Thresholding Estimators for Linear Inverse Problems and Deconvolutions

Jérôme Kalifa,[*] Stéphane Mallat[*†]

Abbreviated title: Thresholding for Inverse Problems

Abstract

Thresholding algorithms in an orthonormal basis are studied to estimate noisy discrete signals degraded by a linear operator whose inverse is not bounded. For signals in a set Θ, sufficient conditions are established on the basis to obtain a maximum risk with minimax rates of convergence. Deconvolutions with kernels having a Fourier transform which vanishes at high frequencies are examples of unstable inverse problems, where a thresholding in a wavelet basis is a suboptimal estimator. A new "mirror wavelet" basis is constructed to obtain a deconvolution risk which is proved to be asymptotically equivalent to the minimax risk over bounded variation signals. This thresholding estimator is used to restore blurred satellite images.

[*]Centre de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau Cedex, France

[†]Courant Institute of Mathematical Science, New York University, New York, NY 10012, USA

[0]AMS Classifications: Primary: 49K35, 34A55, Secondary: 42C40.

[0]Keywords: Ill-posed inverse problems, minimax optimality, wavelet packets, hyperbolic deconvolution.

1. Introduction We consider inverse problems where a signal is degraded by a linear operator and by the addition of a Gaussian white noise. In many statistical settings [22], a function $f(t)$ for $t \in [0,1]$ must be estimated from noisy samples obtained at the output of a linear operator G

$$Y[n] = Gf(n/N) + \nu[n] \quad \text{for } 0 \leq n < N \ , \tag{1.1}$$

where $\nu[n]$ are independent Gaussian random variables of variance σ^2. When G^{-1} does not exist as a bounded linear operator then the problem is said to be ill-posed [22]. Depending upon the properties of G, minimax estimators have been studied for functions f that belong to balls $\overline{\Theta}$ over different types of functional spaces such as Sobolev or Besov spaces [26], which give asymptotic results for N large. In particular, Donoho [8] showed that a nearly minimax estimator is obtained with a thresholding strategy in wavelet-vaguelette bases, if the wavelets are nearly eigenvectors of $G\,G^*$ and if $\overline{\Theta}$ is an orthosymmetric set in this wavelet basis. The discrete data $Y[n]$ is then decomposed and thresholded in discrete wavelet-vaguelette bases. This approach has found applications in several inverse problems including inversion of the Radon transform, density deconvolution, and estimation with fractional Gaussian noise [8, 18, 23, 27].

When inverting the degradation of measuring devices such as cameras, where the diffraction of the optics creates a blur, the situation may be quite different. The sampling interval N^{-1} is adapted to the degradation operator G, and as the technology improves the operator G is modified and N increases. The operator G thus depends upon N. This degradation operator is often calibrated a posteriori, once the system is built and defined directly on the sampled signal. In this framework, the inverse problem can be written

$$Y[n] = U\theta[n] + \nu[n] \quad \text{for } 0 \leq n < N \tag{1.2}$$

where $\theta \in \mathbb{C}^N$ is a vector obtained by discretizing $f(t)$ and U is a calibrated linear operator from $\mathbb{C}^N$ to $\mathbb{C}^N$. If $f \in \overline{\Theta}$ its discretization θ belongs to a subset Θ of $\mathbb{C}^N$. The operator U depends upon N but if the properties of U are known for N large, then an asymptotic study of the minimax estimators can still be performed. The fact that the operator G may change with N opens new asymptotic situations.

Such a problem has been proposed by the spatial agency in France (CNES) to the research community working on linear inverse problems, for the deconvolution of high resolution images obtained with a new generation of satellites. In this case, U is a calibrated convolution operator which blurs the image by suppressing the highest frequencies. A testing procedure was set up by the CNES to compare the results of existing algorithms. Thresholding estimators in wavelet-vaguelette lead to relatively disappointing numerical results, as opposed to a thresholding algorithm in a wavelet packet basis developed by Rougé [24], an engineer in image processing. Surprisingly, this thresholding algorithm clearly outperformed numerically and perceptually all other deconvolution algorithms on all tested satellite images.

The restoration of high resolution satellite images are typically in an asymptotic regime which allows us to compute upper and lower bounds of the minimax risk. To understand this type of problems, Section 3.2 studies general conditions to obtain a nearly minimax estimator with a thresholding algorithm in an arbitrarily chosen basis $\mathcal{B}$ of $\mathbb{C}^N$. In the spirit of the work of Donoho [8], it is shown that a sufficient condition is that $U\,U^*$ is "nearly" diagonal in the basis $\mathcal{B}$, in a sense that is specified. This general result applies to wavelet bases but to other bases as well.

The paper concentrates on application to deconvolutions. If G is a convolution operator then (1.1) can be written as a discrete inverse problem (1.2) where U is a discrete convolution operator. Measuring devices such as cameras introduce a degradation which is a convolution of the input signal f with a kernel whose Fourier transform vanishes at high frequencies. The sampling rate N^{-1} is adapted to satisfy the Shannon sampling theorem. The resulting discrete inverse problem is a *hyperbolic* deconvolution, where wavelet bases do not "nearly diagonalize" $U\,U^*$, and hence lead to relatively poor estimators. For one-dimensional signals, Section 4.3 introduces a new type of orthonormal basis, called *mirror wavelet basis* which nearly diagonalizes $U\,U^*$ for hyperbolic deconvolutions. Section 4.4 computes the minimax risk for bounded variation signals and proves that a thresholding in a mirror wavelet basis has an equivalent asymptotic risk. Section 5 studies a two-dimensional extension for separable deconvolutions of images. For hyperbolic deconvolutions, the bounded variation model presented without proof in [16] is refined with directional total variation conditions, to obtain asymptotic minimax results that are consistent with the observed numerical performances on satellite images.

This paper concentrates on asymptotic computations as opposed to numerical implementation with fast algorithms, which are presented in [17]. The asymptotic behavior of the risk is computed up to a multiplicative constant. If α and β depend upon the parameters of the problem, such as the signal size N, the noise variance σ or any other parameter, we write $\alpha \sim \beta$ and say that they are equivalent, if there exists two constant $A, B > 0$ such that for all values of these parameters $A\,\alpha \leq \beta \leq B\,\alpha$.

2. Minimax Estimation of Inverse Problems A signal $\theta \in \mathbb{C}^N$ must be estimated from the measured data

$$Y[n] = U\theta[n] + \nu[n] \quad \text{for } 0 \leq n < N \ , \tag{2.3}$$

and where ν is a Gaussian white noise of known variance σ^2 and U is a linear operator in $\mathbb{C}^N$. Such inverse problems are studied for large N, depending upon the asymptotic properties of the operator U when N increases.

To invert the degradation, since U may not be invertible we use a pseudo-inverse. We denote by $\mathbf{Im}$U the range of U and by $\mathbf{Ker}$U its null space. Let us also write $\mathbf{W}^\perp$ the orthogonal complement in $\mathbb{C}^N$ of a subspace $\mathbf{W}$. The pseudo inverse $\overline{U}^{-1}$ of U is the operator whose null space is $\mathbf{Im}\mathrm{U}^\perp$ and whose range is $\mathbf{Ker}\mathrm{U}^\perp$. The restriction of $\overline{U}^{-1}$ to $\mathbf{Im}$U is the inverse of the restriction of U to $\mathbf{Ker}\mathrm{U}^\perp$. The restoration is unstable if

$$\lim_{N\to+\infty} \|\overline{U}^{-1}\|_s = +\infty$$

where $\|T\|_s$ is the sup operator norm of a linear operator T defined from the Euclidean norm $\|\,.\,\|$ in $\mathbb{C}^N$ by

$$\|T\|_s = \sup_{g\in\mathbb{C}^N} \frac{\|Tg\|}{\|g\|} \ . \tag{2.4}$$

Following a standard approach, the linear inverse problem (2.3) is transformed into an additive noise estimation problem. Estimating θ from Y in (2.3) is equivalent to estimate it from

$$X = \overline{U}^{-1} Y = \overline{U}^{-1} U\theta + \overline{U}^{-1}\nu. \tag{2.5}$$

The operator $\overline{U}^{-1}U = P_{\mathbf{V}}$ is an orthogonal projection on $\mathbf{V} = \mathbf{Ker}\mathrm{U}^\perp$ so

$$X = P_{\mathbf{V}}\theta + z \quad \text{with} \quad z = \overline{U}^{-1}\nu \ . \tag{2.6}$$

The noise z is not white but remains Gaussian because $\overline{U}^{-1}$ is linear. It is considerably amplified when the problem is unstable. The covariance operator K of z is

$$K = \sigma^2\, \overline{U}^{-1}\, \overline{U}^{-1*} \ , \tag{2.7}$$

where A^* is the adjoint of an operator A.

We suppose that our prior knowledge on signals defines a set Θ such that $\theta \in \Theta$. The inverse problem (2.3) is equivalent to estimating $\theta \in \Theta$ from $X = P_{\mathbf{V}}\theta + z$. The risk of an estimator $\hat{\theta}$ is calculated with an Euclidean norm

$$r(\hat{\theta}, \theta) = \mathsf{E}\{\|\hat{\theta} - \theta\|^2\}.$$

To control the risk for any $\theta \in \Theta$, we compute the maximum risk:

$$r_\Theta(\hat{\theta}, \Theta) = \sup_{\theta \in \Theta} \mathsf{E}\{\|\hat{\theta} - \theta\|^2\}.$$

The *non-linear minimax risk* is

$$r_n(\Theta) = \inf_{\hat{\theta}} r(\hat{\theta}, \Theta).$$

In practice, we must find an estimator that is simple to implement and such that $r(\hat{\theta}, \Theta)$ is close to the minimax risk $r_n(\Theta)$.

A first approach, often used in signal processing, simplifies this problem by choosing an estimator computed with a linear operator. The *linear minimax risk* over Θ is the lower bound:

$$r_l(\Theta) = \inf_{\hat{\theta}\ linear} r(\hat{\theta}, \Theta)$$

This strategy is efficient only if $r_l(\Theta)$ is of the same order as $r_n(\Theta)$. When this is not the case, according to the strategy of Donoho and Johnstone [9] one can still try to simplify the problem by approaching the non-linear minimax risk with non-linear estimators that are diagonal in an orthonormal basis.

3. Diagonal Estimation Since a linear inverse problem can be reduced to an additive noise estimation problem, we study the minimax estimation of $\theta \in \Theta$ from

$$X = \theta + z \ , \tag{3.8}$$

where z is a Gaussian random vector of covariance K. This section first briefly reviews important results concerning the minimax optimality of diagonal estimators in an orthonormal basis $\mathcal{B} = \{g_m\}_{0 \le m < N}$ of $\mathbb{C}^N$. We write $h_{\mathcal{B}}[m] = \langle h, g_m \rangle$ for any $h \in \mathbb{C}^N$. The decomposition coefficients of $X = \theta + z$ in $\mathcal{B}$ are $\langle X, g_m \rangle = \langle \theta, g_m \rangle + \langle z, g_m \rangle$, that we write

$$X_{\mathcal{B}}[m] = \theta_{\mathcal{B}}[m] + z_{\mathcal{B}}[m]. \tag{3.9}$$

A diagonal estimator $\hat{\theta}$ in $\mathcal{B}$ estimates each coefficient $\theta_{\mathcal{B}}[m]$ independently with some decision function $\delta_m(X_{\mathcal{B}}[m])$:

$$\hat{\theta} = \sum_{m=0}^{N-1} \delta_m(X_{\mathcal{B}}[m])\, g_m \ . \tag{3.10}$$

The estimation risk is

$$r(\hat{\theta}, \theta) = \mathsf{E}\{\|\hat{\theta} - \theta\|^2\} = \sum_{m=0}^{N-1} \mathsf{E}\{|\theta_{\mathcal{B}}[m] - \delta_m(X_{\mathcal{B}}[m])|^2\}. \tag{3.11}$$

Observe that $X_{\mathcal{B}}[m] = \theta_{\mathcal{B}}[m] + z_{\mathcal{B}}[m]$ is a Gaussian random variable of mean $\theta_{\mathcal{B}}[m]$ and variance

$$\sigma_m^2 = \mathsf{E}\{|z_{\mathcal{B}}[m]|^2\} = \langle K g_m, g_m \rangle \ . \tag{3.12}$$

If $\delta_m(X_{\mathcal{B}}[m]) = a[m]\, X_{\mathcal{B}}[m]$, one can verify that the minimum risk $r(\hat{\theta}, \theta)$ is achieved by choosing:

$$a[m] = \frac{|\theta_{\mathcal{B}}[m]|^2}{|\theta_{\mathcal{B}}[m]|^2 + \sigma_m^2}, \tag{3.13}$$

and

$$r(\hat{\theta}, \theta) = r_{\mathrm{inf},d}(\theta) = \sum_{m=0}^{N-1} \frac{\sigma_m^2\, |\theta_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |\theta_{\mathcal{B}}[m]|^2} \ . \tag{3.14}$$

Over a signal set Θ, the resulting maximum risk is

$$r_{\mathrm{inf},d}(\Theta) = \sup_{\theta \in \Theta} r_{\mathrm{inf},d}(\theta) \ .$$

The *oracle attenuation* (3.13) cannot be implemented because $a[m]$ depends upon $|\theta_{\mathcal{B}}[m]|$ which is not known in practice. The risk $r_{\inf,d}(\Theta)$ is thus only a lower bound for the minimax risk of diagonal estimators. However, Donoho and Johnstone [9] proved that a simple thresholding estimator can have a maximum risk that is close to $r_{\inf,d}(\Theta)$.

3.1. *Linear Versus Non-Linear* For a linear diagonal estimator $\hat{\theta}$, each $a[m]$ is a constant. Clearly, the *linear diagonal minimax risk* is larger than the linear minimax risk

$$r_{l,d}(\Theta) = \inf_{\hat{\theta}\, linear\ diagonal} r(\hat{\theta}, \Theta) \geq r_l(\Theta) \ .$$

Diagonal estimators that achieve the minimax risk $r_{l,d}(\Theta)$ are characterized by considering the quadratic convex hull $\mathrm{QH}[\Theta]$ of Θ.

The "square" of a set Θ in the basis $\mathcal{B}$ is defined by

$$(\Theta)^2_{\mathcal{B}} = \{\overline{\theta}\ :\ \overline{\theta} = \sum_{m=0}^{N-1} |\theta_{\mathcal{B}}[m]|^2\, g_m \ \text{ with } \theta \in \Theta\} \ . \tag{3.15}$$

We say that Θ is *quadratically convex* in $\mathcal{B}$ if $(\Theta)^2_{\mathcal{B}}$ is a convex set. The *quadratic convex hull* $\mathrm{QH}[\Theta]$ of Θ in the basis $\mathcal{B}$ is such that $(\mathrm{QH}[\Theta])^2_{\mathcal{B}}$ is the convex hull of $(\Theta)^2_{\mathcal{B}}$:

$$\mathrm{QH}[\Theta] = \left\{\theta\ :\ \sum_{m=0}^{N-1} |\theta_{\mathcal{B}}[m]|^2\, g_m \text{ is in the convex hull of } (\Theta)^2_{\mathcal{B}}\right\} \ . \tag{3.16}$$

The risk of an oracle attenuation (3.14) is a lower bound of the minimax linear diagonal risk $r_{l,d}(\Theta)$:

$$r_{l,d}(\Theta) \geq r_{\inf,d}(\Theta) \ . \tag{3.17}$$

If Θ is a closed and bounded set then Donoho et al. [11] proved that

$$r_{l,d}(\Theta) = r_{l,d}(\mathrm{QH}[\Theta]) = r_{\inf,d}(\mathrm{QH}[\Theta]) \ . \tag{3.18}$$

If Θ is not quadratically convex in the basis $\mathcal{B}$ then it may be much smaller than $\mathrm{QH}[\Theta]$. In this case, a non-linear diagonal estimator may have a much smaller risk, that remains of the order of $r_{\inf,d}(\Theta)$.

Among non-linear diagonal estimators, we concentrate on thresholding estimators:

$$\hat{\theta} = \sum_{m=0}^{N-1} \rho_{T_m}(X_{\mathcal{B}}[m])\, g_m \ , \tag{3.19}$$

where $\rho_{T_m}(x)$ is a soft thresholding function

$$\rho_{T_m}(x) = \begin{cases} x - T_m & \text{if } x \geq T_m \\ x + T_m & \text{if } x \leq -T_m \\ 0 & \text{if } |x| \leq T_m \end{cases} . \tag{3.20}$$

For a Gaussian random variable X_0 of mean μ and variance σ^2, Donoho and Johnstone [9] prove that with a threshold $T = \sigma\sqrt{2\log_e M}$ the thresholding risk satisfies

$$r(T,\mu,\sigma) = \mathsf{E}\{|\mu - \rho_T(X_0)|^2\} \leq \frac{(2\log_e M + 1)\sigma^2}{M} + \min\left(T^2 + \sigma^2\ ,\ \mu^2\right) . \tag{3.21}$$

Since $X_{\mathcal{B}}[m]$ is a Gaussian random variable of mean $\theta_{\mathcal{B}}[m]$ and variance σ_m^2

$$r_t(\theta) = r(\hat{\theta},\theta) = \sum_{m=0}^{N-1} \mathsf{E}\{|\theta_{\mathcal{B}}[m] - \rho_{T_m}(X_{\mathcal{B}}[m])|^2\} = \sum_{m=0}^{N-1} r(T_m, \theta_{\mathcal{B}}[m], \sigma_m) . \tag{3.22}$$

To control this risk, Johnstone and Silverman [14] propose to use thresholds $T_m = \sigma_m\sqrt{2\log_e N}$. Taking into account the maximum coefficient amplitude for signals in Θ

$$s_{\mathcal{B}}[m] = \sup_{\theta\in\Theta} |\theta_{\mathcal{B}}[m]| \ ,$$

the thresholds are rather defined by

$$T_m = \begin{cases} \sigma_m\sqrt{2\log_e N} & \text{if } \sigma_m\sqrt{2\log_e N} \leq s_{\mathcal{B}}[m] \\ \infty & \text{otherwise} \end{cases} . \tag{3.23}$$

Setting $T_m = \infty$ guarantees that $\rho_{T_m}(X_{\mathcal{B}}[m]) = 0$ if $\sigma_m \sqrt{2\log_e N} > s_{\mathcal{B}}[m]$ in which case the risk is equal to $|\theta_{\mathcal{B}}[m]|^2$. Inserting (3.21) in (3.22) together with the threshold expressions (3.23) gives

$$r_t(\theta) \leq \sum_{\substack{m=0 \\ T_m \neq \infty}}^{N-1} \frac{\sigma_m^2\,(2\,\log_e N + 1)}{N} + \sum_{m=0}^{N-1} \min\Big(\sigma_m^2 (2\log_e N + 1), |\theta_{\mathcal{B}}[m]|^2\Big) \ . \tag{3.24}$$

Let $r_t(\Theta) = \sup_{\theta\in\Theta} r_t(\theta)$. Since $\min(\alpha a, b) \leq (\alpha+1)ab/(a+b)$ for $a \geq 0$ and $b \geq 0$, and since $\sigma_m \sqrt{2\log_e N} \leq s_{\mathcal{B}}[m]$ if $T_m \neq \infty$, using (3.14) we get

$$\begin{aligned} r_t(\Theta) &\leq \frac{(2\,\log_e N + 1)}{N\,(2\,\log_e N)} \sum_{m=0}^{N-1} \min(\sigma_m^2 2\log_e N\,,\, |s_{\mathcal{B}}[m]|^2) + (2\log_e N + 2)\, r_{\inf,d}(\Theta) \\ &\leq \frac{(2\,\log_e N + 1)^2}{N\,(2\,\log_e N)} \sum_{m=0}^{N-1} \frac{\sigma_m^2\,|s_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |s_{\mathcal{B}}[m]|^2} + (2\log_e N + 2)\, r_{\inf,d}(\Theta) \ . \end{aligned}$$

But $\sigma_m^2\,|s_{\mathcal{B}}[m]|^2/(\sigma_m^2 + |s_{\mathcal{B}}[m]|^2) \leq r_{\inf,d}(\Theta)$ so we get a simple upper bound

$$r_t(\Theta) = \sup_{\theta\in\Theta} r_t(\theta) \leq 4\,(\log_e N + 1)\, r_{\inf,d}(\Theta) \ . \tag{3.25}$$

The risk of this thresholding estimator is thus above $r_{\inf,d}(\Theta)$ by at most a factor proportional to $\log_e N$. In some cases, the $\log_e N$ factor can be reduced by choosing lower amplitude thresholds.

For each index m, we denote j_m the integer such that $2^{j_m} \leq \sigma_m < 2^{j_m+1}$, and N_m the total number of noise coefficients $z_{\mathcal{B}}[n]$ whose variances have same order of magnitude and which satisfy $2^{j_m} \leq \sigma_n < 2^{j_m+1}$. Let us define the thresholds

$$T_m = \begin{cases} \sigma_m \sqrt{2\log_e N_m} & \text{if } \sigma_m\,\sqrt{2\log_e N_m} \leq s_{\mathcal{B}}[m] \\ \infty & \text{otherwise} \end{cases} \tag{3.26}$$

and for $N_m = 1$

$$T_m = \begin{cases} \sigma_m & \text{if } \sigma_m \leq s_{\mathcal{B}}[m] \\ \infty & \text{otherwise} \end{cases} \ . \tag{3.27}$$

The risk upper bound (3.24) then becomes

$$r_t(\theta) \leq \sum_{\substack{m=0 \\ T_m \neq \infty}}^{N-1} \frac{\sigma_m^2\,(2\,\log_e N_m + 1)}{N_m} + \sum_{m=0}^{N-1} \min\Big(\sigma_m^2 (2\log_e N_m + 1), |\theta_{\mathcal{B}}[m]|^2\Big) \ . \tag{3.28}$$

Section 4 computes this thresholding risk for a deconvolution problem.

3.2. *Nearly Diagonal Covariance* It now remains to find conditions so that the risk of diagonal operators in a basis $\mathcal{B}$ can approach the minimax risk. This depends upon the geometry of the set Θ relatively to the vectors of $\mathcal{B}$ and on the covariance matrix of the noise in $\mathcal{B}$.

When z is a Gaussian white noise, Donoho [7, 11] studied the optimality of diagonal estimators in orthosymmetric sets. Any signal θ can be decomposed as $\theta = \sum_{m=0}^{N-1} \theta_{\mathcal{B}}[m] g_m$. A set Θ is *orthosymmetric* in $\mathcal{B}$ if for any $\theta \in \Theta$ and for any $a \in \mathbb{C}^N$ with $|a[m]| \leq 1$ then

$$\sum_{m=0}^{N-1} a[m]\, \theta_{\mathcal{B}}[m]\, g_m \in \Theta \ .$$

This means that the set Θ is elongated along the directions of the vectors g_m of $\mathcal{B}$. In this case, one can prove [10] that

$$r_l(\Theta) = r_{\inf,d}(\mathrm{QH}[\Theta]), \tag{3.29}$$

and we derive from (3.18) that linear diagonal estimators in $\mathcal{B}$ achieve the linear minimax risk. Moreover, we also have [13, 12]

$$\frac{1}{1.25}\, r_{\inf,d}(\Theta) \leq r_n(\Theta) \ . \tag{3.30}$$

When the noise z is not white, to obtain similar near optimality results with diagonal estimators, the noise coefficients $z_{\mathcal{B}}[m]$ should still be “nearly” independent. Since z is Gaussian, to guarantee that the coefficients $z_{\mathcal{B}}[m]$ are nearly independent is equivalent to have them nearly uncorrelated, which means that the covariance K of z is nearly diagonal in $\mathcal{B}$. This approach was studied by Donoho to prove the optimality of wavelet bases for particular inverse problems [8], and these results have been refined by Abramovich and Silverman [1] and Lee and Lucier [19].

Johnstone and Silverman [14] give a different set of conditions that apply to short range dependent noise which have a behavior equivalent to a white noise or for long range dependent stationary noises having a power spectrum that increases like a power law. The following theorem gives a general condition on $\mathcal{B}$ to nearly reach the minimax risk with diagonal estimators for any type of Gaussian noise. We denote by K_d the diagonal operator in the basis $\mathcal{B}$, whose diagonal values are equal to the diagonal values of K: $\sigma_m^2 = \langle K g_m, g_m \rangle$. The square root $K_d^{1/2}$ is the diagonal matrix whose diagonal entries are σ_m.

THEOREM 1. *Suppose that K is invertible. The diagonal preconditioning factor of K^{-1} in the basis $\mathcal{B}$ is defined by*

$$\lambda_{\mathcal{B}} = \| K_d^{1/2} \, K^{-1} \, K_d^{1/2} \|_s \ . \tag{3.31}$$

It satisfies $\lambda_{\mathcal{B}} \geq 1$. If Θ is orthosymmetric in $\mathcal{B}$ then

$$r_l(\Theta) \geq \frac{1}{\lambda_{\mathcal{B}}} \, r_{\inf,d}(\mathrm{QH}[\Theta]) \ , \tag{3.32}$$

and

$$r_n(\Theta) \geq \frac{1}{1.25 \, \lambda_{\mathcal{B}}} \, r_{\inf,d}(\Theta) \ . \tag{3.33}$$

The theorem proof is in Appendix A. The fact that K is invertible means that the noise is not concentrated in a subspace of $\mathbb{C}^N$. Otherwise, we restrict ourself to this subspace. One can verify that $\lambda_{\mathcal{B}} = 1$ if and only if $K = K_d$, which means that K is diagonal in $\mathcal{B}$. The closer $\lambda_{\mathcal{B}}$ is to 1 the more diagonal K.

An upper bound of $r_l(\Theta)$ is computed in (3.18) with linear diagonal operators, and together with (3.32) we get

$$\frac{1}{\lambda_{\mathcal{B}}} \, r_{\inf,d}(\mathrm{QH}[\Theta]) \leq r_l(\Theta) \leq r_{\inf,d}(\mathrm{QH}[\Theta]). \tag{3.34}$$

If $\lambda_{\mathcal{B}}$ is of the order of 1 then $r_l(\Theta)$ is of the order of $r_{\inf,d}(\mathrm{QH}[\Theta])$. For a thresholding estimator with (3.23) since $r_n(\Theta) \leq r_t(\Theta)$, the upper bound (3.25) of the thresholding risk and (3.33) give

$$\frac{1}{1.25\,\lambda_{\mathcal{B}}}\, r_{\inf,d}(\Theta) \leq r_n(\Theta) \leq r_t(\Theta) \leq (4\log_{\mathrm{e}} N + 4)\, r_{\inf,d}(\Theta)\ . \tag{3.35}$$

If $\lambda_{\mathcal{B}}$ is of the order of 1 then $r_n(\Theta)$ and $r_t(\Theta)$ are of the order of $r_{\inf,d}(\Theta)$, up to a factor at most proportional to $\log_{\mathrm{e}} N$.

The set Θ is often not orthosymmetric in any basis $\mathcal{B}$. Yet, if one can find two orthosymmetric sets Θ_1 and Θ_2 such that $\Theta_1 \subset \Theta \subset \Theta_2$, then applying Theorem 1 to Θ_1 and Θ_2 gives upper bounds and lower bounds of $r_l(\Theta)$ and $r_n(\Theta)$. We say that Θ is "nearly" orthosymmetric in $\mathcal{B}$ if the two sets Θ_1 and Θ_2 are close enough so that the upper and lower bounds of the risk are equivalent up to a multiplicative constant. The main difficulty is to find a basis $\mathcal{B}$ where Θ is "nearly" orthosymmetric, and which "nearly" diagonalizes the covariance K of the noise.

3.3. *Inverse Problems* Let us now come back to the estimation of $\theta \in \Theta$ given the degraded data $Y = U\theta + \nu$. This inverse problem is rewritten in (2.6) as

$$X = P_{\mathbf{V}}\theta + z \quad \text{with } z = \overline{U}^{-1}\,\nu\ . \tag{3.36}$$

The range of $\overline{U}^{-1}$ is $\mathbf{V} = \mathbf{Ker}\mathrm{U}^{\perp}$. Let $\mathcal{B} = \{g_m\}_{0\leq m<N}$ be an orthonormal basis such that a subset of its vectors define a basis of $\mathbf{V}$. We consider diagonal estimations $\hat{\theta} \in \mathbf{V}$. Over $\mathbf{V}$, the risk has the same expression as (3.11). There is no data on the projection of θ on $\mathbf{V}^{\perp} = \mathbf{Ker}\mathrm{U}$. We shall suppose that given $P_{\mathbf{V}}\theta$ the minimax estimator of $P_{\mathbf{V}^{\perp}}\theta$ is 0. A sufficient condition is that for

$$\Theta_\theta = \{\phi \in \Theta\ :\ P_{\mathbf{V}}\phi = P_{\mathbf{V}}\theta\}$$

then

$$\arg \min_{y \in \mathbf{V}^\perp} \sup_{\phi \in \Theta_\theta} \|y - P_{\mathbf{V}^\perp} \phi\| = 0 \ .$$

This condition is clearly satisfied if Θ is orthosymmetric in $\mathcal{B}$ since $\theta = P_{\mathbf{V}}\theta + P_{\mathbf{V}^\perp}\theta \in \Theta$ and $P_{\mathbf{V}}\theta - P_{\mathbf{V}^\perp}\theta \in \Theta$. As a result the best estimator over $\mathbf{V}^\perp = \mathbf{Ker}\mathrm{U}$ is 0 and the risk over $\mathbf{V}^\perp$ is equal to the signal energy. Hence

$$(3.37)\quad r(\hat\theta, \theta) = \mathsf{E}\{\|\hat\theta - \theta\|^2\} = \sum_{g_m \in \mathbf{V}} \mathsf{E}\{|\theta_{\mathcal{B}}[m] - \delta_m(X_{\mathcal{B}}[m])|^2\} + \sum_{g_m \in \mathbf{V}^\perp} |\theta_{\mathcal{B}}[m]|^2 \ .$$

An oracle diagonal estimator in $\mathcal{B}$ has a risk calculated like in (3.13)

$$r_{\mathrm{inf},d}(\theta) = \sum_{g_m \in \mathbf{V}} \frac{\sigma_m^2 \, |\theta_{\mathcal{B}}[m]|^2}{\sigma_m^2 + |\theta_{\mathcal{B}}[m]|^2} + \sum_{g_m \in \mathbf{V}^\perp} |\theta_{\mathcal{B}}[m]|^2 \ . \tag{3.38}$$

This oracle risk is equal to the limit of the oracle risk (3.14) computed for the estimation problem $X = \theta + z$ with $\theta \in \Theta$ when letting $\sigma_m^2 = \mathsf{E}\{|z_{\mathcal{B}}[m]|^2\}$ tend to $+\infty$ for all $g_m \in \mathbf{V}^\perp$. This reflects the fact that we have lost all information about $\theta \in \mathbf{V}^\perp$. With this observation, we easily verify that Theorem 1 applies to the estimation problem (3.36), which is interpreted as a limit of estimation problems for a noise which tends to ∞ in $\mathbf{V}^\perp$. The inverse of the covariance K of z over $\mathbf{V}$ is $K^{-1} = \sigma^{-2}\, U^* \, U$. In this case, the diagonal preconditioning factor is

$$\lambda_{\mathcal{B}} = \|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s = \|U\, \sigma^{-1}\, K_d^{1/2}\|_s^2 \ .$$

It corresponds to a diagonal preconditioning factor of the degradation operator U in the basis $\mathcal{B}$.

If $\lambda_{\mathcal{B}}$ remains bounded and Θ is orthosymmetric then a thresholding estimator in $\mathcal{B}$ has a risk that is close to the minimax risk. In practice, we do not need to compute $X = \overline{U}^{-1} Y$ which is numerically unstable since $\|\overline{U}^{-1}\|_s$ is huge. The thresholding estimator $\hat\theta$ in (3.19) requires only to calculate

$$\langle X, g_m \rangle = \langle \overline{U}^{-1} Y, g_m \rangle = \langle Y, \overline{U}^{-1*} g_m \rangle$$

for a threshold $T_m \neq \infty$. Since $\|\overline{U}^{-1*} g_m\|^2 = \sigma_m^2/\sigma^2$ for $T_m \neq \infty$, and (3.23) implies that $\sigma_m \leq s_{\mathcal{B}}[m]$, we get

$$\|\overline{U}^{-1*} g_m\| \leq s_{\mathcal{B}}[m]/\sigma \ .$$

The vector $\overline{g}_m = \overline{U}^{-1*} g_m$ has a bounded norm and can be precalculated to estimate θ with a stable numerical procedure:

$$\hat{\theta} = \sum_{\substack{m=0 \\ T_m \neq \infty}}^{N-1} \rho_{T_m}(\langle Y, \overline{g}_m \rangle)\, g_m \ . \tag{3.39}$$

If $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ is a wavelet basis then Donoho [8] calls $\{\overline{g}_m\}_{0 \leq m < N}$ a *vaguelette* family and the estimator $\hat{\theta}$ is a wavelet-vaguelette thresholding estimator. Abramovich and Silverman [1] showed that inverting the role of wavelets and vaguelettes in (3.39) can improve the constant in the asymptotic behavior of the risk of these estimators.

4. Deconvolution of One-Dimensional Signals In many physical measuring devices the desired analog signal is degraded by a convolution operator to which is added a white noise. Section 4.1 writes this inverse problem as a discrete deconvolution and specifies a bounded variation model for the discretized signal that must be estimated. Section 4.2 reviews the properties of discrete wavelet bases, which are particularly well adapted to power law deconvolutions. For hyperbolic deconvolutions, wavelet bases must be modified. Section 4.3 constructs a new orthonormal basis in which a thresholding estimator can have a minimax convergence rate, as proved in Section 4.4.

4.1. *From Continuous to Discrete Deconvolution* Suppose that the measured discrete signal is a convolution of the desired signal $f(t)$ with a convolution kernel

$g(t)$, which is uniformly sampled and contaminated by a Gaussian white noise:

$$Y[n] = Gf(\frac{n}{N}) + \nu[n] \ \text{ with } \ Gf(t) = \int_{-\infty}^{+\infty} f(x)\, g(t-x)\, dx \ . \tag{4.40}$$

This deconvolution problem is rewritten as a discrete inverse problem, and a bounded variation model is specified for discrete signals.

Let $\widetilde{g}(\omega)$ be the Fourier transform of $g(t)$. If $g \in \mathbf{L^2}(\mathbb{R})$ then Appendix B shows that we can write

$$Gf(\frac{n}{N}) = \theta \star v[n] \ \text{ with } \ \theta[n] = \langle f(t)\,,\, \phi(t - n/N)\rangle \tag{4.41}$$

where $\{\phi(t-n/N)\}_{n\in\mathbb{Z}}$ is an orthonormal family in $\mathbf{L^2}(\mathbb{R})$ and $v \in \mathbf{l^2}(\mathbb{Z})$ has a Fourier transform which satisfies

$$|\widetilde{v}(\omega)|^2 = \sum_{p=-\infty}^{+\infty} |\widetilde{g}(N\omega + 2pN\pi)|^2 \ .$$

We can therefore rewrite (4.40) as a discrete convolution equation

$$Y[n] = \theta \star v[n] + \nu[n] \ . \tag{4.42}$$

We shall simplify border problems with periodic assumptions, by supposing that f is 1 periodic and hence that $\theta[n]$ is N periodic. Appendix B derives that the convolution in (4.41) can then be replaced by a circular convolution with an N periodic filter $u[n]$ whose discrete Fourier transform

$$\widetilde{u}[k] = \sum_{n=0}^{N-1} u[n]\, \mathrm{e}^{-i2\pi kn/N} \tag{4.43}$$

satisfies

$$|\widetilde{u}[k]|^2 = |\widetilde{v}(\frac{2k\pi}{N})|^2 = \sum_{p=-\infty}^{+\infty} |\widetilde{g}(2k\pi + 2pN\pi)|^2 \tag{4.44}$$

and hence

$$Y[n] = U\theta[n] + \nu[n] = \theta \star u[n] + \nu[n] \ \text{ for } 0 \le n < N \ . \tag{4.45}$$

The pseudo inverse of U is $\overline{U}^{-1}\theta = \theta \star \overline{u}^{-1}$ where the discrete Fourier transform of $\overline{u}^{-1}$ is

$$\widetilde{\overline{u}^{-1}}[k] = \begin{cases} 1/\widetilde{u}[k] & \text{if } \widetilde{u}[k] \neq 0 \\ 0 & \text{if } \widetilde{u}[k] = 0 \end{cases}. \tag{4.46}$$

The noise after deconvolution is $z = \nu \star \overline{u}^{-1}$. It is circular stationary and its covariance K is a circular convolution whose kernel is

$$c[n] = \sigma^2 \overline{u}^{-1} \star \overline{u}^{-1*}[n] \quad \text{with } \overline{u}^{-1*}[n] = \overline{u}^{-1}[-n]. \tag{4.47}$$

This covariance is diagonalized in the discrete Fourier basis and the eigenvalues of K (power spectrum) are

$$\sigma_k^2 = \sigma^2 \left|\widetilde{u}[k]\right|^{-2} .$$

If the convolution kernel in (4.40) has a power law decay $|\widetilde{g}(\omega)| \sim 1 + |\omega|^{-\beta}$ with $\beta > 1/2$ then we derive from (4.44) that $|\widetilde{u}[k]| \sim 1 + |k|^{-\beta}$ for $-N/2 \leq k \leq N/2$. This power law deconvolution has been studied by many authors [26, 8]. A simple example is the estimation of derivatives, which can be casted as a power law deconvolution with $\beta = 1$ [1].

A different type of deconvolution problem arises with measuring devices which have a limited resolution and where the sampling rate is adapted to the highest accessible frequencies. Suppose that $\widetilde{g}(\omega) = 0$ for $|\omega| > \xi$ and $|\widetilde{g}(\omega)| \sim |1 - |\omega|/\xi|^p$ for $|\omega| \leq \xi$. The Shannon sampling rate is $N^{-1} = \pi/\xi$ and (4.44) implies that

$$|\widetilde{u}[k]| \sim \left|\frac{2|k|}{N} - 1\right|^p . \tag{4.48}$$

As opposed to the power law example, in this case the convolution kernel g depends upon N. After deconvolution, the power spectrum σ_k^2 has a hyperbolic growth when the frequency k is in the neighborhood of $\pm N/2$. We call it a *hyperbolic deconvolution* problem of order p. If $p = 0$ then the inverse problem is not ill-defined and is equivalent to an estimation in presence of white noise.

The restoration is particularly difficult when the signal includes regular components and sharp transitions such as singularities. Bounded variation functions $f(t)$ have been used to model such signals [6, 8]. Over $[0,1]$ the total variation is defined by

$$\|f\|_{\overline{tv}} = \int_0^1 |f'(t)|\, dt$$

where f' is defined in the sense of distributions. For periodic signals, one must add $|f(1) - f(0)|$ to this integral, to incorporate the jump at the border. The total variation measures the amplitude of all oscillations, and f may include sharp transitions such as discontinuities. Discrete samples are obtained with a local averaging of f:

$$\theta[n] = \langle f\, ,\, \phi(t - n/N)\rangle = \int_{-\infty}^{+\infty} f(t)\, \phi(t - n/N)\, dt\ ,$$

where $\phi(t)$ depends upon N. If $\phi(t) = N^{1/2}\, O(1 + (Nt)^{-2})$ then a standard calculation shows that there exists a constant B such that the discrete total variation satisfies

$$\|\theta\|_{tv} = \sum_{n=0}^{N-1} \Big|\theta[n] - \theta[n-1]\Big| \leq B\, \|f\|_{\overline{tv}}\ . \tag{4.49}$$

If f has a total variation bounded by a constant, then for any N its discretization $\theta[n]$ has a uniformly bounded discrete total variation. We shall therefore consider models of discrete signals corresponding to bounded discrete total variation

$$\Theta_{tv} = \left\{ \theta\ :\ \|\theta\|_{tv} = \sum_{n=0}^{N-1} \Big|\theta[n] - \theta[n-1]\Big| \leq C \right\}\ . \tag{4.50}$$

4.2. *Wavelet Bases for Deconvolution* The properties of wavelet bases are briefly reviewed. Thresholding estimators in wavelet bases are a priori well adapted to estimate bounded variation signals because Θ_{tv} is nearly orthosymmetric in a wavelet basis. Donoho [8] also showed that power law convolution operators are nearly diagonalized in such wavelet bases and thus lead to nearly minimax thresholding estimators. However, this is not the case for hyperbolic deconvolution problems.

A wavelet basis of $\mathbf{L^2}(\mathbb{R})$ is constructed from a single *analog* wavelet $\psi^a(t)$ whose translations and dilations

$$\left\{ \psi^a_{j,m}(t) = \frac{1}{\sqrt{2^j}}\, \psi^a \left(\frac{t-2^j m}{2^j} \right) \right\}_{(j,m)\in\mathbb{Z}^2} \tag{4.51}$$

define an orthonormal basis of $\mathbf{L^2}(\mathbb{R})$. The wavelet $\psi^a(t)$ is obtained with a pair of conjugate mirror filters $h^a[n]$ and $g^a[n] = (-1)^{1-n} h^a[1-n]$ with $|\widetilde{h}^a(\omega)|^2 + |\widetilde{h}^a(\omega + \pi)|^2 = 1$:

$$\widetilde{\psi}^a(\omega) = \widetilde{g}^a\left(\frac{\omega}{2}\right)\, \widetilde{\phi}^a\left(\frac{\omega}{2}\right) \quad \text{with} \quad \widetilde{\phi}^a(\omega) = \prod_{p=1}^{+\infty} \widetilde{h}^a(2^{-p}\omega)\ . \tag{4.52}$$

The wavelet ψ^a has q vanishing moments, meaning that

$$\int_{-\infty}^{+\infty} t^k\, \psi^a(t)\, dt = 0 \quad \text{for any } 0 \le k < q,$$

if and only if $\widetilde{h}^a(\omega+\pi) = O(|\omega|^q)$. It has a compact support if and only if $h^a[n]$ has a compact support [5]. We shall also suppose that $|\widetilde{h}^a(\omega)| \ge 1/2$ for $|\omega| \le \pi/2$ and that $\|\psi^a\|_{\overline{tv}} < +\infty$. Daubechies wavelets [5] are examples of such wavelets.

Let $\psi^{\rm per}_{j,m}(t) = \sum_{k=-\infty}^{+\infty} \psi^a_{j,m}(t+k)$ be a wavelet of period 1, and $\psi^{\rm per}_{1,0}(t) = 1$. One can also prove [21] that

$$\{\psi^p_{j,m}(t)\}_{-\infty<j\le 1\,,\,0\le m<2^{-j}} \tag{4.53}$$

is an orthonormal basis of $\mathbf{L^2}[0,1]$. The corresponding discrete wavelet basis in $\mathbb{C}^N$ is obtained with an N periodic filter $h[n] = \sum_{p=-\infty}^{+\infty} h^a[n+pN]$ whose discrete Fourier transform is $\widetilde{h}[k] = \widetilde{h}(2\pi k/N)$ and with $g[n] = (-1)^{1-n} h[1-n]$ [20]. Let $L = -\log_2 N$. We write $\psi_j[n]$ the N periodic discrete wavelet and $\phi_j[n]$ the scaling signal whose discrete Fourier transforms are

$$(4.54)\qquad \widetilde{\psi}_j[k] = \widetilde{g}[2^{j-L-1}k]\,\widetilde{\phi}_{j-1}[k] \quad \text{with} \quad \widetilde{\phi}_j[k] = \prod_{l=0}^{j-L-1} \widetilde{h}[2^l k] \quad \text{for } j > L,$$

and $\widetilde{\phi}_L[k] = 1$. The Fourier transform $\widetilde{\psi}_j[k]$ is mostly non-negligible in the frequency band $|k| \in [2^{-j-1}, 2^{-j}]$, as illustrated in Figure 1. The wavelet ψ_j is translated: $\psi_{j,m}[n] = \psi_j[n - 2^{j-L}m]$. A constant vector $\psi_{1,0}[n] = N^{-1/2}$ is added to this family. The resulting family

$$(4.55)\qquad \{\psi_{j,m}\}_{L<j\leq 1\,,\,0\leq m<2^{-j}}$$

is an orthonormal basis of $\mathbb{C}^N$. The decomposition of $\theta \in \mathbb{C}^N$ in this basis is calculated with $O(N)$ operations, with a fast filter bank algorithm using the filters h and g [20]. Let $\phi^a_L(t) = 2^{-L/2}\phi^a(2^{-L}t)$ defined from (4.52). One can also verify that

$$(4.56)\qquad \psi^{\rm per}_j(t) = \sum_{n=-\infty}^{+\infty} \psi_j[n]\,\phi^a_L(t - n/N)\ .$$

When the dimension N increases to $+\infty$ the discrete wavelet basis converges to the periodized wavelet basis of $\mathbf{L^2}[0,1]$ and

$$(4.57)\qquad \lim_{N\to+\infty} \frac{\psi_{j,m}[n]}{N^{-1/2}\,\psi^{\rm per}_{j,m}(N^{-1}n)} = 1\ .$$

The properties of $\psi^a(t)$ thus specifies the asymptotic properties of the discrete wavelet basis as N increases.

Wavelet thresholding estimators have been applied by Donoho [8] to deconvolution problems

$$Y[n] = \theta \star u[n] + \nu[n]\ .$$

If θ belongs to a set that is nearly orthosymmetric in a wavelet basis and if this basis nearly diagonalizes the covariance (4.47) of the deconvolved noise then Theorem 1 shows that the resulting thresholding estimator has a nearly minimax risk.

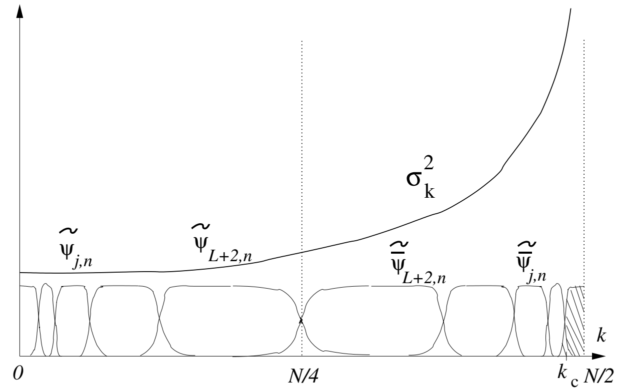

FIG. 1. *In a mirror wavelet basis, standard wavelets $\psi_{j,m}$ have a Fourier transform $\widetilde{\psi}_{j,m}$ that cover the lower frequencies and mirror wavelets $\widetilde{\psi}_{j,m}$ are in the upper half frequency interval. The power spectrum σ_k^2 of the noise has a hyperbolic growth but varies by a bounded factor on the interval where the energy of each mirror wavelet is mostly concentrated. There exists a cut-off frequency k_c above which the variance σ_k^2 of the noise is too high and the thresholding sets all coefficients to zero.*

An upper bound and a lower bound of a total variation norm can be calculated from the absolute value of wavelet coefficients. This result is proved by Meyer [21] for functions decomposed in a wavelet basis of $\mathbf{L^2}[0,1]$. The following proposition states this results for discrete signals and Appendix C adapts the proof of Meyer.

PROPOSITION 1. *There exists $A, B > 0$ such that for all $N > 0$*

$$B^{-1} \|\theta\|_{tv} \leq \sum_{j=L+1}^{1} \sum_{m=0}^{2^{-j}-1} 2^{(L-j)/2} \left|\langle \theta, \psi_{j,m} \rangle\right| \tag{4.58}$$

and

$$A^{-1} \|\theta\|_{tv} \geq \sup_{L<j\leq 1} \left(\sum_{m=0}^{2^{-j}-1} 2^{(L-j)/2} \left|\langle \theta, \psi_{j,m} \rangle\right| \right) . \tag{4.59}$$

This proposition provides an embedding of the bounded variation signal set $\Theta_{tv} = \{\theta \ : \ \|\theta\|_{tv} \leq C\}$, in two orthosymmetric sets defined by

$$\Theta_1 = \left\{ \theta \ : \ B \sum_{j=L+1}^{1} \sum_{m=0}^{2^{-j}-1} 2^{(L-j)/2} \, |\langle \theta, \psi_{j,m} \rangle| \leq C \right\}$$

and

$$\Theta_2 = \left\{ \theta \ : \ A \sup_{j \leq 1} \left(\sum_{m=0}^{2^{-j}-1} 2^{(L-j)/2} \, |\langle \theta, \psi_{j,m} \rangle| \right) \leq C \ \right\} \ .$$

The sets Θ_1 and Θ_2 are clearly orthosymmetric in the wavelet basis and we derive from (4.58) and (4.59) that $\Theta_1 \subset \Theta_{tv} \subset \Theta_2$.

The thresholds in a wavelet basis are chosen depending upon the variance of the deconvolved noise coefficients according to (3.23). As a consequence of (3.12) we have $\sigma_{j,m}^2 = \langle K \psi_{j,m} \, , \, \psi_{j,m} \rangle$. Since the covariance K of the deconvolved noise is diagonal in a the Fourier basis with a power spectrum $\sigma^2 \, |\widetilde{u}[k]|^{-2}$ we get

$$\sigma_{j,m}^2 = \frac{\sigma^2}{N} \sum_{k=-N/2+1}^{N/2} |\widetilde{u}[k]|^{-2} \, |\widetilde{\psi}_{j,m}[k]|^2$$

$$= \frac{\sigma^2}{N} \sum_{k=-N/2+1}^{N/2} |\widetilde{u}[k]|^{-2} \, |\widetilde{\psi}_j[k]|^2 = \sigma_j^2 \ . \tag{4.60}$$

This variance is independent of the position index m of the wavelet.

Using the asymptotic equivalence symbol $\sim$ defined at the end of the introduction, for a power-law deconvolution problem we have $|\widetilde{u}[k]|^{-1} \sim N^{-1} \, |k|^{\beta}$ and since $\widetilde{\psi}_j[k]$ is mostly non-zero in the frequency band $|k| \in [2^{-j-1}, 2^{-j}]$ we get $\sigma_j^2 \sim N^{-2} \, 2^{-2j\beta}$. Donoho [8] as well as Abramovich and Silverman [1] prove that in this case thresholding estimators are nearly minimax for the discretization of functions in Besov spaces or having a bounded variation, because the covariance K is nearly diagonalized in a wavelet basis. Intuitively, the reason is that the power spectrum $\sigma^2 \, |\widetilde{u}[k]|^{-2}$ varies by at most a constant factor over the interval $[2^{-j-1}, 2^{-j}]$

where $\widetilde{\psi}_j[k]$ is mostly non-zero. Johnstone and Silverman [14] also give a thresholding strategy to remove such power-law deconvolved noise, which have a long range dependence.

For hyperbolic deconvolutions where $\widetilde{u}[k]$ satisfies (4.48) the power spectrum is

$$\sigma^2 \, |\widetilde{u}[k]|^{-2} \sim \sigma^2 \left| \frac{2|k|}{N} - 1 \right|^{-2p} .$$

If $2^j > 2^{L+1}$ then this spectrum varies by a bounded factor for $k \in [2^{-j-1}, 2^{-j}]$ which indicates that the wavelets $\psi_{j,m}$ can be approximate eigenvectors of K, as illustrated by Fig. 1. At the finest scale $2^j = 2^{L+1}$, $|\widetilde{\psi}_{L+1}[k]|$ has an energy spread over the higher frequency band $[N/4, N/2]$ where the spectrum varies by a huge factor of the order of N^{2p}. Appendix D proves that as a result a wavelet thresholding estimator has a larger risk and there exists a constant $B > 0$ such that for all C, N and σ with $C^2/\sigma^2 \leq N^{2p+1}$

$$r_t(\Theta_{tv}) \geq B\, C^2 \ . \tag{4.61}$$

We shall see that in this case even a linear estimator can outperform this wavelet thresholding estimator.

4.3. *Mirror Wavelet Basis* For a hyperbolic deconvolution, to construct a basis of approximate eigenvectors of K, the finest scale wavelets $\psi_{L+1,m}$ must be replaced by vectors that have a Fourier transform concentrated in smaller subintervals of $[N/4, N/2]$, where the spectrum σ_k^2 varies by a factor that does not grow with N. Let $\mathbf{W}_{L+1}$ be the space generated by $\{\psi_{L+1,m}\}_{0 \leq m < N/2}$. One must define a new basis of $\mathbf{W}_{L+1}$ whose vectors have a Fourier transform whose energy is mostly concentrated in intervals of the form $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$ for $0 \geq j > L+1$, where $\sigma_k^2 \sim \sigma^2 \, |2|k|/N - 1|^{-2p}$ varies by a bounded factor. We do not want to further reduce the size of these frequency intervals because it would increase the total variation of the basis vectors and thus represent less efficiently signals having a bounded total variation.

Mirror wavelets are defined with a cascade of convolution of the same conjugate mirror filters $h[n]$ and $g[n]$ used to construct the wavelets ψ_j in (4.54) but in a different order. Let $L = -\log_2 N$. The discrete Fourier transform of a mirror wavelet $\widetilde{\overline{\psi}}_j[k]$ is

$$\widetilde{\overline{\psi}}_j[k] = \widetilde{g}[k] \left(\prod_{l=1}^{j-L-2} \widetilde{h}^*[2^l k] \right) \widetilde{g}^*[2^{j-L-1} k] \tag{4.62}$$

where $\widetilde{h}^*[n]$ and $\widetilde{g}^*[n]$ are Fourier transforms of $h^*[n] = h[-n]$ and $g^*[n] = g[-n]$. Since $g[n] = (-1)^{1-n} h[1-n]$ one can verify that

$$\overline{\psi}_{j,m}[n] = (-1)^{n-1} \psi_{j,m}[1-n] \ , \tag{4.63}$$

and

$$|\widetilde{\overline{\psi}}_{j,m}[k]| = |\widetilde{\psi}_{j,m}[N/2-k]| \ .$$

The energy of $\widetilde{\overline{\psi}}_{j,m}[k]$ is thus mostly concentrated in $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$ as illustrated in Fig. 1.

The mirror wavelets (4.62) are obtained with a cascade of convolutions of conjugate mirror filters, beginning with the filter g, and therefore correspond to a particular family of wavelet packets [4], which generate the same space $\mathbf{W}_{L+1}$. From general properties of wavelet packets, we derive that

$$\left\{ \overline{\psi}_{j,m}[n] = (-1)^{n-1} \psi_{j,m}[1-n] \right\}_{L+2 \le j \le 1\,,\, 0 \le m < 2^{-j}}$$

is an orthonormal basis of $\mathbf{W}_{L+1}$. Hence

$$\mathcal{B} = \left\{ \psi_{j,m}, \overline{\psi}_{j,m} \right\}_{0 \le m < 2^{-j}\,,\, L+1 < j \le 1}$$

is an orthonormal basis of $\mathbb{C}^N$, called a *mirror wavelet basis.* Computing the decomposition coefficients of a signal in this basis is performed with a wavelet packet filter bank described in [17], which requires $O(N)$ operations. When N varies, a discrete wavelet basis can be interpreted as a discretization of a single wavelet basis of $\mathbf{L^2}[0,1]$ with a resolution that varies. On the contrary, mirror wavelet bases for different N are totally different bases.

To prove that the covariance K is "almost diagonalized" in the mirror wavelet basis $\mathcal{B}$ with a diagonal preconditioning factor that does not grow with N, one must control the asymptotic behavior of this basis as N increases. We explained that the discrete wavelet basis converges to a periodic basis of $\mathbf{L^2}[0,1]$ defined from a mother wavelet $\psi^a(t) \in \mathbf{L^2}(\mathbb{R})$. The following theorem proves that if ψ^a has enough vanishing moments and is sufficiently regular then the preconditioning factor is bounded by a factor independent of N. The theorem proof is in Appendix E.

THEOREM 2. *Suppose that ψ^a has a compact support is $\mathbf{C}^{q+1}$ and has q vanishing moments. For a hyperbolic deconvolution of order $p < q$, there exists λ such that for all $N > 0$ the covariance K in the mirror wavelet basis $\mathcal{B}$ satisfies*

$$\|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s \le \lambda. \tag{4.64}$$

The theorem proves that the lower frequency wavelets $\{\psi_{j,m}\}_{0\le m<2^{-j}\,,\,L+1<j\le 1}$ are nearly eigenvectors of K because over this range of frequency the eigenvalues of K remain of the same order. The high frequency mirror wavelets $\{\overline{\psi}_{j,m}\}_{0\le m<2^{-j}\,,\,L+1<j\le 1}$ are also nearly eigenvectors of K because they are sufficiently well localized in frequency. The mirror wavelet basis is different from the wavelet packet basis initially proposed by Rougé [24], but it follows his idea to use wavelet packets to refine the frequency resolution of the basis vectors.

4.4. *Thresholding Estimation with Mirror Wavelets* The implementation of a thresholding estimator in a mirror wavelet basis is described and it is proved that the resulting risk is equivalent to the non-linear minimax risk for bounded variation signals.

Let us first consider a subset Θ of $\mathbb{C}^N$ which is orthosymmetric in the mirror wavelet basis. Suppose that the soft thresholding is performed with thresholds computed according to (3.23), which depend upon the variances of the deconvolved noise in the direction of the different wavelets and mirror wavelets. These variances do not depend upon the position index m. The proof of Theorem 2 shows in (E.99) that for $L+2 \leq j \leq 1$ and for the 2^{-j} wavelets corresponding to $0 \leq m < 2^{-j}$

$$\sigma_j^2 = \langle K\psi_{j,m}, \psi_{j,m}\rangle = \frac{1}{N} \sum_{k=-N/2+1}^{N/2} \sigma_k^2 \, |\widetilde{\psi}_j[k]|^2 \sim \sigma^2 \ . \tag{4.65}$$

For the 2^{-j} mirror wavelets corresponding to $0 \leq m < 2^{-j}$, (E.109) proves that

$$\overline{\sigma}_j^2 = \langle K\overline{\psi}_{j,m}, \overline{\psi}_{j,m}\rangle = \frac{1}{N} \sum_{k=-N/2+1}^{N/2-1} \sigma_k^2 \, |\widetilde{\overline{\psi}}_j[k]|^2 \sim \sigma^2 \, 2^{2p(j-L)} \ . \tag{4.66}$$

For a hyperbolic deconvolution, the preconditioning result (4.64) together with (3.35) yields

$$\frac{1}{1.25\,\lambda}\, r_{\inf,d}(\Theta) \leq \, r_{\inf,d}(\Theta) \leq r_n(\Theta) \leq r_t(\Theta) \leq (4 \log_{\mathrm{e}} N + 4)\, r_{\inf,d}(\Theta) \ . \tag{4.67}$$

This proves that for an orthosymmetric set Θ, a thresholding in a mirror wavelet basis reaches the minimax risk up to a factor proportional to $\log_{\mathrm{e}} N$.

To prove that a thresholding estimator is nearly minimax optimal in a set of bounded variation signals Θ_{tv}, it remains to prove that Θ_{tv} is "nearly" orthosymmetric in this basis. For this purpose, the following proposition computes an upper bound and a lower bound of $\|\theta\|_{tv}$ from the absolute value of mirror wavelet coefficients, from which Appendix G derives an embedding of Θ_{tv} in two sets that are orthosymmetric in the mirror wavelet basis.

PROPOSITION 2. *There exists $A, B > 0$ such that for all $N > 0$*

(4.68)

$$B^{-1}\,\|\theta\|_{tv} \leq \sum_{j=L+2}^{1} 2^{(L-j)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m}\rangle| + \sum_{j=L+2}^{1} 2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m}\rangle|$$

and

(4.69)

$$A^{-1}\,\|\theta\|_{tv} \geq \sup_{1\geq j\geq L+2} \left(2^{(L-j)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m}\rangle| \ , \ 2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m}\rangle| \right) \ .$$

The proof is in Appendix F. To avoid having a thresholding risk which is above the minimax risk by a $\log_e N$ factor, the thresholds are chosen according to (3.26). We saw in (4.65) that the deconvolved noise has a variance $\sigma_j \sim \sigma$ for the $N/2$ lower frequency wavelets $\psi_{j,m}$ at scales $2^j > 2^{L+1}$. For each scale 2^j the corresponding threshold is therefore

$$T_j = \sigma_j^2 \sqrt{2\,\log_e(N/2)}\ . \tag{4.70}$$

In (4.66), the noise variance $\overline{\sigma}_j \sim \sigma^2\, 2^{2p(j-L)}$ is the same for the 2^{-j} mirror wavelets $\{\widetilde{\psi}_{j,m}\}_{0\leq m<2^{-j}}$ but it varies greatly with the scale 2^j. The threshold corresponding to a mirror wavelet $\overline{\psi}_{j,m}$ is therefore

$$\overline{T}_j = \begin{cases} \overline{\sigma}_j \sqrt{2\log_e 2^{-j}} & \text{if } \overline{\sigma}_j \sqrt{2\log_e 2^{-j}} \leq \overline{s}_j \\ \infty & \text{otherwise} \end{cases} \tag{4.71}$$

with $\overline{s}_j = \sup_{\theta\in\Theta_{tv}} |\langle \theta, \overline{\psi}_{j,m}\rangle|$, where $\theta \in \Theta_{tv}$ if $\|\theta\|_{tv} \leq C$. Proposition 2 implies that $\overline{s}_j \sim C\,2^{(L-j)/2}$. The thresholding estimation can thus be decomposed in two parts. Over the lower frequency wavelets, it is equivalent to an estimation in presence of nearly white noise whereas the unstable component is carried by the thresholding of the high frequency mirror wavelets. The problem is studied in the range of noise where $C/\sigma \geq N^{1/2}$ to guarantee that the estimation risk is not dominated by the lower frequency noise as opposed to the attenuation of the signal high frequencies. We also impose that $C/\sigma \leq N^{p+1/2}$ otherwise σ is so small that the inverse problem is not truly unstable. The hypothesis $N^{1/2} \leq C/\sigma \leq N^{p+1/2}$ thus corresponds to the range of the signal to noise ratio where the estimation risk is dominated by the degradations of the signal high frequencies, while the noise is non-negligible.

Let 2^c be the critical scale defined as the smallest scale such that $\overline{T}_j = \infty$ for all $2^j > 2^c$. This corresponds to mirror wavelets $\overline{\psi}_{j,m}$ whose Fourier transform have an energy concentrated at high frequencies $|k| \geq k_c = N/2 - 2^{-c}$, illustrated in Figure 1. The thresholding estimator does not recover the signal projection over these higher frequencies. The following theorem proves that over Θ_{tv} the thresholding risk and the non-linear minimax risk are dominated by the risk in the direction of the mirror wavelets $\{\overline{\psi}_{c,m}\}_{0 \leq m < 2^{-c}}$ at the critical scale 2^c. The linear minimax risk is also computed.

THEOREM 3. *Let $\mathcal{B}$ a mirror wavelet basis corresponding to a compactly supported wavelet $\psi^a(t)$ that is $\mathbf{C}^{q+1}$ with q vanishing moments. For a hyperbolic deconvolution of order $p < q$ with $p \geq 1$, if $N^{1/2} \leq C/\sigma \leq N^{p+1/2}$ then*

$$r_l(\Theta_{tv}) \sim C^2 \left(\frac{\sigma^2 N}{C^2}\right)^{1/2p} \tag{4.72}$$

and

$$r_t(\Theta_{tv}) \sim r_n(\Theta_{tv}) \sim C^2 \left(\frac{\sigma^2}{C^2} \log_2 \frac{\sigma N^{p+1/2}}{C}\right)^{1/(2p+1)} . \tag{4.73}$$

The proof is in Appendix G. This theorem proves that a thresholding estimator in a mirror wavelet basis yields a quasi minimax deconvolution estimator for bounded variation signals. The equivalence $\sim$ means that there exists $A, B > 0$ such that for all C, σ and N

$$A\, r_n(\Theta_{tv}) \leq r_t(\Theta_{tv}) \leq B\, r_n(\Theta_{tv}) .$$

As opposed to the result (4.67) there is no loss of a $\log N$ factor between the thresholding and the non-linear minimax risk. If $N^2\sigma^2 = o(C^2)$ then the linear minimax risk (4.72) is much smaller than the lower bound (4.61) obtained for a thresholding estimator in a wavelet basis.

Observe that the set Θ_{tv} is translation invariant in the sense that if $\theta[n] \in \Theta_{tv}$ then $\theta[n-p] \in \Theta_{tv}$ for any $p \in \mathbb{Z}$. Since z is a circular stationary process, the estimation of $\theta \in \Theta_{tv}$ from $X = \theta + z$ is a translation invariant estimation problem. As a result, one can prove [20] that the minimax linear estimator is reached by translation invariant operators and hence by convolution operators. The cut-off frequency of the convolution operator that reaches the minimax risk is smaller than the cut-off frequency of a thresholding estimator, which explains why $r_t(\Theta_{tv})$ is much smaller than the linear minimax risk $r_l(\Theta_{tv})$. If $C^2 \sim N\sigma^2$ then

$$r_l(\Theta_{tv}) \sim C^2 \quad \text{and} \quad r_t(\Theta_{tv}) \sim C^2 \left(\frac{\log_e N}{N}\right)^{1/(2p+1)} . \tag{4.74}$$

The existence of cut-off frequencies for linear and thresholding estimators imply that the pseudo-inverse kernel $\overline{u}^{-1}$ defined in (4.46) can be replaced by a truncated kernel so that the calculation of $X = Y \star \overline{u}^{-1}$ is numerically stable:

$$\widetilde{\overline{u}^{-1}}[k] = \begin{cases} 1/\tilde{u}[k] & \text{for } |k| \leq k_c \\ 0 & \text{for } |k| > k_c \end{cases} .$$

The thresholding estimator in a mirror wavelet basis is not translation invariant because the basis is not translation invariant. It can however be made translation invariant with the shifting procedure of Coifman and Donoho [3]. The degraded data is translated $X_l[n] = X[n-l]$ for $0 \leq l < N$, and the translation-invariant estimate obtained by computing a thresholding estimator $\hat{\theta}_l$ in a mirror wavelet basis from each X_l and averaging the $\hat{\theta}_l[n+l]$ over l. In wavelet and wavelet packet bases, which are partially translation invariant, Coifman and Donoho give a translation-

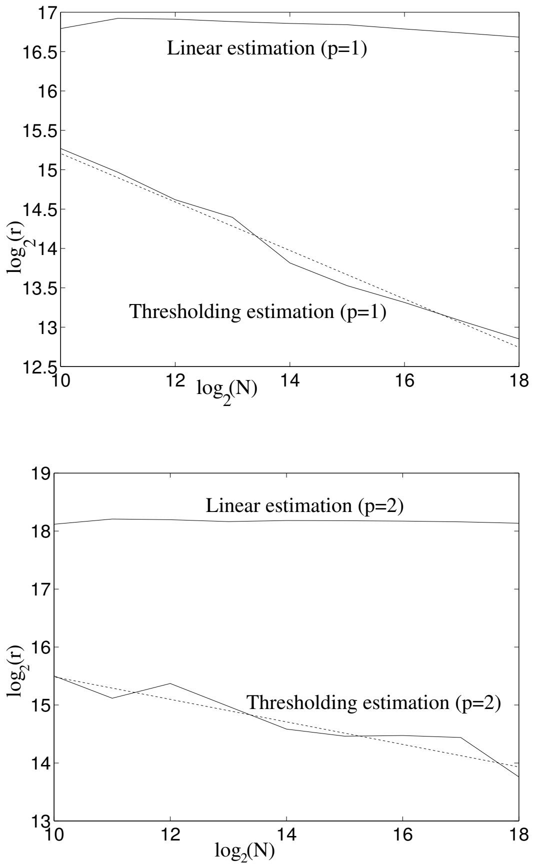

FIG. 2. *When C^2 and $N\sigma^2$ are constant, the risk of a linear estimation remains nearly constant, whereas $\log_2 r_t(\theta)$ decays linearly for a thresholding deconvolution in a mirror wavelet basis, with a slope that depends on the order p of the deconvolution.*

invariant filter bank algorithm which requires $O(N \log N)$ operations. Numerical experiments [3] show that this translation invariant procedure reduces the risk of thresholding estimators, although there is no satisfying theoretical explanation. This improvement also appeared clearly in all our numerical experiments.

Figure 2 shows the risk of the restoration of a piecewise regular signal as a function of $\log_2 N$ for the deconvolution of this same signal, while maintaining C^2

and $N\sigma^2$ constant. The computations are done with two convolution operators whose transfer functions $\widetilde{u}[k]$ have respectively $p = 1$ and $p = 2$ zeros at $k = \pm N/2$. For $p = 1$ and $p = 2$ the value of $N\sigma^2$ is set respectively to 10 and 10^{-1}, with $N \geq 10^3$ and $C = 3\,10^3$. As expected from (4.74) the risk of the linear estimator remains nearly constant. For a thresholding estimator in a mirror wavelet basis, (4.74) shows that $\log_2(r_t(\Theta_{tv}))$ decreases when $\log_2 N$ increases, with a slope equal to $-1/(2p+1)$. The piecewise regular signal that is used has a discontinuity so it behaves like a worst case in Θ_{tv}. The measured slope is approximately -0.32 for $p = 1$ whereas $-1/(2p+1) \simeq -0.33$, and this slope is approximately -0.18 for $p = 2$ whereas $-1/(2p+1) = -0.2$. Numerical results are thus close to theoretical predictions.

5. Deconvolution of Images Nearly optimal deconvolution of bounded variation images can be calculated with a separable extension of the deconvolution estimator in a mirror wavelet basis. Such a restoration algorithm is now used by the French Spatial Agency (CNES) for the production of satellite images [25]. In a satellite observation, the exposition time of the photo-receptors cannot be reduced too much because the light intensity reaching the satellite is small and must not be dominated by electronic noise. The satellite movement thus produces a blur and the optics of the satellite camera introduce another blur. The electronics of the photo-receptors adds a Gaussian white noise. When the satellite is in orbit, a calibration procedure measures the impulse response u of the blur and the noise variance σ^2. The image in Figure 3(b) is a simulated satellite image provided by the CNES, which is calculated from an airplane image shown in Figure 3(a). The impulse response is a separable low-pass filter

$$U\theta[n_1, n_2] = \theta \star u[n_1, n_2] \quad \text{with} \quad u[n_1, n_2] = u_1[n_1]\, u_2[n_2] \ .$$

For square images of $N \times N$ pixels, the discrete Fourier transforms of u_1 and u_2 have respectively a zero of order p_1 and p_2 at $\pm N/2$

$$\widetilde{u}_1[k_1] \sim \left|\frac{2|k_1|}{N} - 1\right|^{p_1} \qquad \text{and} \qquad \widetilde{u}_2[k_2] \sim \left|\frac{2|k_2|}{N} - 1\right|^{p_2} \quad .$$

We shall study the deconvolution of images with such separable kernels.

The deconvolved noise has a covariance K that is diagonalized in a two-dimensional discrete Fourier basis. The eigenvalues (power spectrum) are

$$\sigma^2_{k_1,k_2} = \frac{\sigma^2}{|\widetilde{u}_1[k_1]|^2\,|\widetilde{u}_2[k_2]|^2} \sim \sigma^2 \left|\frac{2|k_1|}{N} - 1\right|^{-2p_1} \left|\frac{2|k_2|}{N} - 1\right|^{-2p_2} \quad . \tag{5.75}$$

Observe that the hyperbolic growth of the eigenvalues at high frequencies depends upon the direction where we move in the frequency plane (k_1, k_2). If we move along the horizontal line $(k_1, k_2) = (k, 0)$, when $|k|$ increases the exponent of the hyperbolic growth is $2p_1$. Along the vertical line $(k_1, k_2) = (0, k)$, this exponent is $2p_2$ and it is $2(p_1 + p_2)$ along the diagonal line $(k_1, k_2) = (k, k)$. The estimation risk thus has a strong dependency upon the directional properties of the two-dimensional signals.

In image processing, a major difficulty is to find relevant image models. For complex images such as satellite images, it is very difficult to find a stochastic model which is consistent with the observed data. This would not be true for more specialized images such as certain types of textures. As a result, it is difficult to compute a Bayes risk for complex images such as satellite images. Recently, the minimax approach has been quite successful to obtain analytical computations of the risk of thresholding estimators which is consistent with numerical results on real images. For image estimation in presence of white noise, bounded variation image models have been used because they allow the presence of discontinuities, which is the case in most images. Yet, we shall see that this model needs to be refined for deconvolution problems.

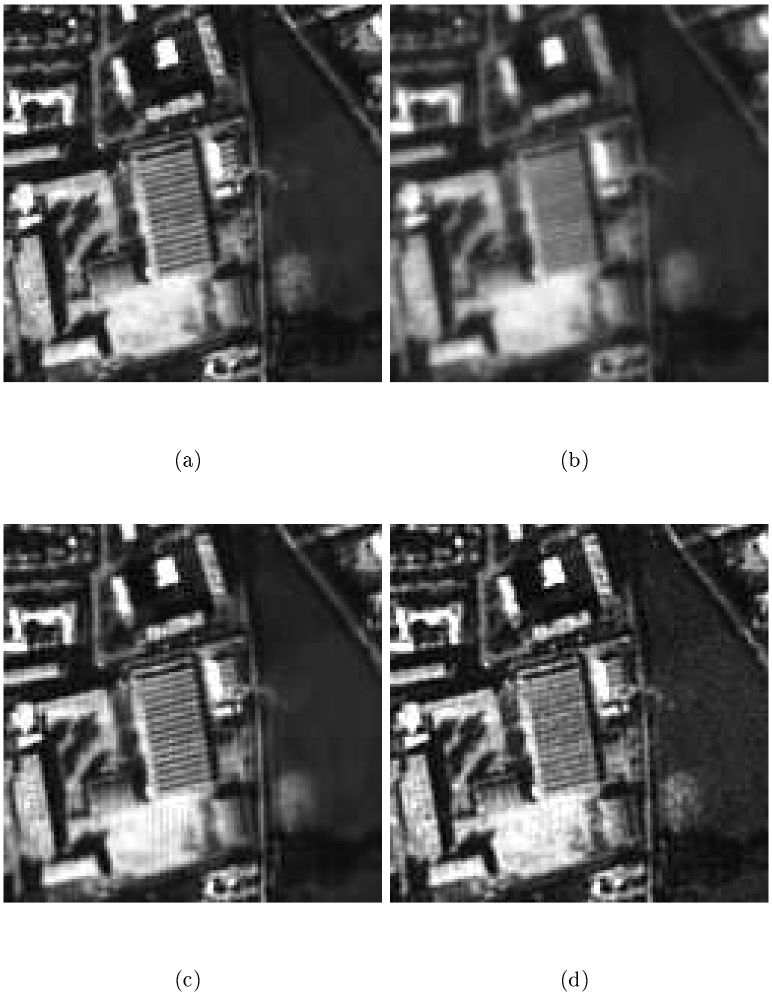

FIG. 3. *(a): Zoom on the original airplane image. (b): Simulation of a satellite image provided by the CNES (SNR = 31.1db). (c): Deconvolution with a translation invariant thresholding in a mirror wavelet basis (SNR = 34.3db). (d): Deconvolution calculated with a circular convolution, which yields a nearly minimax linear risk for bounded variation images (SNR = 32.7db).*

5.1. *Separable Mirror Wavelet Basis* Like in one-dimension, it is tempting to use a wavelet basis to implement a thresholding estimator because large classes of images can be modeled as elements of a set Θ that is nearly orthosymmetric in a wavelet basis. This is the case for bounded variation images. However, a thresholding algorithm in a wavelet basis gives poor results for a hyperbolic deconvolution because the covariance of the noise is not nearly diagonalized in this basis. We thus modify this basis at high frequencies with separable mirror wavelets.

A separable wavelet basis of $\mathbb{C}^{N^2}$ is constructed from the one-dimensional discrete wavelets ψ_j and scaling signals ϕ_j defined in (4.54). Let $L = -\log_2 N$. At each scale $2^L < 2^j \leq 1$ there are three wavelets defined by

(5.76)

$$\psi_j^1[n_1, n_2] = \phi_j[n_1]\,\psi_j[n_2]\ ,\ \psi_j^2[n_1, n_2] = \psi_j[n_1]\,\phi_j[n_2]\ ,\ \psi_j^3[n_1, n_2] = \psi_j[n_1]\,\psi_j[n_2]\ .$$

These wavelets are uniformly translated to define $\psi^\alpha_{j,m_1,m_2}[n_1, n_2] = \psi^\alpha_j[n_1 - 2^{j-L}m_1, n_2 - 2^{j-L}m_2]$ for $\alpha = 1, 2, 3$. The constant wavelet $\psi^0_0[n_1, n_2] = 1/N$ is included so that the separable wavelet family

$$\left\{\psi^0_1\ ,\ \psi^1_{j,m_1,m_2}\ ,\ \psi^2_{j,m_1,m_2}\ ,\ \psi^3_{j,m_1,m_2}\right\}_{L<j\leq 0\ ,\ 0\leq m_1, m_2 < 2^{-j}}$$

is an orthonormal basis of $\mathbb{C}^{N^2}$. Fast wavelet decomposition and reconstruction algorithms are implemented with $O(N^2)$ operations, with a separable filter bank algorithm [20]. The asymptotic properties of this wavelet basis for N large depends upon the asymptotic properties of the corresponding one-dimensional wavelet basis, which is governed by the analog wavelet $\psi^a(t) \in \mathbf{L^2}(\mathbb{R})$.

To nearly diagonalize the noise covariance K, each vector of the basis should have a Fourier transform whose energy is concentrated in a frequency domain where the eigenvalues

$$\sigma^2_{k_1,k_2} \sim \sigma^2 \left|\frac{2|k_1|}{N} - 1\right|^{-2p_1} \left|\frac{2|k_2|}{N} - 1\right|^{-2p_2}$$

vary at most by a constant factor. At low frequencies $(|k_1|, |k_2|) \in [0, N/4]^2$ the variance is nearly constant $\sigma^2_{k_1,k_2} \sim \sigma^2$. At scales $2^j > 2^{L+1}$, the lower frequency wavelets of the family

$$\mathcal{B}_0 = \left\{\psi^0_1\,,\ \psi^1_{j,m_1,m_2}\,,\ \psi^2_{j,m_1,m_2}\,,\ \psi^3_{j,m_1,m_2}\right\}_{L+1<j\leq 0\,,\,0\leq m_1,m_2<2^{-j}} \tag{5.77}$$

have a Fourier transform mostly concentrated in the lower frequency square $[-N/4, N/4]^2$ and hence are approximate eigenvectors of K. However, this is not the case for the $3N^2/4$ finer scale wavelets

$$\left\{\psi^1_{L+1,m_1,m_2}\,,\ \psi^2_{L+1,m_1,m_2}\,,\ \psi^3_{L+1,m_1,m_2}\right\}_{0\leq m_1,m_2<N/2} \tag{5.78}$$

whose Fourier transform have an energy concentrated over high frequency squares where either $|k_1| \in [N/4, N/2]$ or $|k_2| \in [N/4, N/2]$. In these domains $\sigma^2_{k_1,k_2}$ varies by a huge factor. Observe that $\sigma^2_{k_1,k_2} = \sigma^2_{k_1}\,\sigma^2_{k_2}$ with

$$\sigma^2_{k_1} = \sigma \left|\frac{|2k_1|}{N} - 1\right|^{-2p_1} \quad \text{and} \quad \sigma^2_{k_2} = \sigma \left|\frac{2|k_2|}{N} - 1\right|^{-2p_1} .$$

This suggests constructing a separable basis which segments the horizontal and vertical frequency axes into intervals where both σ_{k_1} and σ_{k_2} vary by a bounded factor. This is precisely obtained with one-dimensional mirror wavelets. The one-dimensional results of Section 4.3 imply that the space $\mathbf{W}_{L+1}$ generated by the $3N^2/4$ fine scale wavelets (5.78) also admits a separable basis of mirror wavelets

$$\mathcal{B}_1 = \left\{\overline{\psi}_{j_1,m_1}[n_1]\,\overline{\psi}_{j_2,m_2}[n_2]\right\}_{L+1\leq j_1,j_2\leq 1\,,\,(j_1,j_2)\neq(L+1,L+1)\,,\,0\leq m_1<2^{-j_1}\,,\,0\leq m_2<2^{-j_2}}\ , \tag{5.79}$$

where $\overline{\psi}_{j,m}[n]$ is the one-dimensional mirror wavelet defined in (4.63). The union $\mathcal{B} = \mathcal{B}_0 \cup \mathcal{B}_1$ of low frequency wavelets and high frequency mirror wavelets is there-

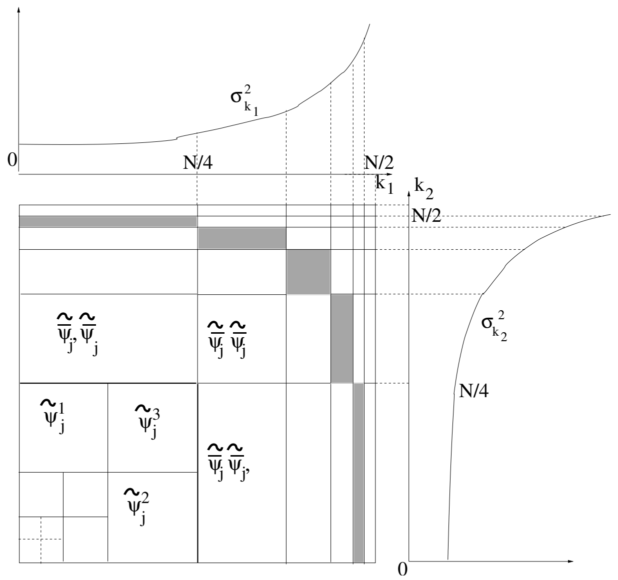

FIG. 4. *The separable mirror wavelet basis segments the frequency plane (k_1, k_2) into rectangles over which the noise variance $\sigma^2_{k_1,k_2} = \sigma^2_{k_1}\,\sigma^2_{k_2}$ varies by a bounded factor. The lower frequencies are covered by separable wavelets ψ^α_j, and the higher frequencies are covered by separable mirror wavelets $\widetilde{\psi}_{j_1}\,\widetilde{\psi}_{j_2}$. The grey rectangles correspond to critical scales beyond which the thresholding sets all coefficients to zero.*

fore an orthonormal basis of $\mathbb{C}^{N^2}$. This two-dimensional mirror wavelet basis segments the Fourier plane into rectangles illustrated in Figure 4. It is an anisotropic wavelet packet basis as defined in [28], because these mirror wavelets have a rectangular support that is generally not square. The decomposition of images in a separable mirror wavelet basis is computed with a filter bank algorithm described in [17], which requires $O(N^2)$ operations. The following theorem formally proves that the covariance K of the deconvolved noise is "nearly" diagonalized in this separable mirror wavelet basis if the analog wavelet $\psi^a(t)$ is sufficiently regular and has enough vanishing moments. The proof is in Appendix H.

THEOREM 4. *Suppose that ψ^a is a compactly supported $\mathbf{C}^{q+1}$ wavelet with q vanishing moments, and that $q > \max(p_1, p_2)$. There exists λ such that for all $N > 0$, the covariance K in the separable mirror wavelet basis $\mathcal{B}$ satisfies*

$$\|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s \leq \lambda\ .$$

5.2. *Thresholding Estimation of Bounded Variation Images* The co-area formula [29] proves that a bounded function f has a bounded total variation

$$\|f\|_{\overline{tv}} = \int_0^1 \int_0^1 |\vec{\nabla} f(x_1, x_2)|\, dx_1\, d_2 \tag{5.80}$$

if and only if the average length of its level set is finite. This class includes functions that are discontinuous along curves of finite length. As a result, image models with a bounded total variation have been used in image processing [2]. Like in one dimension, the discretization of a bounded variation function $f(x_1, x_2)$ leads to a bounded variation discrete image $\theta[n_1, n_2]$, where $\theta[n_1, n_2]$ is a local average of f in the neighborhood of $(n_1/N, n_2/N)$. A Riemann sum approximation of the integral (5.80) and a finite difference approximation of partial derivatives leads to the following definition of the discrete total variation:

$$\|\theta\|_{tv} = \frac{1}{N} \sum_{n_1, n_2 = 0}^{N-1} \left(\Big|\theta[n_1, n_2] - \theta[n_1 - 1, n_2]\Big|^2 + \Big|\theta[n_1, n_2] - \theta[n_1, n_2 - 1]\Big|^2 \right)^{1/2} . \tag{5.81}$$

To incorporate the fact that images have a bounded amplitude, the following set can be used as an image model:

$$\Theta_{tv} = \{\theta \ : \ \|\theta\|_{tv} \leq C_1 \ \ , \ \ \|\theta\|_\infty \leq C_2 \ \ \}\ .$$

Such models have been used to compute the risk of estimators in presence of white noise.

A bounded variation image model was proposed in [16] for a hyperbolic deconvolution, but this model leads to a minimax risk which is much too large compared to what is obtained with typical images. Consider images whose total variation is concentrated along a single row or column, with high frequency oscillations. A typical example is $\theta_\epsilon[n_1, N/2] = \epsilon\, C\, (-1)^{n_1}$ with $\epsilon = \pm 1$ and $\theta_\epsilon[n_1, n_2] = 0$ if $n_2 \neq N/2$. Clearly $\theta_\epsilon \in \Theta_{tv}$. However $U\theta_\epsilon = 0$ so one can verify that the minimax risk to estimate $\theta_\epsilon \in \{\theta_1\,,\,\theta_{-1}\} \subset \Theta_{tv}$ from $Y = U\theta_\epsilon + \nu$ is $\|\theta_\epsilon\|^2 = C\, N^2$. As a result $r_n(\Theta_{tv}) \geq C\, N^2$. In fact, one can verify [15] that $r_n(\Theta_{tv}) \sim C\, N^2$.

The two-dimensional total variation norm (5.81) is equivalent to averaging the one-dimensional total variation of the image along its rows and columns. The images θ_ϵ have their total variation concentrated along a single column. This is not the case for natural images or satellite images, which have a total variation that is nearly uniformly spread across all lines and columns. To incorporate this information, a more precise model is defined with an upper bound of the one-dimensional total variation along each row and column. The resulting *directional total variation* is defined by

$$\|\theta\|_{dv} = \sup_{0 \leq n_1, n_2 < N} \left(\sum_{p=0}^{N-1} |\theta[n_1, p] - \theta[n_1, p-1]|\ ,\ \sum_{p=0}^{N-1} |\theta[p, n_2] - \theta[p-1, n_2]| \right) .$$

Satellite images as well as piece-wise regular images have a bounded directional total variation norm. We thus study the minimax deconvolution over the image set

$$\Theta_{dv} = \left\{ f \in \mathbb{C}^{N^2}\ :\ \|\theta\|_{dv} \leq C \right\} .$$

For a thresholding estimator in a separable mirror wavelet basis, the thresholds are chosen according to (3.26). The variance of the noise in the direction of the low frequency wavelet coefficients satisfy

$$\sigma^2_{j,\alpha} = \langle K \psi^\alpha_{j,m_1,m_2}\,,\, \psi^\alpha_{j,m_1,m_2} \rangle \sim \sigma^2\ . \tag{5.82}$$

Since the noise variance is nearly constant for these $N^2/4$ wavelet coefficients, the thresholds are chosen to be

$$T_{j,\alpha} = \sigma_{j,\alpha} \sqrt{2 \log_e (N^2/4)} \ . \tag{5.83}$$

In the direction of the high frequency mirror wavelets, the noise has a variance

$$\overline{\sigma}^2_{j_1,j_2} = \langle K \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2} \, , \, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2} \rangle \sim \sigma^2 \, 2^{2p_1(j_1-L)} \, 2^{2p_2(j_2-L)} \ . \tag{5.84}$$

It remains constant for the $2^{-j_1-j_2}$ mirror wavelets $\{\overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\}_{0 \le m_1 < 2^{-j_1}, 0 \le m_2 < 2^{-j_2}}$, but it varies greatly with the scales 2^{j_1} and 2^{j_2}. The thresholds also depend upon the maximum amplitude of mirror wavelet coefficients in Θ_{dv}:

$$\overline{s}_{j_1,j_2} = \sup_{\theta \in \Theta_{tv}} |\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2} \rangle| \ .$$

According to (3.26) the thresholds are defined by

$$\overline{T}_{j_1,j_2} = \begin{cases} \overline{\sigma}_{j_1,j_2} \sqrt{2 \log_e 2^{-j_1-j_2}} & \text{if } \overline{\sigma}_{j_1,j_2} \sqrt{2 \log_e 2^{-j_1-j_2}} \le \overline{s}_{j_1,j_2} \\ \infty & \text{otherwise} \end{cases} \ . \tag{5.85}$$

For each scale 2^{j_1} we define a critical scale 2^{c_2} which is the smallest scale such that if $2^{j_2} > 2^{c_2}$ then $\overline{T}_{j_1,j_2} = \infty$. Similarly, for each scale 2^{j_2} we define a critical scale 2^{c_1} which is the smallest scale such that if $2^{j_1} > 2^{c_1}$ then $\overline{T}_{j_1,j_2} = \infty$. These critical scales, illustrated in Figure 4, correspond to cut-off frequencies beyond which the thresholding estimator sets all coefficients to zero. The following theorem proves that the resulting maximum thresholding risk $r_t(\Theta_{dv})$ is equivalent to the non-linear minimax risk and is smaller than the linear minimax risk. The proof is in Appendix I. Like in one dimension, the range of noise $N^{1/2} \le C/\sigma \le N^{p_1+p_2}$ is chosen so that the risk is not dominated by the noise at low frequencies but so that there is enough noise to insure that the problem is indeed unstable.

THEOREM 5. *Suppose that ψ^a is a compactly supported $\mathbf{C}^{q+1}$ wavelet with q vanishing moments, and that $q > \max(p_1, p_2)$. If $\min(p_1,p_2) \ge 1$ and $N^{1/2} \le C/\sigma \le$*

$N^{p_1+p_2}$ *then*

$$r_l(\Theta_{dv}) \sim N\, C^2 \left(\frac{\sigma^2\, N}{C^2}\right)^{1/(2p_1+2p_2-1)} \tag{5.86}$$

and

$$r_t(\Theta_{dv}) \sim r_n(\Theta_{dv}) \sim C^2\, N \left(\frac{\sigma^2}{C^2}\, \log_2 \frac{N^{p_1+p_2}\, \sigma}{C}\right)^{1/(2p_1+2p_2)} . \tag{5.87}$$

Observe that the computed risk $r_l(\Theta_{dv})$ and $r_n(\Theta_{dv})$ is much smaller than the minimax risk $r_n(\Theta_{tv}) \geq C\, N^2$ over a bounded variation set which has no directional condition. Like in one-dimension, since Θ_{dv} is translation invariant, the linear minimax risk is achieved with translation invariant operators, hence convolutions. A translation invariant thresholding algorithm is also implemented with a straightforward extension of the algorithm described in Section 4.4. With the fast filter bank algorithm, the thresholding estimation requires $O(N^2 \log_2 N)$ operations.

Figure 3(a) shows a small part of an airplane image, selected by the CNES for its tests because it includes edges, oscillatory structures and regions with a uniform grey level. The simulated satellite image is in Figure 3(b). It was calculated by the CNES with a low-pass filter having a zero of order $p_1 = p_2 = 1$ in both horizontal and vertical directions. The level of electronic noise in the satellite camera corresponds to $1 \leq \sigma \leq 2$ for image grey levels between 0 and 255. Figure 3(c) gives the results of a deconvolution estimator calculated with a thresholding in the mirror wavelet basis. This can be compared with the linear estimation in Figure 3(d), calculated with a circular convolution estimator, and optimized to minimize the risk over satellite images. This type of algorithm was benchmarked as superior to all competing algorithms by photo-interpreters of the French spatial agency (CNES) for the deconvolution of satellite images.

Acknowledgments: We would like to thank the anonymous reviewer whose detailed questions and suggestions improved considerably this paper.

A. Proof of Theorem 1 The proof considers first the particular case where K is diagonal. If K is diagonal in $\mathcal{B}$ then the coefficients $z_{\mathcal{B}}[m]$ are independent Gaussian random variables of variance σ_m^2. Estimating $\theta \in \Theta$ from $X = \theta + z$ is equivalent to estimating θ^0 from $X^0 = f^0 + z^0$ where

$$z^0 = \sum_{m=0}^{N-1} \frac{z_{\mathcal{B}}[m]}{\sigma_m}\, g_m \quad , \quad X^0 = \sum_{m=0}^{N-1} \frac{X_{\mathcal{B}}[m]}{\sigma_m}\, g_m \quad , \quad f^0 = \sum_{m=0}^{N-1} \frac{\theta_{\mathcal{B}}[m]}{\sigma_m}\, g_m \,. \tag{A.88}$$

The signal θ^0 belongs to an orthosymmetric set Θ^0 and the renormalized noise z^0 is a Gaussian white noise of variance 1. The oracle risk associated to the estimation problem $X^0 = f^0 + z^0$ is

$$r^0_{\inf,d}(\theta) = \sum_n \frac{|\theta^0_{\mathcal{B}}[m]|^2}{1 + |\theta^0_{\mathcal{B}}[m]|^2} \ .$$

Since Θ^0 is orthosymmetric in $\mathcal{B}$, and the noise is Gaussian white, Donoho et. al. [11] proves that

$$r^0_n(\Theta) \geq \frac{1}{1.25}\, r^0_{\inf,d}(\Theta) \ \text{ and } \ r^0_l(\Theta) = r^0_{\inf,d}(\mathrm{QH}[\Theta^0]). \tag{A.89}$$

Since $r_{\inf,d}(\theta) = \sum_m \frac{|\theta_{\mathcal{B}}[m]|^2 \sigma_m^2}{\sigma_m^2 + |\theta_{\mathcal{B}}[m]|^2}$, from (A.88) and (A.89), we derive with a rescaling that

$$r_n(\Theta) \geq \frac{1}{1.25}\, r_{\inf,d}(\Theta) \ \text{ and } \ r_l(\Theta) = r_{\inf,d}(\mathrm{QH}[\Theta]) \ . \tag{A.90}$$

The general case is proved using inequalities over symmetric matrices. If A and B are two symmetric matrices, we write $A \geq B$ if the eigenvalues of $A - B$ are positive, which means that $\langle Af, f\rangle \geq \langle Bf, f\rangle$ for all $\theta \in \mathbb{C}^N$. Since $\lambda_{\mathcal{B}}$ is the largest eigenvalue of $K_d^{1/2}\, K^{-1}\, K_d^{1/2}$, the inverse $\lambda_{\mathcal{B}}^{-1}$ is the smallest eigenvalue of the inverse $K_d^{-1/2}\, K\, K_d^{-1/2}$. It follows that $\langle K_d^{-1/2}\, K\, K_d^{-1/2} f, f\rangle \geq \lambda_{\mathcal{B}}^{-1}\, \langle \theta, f\rangle$. By setting $g = K_d^{-1/2}\theta$ we get $\langle Kg, g\rangle \geq \lambda_{\mathcal{B}}^{-1} \langle K_d^{1/2} g, K_d^{1/2} g\rangle$. Since this is valid for all $g \in \mathbb{C}^N$, we derive that

$$K \geq \lambda_{\mathcal{B}}^{-1} K_d \; . \tag{A.91}$$

Observe that $\lambda_{\mathcal{B}} \geq 1$ because $\langle K g_m, g_m\rangle = \langle K_d \, g_m \, , \, g_m\rangle$. Lower bounds for the minimax risks are proved as a consequence of the following lemma.

Lemma 1 Noise augmentation. *Consider the two estimation problems $X_i = \theta + z_i$ for $i = 1, 2$, where K_i is the covariance of the Gaussian noise z_i. We denote by $r_{i,n}(\Theta)$ and $r_{i,l}(\Theta)$ the non-linear and linear minimax risks for each estimation problem $i = 1, 2$. If $K_1 \geq K_2$ then*

$$r_{1,n}(\Theta) \geq r_{2,n}(\Theta) \quad \textit{and} \quad r_{1,l}(\Theta) \geq r_{2,l}(\Theta) \; . \tag{A.92}$$

Since $K_1 \geq K_2$ one can write $z_1 = z_2 + z_3$ where z_2 and z_3 are two independent Gaussian random vectors and the covariance of z_3 is $K_3 = K_1 - K_2 \geq 0$. Both estimation problems may thus be written $X = \theta + z_2 + z_3$ where z_3 is known, and equal to zero, if $i = 2$. The decision rule takes the form $\hat{\theta}_1 = \delta(X)$ and $\hat{\theta}_2 = \delta(X, z_3)$. The set of decision rules for $i = 1$ is a subset of the set of decision rules for $i = 2$ and hence the minimax risk cannot be smaller.

Since $K \geq \lambda_{\mathcal{B}}^{-1} K_d$, Lemma 1 proves that the estimation problem with the noise z of covariance K has a minimax risk that is larger than the minimax risk of the estimation problem with a noise of covariance $\lambda_{\mathcal{B}}^{-1} K_d$. But since this covariance is diagonal we can apply (A.90). The definition of $r_{\inf,d}(\Theta)$ is the same for a noise of covariance K and for a noise of covariance K_d because $\sigma_m^2 = \langle K g_m, g_m\rangle = \langle K_d g_m, g_m\rangle$. When multiplying K_d by a constant $\lambda_{\mathcal{B}}^{-1} \leq 1$, the value $r_{\inf,d}(\Theta)$ that appears in (A.90) is modified into $r'_{\inf,d}(\Theta)$ with $r'_{\inf,d}(\Theta) \geq \lambda_{\mathcal{B}}^{-1} \, r_{\inf,d}(\Theta)$. We thus derive (3.32) and (3.33).

B. Proof of (4.41) and (4.44) The Fourier transform of $b[n] = Gf(n/N) = f \star g(n/N)$ is

$$\widetilde{b}(\omega) = N \sum_{p=-\infty}^{+\infty} \widetilde{g}(N\,\omega + 2pN\pi)\, \widetilde{f}(N\,\omega + 2pN\pi)$$

$$= \widetilde{v}(\omega) \sum_{p=-\infty}^{+\infty} \widetilde{\phi}^*(N\,\omega + 2pN\pi)\, \widetilde{f}(N\,\omega + 2pN\pi) \tag{B.93}$$

with $\widetilde{v}(\omega)$ 2π periodic which satisfies

$$|\widetilde{v}(\omega)|^2 = \sum_{p=-\infty}^{+\infty} |\widetilde{g}(N\,\omega + 2pN\pi)|^2 \ .$$

and

$$\widetilde{\phi}(\omega) = \frac{\widetilde{g}^*(\omega)}{N^{-1}\,\widetilde{v}^*(N^{-1}\omega)} \ .$$

We verify that $\widetilde{v} \in \mathbf{L^2}[0, 2\pi]$ because $g \in \mathbf{L^2}(\mathbb{R})$ and hence $\widetilde{g} \in \mathbf{L^2}(\mathbb{R})$, so $v \in \mathbf{l^2}(\mathbb{Z})$. Since

$$N \sum_{p=-\infty}^{+\infty} |\widetilde{\phi}(\omega + 2pN\pi)|^2 = 1$$

one can derive [20] that $\{\phi(t - n/N)\}_{n\in\mathbb{Z}}$ is an orthonormal family. Computing the inverse Fourier transform or (B.93) proves (4.41).

If $\theta[n]$ is N periodic then

$$\theta \star v[n] = \sum_{p=-\infty}^{+\infty} \theta[p]\, v[n-p] = \sum_{p=0}^{N-1} \theta[p]\, u[n-p]$$

where $u[n] = \sum_{p=-\infty}^{+\infty} v[n-pN]$ is N periodic. Inserting this in (4.43) proves $\widetilde{u}[k] = \widetilde{v}(\frac{2k\pi}{N})$ which implies (4.44).

C. Proof of Proposition 1 Since $\theta = \sum_{j,m} \langle \theta, \psi_{j,m}\rangle\, \psi_{j,m}$ we have

$$\|\theta\|_{tv} \le \sum_{j=L+1}^{1} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m}\rangle|\, \|\psi_{j,m}\|_{tv} \ .$$

When N increases, we saw in (4.57) that $\psi_{j,m}[n]$ converges to $2^{L/2}\psi_j^{\rm per}(t - 2^j m)$ evaluated at $t = n/N$ so one can derive that there exists B_1 such that

$\|\psi_{j,m}\|_{tv} \leq B_1 2^{L/2}\, \|\psi_j^{\mathrm{per}}\|_{\overline{tv}}$. But $\psi_j^{\mathrm{per}}(t) = \sum_{k=-\infty}^{+\infty} \psi_j^a(t+k)$ so $\int_0^1 |\psi_j^{\mathrm{per}}(t)|' dt \leq \int_{-\infty}^{+\infty} |\psi_j^a(t)| dt = \|\psi_j^a\|_{\overline{tv}}$. Moreover $\psi_j^a(t) = 2^{-j/2}\psi^a(2^{-j}t)$ so $\|\psi_j^a\|_{\overline{tv}} = 2^{-j/2}\|\psi^a\|_{\overline{tv}}$. Hence, for any $N > 0$ and all for all j $\|\psi_{j,m}\|_{tv} \leq B_1\, \|\psi^a\|_{\overline{tv}} 2^{(L-j)/2}$, from which we derive (4.58).

To prove (4.59) we write $b[n] = \theta[n] - \theta[n-1]$. Since $\psi_j[n]$ has at least one vanishing moment it can be rewritten $\psi_{j,m}[n] = \phi_{j,m}[n] - \phi_{j,m}[n+1]$. Using the periodicity of θ and of $\psi_{j,m}[n] = \psi_j[n - 2^{L-j}m]$ we get $\langle \theta, \psi_{j,m}\rangle = \langle b, \phi_{j,m}\rangle$. Hence

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m}\rangle| = \sum_{m=0}^{2^{-j}-1} |\langle b, \phi_{j,m}\rangle| \leq \sum_{n=0}^{N-1} |b[n]| \sup_{0\leq n<N} \left| \sum_{m=0}^{2^{-j}-1} |\phi_{j,m}[n]| \right| .$$

Since $\psi_{j,m}$ has a support of size $A_1 2^{j-L}$ and is bounded by $A_2 2^{(L-j)/2}$ it results that

(C.94)

$$\sup_n |\phi_{j,m}[n]| \leq \|\psi_{j,m}\|_1 = \sum_n |\psi_{j,m}[n]| \leq A_1 2^{j-L}\, A_2 2^{(L-j)/2} = B_2\, 2^{(j-L)/2} .$$

But $\phi_{j,m}$ also has a support of size $A_1 2^{j-L}$ and $\|\theta\|_{tv} = \sum_{n=0}^{N-1} |b[n]|$ hence for all j

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m}\rangle| \leq \|\theta\|_{tv}\, A_1\, A_2\, 2^{(j-L)/2} ,$$

which proves (4.59).

D. Proof of (4.61) Since $|\widetilde{u}[k]| \sim |2|k|/N - 1|^p$ we derive from (4.60) that

$$\sigma_{L+1,m}^2 = \sigma_{L+1}^2 \sim \sigma^2\, N^{2p+1} .$$

because $|\widetilde{\psi}_{L+1}[N/2]| = \sqrt{2}$ and $\sqrt{2} \geq |\widetilde{\psi}_{L+1}[k]| \geq 1$ for $|k| \in [N/4, N/2]$.

If $\theta \in \Theta_{tv}$ then (4.59) proves that

$$A^{-1}\, C \geq \sum_{m=0}^{2^{-L-1}-1} 2^{-1/2}\, |\langle \theta, \psi_{L+1,m}\rangle|$$

so $|\langle \theta, \psi_{L+1,m}\rangle| \leq C\, 2^{1/2}\, A^{-1}$. If $C/\sigma^2 \leq N^{2p+1}$ then for N large the maximum amplitude of wavelet coefficients $|\langle \theta, \psi_{L+1,m}\rangle|$ is below $\sigma_{L+1}\sqrt{2\log N}$ so according to (3.23) the thresholding sets all coefficients to zero. The deconvolution thus do not recover any high frequency data. The discrete total variation of the finest scale wavelet $\|\psi_{L+1}\|_{tv} = \beta$ is a constant independent of N. Hence $\theta = (C/\beta)\,\psi_{L+1} \in \Theta_{tv}$ and since the thresholding estimator sets to zero all finest scale coefficients the resulting thresholding risk satisfies

$$r_t(\Theta_{tv}) \geq r_t(\theta) \geq \frac{C^2}{\beta^2}$$

which proves (4.61) for $B = 1/\beta^2$.

E. Proof of Theorem 2 The operator $K = \sigma^2\, \overline{U}^{-1*}\, \overline{U}^{-1}$ is a circular convolution whose eigenvalues are $\sigma^2\, |\widetilde{\overline{u}}^{-1}[k]|^2 \sim \sigma^2\, |2|k|/N - 1|^{-2p}$. We study separately the restriction of K over the space $\mathbf{V}_{L+1}$ generated by the $N/2$ lower frequency wavelets $\{\psi_{j,m}\}_{j>L+1,m}$ and over the orthogonal complement $\mathbf{W}_{L+1}$ generated by the $N/2$ higher frequency mirror wavelets $\{\overline{\psi}_{j,m}\}_{j>L+1,m}$.

Over $\mathbf{V}_{L+1}$ we show that there exists $\lambda_1 > 0$ and λ_2 such that

$$\lambda_1\, \sigma^2\, Id \leq K \leq \lambda_2\, \sigma^2\, Id\ . \tag{E.95}$$

There exists $\lambda_1 > 0$ such that the eigenvalues of K are above $\lambda_1\, \sigma^2$ which proves the first inequality. The second inequality means that over $\mathbf{V}_{L+1}$ the norm of K satisfies $\|K\|_s = O(\sigma^2)$. The theorem supposes that the asymptotic wavelet $\psi^a(t)$ which defines the discrete wavelet basis has $q > p$ vanishing moments. So $\widetilde{h}^a(\omega + \pi) = O(|\omega|^q)$ and hence $\widetilde{h}[k] = \widetilde{h}^a(2\pi k/N) = O(|2|k|/N - 1|^q)$. It results that $|\widetilde{h}[k]\, \widetilde{\overline{u}}^{-1}[k]| = O(1)$. For $j > L+1$ we derive from (4.54) that

$$\widetilde{\psi}_j[k] = \widetilde{h}[k]\, \widetilde{\xi}_j[k] \tag{E.96}$$

with

$$\widetilde{\xi}_j[k] = \widetilde{g}[2^{j-L-1}k] \prod_{l=1}^{j-L-1} \widetilde{h}[2^l k]\ . \tag{E.97}$$

If $\xi_{j,m}[n] = \xi_j[n - 2^{j-L}m]$ then $\xi_{j,m}[n] = \psi_{j,m} \star h[n]$. From (E.97) and the fact that $h[n]$ and $g[n]$ are conjugate mirror filters, we know that $\{\xi_{j,m}\}_{j,m}$ is an orthonormal family [20]. Let R be the convolution operator defined by $R\theta = K\theta \star h \star h^*$ with $h^*[n] = h[-n]$. Its eigenvalues are $\sigma^2 |\widetilde{\overline{u}}^{-1}[k]\, \widetilde{h}[k]|^2 = O(\sigma^2)$ so $\|R\|_s = O(\sigma^2)$. Moreover,

$$\langle R\xi_{j,m}\, ,\, \xi_{j',m'} \rangle = \langle K\psi_{j,m}\, ,\, \psi_{j',m'} \rangle \tag{E.98}$$

so the norm of K over the space $\mathbf{V}_{L+1}$ generated by $\{\psi_{j,m}\}_{j>L+1,m}$ is equal to the norm of R over the space generated by $\{\xi_{j,m}\}_{j>L+1,m}$ and it is therefore $O(\sigma^2)$. It proves that there exists λ_2 such that $\lambda_1 \sigma^2 Id \leq K \leq \lambda_2 \sigma^2 Id$ over $\mathbf{V}_{L+1}$. As a result, for any $j > L+1$ the diagonal entries of K satisfy

$$\langle K\psi_{j,m}, \psi_{j,m} \rangle \sim \sigma^2 \ , \tag{E.99}$$

and there exists λ_3 such that $\|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s \leq \lambda_3$ over $\mathbf{V}_{L+1}$.

Over the space $\mathbf{W}_{L+1}$ generated by the high frequency mirror wavelets $\{\overline{\psi}_{j,m}\}_{j>L+1,m}$, the proof proceeds in two steps. It first shows that the matrix of K^{-1} in the mirror wavelet basis of $\mathbf{W}_{L+1}$ is identical to the matrix of a circular convolution operator K_1 over low frequency wavelets in $\mathbb{C}^N$. To obtain an asymptotic result as N increases, this matrix is related to the restriction in an N dimensional space of a homogeneous convolution operator K_2 in $\mathbf{L^2}(\mathbb{R})$.

Let us first verify that for all $0 \geq j > L+1$ and $0 \leq m < 2^{-j}$

$$\langle K^{-1}\overline{\psi}_{j,m}, \overline{\psi}_{j',m'} \rangle = \langle K_1 \psi_{j,m}, \psi_{j',m'} \rangle \tag{E.100}$$

where

$$K_1\theta[n] = \theta \star c_1[n] \ \text{ with } \ \widetilde{c_1}[k] = \sigma^{-2}\, |\widetilde{u}[k + N/2]|^2 \ . \tag{E.101}$$

To verify (E.100), we use the relation (4.63) between wavelets and mirror wavelets

$$\overline{\psi}_{j,m}[n] = (-1)^{n-1}\, \psi_{j,m}[1-n]\ .$$

The operator K^{-1} is a convolution with a real and symmetric kernel $c^{-1}[n]$ so $\langle K^{-1}\overline{\psi}_{j,m}, \overline{\psi}_{j',m'}\rangle$ is not modified if $\overline{\psi}_{j,m}$ and $\overline{\psi}_{j',m'}$ are transformed by a symmetry and a translation. Multiplying these vectors by $(-1)^n$ translates their discrete Fourier transform by $N/2$. This is compensated by translating the Fourier transform $\widetilde{c}^{-1}[k] = \sigma^{-2}|\widetilde{u}[k]|^2$ of $c^{-1}[n]$ by $N/2$, which proves (E.101). Since $|\widetilde{u}[k]| \sim |1-2|k|/N|^p$ it implies that $|\widetilde{c}_1[k]| \sim \sigma^{-2}|2k/N|^{2p}$.

We are now going to relate K_1 to a continuous variable convolution operator K_2 defined by $K_2 f(t) = f \star c_2(t)$ with $\widetilde{c}_2(\omega) = |\omega|^{2p}$. Since ψ^a is a compactly supported wavelet which is $\mathbf{C}^{q+1}$ and which has $q > p$ vanishing moments, we know that K_2 is preconditioned by its diagonal in the wavelet basis of $\mathbf{L^2}(\mathbb{R})$ generated by ψ^a, and this results remains valid in the corresponding periodic wavelet basis of $\mathbf{L^2}[0,1]$. Let $K_{2,d}^{-1/2}$ be the diagonal operator in the periodic wavelet basis, whose diagonal entries are the inverse square root of the non-zero diagonal entries of K_2 and are equal to zero otherwise. This means that exists λ_4 such that over $\mathbf{L^2}[0,1]$

$$\|K_{2,d}^{-1/2}\, K_2\, K_{2,d}^{-1/2}\|_s \leq \lambda_4\ . \tag{E.102}$$

Moreover the diagonal entries satisfy for $j \leq 0$ [21]

$$\langle K_2 \psi^{\rm per}_{j,m}\, ,\, \psi^{\rm per}_{j,m}\rangle \sim 2^{-2pj}\ . \tag{E.103}$$

Let us show that the matrix of K_2 on the subspace $\mathbf{V}^{\rm per}_{L+1}$ of $\mathbf{L^2}[0,1]$ generated by the periodic wavelets $\{\psi^{\rm per}_{j,m}\}_{0\geq j>L+1, 0\leq m<2^{-j}}$ is equal to the matrix of a discrete circular convolution operator in the discrete wavelet basis $\{\psi_{j,m}\}_{0\geq j>L+1,0\leq m<2^{-j}}$ of $\mathbf{V}_{L+1}$. The discrete circular convolution operator is computed in two steps. First we find the corresponding discrete convolution operator and then its periodized version.

We saw in (4.56) that

$$\psi_j^{\mathrm{per}}(t) = \sum_{n=-\infty}^{+\infty} \psi_j[n]\, \phi_L^a(t - n/N)\ . \tag{E.104}$$

Let $c_3[n]$ be the discrete kernel obtained by sampling the continuous variable convolutions:

$$c_3[n] = c_2 \star \phi_L^a \star \phi_L^{a*}(n/N) \tag{E.105}$$

with $\phi_L^{a*}(t) = \phi_L^{a*}(-t)$. Let $K_3\theta[n] = c_3 \star \theta[n]$ be the corresponding discrete convolution operator. One can derive from (E.104) that for all j, j', m, m'

$$\langle K_2 \psi_{j,m}^{\mathrm{per}}\, ,\, \psi_{j',m'}^{\mathrm{per}} \rangle = \langle K_3 \psi_{j,m}\, ,\, \psi_{j',m'} \rangle\ . \tag{E.106}$$

Moreover, (E.105) shows that Fourier transform of c_3 is

$$\widetilde{c_3}(\omega) = N \sum_{l=-\infty}^{+\infty} \widetilde{c}_2(N\omega + 2Nl\pi)\, |\widetilde{\phi}_L^a(N\omega + 2Nl\pi)|^2\ .$$

Since $\phi_L^a(t) = N^{1/2}\phi^a(Nt)$ we have $\widetilde{\phi}_L^a(\omega) = N^{-1/2}\widetilde{\phi}^a(N^{-1}\omega)$ and since $\widetilde{c}_2(\omega) = |\omega|^{2p}$

$$\widetilde{c}_3(\omega) = N^{2p} \sum_{l=-\infty}^{+\infty} (\omega + 2l\pi)^{2p}\, |\widetilde{\phi}^a(\omega + 2l\pi)|^2\ . \tag{E.107}$$

Because ψ^a is a $\mathbf{C}^{q+1}$ compactly supported function with q vanishing moments, we know [5] that for any $l \neq 0$

$$\forall \omega \in [-\pi, \pi]\ \ ,\ \ |\phi^a(\omega + 2l\pi)| = O(|\omega|^q)\, O(|l|^{-q-1})\ .$$

Since $|\widetilde{h}^a(\omega)| \geq 1/2$ for $|\omega| \leq \pi/2$ there exists $\eta > 0$ such that $1 \geq |\phi^a(\omega)|^2 \geq \eta$ for $|\omega| < \pi$. Since $q > p$, (E.107) gives

$$\forall \omega \in [-\pi, \pi]\ \ ,\ \ |\widetilde{c}_3(\omega)| \sim N^{2p}\, |\omega|^{2p}\, .$$

Let us now consider the circular convolution operator $K_4\theta[n] = c_4 \star \theta[n]$ where $c_4[n] = \sum_{l=-\infty}^{+\infty} c_3[n + lN]$. If θ is N periodic then $K_3\theta = K_4\theta$ so (E.106) implies that

$$\langle K_2 \psi^{\rm per}_{j,m} \, , \, \psi^{\rm per}_{j',m'} \rangle = \langle K_4 \psi_{j,m} \, , \, \psi_{j',m'} \rangle \; . \tag{E.108}$$

The discrete Fourier transform of c_4 is

$$\widetilde{c}_4[k] = \widetilde{c}_3(2k\pi/N) \sim |k|^{2p} \sim \sigma^2 \, N^{2p} \, \widetilde{c}_1[k] \; .$$

This means that there exists a convolution operator K_5 with constants $A > 0$ and B such that $AId \leq K_5 \leq BId$ such that $K_5 \, K_1 = \sigma^{-2} \, N^{-2p} \, K_4$. Since all convolution operators commute, (E.103) and (E.108) imply that

$$\langle K_1 \psi_{j,m} \, , \, \psi_{j,m} \rangle \sim \sigma^{-2} \, N^{-2p} \, 2^{-2pj} = \sigma^{-2} \, 2^{2p(L-j)} \; .$$

With (E.100) it yields

$$\langle K^{-1} \overline{\psi}_{j,m}, \overline{\psi}_{j',m'} \rangle \sim \sigma^{-2} \, 2^{2p(L-j)} \; . \tag{E.109}$$

Moreover (E.108) proves that the operator $\sigma^2 \, N^{2p} \, K_5 \, K_1$ is represented by a matrix in a discrete wavelet basis of $\mathbf{V}_{L+1}$ which is the same as the matrix of K_2 in the continuous variable periodic wavelet basis of $\mathbf{V}^{\rm per}_{L+1}$. We thus derive from (E.102) that over $\mathbf{V}_{L+1}$

$$\| K_{1,d}^{-1/2} \, K_1 \, K_{1,d}^{-1/2} \|_s \leq \lambda_4 \, \frac{B}{A} \; .$$

Finally (E.100) shows that K^{-1} is represented by a matrix in a mirror wavelet basis of $\mathbf{W}_{L+1}$ which is the same as the matrix of K_1 in a wavelet basis of $\mathbf{V}_{L+1}$ so over $\mathbf{W}_{L+1}$

$$\| K_d^{1/2} \, K^{-1} \, K_d^{1/2} \|_s \leq \lambda_4 \, \frac{B}{A} \; ,$$

which finishes the proof.

F. Proof of Proposition 2 Proposition 2 is proved by verifying that there exists $A_2, B_2 > 0$ such that

$$\text{(F.110)} \qquad \sum_{m=0}^{2^{-L-1}-1} |\langle \theta, \psi_{L+1,m} \rangle| \leq B_2 \sum_{j=L+2}^{1} 2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle| \ ,$$

and

$$\text{(F.111)} \qquad \sum_{m=0}^{2^{-L-1}-1} |\langle \theta, \psi_{L+1,m} \rangle| \geq A_2 \sup_{1 \geq j \geq L+2} \left(2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle| \right) \ .$$

Inserting these inequalities in Proposition 1 implies (4.68) and (4.69). These inequalities are derived from upper and lower bounds for the $\mathbf{l^1}$ norm $\|\theta\|_1 = \sum_{n=0}^{N-1} |\theta[n]|$ provided by the following lemma.

LEMMA 2. *There exists $A_1, B_1 > 0$ such that*

$$\text{(F.112)} \qquad \|\theta\|_1 \leq B_1 \sum_{j=L+1}^{1} 2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m} \rangle| \ ,$$

and

$$\text{(F.113)} \qquad \|\theta\|_1 \geq A_1 \sup_{1 \geq j \geq L+1} \left(2^{(j-L)/2} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m} \rangle| \right) \ .$$

This lemma is proved with similar calculations as in the proof of Proposition 1, by using the fact that $\|\psi_{j,m}\|_1 = O(2^{(j-L)/2})$ which is proved in (C.94). Let us now consider $b[p] = \langle \theta, \psi_{L+1,p} \rangle$ and $\psi^*_{j,m}[n] = \psi_{j,m}[-n]$. As a consequence of the Fourier transform expressions (4.54) and (4.62) we verify that

$$\overline{\psi}_{j,m}[n] = \sum_{p=0}^{N/2-1} \psi^*_{j-1,m}[p]\, \psi_{L+1}[n-2p] \ .$$

It results that

$$\forall L+2 \leq j \leq 1 \quad \langle \theta, \overline{\psi}_{j,m} \rangle = \langle b, \psi^*_{j-1,m} \rangle = \langle b^*, \psi_{j-1,m} \rangle$$

with $b^*[n] = b[-n]$. Lemma 2 applied to b^* gives (F.110) and (F.111), which finishes the proof.

G. Proof of Theorem 3 The proof constructs two sets Θ_1 and Θ_2 that are orthosymmetric in the mirror wavelet basis and such that $\Theta_1 \subset \Theta_{tv} \subset \Theta_2$. Clearly

$$r_n(\Theta_1) \le r_n(\Theta_{tv}) \le r_t(\Theta_{tv}) \le r_t(\Theta_2) \ , \tag{G.114}$$

and equation (3.34) together with Theorem 2 proves that there exists $\lambda > 0$ such that

$$\frac{1}{\lambda}\, r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_1]) \le r_l(\Theta_1) \le r_l(\Theta_{tv}) \le r_l(\Theta_2) \le r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2]) \ . \tag{G.115}$$

The set Θ_2 is derived from Proposition 2. If $\theta \in \Theta_{tv}$ then $\|\theta\|_{tv} \le C$ and (4.69) proves that it belongs to the set Θ_2 of all θ such that for $1 \ge j > L+1$

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m} \rangle| \le A^{-1}\, C\, 2^{(j-L)/2} \quad , \tag{G.116}$$

and

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle| \le A^{-1}\, C\, 2^{(L-j)/2} \ . \tag{G.117}$$

So $\Theta_{tv} \subset \Theta_2$ and Θ_2 is orthosymmetric in the mirror wavelet basis.

Let us now compute an upper bound of $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2])$. The quadratic envelope of an $\mathbf{l^1}$ ball of sequences $a[m]$ such that $\sum_{m=0}^{M-1} |a[m]| \le B$ is the $\mathbf{l^2}$ ball defined by $\sum_{m=0}^{M-1} |a[m]|^2 \le B^2$. It results from (G.116) and (G.117) that $\mathrm{QH}[\Theta_2]$ is the set of θ such that

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m} \rangle|^2 \le A^{-2}\, C^2\, 2^{j-L} \quad , \tag{G.118}$$

and

$$\sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle|^2 \le A^{-2}\, C^2\, 2^{L-j} \ . \tag{G.119}$$

Let us compute $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2])$. The general expression of $r_{\mathrm{inf},d}(\theta)$ in a basis $\mathcal{B}$ is given in (3.14). Using the fact that $ab/(a+b) \le \min(a,b)$ for $a, b > 0$, in the mirror wavelet basis we get

$$r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2]) \leq \sup_{\theta \in \mathrm{QH}[\Theta_2]} \sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min(|\langle \theta, \psi_{j,m}\rangle|^2, \sigma_j^2)$$

$$+ \sup_{\theta \in \mathrm{QH}[\Theta_2]} \sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min(|\langle \theta, \overline{\psi}_{j,m}\rangle|^2, \overline{\sigma}_j^2) = I + II \ .$$

Since $\sigma_j \sim \sigma$

$$I \sim \sup_{\theta \in \mathrm{QH}[\Theta_2]} \sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min(|\langle \theta, \psi_{j,m}\rangle|^2, \sigma^2) \ . \tag{G.120}$$

For a fixed σ^2 the function $\min(a, \sigma^2)$ is a concave function of a. Using this concavity we verify that

$$\sup_{\sum_{m=0}^{M-1} |a[m]|^2 \leq B^2} \sum_{m=0}^{M-1} \min(|a[m]|^2, \sigma^2) = \min(B^2, M\sigma^2) \ . \tag{G.121}$$

Applying (G.121) to (G.120) for $B^2 = A^{-2}\, C^2\, 2^{j-L}$ gives

$$I \sim \sum_{j=L+2}^{1} \min(A^{-2}\, C^2\, 2^{j-L} \ , \ 2^{-j}\sigma^2) \ .$$

Since $C/\sigma \geq N^{1/2}$ it results that $I \sim N\sigma^2$. The second sum is computed similarly, with $\overline{\sigma}_j \sim \sigma\, 2^{p(j-L)}$:

$$II \sim \sup_{\theta \in \mathrm{QH}[\Theta_2]} \sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min\Big(|\langle \theta, \overline{\psi}_{j,m}\rangle|^2, \sigma^2 2^{2p(j-L)}\Big) \tag{G.122}$$

$$\sim \sum_{j=L+2}^{1} \min(C^2\, 2^{L-j} \ , \ 2^{-j} 2^{2p(j-L)}\, \sigma^2) \ . \tag{G.123}$$

Let 2^l be the critical scale of this linear estimation, where l is the closest integer to y such that

$$C^2\, 2^{L-y} = 2^{-y} 2^{2p(y-L)}\, \sigma^2 \ . \tag{G.124}$$

Since $N^{1/2} \leq C/\sigma \leq N^{p+1/2}$ we verify that $1 \geq 2^l \geq 2^L$. A direct calculation shows that

$$II \sim C^2\, 2^{L-l} \sim C^2 \, \left(\frac{\sigma\, N^{1/2}}{C}\right)^{1/p} \ .$$

Since $C/\sigma \geq N^{1/2}$, the sum I is smaller than II and hence

$$r_{\inf,d}(\mathrm{QH}[\Theta_2]) \sim C^2 \left(\frac{\sigma\, N^{1/2}}{C}\right)^{1/p} . \tag{G.125}$$

To compute a lower bound of $r_l(\Theta_{tv})$ we now construct an orthosymmetric set $\Theta_1 \subset \Theta_{tv}$. Let $\overline{\mathbf{W}}_l$ be the space generated by the mirror wavelets $\{\overline{\psi}_{l,m}\}_{0\leq m<2^{-l}}$ at the critical scale 2^l defined by (G.124). We define Θ_1 as the set of all $\theta \in \overline{\mathbf{W}}_l$ such that

$$\sum_{m=0}^{2^{-l}-1} |\langle \theta, \overline{\psi}_{l,m}\rangle| \leq B^{-1}\, C\, 2^{(L-l)/2} .$$

Since $\Theta_1 \subset \overline{\mathbf{W}}_l$ if $\theta \in \Theta_1$ then $\langle \theta, \overline{\psi}_{j,m}\rangle = 0$ if $j \neq l$. Proposition 2 proves in (4.68) that there exists an appropriate B such that $\Theta_1 \subset \Theta_{tv}$. The quadratic envelop $\mathrm{QH}[\Theta_1]$ is the set of all $\theta \in \overline{\mathbf{W}}_l$ such that

$$\sum_{m=0}^{2^{-l}-1} |\langle \theta, \overline{\psi}_{l,m}\rangle|^2 \leq B^{-2}\, C^2\, 2^{L-l} ,$$

and

$$r_{\inf,d}(\mathrm{QH}[\Theta_1]) \sim \sup_{\theta\in \mathrm{QH}[\Theta_1]} \sum_{m=0}^{2^{-l}-1} \min\Big(|\langle \theta, \overline{\psi}_{j,m}\rangle|^2, \sigma^2 2^{2p(l-L)}\Big) \tag{G.126}$$

$$\sim \min\Big(C^2\, 2^{L-l}, 2^{-l} 2^{2p(l-L)}\, \sigma^2 \Big) \sim C^2 \left(\frac{\sigma\, N^{1/2}}{C}\right)^{1/p} . \tag{G.127}$$

It results from (G.115) and (G.125) that $r_l(\Theta_{tv}) \sim C^2\, ((\sigma\, N^{1/2})/C)^{1/p}$, which proves (4.72).

Let us now compute the thresholding risk $r_t(\Theta_2)$. The thresholds T_j for the wavelets $\psi_{j,m}$ and $\overline{T}_j$ for the mirror wavelets $\overline{\psi}_{j,m}$ are defined by (4.70) and (4.71). We derive from (4.69) in Proposition 2 that

$$\overline{s}_j = \sup_{\theta\in\Theta_{tv}} |\langle \theta, \overline{\psi}_{j,m}\rangle| \sim C\, 2^{(L-j)/2} .$$

Let 2^c be the critical scale which is the smallest scale such that $\overline{T}_j = \infty$ and hence $\overline{\sigma}_j \sqrt{2 \log_e 2^{-j}} > \overline{s}_j$ for $2^j > 2^c$. Since $\overline{s}_j \sim C\, 2^{(L-j)/2}$ which decreases when

2^j increases whereas $\overline{\sigma}_j \sqrt{2\log_e 2^{-j}} \sim \sigma\, 2^{p(j-L)} \sqrt{2\log_e 2^{-j}}$ which increases when when 2^j increases, it results that

$$\overline{\sigma}_c \sqrt{2\log_e 2^{-c}} \sim \overline{s}_c \tag{G.128}$$

and hence

$$C\, 2^{(L-c)/2} \sim \sigma\, 2^{p(c-L)} \sqrt{\log_e 2^{-c}} \tag{G.129}$$

which implies that

$$2^{L-c} \sim \left(\frac{\sigma}{C}\right)^{1/(p+1/2)} \left(\log_2 \frac{\sigma\, N^{p+1/2}}{C}\right)^{1/(2p+1)} . \tag{G.130}$$

We have $2^L \leq 2^c \leq 1$ because $N^{1/2} \leq C/\sigma \leq N^{p+1/2}$. An upper bound of $r_t(\Theta_2)$ is calculated with (3.28):

$$r_t(\Theta_2) \leq 2\, \log_e(N/2) \sup_{L+2 \leq j \leq 1} \sigma_j^2 + \sum_{j=L+2}^{c} (1 + 2\log_e 2^{-j}) \overline{\sigma}_j^2 \tag{G.131}$$

$$+ \sup_{\theta \in \Theta_2} \left(\sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min\Big(|\langle \theta, \psi_{j,m} \rangle|^2, \sigma_j^2 (2\log_e N + 1) \Big) \right) \tag{G.132}$$

$$+ \sup_{\theta \in \Theta_2} \left(\sum_{j=L+2}^{c} \sum_{m=0}^{2^{-j}-1} \min\Big(|\langle \theta, \overline{\psi}_{j,m} \rangle|^2, \overline{\sigma}_j^2 (2\log_e 2^{-j} + 1) \Big) \right) \tag{G.133}$$

$$+ \sup_{\theta \in \Theta_2} \left(\sum_{j=c+1}^{1} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle|^2 \right) , \tag{G.134}$$

which we write $r_t(\Theta_2) \leq I + II + III + IV$. Since $\sigma_j \sim \sigma$ and $\overline{\sigma}_j \sim \sigma\, 2^{p(j-L)}$, the calculation of I in (G.131) gives

$$I \sim 2\log_e N\, \sigma^2 + |c|\, \sigma^2\, 2^{2p(c-L)} \sim |c|\, \sigma^2\, 2^{2p(c-L)} \sim C^2 \left(\frac{\sigma^2}{C^2} \log_2 \frac{\sigma\, N^{p+1/2}}{C} \right)^{1/(2p+1)} . \tag{G.135}$$

The second sum in (G.132) satisfies

$$II \sim \sup_{\theta \in \Theta_2} \left(\sum_{j=L+2}^{1} \sum_{m=0}^{2^{-j}-1} \min\Big(|\langle \theta, \psi_{j,m} \rangle|^2, \log_e N\, \sigma^2 \Big) \right) \tag{G.136}$$

and we saw in (G.116) that $\theta \in \Theta_2$ if

$$\forall\ L+2 \le j \le 1 \quad \sum_{m=0}^{2^{-j}-1} |\langle \theta, \psi_{j,m} \rangle| \le A^{-1}\, C\, 2^{(j-L)/2} \ . \tag{G.137}$$

One can verify that

$$\sup_{\sum_{m=0}^{M-1} |a_m| \le A} \sum_{m=0}^{M-1} \min\Big(|a_m|^2, \beta^2\Big) \sim \min\Big(M\beta^2, A\beta, A^2\Big) \ . \tag{G.138}$$

Applying (G.138) to each scale 2^j of (G.136) for $M = 2^{-j}$, $a_m = \langle \theta, \psi_{j,m} \rangle$, $A = C\, 2^{(j-L)/2}$ and $\beta^2 = \sigma^2 \log_e N$, and using the fact that $C/\sigma \ge 1$, we get

$$\begin{aligned} II &\sim \sum_{j=L+2}^{1} \min\Big(2^{-j}\sigma^2 \log_e N \ , \ C\, 2^{(j-L)/2}\sigma \sqrt{\log_e N}\Big) \\ &\sim \sigma^2 \log_e N \sum_{j=L+2}^{1} \min\Big(2^{-j} \ , \ \frac{C\, N^{1/2}\, 2^{j/2}}{\sigma \sqrt{\log_e N}}\Big) \ . \end{aligned}$$

Setting $a = C\, N^{1/2} / (\sigma \sqrt{\log_e N})$ we get

$$\frac{II}{\sigma^2 \log_e N} \sim \max_{j \le 1} \min(2^{-j} \ , \ a\, 2^{j/2}) \sim a^{3/2}$$

and hence

$$II \sim \log_e N\, \sigma^2 \left(\frac{C^2 N}{\sigma^2 \log_e N} \right)^{1/3} = C^2 \left(\frac{\sigma^4\, (\log_e N)^2 N}{C^4} \right)^{1/3} \ . \tag{G.139}$$

The third sum in (G.133) satisfies

$$III \sim \sup_{\theta \in \Theta_2} \left(\sum_{j=L+2}^{c} \sum_{m=0}^{2^{-j}-1} \min\Big(|\langle \theta, \overline{\psi}_{j,m} \rangle|^2 \ , \ 2^{2p(j-L)}\, |j|\, \sigma^2\Big) \right)$$

and we saw in (G.116) that $\theta \in \Theta_2$ if

$$\forall\ L+2 \le j \le 1 \quad \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m} \rangle| \le A^{-1}\, C\, 2^{(L-j)/2} \ .$$

Let us apply (G.138) to each scale 2^j for $M = 2^{-j}$, $a_m = \langle \theta, \overline{\psi}_{j,m} \rangle$, $A = C\, 2^{(L-j)/2}$ and $\beta^2 = \sigma^2\, |j|\, 2^{2p(j-L)}$. The property (G.129) of the critical scale 2^c implies that if $L+2 \le j \le c$ then A multiplied by a constant factor is larger than β so $\min(M\beta^2, A\beta, A^2) \sim \min(M\beta^2, A\beta)$. At the critical 2^c we have $A \sim \beta$. Writing the sum we obtain a geometric series whose value is dominated by the final term at $j = c$ where $\beta \sim A \sim C\,2^{(L-c)/2}$. It results that

$$\text{(G.140)} \qquad III \sim C^2\, 2^{L-c} \sim C^2 \left(\frac{\sigma}{C}\right)^{1/(p+1/2)} \left(\log_2 \frac{\sigma\, N^{p+1/2}}{C}\right)^{1/(2p+1)} .$$

Finally we compute the upper bound

$$\text{(G.141)} \quad IV \sim \sup_{\theta\in\Theta_2} \sum_{j=c+1}^{1} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m}\rangle|^2 \le A^{-1} \sum_{j=c+1}^{1} C^2\, 2^{L-j} \le A^{-1} C^2\, 2^{L-c} .$$

Moreover, (G.117) shows that the signal proportional to the mirror wavelet $\theta = 2^{(L-c-1)/2} A^{-1} C\, \overline{\psi}_{c+1,0}$ belongs to Θ_2. Hence

$$IV \sim \sup_{\theta\in\Theta_2} \sum_{j=c+1}^{1} \sum_{m=0}^{2^{-j}-1} |\langle \theta, \overline{\psi}_{j,m}\rangle|^2 \ge 2^{-2}\, A^{-1} C^2\, 2^{L-c}$$

so $IV \sim C^2\, 2^{L-c}$.

Since $p \ge 1$ and $C/\sigma \ge N^{1/2}$, (G.139) and (G.140) imply that II is smaller than III and hence that

$$\text{(G.142)} \qquad r_t(\Theta_2) \le I + II + III + IV \sim C^2 \left(\frac{\sigma^2}{C^2} \log_2 \frac{\sigma\, N^{p+1/2}}{C}\right)^{1/(2p+1)} .$$

We now compute a lower bound of $r_n(\Theta_{tv})$ by finding an appropriate orthosymmetric set $\Theta_1 \subset \Theta_{tv}$ and by computing $r_n(\Theta_1)$. Let $\overline{\mathbf{W}}_c$ be the space generated by the 2^{-c} mirror wavelets $\{\overline{\psi}_{c,m}\}_{0\le m<2^{-c}}$ at the critical scale 2^c which satisfies (G.130). We define Θ_1 as the set of all signals $\theta \in \overline{\mathbf{W}}_c$ such that

$$\text{(G.143)} \qquad \sum_{m=0}^{2^{-c}-1} |\langle \theta, \overline{\psi}_{c,m}\rangle| \le B^{-1}\, C\, 2^{(L-c)/2} .$$

Since $\Theta_1 \subset \overline{\mathbf{W}}_c$ if $\theta \in \Theta_1$ then $\langle \theta, \overline{\psi}_{j,m}\rangle = 0$ if $j \ne c$. The upper bound (4.68) in Proposition 2 proves that there exists a constant B such that $\Theta_1 \subset \Theta_{tv}$. We need a lower bound of the minimax risk $r_n(\Theta_1)$ when estimating $\theta \in \Theta_1$ from $X = \theta + z$.

Theorem 2 proves that there exists λ such that over $\overline{\mathbf{W}}_c$ the covariance K of z satisfies $K \ge \lambda\, \overline{\sigma}_c^2\, Id$. Since $\Theta_1 \subset \overline{\mathbf{W}}_c$, the noise augmentation lemma 1 implies that $r_n(\Theta_1) \ge r_n^0(\Theta_1)$ where $\overline{r}_n^0(\Theta_1)$ is the minimax risk to estimate $\theta \in \Theta_1$ from

$X = \theta + z^0$, where z^0 is a Gaussian white noise of variance $\lambda\,\overline{\sigma}_c^2$ in the space $\overline{\mathbf{W}}_c$. Let $\delta[m]$ be a discrete Dirac. A lower bound of this minimax risk is calculated from the Bayes risk corresponding to a random vector $F \in \overline{\mathbf{W}}_c$ such that

$$\langle F, \overline{\psi}_{c,m}\rangle = A\,C\,2^{(L-c)/2}\,\delta[m-P] \tag{G.144}$$

where P is a random variable uniformly distributed over $\{0,..,2^{-c}-1\}$. The following lemma proved by Donoho et. al. [10] computes a lower bound of the Bayes risk $r(\pi)$ to estimate F from $X = F + z^0$.

Lemma 3. *Let P be a random variable uniformly distributed over $\{0,..,M-1\}$ and $G[m] = K\,\delta[m-P]$ for $0 \leq m < M$. Let $\nu[m]$ be a Gaussian white noise of variance β^2. If $K \leq \beta\,\sqrt{\log_e M}$ then the Bayes estimator $\hat{G}$ of G calculated from $Y = G + \nu$ has a risk which satisfies for $M \geq 4$*

$$r(\pi) = \mathsf{E}\{\|G - \hat{G}\|^2\} \geq K^2/2\ . \tag{G.145}$$

This lemma is applied to the process F with $A \leq B^{-1}$ sufficiently small so that

$$A^2\,C^2\,2^{L-c} \leq \lambda^2\,\overline{\sigma}_c\,\log_e 2^{-c}\ .$$

The lower bound (G.145) proves that

$$r_n(\Theta_1) \geq r_n^0(\Theta_1) \geq r(\pi) \geq A^2\,C^2\,2^{L-c-1} \sim C^2\,\left(\frac{\sigma^2}{C^2}\,\log_2\frac{\sigma\,N^{p+1/2}}{C}\right)^{1/(2p+1)}\ . \tag{G.146}$$

Since

$$r_n(\Theta_1) \leq r_n(\Theta_{tv}) \leq r_t(\Theta_{tv}) \leq r_t(\Theta_2)$$

(G.142) and (G.146) imply that

$$r_n(\Theta_{tv}) \sim r_t(\Theta_{tv}) \sim C^2\,\left(\frac{\sigma^2}{C^2}\,\log_2\frac{\sigma\,N^{p+1/2}}{C}\right)^{1/(2p+1)}\ ,$$

which finishes the proof.

H. Proof of Theorem 4 Theorem 4 is a separable extension of Theorem 2. In two dimensions, the covariance operator K of z is a circular convolution

$$K\theta[n_1, n_2] = \theta \star c[n_1, n_2]$$

with a separable kernel $c[n_1, n_2] = c_1[n_1]\, c_2[n_2]$. The Fourier transforms satisfy $|\widetilde{c}_1[k]| \sim |1 - 2|k|/N|^{-2p_1}$ and $|\widetilde{c}_2[k]| \sim |1 - 2|k|/N|^{-2p_2}$. Let $K_1\theta[n] = \theta \star c_1[n]$ and $K_2\theta[n] = \theta \star c_2[n]$. Like in the proof of Theorem 2 in Appendix E the preconditioning of K by its diagonal is separated in two parts. We consider the space $\mathbf{V}_{L+1}$ generated by the lower frequency wavelets $\mathcal{B}_0$ in (5.77) and its orthogonal complement $\mathbf{W}_{L+1}$ in $\mathbb{C}^{N^2}$ generated by the high frequency mirror wavelets $\mathcal{B}_1$ in (5.79).

There exists λ_1 such that all eigenvalues of K are above $\lambda_1\, \sigma^2$ and hence $K \geq \lambda_1\, \sigma^2\, Id$. We must prove that there exists $\lambda_2 > 0$ such that over $\mathbf{V}_{L+1}$ $\|K\|_s = O(\sigma^2)$. Although $\mathcal{B}_0$ is not a separable product of two one-dimensional wavelet bases, we saw in (5.76) that each of the wavelet of $\mathcal{B}_0$ can be separated as a product of one-dimensional wavelets ψ_j and scaling signals ϕ_j. Moreover like in (E.96) we verify that $\psi_j = h \star \xi_j$ and $\phi_j = h \star \gamma_j$. Let $h^2[n_1, n_2] = h[n_1]\, h[n_2]$. Let $R\theta = \theta \star c \star h^2$. Like in Appendix E we verify that $\|R\|_s = O(\sigma^2)$. Let $\mathcal{B}_2$ the family constructed like $\mathcal{B}_0$ from translated separable vectors obtained by replacing ψ_j and ϕ_j respectively by ξ_j and γ_j. The properties of conjugate mirror filters imply that $\mathcal{B}_2$ is still an orthonormal family and the matrix of K in the basis $\mathcal{B}_0$ of $\mathbf{V}_{L+1}$ is the same as the matrix of R in $\mathcal{B}_2$. Since $\|R\|_s = O(\sigma^2)$ we derive that $\|K\|_s = O(\sigma^2)$ over $\mathbf{V}_{L+1}$.

Over $\mathbf{W}_{L+1}$ the family of two-dimensional mirror wavelets $\mathcal{B}_1$ in (5.79) is a separable product of two one-dimensional mirror wavelet families. The operator norm of K in this space can thus be written as a product of the operator norms for K_1 and K_2 in one-dimensional mirror wavelet bases:

$$\|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s = \|K_{1,d}^{1/2}\, K_1^{-1}\, K_{1,d}^{1/2}\|_s\; \|K_{2,d}^{1/2}\, K_2^{-1}\, K_{2,d}^{1/2}\|_s\ .$$

Theorem 2 proves that there exist lower bounds $\lambda_1 > 0$ and $\lambda_2 > 0$ of the two norms on the right and hence that $\|K_d^{1/2}\, K^{-1}\, K_d^{1/2}\|_s > \lambda_1\, \lambda_2$.

I. Proof of Theorem 5 The proof proceeds like for Theorem 3, by finding two appropriate sets Θ_1 and Θ_2 that are orthosymmetric in the separable mirror wavelet basis $\mathcal{B}$ and such that $\Theta_1 \subset \Theta_{dv} \subset \Theta_2$. It results from Theorem 1 and Theorem 4 that there exists $\lambda > 0$ such that

$$\frac{1}{\lambda}\, r_{\inf,d}(\mathrm{QH}[\Theta_1]) \leq r_l(\Theta_{dv}) \leq r_{\inf,d}(\mathrm{QH}[\Theta_2]) \tag{I.147}$$

and

$$r_n(\Theta_1) \leq r_n(\Theta_{dv}) \leq r_t(\Theta_{dv}) \leq r_t(\Theta_2)\ . \tag{I.148}$$

The set Θ_2 is defined with the following lemma.

Lemma 4. *There exists $A > 0$ such that if $N > 0$ and $\theta \in \mathbb{C}^{N^2}$ then for all $L \leq j, j_1, j_2 \leq 1$ and $1 \leq \alpha \leq 3$*

$$\begin{aligned} A^{-1}\, \|\theta\|_{dv} &\geq \sup_{0 \leq m < 2^{-j}} \left(2^{L-j} \sum_{m_1=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m_1,m}\rangle|\ ,\ 2^{L-j} \sum_{m_2=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m,m_2}\rangle| \right) \\ &\geq \sup_{0 \leq m < 2^{-j_2}} \left(2^{(j_1-j_2)/2} \sum_{m_1=0}^{2^{-j_1}-1} |\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m}\rangle| \right) \\ &\geq \sup_{0 \leq m < 2^{-j_1}} \left(2^{(j_2-j_1)/2} \sum_{m_2=0}^{2^{-j_2}-1} |\langle \theta, \overline{\psi}_{j_1,m} \overline{\psi}_{j_2,m_2}\rangle| \right)\ . \end{aligned} \tag{I.149}$$

Since $\|\theta\|_{dv}$ is computed from the one-dimensional total variation of the rows and columns of θ, these inequalities are derived from (4.69), using the fact two-dimensional wavelets (5.76) are separable and the $\mathbf{l^1}$ norm of one-dimensional wavelets and mirror wavelets satisfy $\|\psi_{j,m}\|_1 \sim \|\overline{\psi}_{j,m}\|_1 \sim 2^{(j-L)/2}$.

Since $\Theta_{dv} = \{\theta \ : \ \|\theta\|_{dv} \leq C\}$, Θ_2 is defined as the set of all θ such that

$$
\begin{aligned}
A^{-1}\,C &\geq \sup_{0\leq m<2^{-j}} \left(2^{L-j} \sum_{m_1=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m_1,m}\rangle| \ , \ 2^{L-j} \sum_{m_2=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m,m_2}\rangle| \right) \\
&\geq \sup_{0\leq m<2^{-j_2}} \left(2^{(j_1-j_2)/2} \sum_{m_1=0}^{2^{-j_1}-1} |\langle \theta, \overline{\psi}_{j_1,m_1}\overline{\psi}_{j_2,m}\rangle| \right) \\
&\geq \sup_{0\leq m<2^{-j_1}} \left(2^{(j_2-j_1)/2} \sum_{m_2=0}^{2^{-j_2}-1} |\langle \theta, \overline{\psi}_{j_1,m}\overline{\psi}_{j_2,m_2}\rangle| \right) \ .
\end{aligned}
$$

Lemma 4 implies that $\Theta_{dv} \subset \Theta_2$, and Θ_2 is clearly orthosymmetric in the separable mirror wavelet basis.

Let us first compute $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2])$ to obtain an upper bound of $r_l(\Theta_{dv})$. The quadratic envelope of an $\mathbf{l^1}$ ball of sequences $a[m]$ such that $\sum_{m=0}^{M-1} |a[m]| \leq B$ is the $\mathbf{l^2}$ ball defined by $\sum_{m=0}^{M-1} |a[m]|^2 \leq B^2$. Hence $\mathrm{QH}[\Theta_2]$ is the set of all θ such that

$$
\begin{aligned}
A^{-2}\,C^2 &\geq \sup_{0\leq m<2^{-j}} \left(2^{2(L-j)} \sum_{m_1=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m_1,m}\rangle|^2 \ , \ 2^{2(L-j)} \sum_{m_2=0}^{2^{-j}-1} |\langle \theta, \psi^{\alpha}_{j,m,m_2}\rangle|^2 \right) \\
&\geq \sup_{0\leq m<2^{-j_2}} \left(2^{j_1-j_2} \sum_{m_1=0}^{2^{-j_1}-1} |\langle \theta, \overline{\psi}_{j_1,m_1}\overline{\psi}_{j_2,m}\rangle|^2 \right) \\
&\geq \sup_{0\leq m<2^{-j_1}} \left(2^{j_2-j_1} \sum_{m_2=0}^{2^{-j_2}-1} |\langle \theta, \overline{\psi}_{j_1,m}\overline{\psi}_{j_2,m_2}\rangle|^2 \right) \ .
\end{aligned}
$$

Since $1/2\min(a,b) \leq a\,b/(a+b) \leq \min(a,b)$, we rewrite $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2]) \sim I + II$ with

$$
I = \sup_{\theta\in\mathrm{QH}[\Theta_2]} \sum_{j=L+2}^{1} \sum_{\alpha=1}^{3} \sum_{m_1,m_2=0}^{2^{-j}-1} \min\left(\left|\langle \theta, \psi^{\alpha}_{j,m_1,m_2}\rangle\right|^2, \sigma^2 \right)
$$

and

$$
II = \sup_{\theta\in\mathrm{QH}[\Theta_2]} \sum_{\substack{j_1,j_2=L+1 \\ (j_1,j_2)\neq(L+1,L+1)}}^{1} \ \sum_{\substack{0\leq m_1<2^{-j_1} \\ 0\leq m_2<2^{-j_2}}} \min\left(\left|\langle \theta, \overline{\psi}_{j_1,m_1}\overline{\psi}_{j_2,m_2}\rangle\right|^2, \sigma^2 2^{2p_2(j_2-L)} 2^{2p_1(j_1-L)} \right) \ .
$$

These two sums are computed by applying (G.121):

$$II \sim \sum_{j_2=L+1}^{1} \sum_{j_1=j_2}^{1} \min(C^2\, 2^{-j_1} \;,\; 2^{-j_1-j_2}\, \sigma^2 2^{2p_2(j_2-L)} 2^{2p_1(j_1-L)}) +$$
$$\sum_{j_1=L+1}^{1} \sum_{j_2=j_1+1}^{1} \min(C^2\, 2^{-j_2} \;,\; 2^{-j_1-j_2}\, \sigma^2 2^{2p_2(j_2-L)} 2^{2p_1(j_1-L)}) \;.$$

A dominating term appears in II at a critical scale $2^l = 2^{j_1} = 2^{j_2}$ where l is the closest integer to y such that

$$C^2\, 2^{-y} = 2^{-2y}\, \sigma^2 2^{2p_2(y-L)} 2^{2p_1(y-L)}$$

and hence

$$2^l \sim 2^y = 2^L \left(\frac{C}{\sigma\, 2^{-L/2}} \right)^{1/(p_1+p_2-1/2)} \quad .$$

We verify that $II \sim C^2\, 2^{-l}$. Since $N^{1/2} \leq C/\sigma \leq N^{p_1+p_2}$ we are guaranteed that $2^L \leq 2^l \leq 1$. A direct calculation shows that I is smaller than II so that

$$r_{\inf,d}(\mathrm{QH}[\Theta_2]) \sim NC^2 \left(\frac{\sigma\, N^{1/2}}{C} \right)^{1/(p_1+p_2-1/2)} \quad . \tag{I.150}$$

We now compute a lower bound of $r_n(\Theta_{dv})$ by finding an appropriate orthosymmetric set $\Theta_1 \subset \Theta_{dv}$ and computing $r_{\inf,d}(\mathrm{QH}[\Theta_1])$. Let $\overline{\mathbf{W}}_l$ be the space generated by the 2^{-2l} mirror wavelets $\{\overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2}\}_{0\leq m_1,m_2<2^{-l}}$ at the critical scale 2^l. We define Θ_1 as the set of all signals $\theta \in \overline{\mathbf{W}}_l$ such that

$$|\langle \theta, \overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2} \rangle| \leq B^{-1}\, C \tag{I.151}$$

and for any m_1 there is at most one m_2 for which $\langle \theta, \overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2} \rangle \neq 0$ and for any m_2 there is at most one m_1 for which $\langle \theta, \overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2} \rangle \neq 0$. Since $\Theta_1 \subset \overline{\mathbf{W}}_l$ if $\theta \in \Theta_1$ then $\langle \theta, \overline{\psi}_{j,m_1}\overline{\psi}_{j',m_2} \rangle = 0$ if $j \neq l$ or $j' \neq l$. This set is orthosymmetric and $\Theta_1 \subset \Theta_{dv}$ for an appropriate B, because $\|\overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2}\|_{dv} \sim 1$ and the coefficients are distributed so that the total variation does not add-up along each row and column.

The quadratic convex envelop $\mathrm{QH}[\Theta_1]$ is the set of all θ such that

$$\sum_{m_1,m_2=0}^{2^{-l}-1} |\langle \theta, \overline{\psi}_{l,m_1} \overline{\psi}_{l,m_2} \rangle|^2 \leq B^{-2}\, C^2\, 2^{-l} \ .$$

Over $\overline{\mathbf{W}}_l$ the noise z has a variance $\overline{\sigma}_l^2 \sim C^2\, 2^{-l}$ so a direct calculation shows that $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_1]) \sim C^2\, 2^{-l}$. We saw in (I.150) that $r_{\mathrm{inf},d}(\mathrm{QH}[\Theta_2]) \sim C^2\, 2^{-l}$ and (I.147) thus implies that

$$r_l(\Theta_{dv}) \sim C^2\, 2^{-l} \sim N C^2 \left(\frac{\sigma\, N^{1/2}}{C} \right)^{1/(p_1+p_2-1/2)} \ ,$$

which proves (5.86).

Let us now concentrate on the non-linear thresholding and minimax risks over Θ_{dv}. The thresholds $T_{j,\alpha}$ and $\overline{T}_{j_1,j_2}$ are chosen according to (5.83) and (5.85). Two critical scales $2^{c_2(j_1)}$ and $2^{c_1(j_2)}$ are respectively defined as the smallest scales such that $\overline{T}_{j_1,j_2} = \infty$ if $j_1 > c_1(j_2)$ or $j_2 > c_2(j_1)$. They satisfy

$$\text{(I.152)} \qquad \overline{\sigma}_{c_1,j_2} \sqrt{2 \log_e 2^{-c_1-j_2}} \sim \overline{s}_{c_1,j_2}$$

and

$$\text{(I.153)} \qquad \overline{\sigma}_{j_1,c_2} \sqrt{2 \log_e 2^{-j_1-c_2}} \sim \overline{s}_{j_1,c_2} \ .$$

The diagonal critical scale 2^c is defined as the smallest scale such that $\overline{T}_{j_1,j_2} = \infty$ if $j_1 > c$ and $j_2 > c$ and thus satisfies

$$\text{(I.154)} \qquad \overline{\sigma}_{c,c} \sqrt{2 \log_e 2^{-2c}} \sim \overline{s}_{c,c} \ .$$

If $2^{j_1} \geq 2^c$ and $2^{j_2} \geq 2^c$ then $\overline{T}_{j_1,j_2} = \infty$. Using (I.149) one can verify that $\overline{s}_{c,c} \sim C$ and since $\overline{\sigma}_{c,c} \sim \sigma\, 2^{p_1(c-L)}\, 2^{p_2(c-L)}$ we get

$$\sigma\, 2^{(p_1+p_2)(c-L)}\, |c|^{1/2} \sim C$$

and hence

$$\text{(I.155)} \qquad 2^{-c} \sim N \left(\frac{\sigma^2}{C^2} \log_2 \frac{N^{p_1+p_2}\, \sigma}{C} \right)^{1/(2p_1+2p_2)} \ .$$

An upper bound of the thresholding risk $r_t(\Theta_2)$ is computed with (3.28):

$$r_t(\Theta_2) \leq I + II + III_1 + III_2 + IV$$

with

$$I \sim \log_e(N^2/4)\,\sigma^2 + \sum_{j_1=L+1}^{c} \sum_{j_2=j_1}^{c_2(j_1)} \log_e 2^{-j_1-j_2}\sigma^2\, 2^{2p_1(j_1-L)}\, 2^{2p_2(j_2-L)}$$

$$+ \sum_{j_2=L+1}^{c} \sum_{j_1=j_2}^{c_1(j_2)} \log_e 2^{-j_1-j_2}\sigma^2\, 2^{2p_1(j_1-L)}\, 2^{2p_2(j_2-L)} \;,$$

$$II \sim \sup_{\theta\in\Theta_2} \left(\sum_{\alpha=1}^{3} \sum_{j=L+1}^{1} \sum_{m_1,m_2=0}^{2^{-j}-1} \min\Big(|\langle \theta, \psi^{\alpha}_{j,m_1,m_2}\rangle|^2, \sigma^2 \log_e N \Big) \right)$$

$$III_1 \sim \sup_{\theta\in\Theta_2} \left(\sum_{j_1=L+1}^{c} \sum_{m_1=0}^{2^{-j_1}-1} \left(\sum_{j_2=j_1}^{c_2(j_1)} \sum_{m_2=0}^{2^{-j_2}-1} \right.\right.$$

$$\min\Big(|\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\rangle|^2 \;,\; \log_e 2^{-j_1-j_2}\sigma^2\, 2^{2p_1(j_1-L)}\, 2^{2p_2(j_2-L)} \Big)$$

$$\left.\left. + \sum_{j_2=c_2(j_1)+1}^{1} \sum_{m_2=0}^{2^{-j_2}-1} |\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\rangle|^2 \right)\right)$$

$$III_2 \sim \sup_{\theta\in\Theta_2} \left(\sum_{j_2=L+1}^{c} \sum_{m_2=0}^{2^{-j_2}-1} \left(\sum_{j_1=j_2+1}^{c_1(j_2)} \sum_{m_1=0}^{2^{-j_1}-1} \right.\right.$$

$$\min\Big(|\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\rangle|^2 \;,\; \log_e 2^{-j_1-j_2}\sigma^2\, 2^{2p_1(j_1-L)}\, 2^{2p_2(j_2-L)} \Big)$$

$$\left.\left. + \sum_{j_1=c_1(j_2)+1}^{1} \sum_{m_1=0}^{2^{-j_1}-1} |\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\rangle|^2 \right)\right)$$

and

$$IV \sim \sup_{\theta\in\Theta_2} \sum_{\substack{j_1>c \\ j_2>c}} \sum_{m_2=0}^{2^{-j_2}-1} \sum_{m_1=0}^{2^{-j_1}-1} |\langle \theta, \overline{\psi}_{j_1,m_1} \overline{\psi}_{j_2,m_2}\rangle|^2 \;.$$

The sum I is a geometrical series which is of the order of the dominating term for $j_1 = j_2 = c$

$$I \sim \log_e 2^{-2c}\,\sigma^2\, 2^{2(p_1+p_2)(c-L)} \sim C^2 \;.$$

The sum II is calculated with (G.138)

$$II \sim \sum_{j=L+1}^{1} 2^{-j} \min(2^{-j}\sigma^2 \log_e N \ , \ C 2^{j-L}\sigma\sqrt{\log_e N}) \sim \log_2 N \sqrt{\log_e N}\, N\, C\, \sigma \ .$$

The sum III_1 is also computed with (G.138):

$$III_1 \sim \sum_{j_1=L+1}^{c} 2^{-j_1} \left(\sum_{j_2=j_1}^{c_2(j_1)} \min\Big(C^2\, 2^{j_1-j_2} \ , \ C\, 2^{(j_1-j_2)/2} \log_e 2^{-j_1-j_2} \sigma\, 2^{p_1(j_1-L)}\, 2^{p_2(j_2-L)}\Big) + \sum_{j_2=c_2(j_1)+1}^{1} C^2\, 2^{j_1-j_2} \right)$$

Again the dominating term corresponds to $j_1 = j_2 = c$ and hence $III_1 \sim C^2\, 2^{-c}$. We verify similarly that $III_2 \sim C^2\, 2^{-c}$. The sum IV also satisfies

$$IV \sim \sum_{\substack{j_1 > c \\ j_2 > c}} C^2\, 2^{-\max(j_1,j_2)} \sim C^2\, 2^{-c} \ .$$

Inserting (I.155) gives

(I.156)
$$r_t(\Theta_2) \le I + II + III_1 + III_2 + IV \sim C^2\, N \left(\frac{\sigma^2}{C^2} \log_2 \frac{N^{p_1+p_2}\, \sigma}{C} \right)^{1/(2p_1+2p_2)} .$$

We now compute a lower bound of $r_n(\Theta_{dv})$ with an appropriate orthosymmetric set $\Theta_1 \subset \Theta_{dv}$. Let $\overline{\mathbf{W}}_c$ be the space generated by the 2^{-2c} mirror wavelets $\{\overline{\psi}_{c,m_1}\overline{\psi}_{c,m_2}\}_{0\le m_1,m_2<2^{-c}}$. The set Θ_1 is defined like in the linear case as the set of all signals $\theta \in \overline{\mathbf{W}}_c$ such that

(I.157)
$$|\langle \theta, \overline{\psi}_{c,m_1}\overline{\psi}_{c,m_2}\rangle| \le B^{-1}\, C$$

and for any m_1 there is at most one m_2 for which $\langle \theta, \overline{\psi}_{c,m_1}\overline{\psi}_{c,m_2}\rangle \neq 0$ and for any m_2 there is at most one m_1 for which $\langle \theta, \overline{\psi}_{l,m_1}\overline{\psi}_{l,m_2}\rangle \neq 0$. Since $\Theta_1 \subset \overline{\mathbf{W}}_c$ if $\theta \in \Theta_1$ then $\langle \theta, \overline{\psi}_{j,m_1}\overline{\psi}_{j',m_2}\rangle = 0$ if $j \neq c$ or $j' \neq c$. We have $\Theta_1 \subset \Theta_{dv}$ for an appropriate $B > 0$.

Let us compute a lower bound of the minimax risk $r_n(\Theta_1)$ to estimate $\theta \in \Theta_1$ from $X = \theta + z$. Theorem 4 proves that there exists λ such that over $\overline{\mathbf{W}}_c$ the covariance K of z satisfies for all $N > 0$

$$K \geq \lambda \, \overline{\sigma}_{c,c} \, Id \ .$$

Since $\Theta_1 \subset \overline{\mathbf{W}}_c$, the noise augmentation lemma 1 implies that $r_n(\Theta_1) \geq r_n^0(\Theta_1)$ where $r_n^0(\Theta_1)$ is the minimax risk to estimate $\theta \in \Theta_1$ from $X = \theta + z^0$, with a Gaussian white noise z^0 of variance $\beta^2 = \lambda \, \overline{\sigma}_{c,c}^2$ in the space $\overline{\mathbf{W}}_c$. Let $\delta[m]$ be a discrete Dirac. A lower bound of $r_n^0(\Theta_1)$ is obtained by calculating the Bayes risk for a random process $F \in \overline{\mathbf{W}}_c$ defined by

$$\langle F, \overline{\psi}_{l,m_1} \overline{\psi}_{l,m_2} \rangle = \begin{cases} A \, C \, \delta[m_1 - P[m_2]] & \text{for } 0 \leq m_2 < 2^{-c-1} \\ 0 & \text{for } 2^{-c-1} \leq m_2 < 2^{-c} \end{cases} .$$

To make sure that $F \in \Theta_1$, two non-zero coefficients should not be on the same row. This is done by letting $P[0]$ be a random variable uniformly distributed over $\{0, .., M-1\}$ and the probability distribution of $P[m_2]$ given $P[k]$ for $k < m_2$ is uniformly distributed over $\{0, .., M-1\} - \{P[k]\}_{0 \leq k < m_2}$. If $A \leq B^{-1}$ then $F \in \Theta_1$. The constant A is chosen small enough so that

$$A \, C \leq \lambda \, \overline{\sigma}_{c,c} \sqrt{\log_e 2^{-c-1}} \, .$$

Using Lemma 3 we verify that the Bayes risk $r(\pi)$ to estimate F from $X = F + z^0$ satisfies

$$r(\pi) \geq \sum_{m_2=0}^{2^{-c-1}-1} A^2 \, C^2 / 2 \sim C^2 \, 2^{-c} \ . \tag{I.158}$$

Since $r_n(\Theta_1) \geq r_n^0(\Theta_1) \geq r(\pi)$ and

$$r_n(\Theta_1) \leq r_n(\Theta_{dv}) \leq r_t(\Theta_{dv}) \leq r_t(\Theta_2)$$

we derive from (I.156) and (I.158) that

$$r_n(\Theta_{dv}) \sim r_t(\Theta_{dv}) \sim C^2\, N\, \left(\frac{\sigma^2}{C^2}\log_2\frac{\sigma\, N^{p_1+p_2}}{C}\right)^{1/(2p_1+2p_2)},$$

which finishes the proof.

REFERENCES

[1] F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129, 1998.

[2] L. Alvarez and J. M. Morel. Formalization and computational aspects of image analysis. *Acta Numerica*, pages 1–59, 1994.

[3] R. R. Coifman and D. Donoho. Translation invariant de-noising. Technical Report 475, Dept. of Statistics, Stanford University, May 1995.

[4] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser. Wavelet analysis and signal processing. In *Wavelets and their Applications*, pages 153–178, Boston, 1992. Jones and Barlett. B. Ruskai et al. eds.

[5] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. on Pure and Appl. Math.*, 41:909–996, November 1988.

[6] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Info. Theory*, 38(2):719–746, March 1992.

[7] D. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *J. of Appl. and Comput. Harmonic Analysis*, 1:100–115, 1993.

[8] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions. *J. of Appl. and Comput. Harmonic Analysis*, 2(2):101–126, 1995.

[9] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, December 1994.

[10] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. of Royal Stat. Soc. B.*, 57(2):301–369, 1995.

[11] D. Donoho, R.C. Liu, and K.B. Gibbon. Minimax risk over hyperrectangles and implications. *Annals of statistics*, 18:1416–1437, 1990.

[12] D. Donoho, S. Mallat, and R. von Sachs. Estimating covariances of locally stationary processes: consistency of best basis methods. In *Proc. of Time-Freq. and Time-Scale Symp.*, Paris, July 1996.

[13] I. A. Ibragimov and R. Z. Khas'minskii. Bounds for the risks of non parametric regression estimates. *Theory of Probability and its Applications*, 27:84–89, 1984.

[14] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. Technical report, Dept. of Statistics, Stanford University, December 1995.

[15] J. Kalifa. *Restauration minimax et déconvolution dans une base d'ondelettes miroirs.* PhD thesis, Ecole Polytechnique, 1999.

[16] J. Kalifa and S. Mallat. *Bayesian Inference in Wavelet Based Models*, chapter "Minimax restoration and deconvolution". Springer-Verlag, 1999. P. Müller and B. Vidakovic, eds.

[17] J. Kalifa, S. Mallat, and B. Rougé. Minimax deconvolution in mirror wavelet bases. 1999. Submitted to *IEEE Transactions on Signal Processing*, available at http://www.cmap.polytechnique.fr/˜ kalifa/.

[18] E. Kolaczyk. A wavelet shrinkage approach to tomographic image reconstruction. *J. Amer. Statist. Assoc.*, 91:1079–1090, 1996.

[19] N.M Lee and B. Lucier. Wavelet methods for inverting the Radon transform with noisy data. *IEEE Trans. on IP, to appear*, 2000.

[20] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 2nd edition, 1999.

[21] Y. Meyer. *Wavelets and Operators.* Advanced mathematics. Cambridge university press, 1992.

[22] F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, pages 502–527, 1986.

[23] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, 27(6), December 1999.

[24] B. Rougé. Remarks about space-frequency and space-scale representations to clean and restore noisy images in satellite frameworks. In *Progress in wavelet analysis and applications*, pages 433–442, Paris, 1993. Frontières. edts. Y Meyer and S. Roques.

[25] B. Rougé. *Théorie de la chaine image optique et restauration.* PhD thesis, Université Paris-Dauphine, 1997. Thèse d'habilitation.

[26] A.B. Tsybakov and L. Cavalier. Sharp adaptation for inverse problems with random noise. *Proba. Theory and Related Fields*, 2001. to appear.

[27] Y. Wang. Minimax estimation via wavelets for indirect long-memory data. *J. Statist. Plann. Inference*, 64:45–55, 1997.

[28] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software.* AK Peters, 1994.

[29] W.P. Ziemer. *Weakly differentiable functions : Sobolev spaces and functions of bounded variations*, volume 120. Springer-Verlag, 1989.