A.1, we use Lemma D.1 below.

Lemma D.1 ([9]) *Let $(v_i)_{i\in\mathbb{N}^*}$ be a sequence of real numbers, $(w_i)_{i\in\mathbb{N}^*}$ be i.i.d. $\mathcal{N}(0,1)$ and $\sigma \in \mathbb{R}^*$. Set, for any $i \in \mathbb{N}^*$,*

$$u_i = v_i + \sigma w_i.$$

Then, for any $m \in \mathbb{N}^*$ and any $\gamma > 1$, the sequence of estimates $(\tilde{u}_i)_{i=1,\dots,m}$ defined by $\tilde{u}_i = u_i \left(1 - \gamma m \sigma^2 \left(\sum_{i=1}^m u_i^2\right)^{-1}\right)_+$ satisfies

$$\mathbb{E}\left(\sum_{i=1}^m (\tilde{u}_i - v_i)^2\right) \le 2\sigma^2 \pi^{-1/2}(\gamma-1)^{-1}m^{-1/2}e^{-(m/2)(\gamma-\log\gamma-1)} + \gamma \min\left(\sum_{i=1}^m v_i^2, \sigma^2 m\right).$$

To clarify, if the variables $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are i.i.d. $\mathcal{N}(0,1)$ then Lemma D.1 improves the bound of the term B_1 appearing in the proof of Theorem 3.1.

If we analyze the proof of Theorem 3.1 and we use Lemma A.1 instead of Lemma D.1, we see that it is enough to determine λ_* such that there exists a constant $Q_5 > 0$ satisfying

$$\sum_{j=j_0}^{J_*} \text{Card}(B_j)\text{Card}(\mathcal{A}_j)e^{-(L^q/2)(\lambda_* - \log\lambda_* - 1)} \le Q_5.$$

(It corresponds to the bound of the term B_1 that appears in (A.5)). If λ_* is the root of $x - \log x = 3$, it comes that

$$\sum_{j=j_0}^{J_*} \text{Card}(B_j)\text{Card}(\mathcal{A}_j)e^{-(L^q/2)(\lambda_* - \log\lambda_* - 1)} = c_* e^{-(L^q/2)(\lambda_* - \log\lambda_* - 1)} 2^{J_*(d_*+v)}$$

$$\le C e^{-(L^q/2)(\lambda_* - \log\lambda_* - 1)} n^r \le Q_5.$$

Proposition 3.2 is proved.

References

[1] F. Abramovich, T. Besbeas, and T. Sapatinas. Empirical Bayes approach to block wavelet function estimation. *Computational Statistics and Data Analysis*, 39:435–451, 2002.

[2] R. J. Adler. *An introduction to continuity, extrema, and related topics for general Gaussian processes.* Institute of Mathematical Statistics, Hayward, CA, 1990.

[3] A. Antoniadis, J. Bigot, and T. Sapatinas. Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study. *Journal of Statistical Software*, 6(6), 2001.

[4] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *Journal of Fourier Analysis and Applications*, 13(1):39–70, 2007.

[5] J. Buckheit and D.L. Donoho. Wavelab and reproducible research. In A. Antoniadis, editor, *Wavelets and Statistics.* Springer, 1995.

[6] T. Cai. On adaptivity of blockshrink wavelet estimator over Besov spaces. Technical Report 97- 05, Department of Statistics, Purdue University, 1997.

[7] T. Cai. Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of Statistics*, 27:898–924, 1999.

[8] T. Cai. On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. *Statistica Sinica*, 12(4):1241–1273, 2002.

[9] T. Cai and B. W. Silverman. Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya*, 63:127–148, 2001.

[10] T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. *Annals of Statistics*, 2007. to appear.

[11] E. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math*, 58(11):1472–1528, 2005.

[12] E. J. Candès. Ridgelets: Estimating with Ridge Functions. *Annals of Statistics*, 31:1561–1599, 1999.

[13] E. J. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Modeling and Simulation*, 5(3):861–899, 2006.

[14] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L.L. Schumaker, editors, *Curve and Surface Fitting: Saint-Malo 1999*, Nashville, TN, 1999. Vanderbilt University Press.

[15] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math*, 57(2):219–266, 2004.

[16] E. J. Candès and D.L. Donoho. Ridgelets: the key to high dimensional intermittency? *Philosophical Transactions of the Royal Society of London A*, 357:2495–2509, 1999.

[17] L. Cavalier and A.B. Tsybakov. Penalized blockwise stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods of Stat.*, 10:247–282, 2001.

[18] C. Chaux, A. Benazza-Benyahia, and J.-C. Pesquet. A block-thresholding method for multispectral image denoising. In *SPIE Conference*, volume 5914, pages 1H–1–1H–13, San Diego, CA, 2005.

[19] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear stein based estimator for multichannel image denoising. *IEEE Trans. Signal Processing*, 56(8):3855–3870, 2008.

[20] C. Chesneau. Wavelet estimation via block thresholding : a minimax study under l_p risk. *Statistica Sinica*, 2008. to appear.

[21] E. Chicken. Asymptotic rates for coefficient-dependent and block-dependent thresholding in wavelet regression. Technical Report M960, Department of Statistics, Florida State University, 2005.

[22] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

[23] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods, II. *Math. Nachr.*, 132:207–237, 1987.

[24] H.-Y. Gao. Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4):469–488, 1998.

[25] H.-.Y Gao and A. G. Bruce. Waveshrink with firm shrinkage. *Statistica Sinica*, 7:855–874, 1997.

[26] P. Hall, G. Kerkyacharian, and D. Picard. Block threshold rules for curve estimation using kernel and wavelet methods. *Annals of Statistics*, 26(3): 922–942, 1998.

[27] P. Hall, G. Kerkyacharian, and D. Picard. On the minimax optimality of block thresholded wavelet estimators. *Staistica Sinica*, 9:33–50, 1999.

[28] W. Härdle, G. Kerkyacharian, D. Picard, and A. B. Tsybakov. *Wavelets, Approximation and Statistical Applications*. Lecture Notes in Statistics. Springer, New York, 1998.

[29] I. Johnstone. Wavelets and the theory of non-parametric function. *Phil. Trans. Roy. Soc. Lond. A.*, 357:2475–2494, 1999.

[30] I. Johnstone. Function estimation and gaussian sequence models. Draft of Monograph, 2002. URL `http://www-stat.stanford.edu/$\sim$imj/`.

[31] I. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 6(3):547–573, 2004.

[32] A. P. Korostelev and A. B. Tsybakov. *MinimaxTheory of ImageReconstruction*, volume 82. Springer, 1993.

[33] F. Luisier, T. Blu, and M. Unser. A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Trans. Image Processing*, 16(3):593–606, March 2007.

[34] Y. Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. ISBN 0-521-42000-8; 0-521-45869-2. Translated from the 1990 French original by D. H. Salinger.

[35] E. Le Pennec, C. Dossal, G. Peyré, and S. Mallat. Débruitage géométrique d'images dans des bases orthonormées de bandelettes. In *GRETSI Conference*, Troyes, France, 2007.

[36] E. Le Pennec and S. Mallat. Bandelet image approximation and compres-

sion. *SIAM Multiscale Modeling and Simulation*, 4(3):992–1039, 2005.

[37] A. Pizurica, W. Philips, I. Lemahieu, and M. Achenoy. Joint inter- and intrascale statistical model for Bayesian wavelet-based image denoising. *IEEE Trans. Image Processing*, 11(5):545–557, 2002.

[38] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Processing*, 12(11):1338–1351, November 2003.

[39] L. Sendur and I.W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Trans. Signal Processing*, 50(11):2744–2756, November 2002.

[40] C. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 10:1135–1151, 1990.

[41] G. Yu, S. Mallat, and E. Bacry. Audio denoising by time-frequency block thresholding. *IEEE Trans. Signal Processing*, 56(5):1830–1839, 2008.