Stein Block Thresholding For Image Denoising

Christophe Chesneau, Jalal M. Fadili, Jean-Luc Starck

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 [5] of family of smoothness spaces on $\mathbb{R}^d$, a subclass of so-called decomposition spaces [28]. 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 noise sequence model in the transform coefficient domain satisfying some mild assumptions. This covers for instance the Gaussian case with frames where the noise is not white in the coefficient domain. 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

1 Introduction

Consider the nonparametric regression model:

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

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 [31, 32, 33]. 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, let's consider the one-dimensional case $d = 1$. The most standard of wavelet shrinkage methods is VisuShrink of [27]. 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. Other term-by-term shrinkage rules have been developed in the literature. An exhaustive account is provided in [3] that the interested reader may refer to.

The individual thresholding 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. 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 [29, 30] 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 [29, 30] is defined with the block size $(\log n)^2$. BlockShrink of [9, 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 by removing the extra logarithmic factor. The minimax properties of BlockShrink under the L_p risk have been studied in [22]. Other local block thresholding rules have been developed. Among them, there is BlockJS of [10, 8] which combines James-Stein rule (see [43]) with the wavelet methodology. In particular, it is minimax optimal but improves the constant in the rate. From a practical point view, it is better than BlockShrink. Further details about the theoretical performances of BlockJS can be found in [19]. We refer to [3] and [11] for a comprehensive simulation study. Variations of BlockJS are BlockSure of [23] and SureBlock of [12]. The distinctive aspect of these block thresholding procedures is to provide data-driven algorithms to choose 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 [27], 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 [17], curvelets [15, 18] or bandelets [38], 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. 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 [41, 42, 40, 36] which are amongst the most efficient in the literature. Except [36], all cited approaches use usual 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 [14] 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 [16] [2]. The wedgelet estimator of Donoho [26] adapts to the anisotropic smoothness of the image by finding the best edgelet-decorated recursive partition of the image which minimizes a complexity-penalized sum of squares. This wedgelet estimator is nearly-minimax with the rate $O(n^{-2\alpha/(\alpha+1)}), \alpha \in [1, 2]$ for C^α functions away from C^α edges. Similarly, Le Pennec et al. [39] have recently proved that individual thresholding in an adaptively selected best bandelet orthobasis achieves nearly this minimax rate over the C^α-C^α image class.

In the image processing community, block thresholding/shrinkage in a non-Bayesian framework has been used very little. In [20, 21] 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. [45] advocated the use of BlockJS [10] 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.

1.3 Contributions

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 (2.2). The choice of this space is inspired by the characterization provided in [5] of family of smoothness spaces on $\mathbb{R}^d$, a subclass of so-called decomposition spaces [5, 28]. We will elaborate more on these (sparsity) smoothness spaces later in subsection 2.2. 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.

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

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 [41]. Our procedure is however much simpler to implement and has a much lower computational cost than usual 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. In Section 2, we describe the multi-dimensional BlockJS under a fairly general noise model beyond the i.i.d. Gaussian case. This section also contains our main theoretical results. In Section 3, a comprehensive experimental study is reported and discussed. We finally conclude in Section 4 and point to some perspectives. The proofs of the results are deferred to the appendix awaiting inspection by the interested reader.

2 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.

Let $(\psi_{j,\ell,\mathbf{k}}(\mathbf{x}))_{j,\ell,\mathbf{k}}$, $\mathbf{x} \in [0,1]^d$ be a collection of unit-norm functions forming a (tight) frame of $L^2([0,1]^d)$. Set $\theta_{j,\ell,\mathbf{k}} = \langle f, \psi_{j,\ell,\mathbf{k}} \rangle$ the unknown frame coefficients of f, $y_{j,\ell,\mathbf{k}} = \langle Y, \psi_{j,\ell,\mathbf{k}} \rangle$ and $z_{j,\ell,\mathbf{k}}$ is a sequence of noise random variables. We then observe a multi-dimensional sequence of coefficients $(y_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ defined by

$$y_{j,\ell,\mathbf{k}} = \theta_{j,\ell,\mathbf{k}} + n^{-r/2} z_{j,\ell,\mathbf{k}}, \quad j = 0, ..., J, \quad \ell \in B_j, \quad \mathbf{k} \in D_j, \tag{2.1}$$

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

A wide variety of statistical models fall within the scope of the sequence model (2.1). For instance, consider the d-dimensional Gaussian model in zero-mean white noise (1.1). Then we have $y_{j,\ell,\mathbf{k}} = \theta_{j,\ell,\mathbf{k}} + \sigma z_{j,\ell,\mathbf{k}}$. Taking $\sigma = n^{-d/2}$ as the classical noise level for a dataset defined on a d-dimensional discrete grid of equally-spaced samples, (2.1) is a projection of (1.1) onto the frame $(\psi_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$. In general, setting $\sigma = n^{-r/2}$, the sequence model (2.1) is equivalent, in some sense, to (1.1) (see, for instance, [6]). Other models can be re-expressed in this form; see the fractional Gaussian noise model discussed in Remark 2.1.

The notations of the indices are those of multiscale transforms (wavelets, curvelets, etc), generally corresponding to tight frames or orthobases. Here, 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, and B_j is the set of subbands at scale j. 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 $\upsilon = 0$, $c^* = 3$, and $\mu_1 = \mu_2 = 1$. Thus, as expected, we get three isotropic subbands at each scale. For the second generation curvelet transform [18], we have $\upsilon = 1/2$, $\mu_1 = 1$ and $\mu_2 = 1/2$ which corresponds to the parabolic scaling of curvelets.

Remark 2.1 *(Comment on r) r is a tuning parameter depending on the model. For standard statistical models (e.g. Gaussian white noise), we have $r = d$. However, there exist sophisticated models where $r \in (0, d)$. This is case for the d-dimensional fractional Gaussian noise defined by $dY(\mathbf{x}) = f(\mathbf{x})d\mathbf{x} + n^{-d(1-H)} dW_H(\mathbf{x})$, $\mathbf{x} \in [0,1]^d$, where $\{W_H(\mathbf{x});\ \mathbf{x} \in [0,1]^d\}$ is a fractional Brownian motion and $H \in (1/2, 1)$. This model can be reexpressed as (2.1) with $r = 2d(1-H)$ in a tight frame domain. We refer to [44].*

Remark 2.2 *Our study is formulated in the coefficient domain according to the sequence model (2.1). However, as the transforms that we deal with correspond to frames, thanks to the generalized Parseval relation, the convergence*

rates on the function estimator can be easily deduced from those stated in terms on the coefficients.

2.1.1 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 $J_* = \lfloor (r/(d_* + \delta + \upsilon)) \log_2 n \rfloor$. For any scale $j \in \{j_0, ..., J_*\}$, let

- $\mathcal{A}_j = \prod_{i=1}^{d} \{1, ..., \lfloor 2^{\mu_i j} L^{-1} \rfloor\}$ be the set indexing the blocks at scale j.
- For each block index $\mathbf{K} = (K_1, ..., K_d) \in \mathcal{A}_j$, $U_{j,\mathbf{K}} = \{\mathbf{k} \in D_j;\ (K_1 - 1)L \leq k_1 \leq K_1 L - 1, ..., (K_d - 1)L \leq k_d \leq K_d L - 1\}$ 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 $\delta \geq 0$, $\lambda_* > 0$, $Q_1 > 0$ and $Q_2 > 0$ independent of n such that

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

(A2)

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

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 2.1 and 2.2 below.

Remark 2.3 (*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 3), 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 2.3) to other inverse problems such as deconvolution. The role of δ becomes explicit 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 [34].*

2.2 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 (2.1). 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 adopt 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

$$\mathbf{\Theta}^s_{p,q}(M)=\left\{\theta=(\theta_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}};\quad\left(\sum_{j=0}^{\infty}\sum_{\ell\in B_j}\left(2^{j(s+d_*/2-d_*/p)}\left(\sum_{\mathbf{k}\in D_j}|\theta_{j,\ell,\mathbf{k}}|^p\right)^{1/p}\right)^q\right)^{1/q}\leq M\right\},\tag{2.2}$$

with a smoothness parameter s, $0<p\leq+\infty$ and $0<q\leq+\infty$ are norm parameters [3], and $M\in(0,\infty)$ denotes the radius of the ball. We recall that $d_*=\sum_{i=1}^d\mu_i$.

The definition of these smoothness spaces is motivated by the work of [5]. 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 (2.2). 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 $\mathbf{B}^s_{p,q}$, for which suitable wavelet expansion [4] is known to provide a sparse representation [37]. In this case, from subsection 2.1 we have $d^*=d$, and $\mathbf{\Theta}^s_{p,q}(M)$ 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 [13]. For $d=2$, the angular wedges obey the parabolic scaling law [15]. 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 [5, Lemma 10] showed that the smoothness/sparseness space (2.2) and the smoothness/sparseness of the second-generation curvelets [18] are the same, in which case $d^*=3/2$. Embedding results for curvelet-type decomposition spaces relative to Besov spaces were also provided in [5]. Furthermore, it was shown that piecewise C^2 images away from piecewise-C^2 singularities, which are sparsely represented in the curvelet tight frame [15], are contained

[3] This is a slight abuse of terminology as for $0<p,q<1$, we have quasi-normed spaces.

[4] With a wavelet having sufficient regularity and number of vanishing moments [37].

in $\mathbf{\Theta}^{3/2+\beta}_{2/3,2/3}$, $\forall \beta > 0$. Even though the role and the range of β has not been clarified by the authors in [5].

2.3 Multi-dimensional block estimator

We wish to construct an adaptive estimator $\widehat{\theta} = (\widehat{\theta}_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ such that $\sup_{\theta \in \mathbf{\Theta}^s_{p,q}(M)} R\left(\widehat{\theta}, \theta\right)$ is as small as possible. To reach this goal, we propose a multi-dimensional version of the BlockJS procedure introduced in [10].

From subsection 2.1.1, recall the definitions of L, j_0, J_*, $\mathcal{A}_j$ and $U_{j,\mathbf{K}}$. We 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 $\mathbf{k} \in U_{j,\mathbf{K}}$ in the block $\mathbf{K} \in \mathcal{A}_j$ and subband $\ell \in B_j$,

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

$(x)_+ = \max(x, 0)$. In this definition, δ and λ_* denote the constants involved in (A1) and (A2). Thus, at the coarsest scales $j \in \{0, ..., j_0\}$, the observed coefficients $(y_{j,\mathbf{k}})_{\mathbf{k} \in U_{j,\mathbf{K}}}$ are left intact as usual. For each block $U_{j,\mathbf{K}}$ in the scales $j \in \{j_0, ..., J\}$, if the mean energy within the block $\sum_{\mathbf{k} \in U_{j,\mathbf{K}}} y^2_{j,\ell,\mathbf{k}}/L^d$ is larger than $\lambda_* n^{-r} 2^{\delta j}$ then $y_{j,\mathbf{k}}$ is shrunk by the amount $y_{j,\ell,\mathbf{k}} \frac{\lambda_* n^{-r} 2^{\delta j}}{\frac{1}{L^d} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} y^2_{j,\ell,\mathbf{k}}}$; otherwise, $\theta_{j,k}$ is estimated by zero. In fact, $\frac{\frac{1}{L^d} \sum_{\mathbf{k} \in U_{j,\mathbf{K}}} y^2_{j,\ell,\mathbf{k}}}{n^{-r}}$ can be interpreted as a local measure of signal-to-noise ratio in the block $U_{j,\mathbf{K}}$. Such a block thresholding originates from the James-Stein rule introduced in [43]. 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 2.5 below.

2.4 Minimax theorem

Theorem 2.1 below investigates the minimax rate of (2.3) over $\mathbf{\Theta}^s_{p,q}$.

Theorem 2.1 *Consider the model* (2.1) *for n large enough. Suppose that* (A1) *and* (A2) *are satisfied. Let* $\widehat{\theta}^*$ *be given as in* (2.3).

- *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

$$\rho_n = \begin{cases} n^{-2r(s+\upsilon(1/q-1/2))/(2s+\delta+d_*+2\upsilon/q)} & \text{for } p \geq 2,\ q \geq 2,\ s > \upsilon(1/2-1/q), \\ n^{-2sr/(2s+\delta+d_*+\upsilon)}, & \text{for } q \leq 2 \leq p, \\ (\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)}, & \text{for } q \leq p < 2,\ sp > d_* \vee (1-p/2)(\delta+d_*+\upsilon). \end{cases} \tag{2.4}$$

- *If $\upsilon = 0$, in addition to (2.4), we have $\rho_n = (\log n/n)^{2sr/(2s+\delta+d_*)}$ for $p \leq q < 2$ and $sp > d_* \vee (1-p/2)(\delta+d_*)$.*

The rates of convergence (2.4) are optimal for a wide class of variables $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ satisfying (A1) and (A2). 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 [10]. There is only a minor difference on the power of the logarithmic term for $p < 2$. Thus, Theorem 2.1 can be viewed as a generalization of that result.

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 2.1 gives two distinct rates depending whether $p \geq 2$ or $p < 2$, $\forall q$. Therefore, for $p \geq 2$, Theorem 2.1 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.

Note that the condition $sp > (1-p/2)(\delta + d_* + \upsilon)$ is only technical but seems inevitable. Such condition is usual when we deal with minimax rates of convergence over Besov spaces. See [25] for further details on the multidimensional Besov balls with the tensor-product wavelet basis ($\delta = 0$, $d_* = d$ and $\upsilon = 0$).

As far as curvelet-type decomposition spaces are concerned, from section 2.1 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

$$\rho_n = \begin{cases} n^{-(4s+2/q-1)/(2s+1+1/q)}, & \text{for } q \geq 2 \text{ and } p \geq 2, \\ n^{-2s/(s+1)}, & \text{for } q \leq 2 \leq p, \\ (\log n/n)^{2s/(s+1)}, & \text{for } q \leq p < 2,\ sp > \dfrac{3}{2} \vee (2-p). \end{cases}$$

where the logarithmic factor disappears only for $q \leq 2 \leq p$. Following the discussion of section 2.2, C^2-C^2 images correspond to a smoothness space $\mathbf{\Theta}^s_{p,q}$ with $p = q = 2/3$. Moreover, $\exists \kappa > 0$ such that taking $s = 2 + \kappa$ satisfies the condition of Theorem 2.1, and C^2-C^2 images are contained in

$\mathbf{\Theta}^s_{2/3,2/3}$ with such a choice. We then arrive at the rate $O(n^{-4/3})$ (ignoring the logarithmic factor). This is clearly consistent with our rate for $\alpha = 2$ up to a logarithmic factor which is we believe to be the the price to pay for the estimator adaptivity.

For the class of C^α-C^α geometrically regular 2D images, a non-adaptive estimation procedure– it depends on a priori knowledge of α– was proposed in [35]. These authors proved that the minimax rate uniformly over this class is $O(n^{-2\alpha/(\alpha+1)})$ and no better. Both the adaptive wedgelet and bandelet estimators of [26] and [39] reviewed in subsection 1.2 achieve nearly this minimax rate over the C^α-C^α image class. Both these estimators have the same rate as ours (with the fixed curvelet transform) for C^2-C^2 images. Individual thresholding in the curvelet tight frame has also the nearly-minimax rate $O(n^{-4/3})$ [16] uniformly over 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.

2.5 On the (theoretical) choice of the threshold

To apply Theorem 2.1, 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 and is crucial towards good performance of the estimator $\widehat{\theta}^*$. From a theoretical point of view, since the constant C of the bound (2.4) increases with growing threshold, the optimal threshold value 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 2.1 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 2.1 *Consider the model* (2.1) *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)*: $\sup_{j\in\{0,...,J\}} \sup_{\ell\in B_j} \sup_{\mathbf{k}\in D_j} 2^{-2\delta j}\mathbb{E}\left(z^4_{j,\ell,\mathbf{k}}\right) \leq Q_3$.*
- (A4)*: for any $a = (a_\mathbf{k})_{\mathbf{k}\in D_j}$ such that $\sup_{j\in\{0,...,J\}} \sup_{\mathbf{K}\in\mathcal{A}_j} \sum_{\mathbf{k}\in U_{j,\mathbf{K}}} a^2_\mathbf{k} \leq 1$, 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 2.1 can be applied to* $\widehat{\theta}^*$ *defined by* (2.3) *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 whose covariance structure is given by the Gram matrix of the frame. However, the estimate provided by Proposition 2.1 is clearly not optimal for arbitrary frames. This remains a challenging open question.

Remark 2.4 (*Comments on* (A4)*)** *Assumption* (A4) *can be re-expressed using the covariance of the noise in the coefficient domain. Denote such a covariance* $c_{j,\ell,\mathbf{k},\mathbf{k}'} = \mathbb{E}\left(z_{j,\ell,\mathbf{k}} z_{j,\ell,\mathbf{k}'}\right)$, *then* (A4) *is satisfied if and only if, there exists a constant* $Q_4 > 0$ *such that*

$$\sup_{j\in\{0,\dots,J\},\ell\in B_j,\mathbf{K}\in\mathcal{A}_j} 2^{-\delta j} \sum_{(\mathbf{k},\mathbf{k}')\in U_{j,\mathbf{K}}\times U_{j,\mathbf{K}}} a_{\mathbf{k}} a_{\mathbf{k}'} c_{j,\ell,\mathbf{k},\mathbf{k}'} \leq Q_4.$$

In particular, one can prove that it is satisfied if $|c_{j,\ell,\mathbf{k},\mathbf{k}'}| \leq C 2^{\delta j} b_{\|\mathbf{k}-\mathbf{k}'\|}$, $(b_{\mathbf{u}})_{\mathbf{u}\in\mathbb{N}}$ *is a positive summable sequence. For example, with curvelets and white noise in the original domain, one can show that* $|c_{j,\ell,\mathbf{k},\mathbf{k}'}| \leq C_N(1 + \|\mathbf{k} - \mathbf{k}'\|^2)^{-N/2}$, $N \geq 3$ *is a regularity parameter, so that* $\sum_{\mathbf{u}\in\mathbb{N}^2} b_{\|\mathbf{u}\|} = \sum_{\mathbf{u}\in\mathbb{N}^2}(1 + \|\mathbf{u}\|^2)^{-N/2} \leq (1/2)\left(\sum_{u\in\mathbb{N}}(1+u^2)^{-N/4}\right)^2 < \infty$.

If additional information is considered on $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$, the threshold constant λ_* defined in Proposition 2.1 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 orthonormal (e.g. orthogonal wavelet transform). The statement is made formal in the following proposition.

Proposition 2.2 *Consider the model* (2.1) *for* n *large enough. Suppose that, for any* $j \in \{0,\dots,J\}$ *and any* $\ell \in B_j$, $(z_{j,\ell,\mathbf{k}})_{\mathbf{k}}$ *are i.i.d.* $\mathcal{N}(0,1)$. *Theorem 2.1 can be applied with the estimator* $\widehat{\theta}^*$ *defined by* (2.3) *with* $\delta = 0$ *and* λ_* *the root of* $x - \log x = 3$, *i.e.* $\lambda_* = 4.50524\dots$.

The optimal threshold constant λ_* described in Proposition 2.2 corresponds to the one isolated in [10].

3 Application to image block denoising

3.1 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 [18]. 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. This value is in a good agreement with the choice dictated by our theoretical procedure, although derived in an asymptotic setting. 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 2.2. Even with the other redundant transforms, that correspond to tight frames for which Proposition 2.2 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 2.1. Even though the estimate of Proposition 2.1 was expected to be rather crude. To summarize, the value 4.50524... intended to work for orthonormal bases seems to yield good results also with redundant transforms.

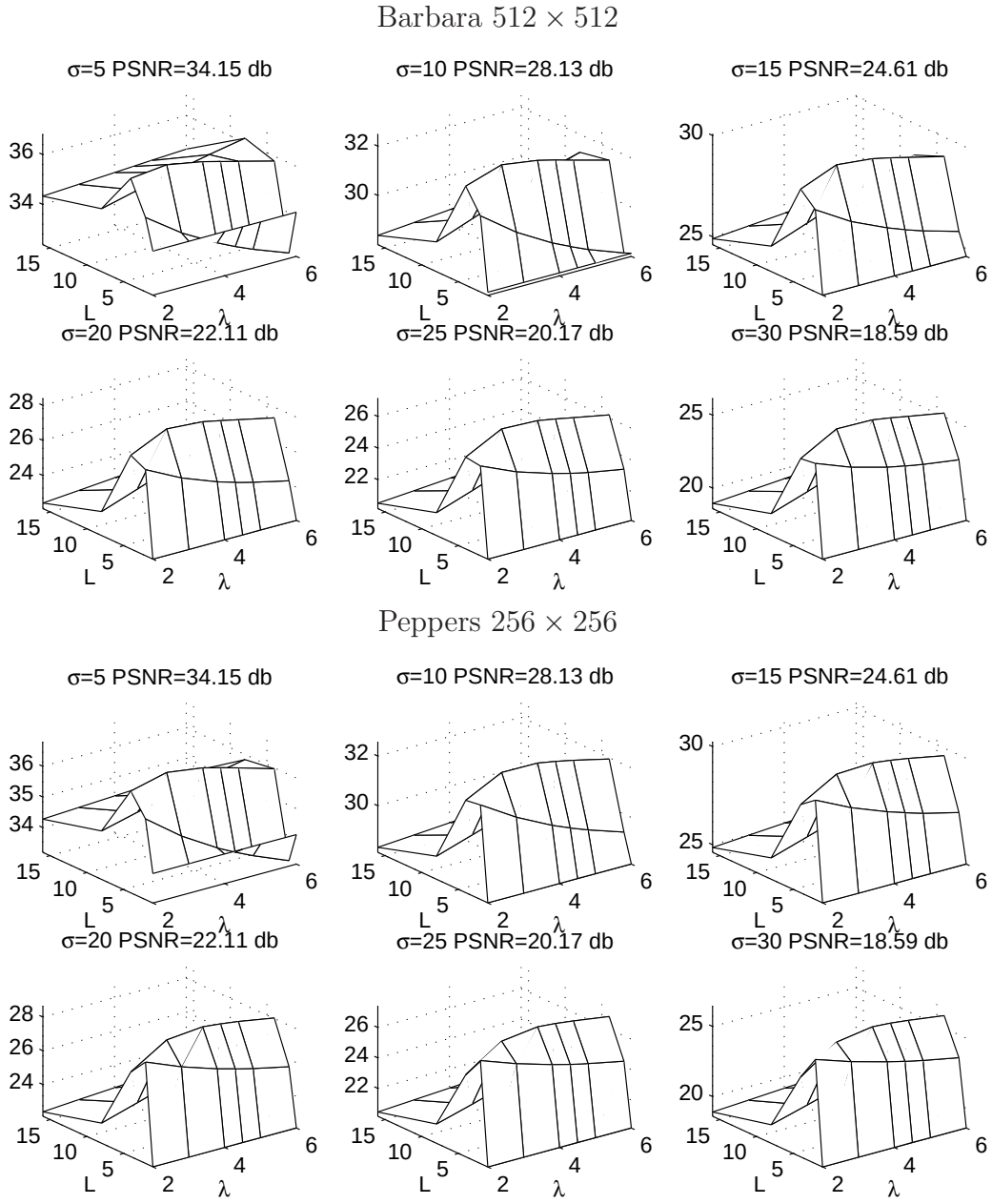

Fig. 1. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the DWT domain.

3.2 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 to a maximum grayscale value 255. The images were corrupted by a zero-

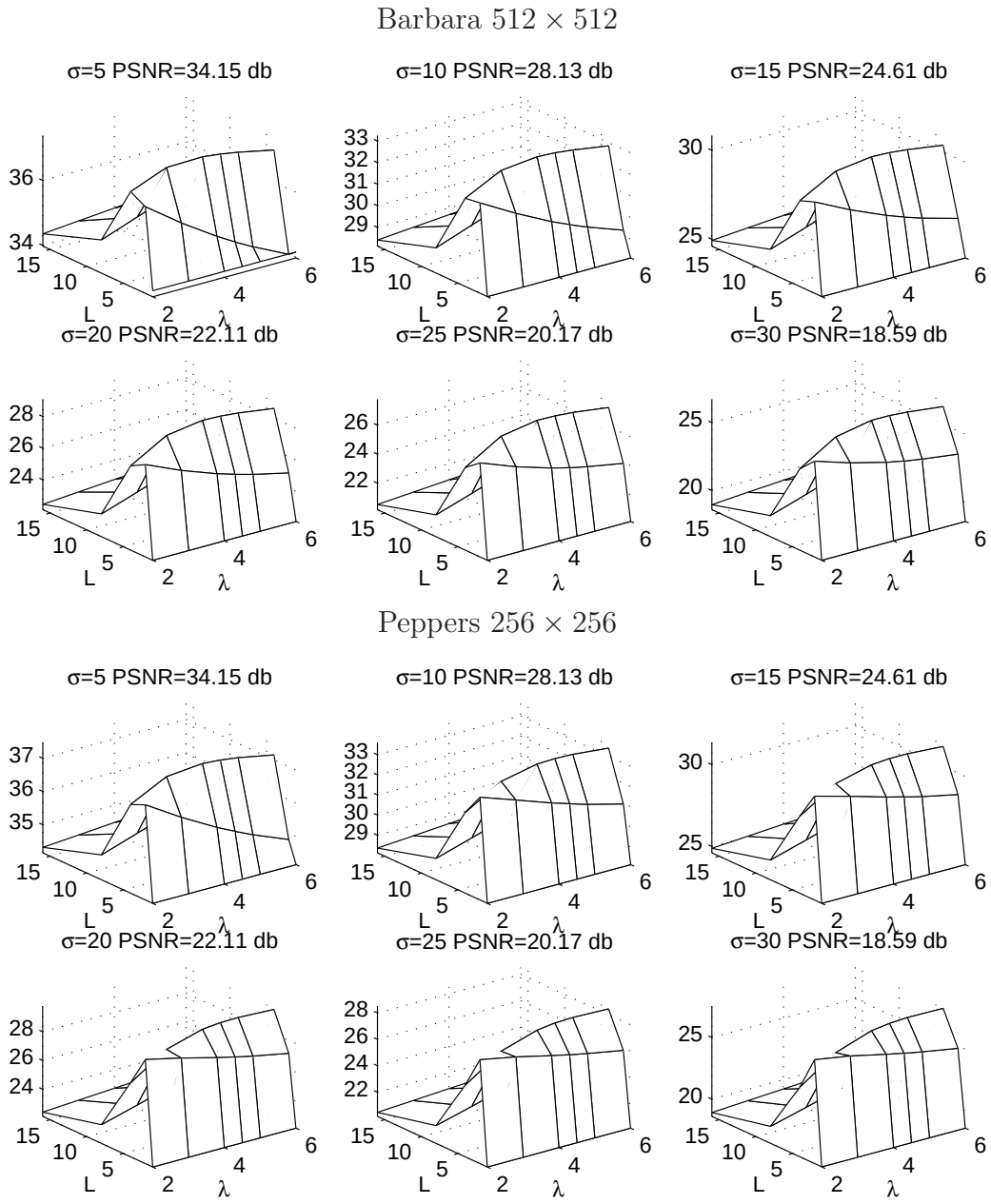

Fig. 2. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the UDWT domain.

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,

- Block shrinkage improves the denoising results in general compared to in-

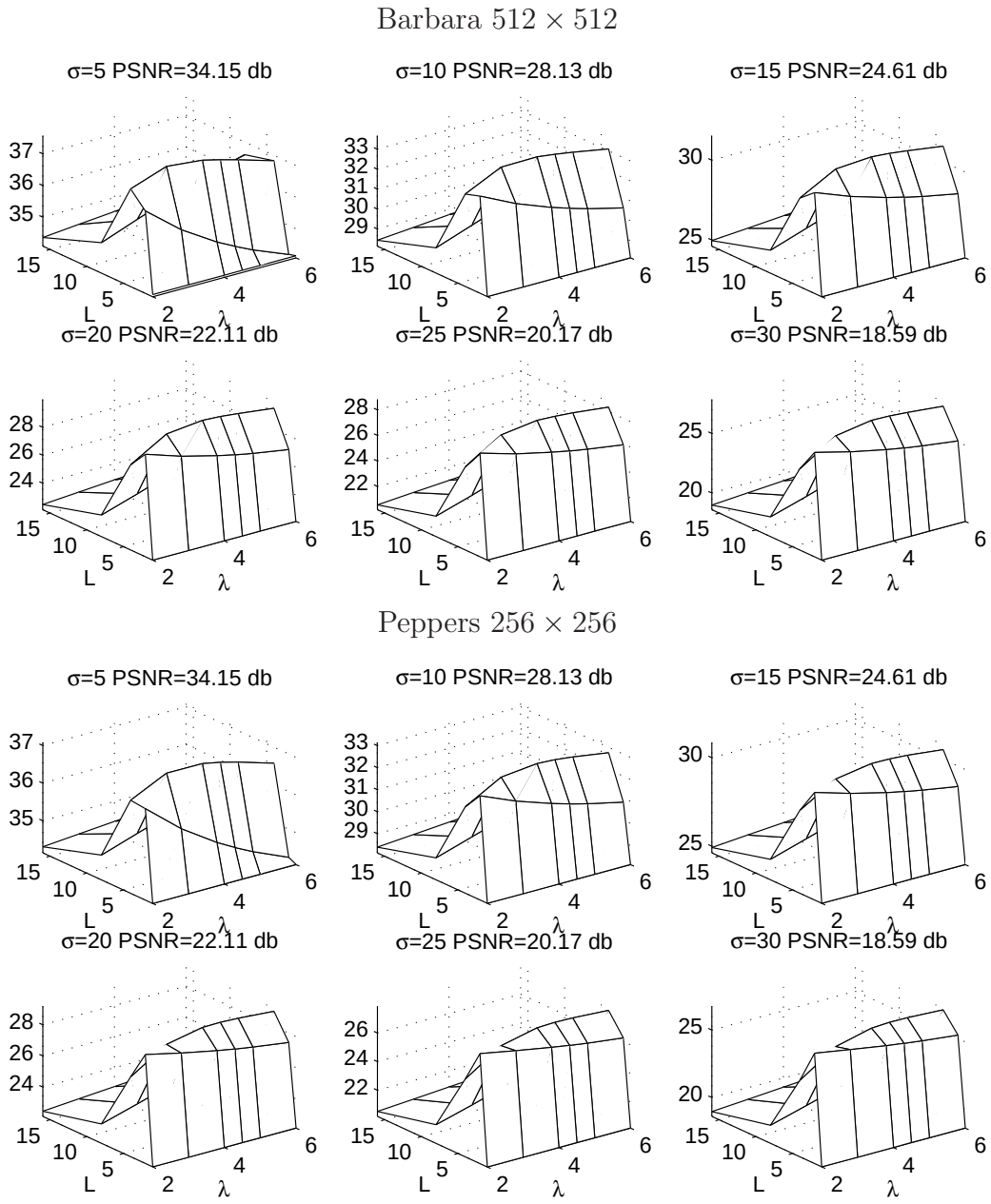

Fig. 3. Output PSNR as a function of the block size and the threshold λ at different noise levels $\sigma \in \{5, 10, 15, 20, 25, 30\}$. Block denoising was applied in the FDCT domain.

dividual 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. The reason is that block thresholding better preserves the textured part.
- 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 marginal: 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 [41]. 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.

3.3 Reproducible research

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

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

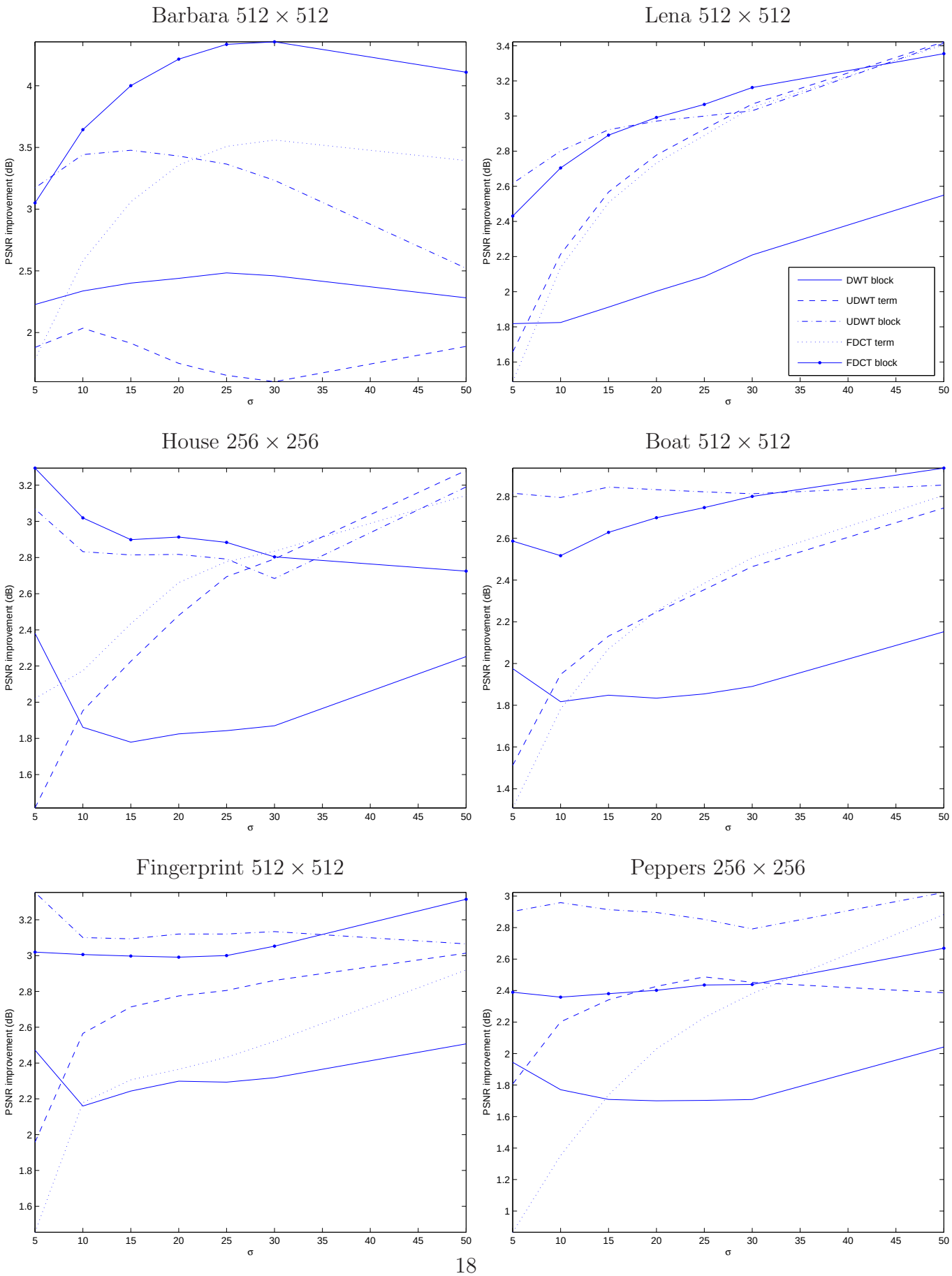

Fig. 4. Block vs term-by-term thresholding. Each plot corresponds to PSNR improvement over DWT term-by-term thresholding as a function of σ.

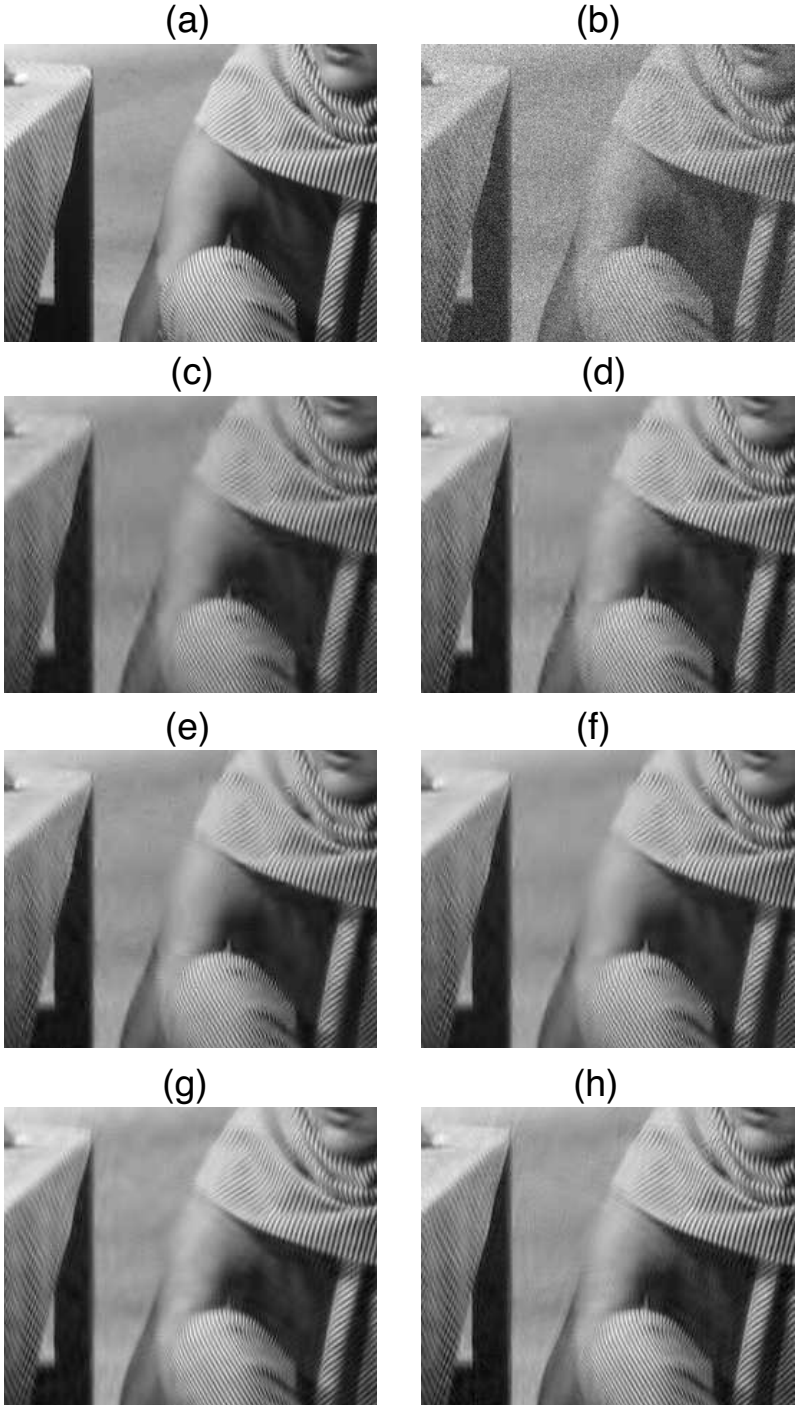

Fig. 5. Visual comparison of our block denoising to BLS-GSM on Barbara 512×512. (a) original. (b) noisy $\sigma = 20$. (c), (e) and (g) block denoising with respectively DWT (28.04 dB), UDWT (29.01 dB) and FDCT (30 dB). (d), (f) and (h) BLS-GSM with respectively DWT (28.6 dB), UDWT (29.3 dB) and FDCT (30.07 dB).

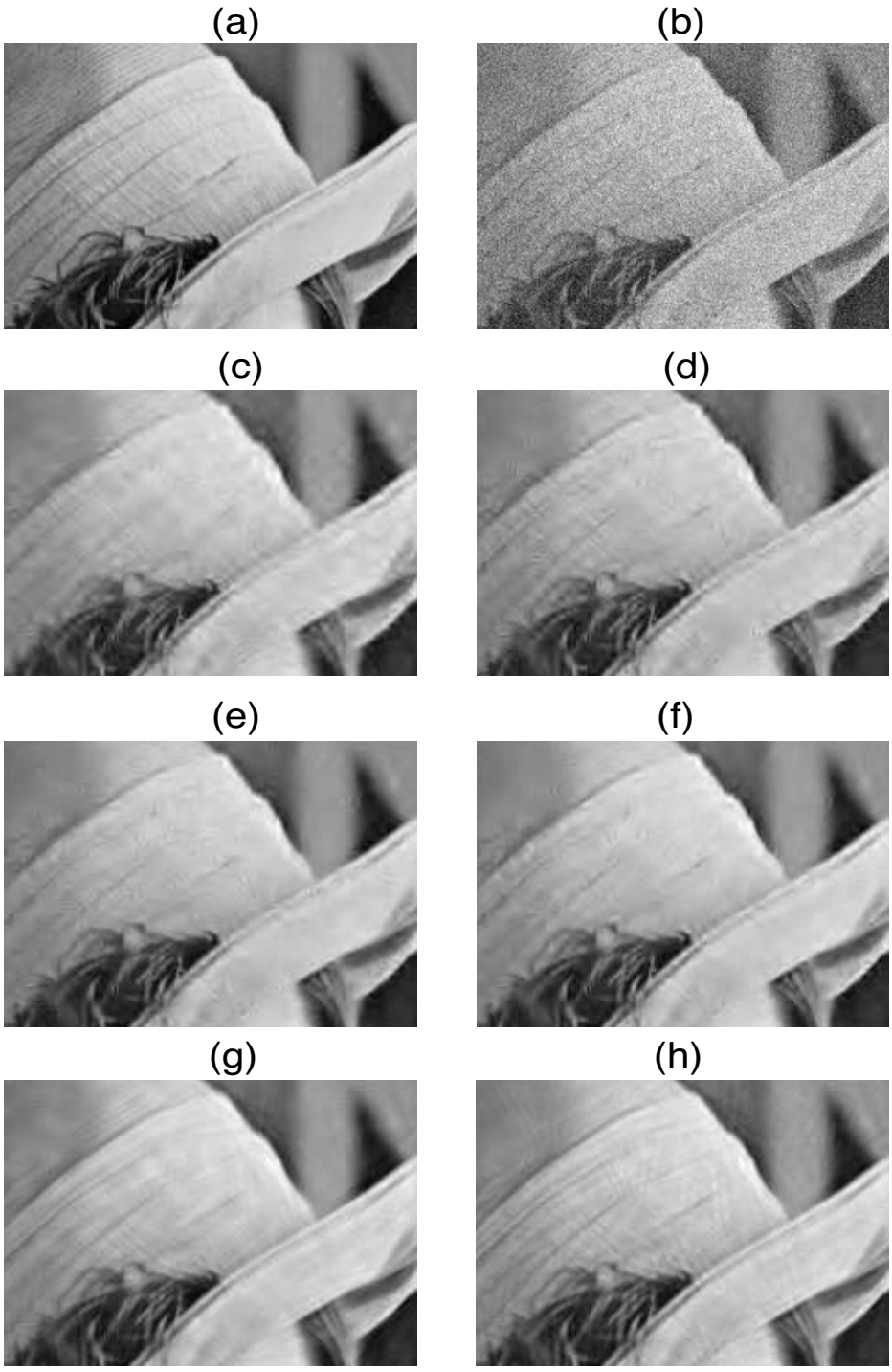

Fig. 6. Visual comparison of our block denoising to BLS-GSM on Lena 512×512. (a) original. (b) noisy $\sigma = 20$. (c), (e) and (g) block denoising with respectively DWT (30.51 dB), UDWT (31.47 dB) and FDCT (31.48 dB). (d), (f) and (h) BLS-GSM with respectively DWT (30.62 dB), UDWT (32 dB) and FDCT (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								Lena 512 × 512						
σ	5	10	15	20	25	30	50	5	10	15	20	25	30	50
$\mathrm{PSNR}_{\mathrm{in}}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15	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	37.61	34.05	31.99	30.62	29.58	28.71	26.36
BLS-GSM DWT	36.87	32.65	30.26	28.61	27.40	26.40	23.90	37.41	33.97	31.68	30.62	29.62	28.70	26.36
Block UDWT	37.37	33.24	30.80	29.09	27.77	26.70	24.01	38.02	34.75	32.85	31.48	30.41	29.53	27.16
BLS-GSM UDWT	37.44	33.43	31.06	29.40	28.16	27.13	24.49	38.16	35.15	33.34	32.02	30.97	30.13	27.78
Block FDCT	37.57	33.68	31.52	30.00	28.83	27.86	25.38	38.09	34.78	32.86	31.45	30.43	29.55	27.12
BLS-GSM FDCT	37.63	33.82	31.64	30.08	28.93	28.01	25.36	38.10	34.93	33.03	31.60	30.53	29.65	27.02
House 256 × 256								Boat 512 × 512						
σ	5	10	15	20	25	30	50	5	10	15	20	25	30	50
$\mathrm{PSNR}_{\mathrm{in}}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15	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	36.41	32.52	30.41	28.93	27.81	26.97	24.83
BLS-GSM DWT	37.43	33.97	31.77	29.88	29.17	28.43	26.12	36.06	32.36	30.36	29.04	27.35	26.76	24.86
Block UDWT	38.10	34.31	32.31	30.86	29.75	28.80	26.35	36.89	33.15	31.11	29.67	28.59	27.71	25.45
BLS-GSM UDWT	38.17	34.79	32.95	31.52	30.41	29.49	27.00	36.85	33.46	31.52	30.14	29.09	28.22	26.00
Block FDCT	38.35	34.36	32.04	30.32	29.70	28.71	25.90	36.89	33.07	31.03	29.65	28.59	27.70	25.49
BLS-GSM FDCT	38.47	34.69	32.47	30.92	29.71	28.72	25.93	36.74	33.17	31.20	29.80	28.77	27.88	25.52
Fingerprint 512 × 512								Peppers 256 × 256						
σ	5	10	15	20	25	30	50	5	10	15	20	25	30	50
$\mathrm{PSNR}_{\mathrm{in}}$	34.15	28.13	24.61	22.11	20.17	18.59	14.15	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	36.81	32.56	30.28	28.64	27.42	26.42	23.77
BLS-GSM DWT	35.53	31.08	28.82	27.08	26.01	25.11	22.72	36.69	32.50	30.38	28.90	27.65	26.70	23.55
Block UDWT	36.22	31.89	29.62	28.06	26.87	25.90	23.37	37.48	33.60	31.37	29.74	28.52	27.52	24.71
BLS-GSM UDWT	36.54	32.23	29.91	28.36	27.20	26.30	23.85	37.59	33.96	31.78	30.17	28.99	27.97	25.16
Block FDCT	36.13	31.98	29.66	28.03	26.84	25.92	23.51	37.09	33.14	30.86	29.17	28.01	27.09	24.38
BLS-GSM FDCT	36.34	32.14	29.82	28.21	27.05	26.14	23.70	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.

This toolbox is a collection of Matlab functions, scripts and datasets for image block denoising. It requires at least WaveLab 8.02 [7] 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.

4 Conclusion

In this paper, a 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 than 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. A possible alternative to our minimax study is the maxiset point of view introduced by [24] and recently studied for μ-thresholding rules, including block thresholding in the one-dimensional case, by [4]. This approach might allow to better explain the reason behind the fact that in the curvelet case, BlockJS is better in practice than hard thresholding for image denoising. As remarked in subsection 2.1.1, 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.

I Proof of Theorem 2.1

We have the decomposition:

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

where

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

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

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

$$\begin{aligned}
R_1 &= n^{-r} \sum_{j=0}^{j_0-1} \sum_{\ell\in B_j} \sum_{\mathbf{k}\in D_j} \mathbb{E}\left(z^2_{j,\ell,\mathbf{k}}\right) \le Q_1 n^{-r} \sum_{j=0}^{j_0-1} 2^{j(d_*+\delta)} \mathrm{Card}(B_j) \\
&= c_* Q_1 n^{-r} \sum_{j=0}^{j_0-1} 2^{j(d_*+\delta+\upsilon)} \le C 2^{j_0(d_*+\delta+\upsilon)} n^{-r} \\
&\le C L^{(1/\min_{i=1,\dots,d}\mu_i))(d_*+\delta+\upsilon)} n^{-r} \le C(\log n)^{(1/(d\min_{i=1,\dots,d}\mu_i))(d_*+\delta+\upsilon)} n^{-r} \\
&\le C\rho_n. \qquad \text{(I.2)}
\end{aligned}$$

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

Let $q \ge 2$ and $p \ge 2$. Since $p \ge 2$, we have $\theta \in \mathbf{\Theta}^s_{p,q}(M) \subseteq \mathbf{\Theta}^s_{2,q}(M)$. The Hölder inequality applied with the exponent $q/2 \ge 1$, and the fact that $s > \upsilon(1/2 - 1/q)$ imply

$$\begin{aligned}
R_3 &\le \sum_{j=J_*+1}^{\infty} \left(\sum_{\ell\in B_j} \left(\sum_{\mathbf{k}\in D_j} \theta^2_{j,\ell,\mathbf{k}} \right)^{q/2} \right)^{2/q} \mathrm{Card}(B_j)^{(1-2/q)} \\
&\le C \sum_{j=J_*+1}^{\infty} 2^{-2js} 2^{j\upsilon(1-2/q)} \le C 2^{-2J_*(s+\upsilon(1/q-1/2))} \\
&\le C n^{-2r(s+\upsilon(1/q-1/2))/(d_*+\delta+\upsilon)} \le C n^{-2r(s+\upsilon(1/q-1/2))/(2s+d_*+\delta+2\upsilon/q)} \\
&= C\rho_n. \qquad \text{(I.3)}
\end{aligned}$$

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

$$\begin{aligned}
R_3 &\le M^2 \sum_{j=J_*+1}^{\infty} 2^{-2js} \le C 2^{-2J_* s} \le C n^{-2sr/(d_*+\delta+\upsilon)} \le C n^{-2sr/(2s+\delta+d_*+\upsilon)} \\
&= C\rho_n. \qquad \text{(I.4)}
\end{aligned}$$

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

$$\begin{aligned}
R_3 &\le M^2 \sum_{j=J_*+1}^{\infty} 2^{-2j(s-d_*/p+d_*/2)} \le C 2^{-2J_*(s-d_*/p+d_*/2)} \\
&\le C n^{-2r(s-d_*/p+d_*/2)/(d_*+\delta+\upsilon)} \le C n^{-2sr/(2s+\delta+d_*+\upsilon)} \le C\rho_n. \qquad \text{(I.5)}
\end{aligned}$$

Putting (I.3), (I.4) and (I.5) together, we obtain the desired upper bound.

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

Lemma I.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 I.1 yields

$$R_2 = \sum_{j=j_0}^{J_*} \sum_{\ell\in B_j} \sum_{\mathbf{K}\in\mathcal{A}_j} \sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \mathbb{E}\left((\widehat{\theta}^*_{j,\ell,\mathbf{k}} - \theta_{j,\ell,\mathbf{k}})^2\right) \le 10(B_1+B_2), \qquad \text{(I.6)}$$

where

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

and

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

Using (A2), we bound B_1 by

$$B_1 \leq Q_2 n^{-r} \leq C\rho_n. \tag{I.7}$$

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

Let $p \geq 2$ and $q \geq 2$. Since $p \geq 2$, we have $\theta \in \mathbf{\Theta}^s_{p,q}(M) \subseteq \mathbf{\Theta}^s_{2,q}(M)$. Let j_s be the integer $j_s = \lfloor (r/(2s+\delta+d_*+2\upsilon/q)) \log_2 n \rfloor$. It follows from the Hölder inequality applied with the exponent $q/2 \geq 1$ that

$$\begin{aligned}
B_2 &\leq 4^{-1}\lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j\delta} \mathrm{Card}(\mathcal{A}_j)\mathrm{Card}(B_j) + \sum_{j=j_s+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta^2_{j,\ell,\mathbf{k}} \\
&\leq 4^{-1} c_* \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j(d_*+\delta+\upsilon)} L^{-d} \\
&\quad + \sum_{j=j_s+1}^{J_*} \left(\sum_{\ell \in B_j} \left(\sum_{\mathbf{k} \in D_j} \theta^2_{j,\ell,\mathbf{k}} \right)^{q/2} \right)^{2/q} \mathrm{Card}(B_j)^{(1-2/q)} \\
&\leq Cn^{-r} 2^{j_s(d_*+\delta+\upsilon)} + C \sum_{j=j_s+1}^{J_*} 2^{-2js} 2^{\upsilon(1-2/q)} \\
&\leq Cn^{-r} 2^{j_s(d_*+\delta+\upsilon)} + C 2^{-2j_s(s+\upsilon(1/q-1/2))} \\
&\leq Cn^{-2r(s+\upsilon(1/q-1/2))/(2s+d_*+\delta+2\upsilon/q)} = C\rho_n.
\end{aligned} \tag{I.8}$$

Putting (I.6), (I.7) and (I.8) together, it follows immediately that

$$R_2 \leq C\rho_n. \tag{I.9}$$

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

$$\begin{aligned}
B_2 &\leq 4^{-1}\lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j\delta} \mathrm{Card}(\mathcal{A}_j)\mathrm{Card}(B_j) + \sum_{j=j_s+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta^2_{j,\ell,\mathbf{k}} \\
&\leq 4^{-1} c_* \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s} 2^{j(d_*+\delta+\upsilon)} L^{-d} + \sum_{j=j_s+1}^{J_*} \sum_{\ell \in B_j} \sum_{\mathbf{k} \in D_j} \theta^2_{j,\ell,\mathbf{k}} \\
&\leq Cn^{-r} 2^{j_s(d_*+\delta+\upsilon)} + M^2 \sum_{j=j_s+1}^{J_*} 2^{-2js} \\
&\leq Cn^{-r} 2^{j_s(d_*+\delta+\upsilon)} + C 2^{-2j_s s} \leq Cn^{-2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n.
\end{aligned} \tag{I.10}$$

Putting (I.6), (I.7) and (I.10) together, it follows immediately that

$$R_2 \leq Cn^{-2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n. \tag{I.11}$$

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

$$B_2 \leq D_1 + D_2 + D_3, \tag{I.12}$$

where

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

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

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

$$\begin{aligned} D_1 &\leq 4^{-1} c_* \lambda_* L^d n^{-r} \sum_{j=j_0}^{j_s^*} 2^{j(d_*+\delta+\upsilon)} L^{-d} \leq Cn^{-r} 2^{j_s^*(d_*+\delta+\upsilon)} \\ &\leq C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n. \end{aligned} \tag{I.13}$$

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

$$\begin{aligned} D_2 &\leq CL^d n^{-r} (L^d n^{-r})^{-p/2} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} \sum_{\ell\in B_j} \sum_{\mathbf{K}\in\mathcal{A}_j} \left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \theta_{j,\ell,\mathbf{k}}^2 \right)^{p/2} \\ &\leq C(\log n/n)^{r(1-p/2)} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} \sum_{\ell\in B_j} \sum_{\mathbf{k}\in D_j} |\theta_{j,\ell,\mathbf{k}}|^p. \end{aligned} \tag{I.14}$$

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

$$\begin{aligned} D_2 &\leq C(\log n/n)^{r(1-p/2)} \sum_{j=j_s^*+1}^{J_*} 2^{\delta j(1-p/2)} 2^{-j(s+d_*/2-d_*/p)p} \\ &\leq C(\log n/n)^{r(1-p/2)} 2^{-j_s^*(s+d_*/2-d_*/p-\delta/p+\delta/2)p} \\ &\leq C(\log n/n)^{(2s+\upsilon(1-p/2))r/(2s+\delta+d_*+\upsilon)} \\ &\leq C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n. \end{aligned} \tag{I.15}$$

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

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

which is the same bound as for D_2 in (I.14). Then using similar arguments as those used for in (I.15), we arrive at

$$D_3 \leq C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n. \tag{I.16}$$

Inserting (I.13), (I.15) and (I.16) into (I.12), it follows that

$$R_2 \leq C(\log n/n)^{2sr/(2s+\delta+d_*+\upsilon)} = C\rho_n. \tag{I.17}$$

Finally, bringing (I.1), (I.2), (I.3), (I.4), (I.5), (I.9), (I.11) and (I.17) together we obtain

$$\sup_{\theta\in\mathbf{\Theta}^s_{p,q}(M)} R(\widehat{\theta}^*, \theta) \leq R_1 + R_2 + R_3 \leq C\rho_n,$$

where ρ_n is defined by (2.4). This ends the proof of Theorem 2.1.

II Proof of Lemma I.1

We have

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

where

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

Let's bound A and B, in turn.

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

$$
\begin{aligned}
A &\le 2\sum_{i=1}^m \left(w_i^2 + \lambda^4 u_i^2 \left(\sum_{i=1}^m u_i^2 \right)^{-2} \right) \mathbf{1}_{\left\{ \left(\sum_{i=1}^m u_i^2\right)^{1/2} > \lambda \right\}} \\
&= 2\left(\sum_{i=1}^m w_i^2 + \lambda^4 \left(\sum_{i=1}^m u_i^2 \right)^{-1} \right) \mathbf{1}_{\left\{ \left(\sum_{i=1}^m u_i^2\right)^{1/2} > \lambda \right\}} \\
&\le 2\left(\sum_{i=1}^m w_i^2 + \lambda^2 \right) \mathbf{1}_{\left\{ \left(\sum_{i=1}^m u_i^2\right)^{1/2} > \lambda \right\}}. \qquad \text{(II.2)}
\end{aligned}
$$

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

$$
D = D_1 + D_2, \qquad \text{(II.3)}
$$

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

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

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

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

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

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

The upper bound for B. We have the decomposition

$$
B = G_1 + G_2 \qquad \text{(II.7)}
$$

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

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

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

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

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

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

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

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

$$\begin{aligned} \sum_{i=1}^m (\tilde{u}_i - v_i)^2 &= \max(A, B) \\ &\leq 10 \sum_{i=1}^m w_i^2 \mathbf{1}_{\left\{\left(\sum_{i=1}^m w_i^2\right)^{1/2} > \lambda/2\right\}} + 10 \min\left(\sum_{i=1}^m v_i^2, \lambda^2/4\right). \end{aligned}$$

Lemma I.1 is proved.

III Proof of Proposition 2.1

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 (A2). 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

$$\sum_{j=j_0}^{J_*}\sum_{\ell\in B_j}\sum_{\mathbf{K}\in\mathcal{A}_j}\sum_{\mathbf{k}\in U_{j,\mathbf{K}}}\left(\mathbb{E}\left(z_{j,\ell,\mathbf{k}}^4\right)\right)^{1/2}\left(\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}}z_{j,\ell,\mathbf{k}}^2\right)^{1/2}>(\lambda_*2^{\delta j}L^d)^{1/2}/2\right)\right)^{1/2}$$
$$\leq C2^{J_*(d_*+\delta+\upsilon)}Q_3^{1/2}\sup_{j\in\{j_0,\dots,J_*\}}\sup_{\ell\in B_j}\sup_{\mathbf{K}\in\mathcal{A}_j}\left(\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}}z_{j,\ell,\mathbf{k}}^2\right)^{1/2}>(\lambda_*2^{\delta j}L^d)^{1/2}/2\right)\right)^{1/2}$$
$$\leq Cn^rQ_3^{1/2}\sup_{j\in\{j_0,\dots,J_*\}}\sup_{\ell\in B_j}\sup_{\mathbf{K}\in\mathcal{A}_j}\left(\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}}z_{j,\ell,\mathbf{k}}^2\right)^{1/2}>(\lambda_*2^{\delta j}L^d)^{1/2}/2\right)\right)^{1/2}. \tag{III.1}$$

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

Lemma III.1 (The Cirelson-Ibragimov-Sudakov inequality) *Let $(\eta_t)_{t\in\mathcal{D}}$ 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 III.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

$$\begin{aligned}
\mathbb{E}\left(\sup_{a\in\mathcal{S}_2} \mathcal{Z}(a)\right) &= \mathbb{E}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2}\right) \leq \left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \mathbb{E}(z_{j,\ell,\mathbf{k}}^2)\right)^{1/2} \\
&\leq \left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} \left(\mathbb{E}(z_{j,\ell,\mathbf{k}}^4)\right)^{1/2}\right)^{1/2} \leq Q_3^{1/4} 2^{\delta j/2} L^{d/2} \\
&\leq Q_3^{1/4} 2^{\delta j/2} (\log n)^{1/2}.
\end{aligned}$$

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 III.1, for any $j \in \{j_0, ..., J_*\}$, $\mathbf{K} \in \mathcal{A}_j$ and $\mathbf{k} \in U_{j,\mathbf{K}}$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} \geq (\lambda_* 2^{\delta j} L^d)^{1/2}/2\right) \\
&= \mathbb{P}\left(\left(\sum_{\mathbf{k}\in U_{j,\mathbf{K}}} z_{j,\ell,\mathbf{k}}^2\right)^{1/2} \geq (\lambda_*^{1/2}/2 - Q_3^{1/4})(2^{\delta j} L^d)^{1/2} + Q_3^{1/4}(2^{\delta j} L^d)^{1/2}\right) \\
&= \mathbb{P}\left(\sup_{a\in\mathcal{S}_2} \mathcal{Z}(a) \geq (\lambda_*^{1/2}/2 - Q_3^{1/4})(2^{\delta j} L^d)^{1/2} + N\right) \\
&\leq \exp\left(-(\lambda_*^{1/2}/2 - Q_3^{1/4})^2 2^{\delta j} L^d/(2Z)\right) \leq n^{-r(\lambda_*^{1/2}/2 - Q_3^{1/4})^2/(2Q_4)}.
\end{aligned}
$$

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

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

Putting (III.1) and (III.2) together, we have proved (A2). This ends the proof of Proposition 2.1.

IV Proof of Proposition 2.2

The proof of this proposition is similar to the one of Theorem 2.1. The only difference is that, instead of using Lemma I.1, we use Lemma IV.1 below.

Lemma IV.1 (Cai and Silverman [11]) *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

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

To clarify, if the variables $(z_{j,\ell,\mathbf{k}})_{j,\ell,\mathbf{k}}$ are i.i.d. $\mathcal{N}(0,1)$ then Lemma IV.1 improves the bound of the term B_1 appearing in the proof of Theorem 2.1.

If we analyze the proof of Theorem 2.1 and we use Lemma I.1 instead of Lemma IV.1, we see that it is enough to determine λ_* such that there exists

a constant $Q_5 > 0$ satisfying

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

(It corresponds to the bound of the term B_1 that appears in (I.6)). If λ_* is the root of $x - \log x = 3$, it comes that

$$\begin{aligned}\sum_{j=j_0}^{J_*} \mathrm{Card}(B_j)\mathrm{Card}(\mathcal{A}_j)e^{-(L^d/2)(\lambda_*-\log\lambda_*-1)} &= c_* e^{-(L^d/2)(\lambda_*-\log\lambda_*-1)} 2^{J_*(d_*+\upsilon)} \\ &\leq C e^{-(L^d/2)(\lambda_*-\log\lambda_*-1)} n^r \leq Q_5.\end{aligned}$$

Proposition 2.2 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] F. Autin. Maxisets for μ-thresholding rules. *TEST*, 17(2):332–349, 2008.

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

[6] L. D. Brown and M. Low. Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, 24:2384–2398, 1996.

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

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

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

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

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

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

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

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

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

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

[24] A. Cohen, R. Devore, G. Kerkyacharian, and D. Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(1):167–191, 2000.

[25] B. Delyon and A. Juditsky. Wavelet estimators, global error measures: revisited. *Preprint IRISA*, 1993.

[26] D. L. Donoho. Wedgelets: Nearly-minimax estimation of edges. *Annals of Statistics*, 27:353–382, 1999.

[27] 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.

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

[29] 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.

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

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

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

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

[34] 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.

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

[36] 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.

[37] 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.

[38] E. Le Pennec and S. Mallat. Bandelet image approximation and compression. *SIAM Multiscale Modeling and Simulation*, 4(3):992–1039, 2005.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

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

[44] Y. Wang. Function estimation via wavelet shrinkage for long-memory data. *Annals of Statistics*, 24(2):466–484, 1996.

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