Deconvolution by Thresholding in Mirror Wavelet Bases

Jérôme Kalifa, Stéphane Mallat, and Bernard Rougé

Abstract—The deconvolution of signals is studied with thresholding estimators that decompose signals in an orthonormal basis and threshold the resulting coefficients. A general criterion is established to choose the orthonormal basis in order to minimize the estimation risk. Wavelet bases are highly sub-optimal to restore signals and images blurred by a low-pass filter whose transfer function vanishes at high frequencies. A new orthonormal basis called *mirror wavelet basis* is constructed to minimize the risk for such deconvolutions. An application to the restoration of satellite images is shown.

Index Terms—Deconvolution, inverse problem, thresholding, wavelet packets.

I. INTRODUCTION

IN MANY imaging devices, the diffraction of the optics creates a blur and the electronics produce an additive noise. When the blur is uniform over the image, it can be modeled as a low-pass filtering. Inverting this degradation is a well known ill-posed deconvolution problem, which requires using prior information on the signal and the noise to optimize the estimation [1]. This problem is equivalent to the removal of a colored stationary noise, whose power spectrum is huge at certain frequencies. The main difficulty is to restore the sharp transitions and edges in images, which often requires using nonlinear estimators [2], [3]. A classical approach to deconvolution is to write the estimation as a minimization problem that incorporates a fidelity term to the observed data and an a priori measure that regularizes the estimation. This approach can also be casted as a Bayesian estimation which minimizes a posterior distribution computed from the observed data and the joint prior distribution of the signal and the noise [4]. A large body of literature is entirely devoted to such approaches that we shall not further develop [5]–[9].

Donoho and Johnstone [10] have introduced a different class of nonlinear estimators that decompose the observed data in an orthonormal basis and threshold the resulting coefficients. In this case, the prior information should be incorporated in the choice of the orthonormal basis and the values of thresholds. To suppress additive Gaussian white noise on piecewise regular signals, Donoho and Johnstone [10] have proved that a thresholding in a wavelet basis restores efficiently the sharp transitions and produces a nearly minimax risk over large classes of signals. These results have been extended to specific deconvolution problems [11]. However, wavelet bases are not always well adapted to construct efficient thresholding estimators for deconvolution problems. We thus establish a general criterion to adapt the choice of basis to the properties of the deconvolution kernel and minimize the resulting risk.

Section II introduces thresholding estimators in the context of deconvolution problems, for signals of arbitrary dimension. Lower and upper bounds of the resulting risk are calculated. To minimize the estimation risk, it is shown that the basis vectors should have an energy sufficiently concentrated in the frequency domain. Applied to wavelet bases, this general result gives the particular class of "mild" deconvolution problems studied by Donoho [11], where these bases yield efficient thresholding estimators. However, when a one-dimensional signal is blurred by a low-pass filter having a zero at high frequencies, Section III explains that it is necessary to use a basis of vectors having a better frequency resolution than wavelets. This leads to the construction of a new type of bases, called *mirror wavelet bases*. Thresholding estimators in mirror wavelet bases are studied, with an emphasis on fast algorithms and applications. We prove in [12] that the risk of mirror wavelet thresholding estimators is asymptotically minimax for bounded variation signals.

Mirror wavelet thresholding estimators are extended in two dimensions for the restoration of images blurred by low-pass filters. Algorithms and numerical results on satellite images are presented in Section IV. A thresholding estimator in such a basis has been selected by the French spatial agency (Centre National d'Études Spatiales, CNES) for the deconvolution of images obtained by a new generation of satellites. This decision was the result of an extensive numerical comparison between competing deconvolution procedures developed by different French laboratories, including different types of energy based minimization approaches [5], [6], [13] and thresholding algorithms in wavelet bases [14].

II. THRESHOLDING ESTIMATORS FOR DECONVOLUTION

This section reviews the properties of thresholding estimators for the deconvolution of signals, and finds conditions to build estimators that are close to minimax. The analysis applies to d-dimensional signals $f[n]$, with $n = (n_1, n_2, \ldots, n_d)$, of size N^d. The signal is degraded by a convolution with a filter $u[n]$

and by an additive random noise $W[n]$, which is assumed to be wide-sense stationary, and, thus, we measure

$$Y[n] = f \star u[n] + W[n]. \tag{1}$$

In the following, we shall use capital letters for random vectors as opposed to deterministic vectors. For simplicity, we assume periodic boundary conditions, which means that the convolution is circular and that the noise $W[n]$ is circular stationary. The covariance of W is therefore diagonalized by the discrete Fourier basis. The filter u and the power spectrum of W are supposed to be known *a priori*, through a calibration procedure. All convolutions are assumed to be circular convolutions in the paper.

A naive deconvolution would consist in inverting directly the filtering by u. Let $\widehat{u}[k]$ be the d-dimensional discrete Fourier transform of u, where $k = (k_1, k_2, \ldots, k_d)$ is a d-dimensional frequency index. The pseudo-inverse filter u^{-1} is defined by

$$\widehat{u}^{-1}[k] = \begin{cases} \frac{1}{\widehat{u}[k]} & \text{if } \widehat{u}[k] \neq 0 \\ 0 & \text{if } \widehat{u}[k] = 0 \end{cases}. \tag{2}$$

Applying u^{-1} gives

$$X = Y \star u^{-1} = f \star c[n] + W \star u^{-1}[n] \tag{3}$$

where the filter $c = u \star u^{-1}$ projects f over frequencies where $\widehat{u}[k]$ does not vanish

$$\widehat{c}[k] = \begin{cases} 1, & \text{if } \widehat{u}[k] \neq 0 \\ 0, & \text{if } \widehat{u}[k] = 0 \end{cases}. \tag{4}$$

Since W is wide-sense circular stationary, the deconvolved noise $Z[n] = W \star u^{-1}[n]$ is also circular stationary. Its power spectrum is related to the power spectrum $P_W[k]$ of W by

$$P_Z[k] = P_W[k] \, |\widehat{u}^{-1}[k]|^2. \tag{5}$$

In the neighborhood of frequencies k where $\widehat{u}[k]$ vanishes, $\widehat{u}^{-1}[k]$ is huge so the power spectrum $P_Z[k]$ is considerably amplified.

A thresholding estimator attenuates the amplified noise Z by decomposing $X = Y \star u^{-1}$ in an orthonormal basis $\mathcal{B} = \{g_m\}_{1 \leq m \leq N^d}$ of $\mathbb{C}^{N^d}$, and by thresholding the resulting coefficients. Let $\langle X, g_m \rangle$ be the inner product of X with g_m. A thresholding estimator $\widetilde{F}$ of f in $\mathcal{B}$ is defined as

$$\widetilde{F} = D_t X = \sum_{m=1}^{N^d} \rho_{T_m} \left(\langle X, g_m \rangle \right) g_m \tag{5}$$

where ρ_T is a hard thresholding function that sets to zero a coefficient whose amplitude is below T

$$\rho_T(x) = \begin{cases} x, & \text{if } |x| > T \\ 0, & \text{if } |x| \leq T \end{cases}. \tag{6}$$

We shall compute the risk with the Euclidean norm of $\mathbb{C}^{N^d}$

$$r_t(f) = \mathbf{E}\{\|f - D_t X\|^2\} \tag{7}$$

with an expected value calculated with respect to the probability distribution of the noise W.

A. Thresholding Risk Versus Minimax Risk

The goal of this paper is to optimize the thresholding estimator given some prior information on the signal f and on the noise W. A stochastic prior model considers f as the realization of a stochastic process whose probability distribution is known. However, it is rare that we know the probability distribution of complex signals such as images. Weaker but often more realistic models define a prior set Θ that includes f, without specifying a probability distribution in Θ. The more prior information the smaller the set Θ. For example, for images that do not include fractal textures, Θ may correspond to the set of images whose edges have an average length bounded by a given constant. Such a model can be formalized by imposing an upper bound on the total variation of the signal [12].

The expected risk over Θ cannot be computed because we do not know the probability distribution of signals in Θ. To control the risk for any $f \in \Theta$, a minimax approach tries to minimize the maximum risk. Let DY be an estimator of f computed by applying an operator D to the original data Y. The maximum risk over Θ is

$$r(D, \Theta) = \sup_{f \in \Theta} \mathbf{E}\{\|DY - f\|^2\}.$$

The *minimax risk* is the lower bound computed over all operators D

$$r_{\min}(\Theta) = \inf_D r(D, \Theta).$$

For the thresholding estimator (5) applied to $X = Y \star u^{-1}$, since $\mathcal{B}$ is an orthonormal basis the risk can be written

$$\begin{aligned} r_t(f) &= \mathbf{E}\{\|f - D_t X\|^2\} \\ &= \sum_{m=1}^{N^d} \mathbf{E}\{|\langle f, g_m \rangle - \rho_{T_m}(\langle X, g_m \rangle)|^2\} \end{aligned} \tag{8}$$

and the maximum risk over Θ is $r_t(\Theta) = \sup_{f \in \Theta} r_t(f)$. The main difficulty is to choose the thresholds T_m and the basis $\mathcal{B}$ so that $r_t(\Theta)$ is as close as possible to the minimax risk $r_{\min}(\Theta)$.

To choose the thresholds T_m, we first recall the lower bound of the risk $r_t(f)$ established by Donoho and Johnstone [10], and then give values of thresholds to nearly reach this lower bound. The thresholding estimator (5) belongs to a class of *diagonal estimators*

$$\widetilde{F}_d = \sum_{m=0}^{N^d - 1} d_m \left(\langle X, g_m \rangle \right) g_m$$

where d_m is a decision function that sets to zero the coefficient $\langle X, g_m \rangle$ or keeps its value as is. Let $\sigma_m^2 = \mathbf{E}\{|\langle Z, g_m \rangle|^2\}$ be the variance of the noise in the direction of g_m. If f is known then one can verify that the decision that minimizes the risk $\mathbf{E}\{\|f - \widetilde{F}_d\|^2\}$ is

$$d_m \left(\langle X, g_m \rangle \right) = \begin{cases} \langle X, g_m \rangle & \text{if } |\langle f, g_m \rangle| > \sigma_m \\ 0 & \text{if } |\langle f, g_m \rangle| > \sigma_m \end{cases} \tag{9}$$

in which case

$$r_d(f) = \mathbf{E}\{\|f - \widetilde{F}_d\|^2\} = \sum_{m=1}^{N^d} \min(\sigma_m^2, |\langle f, g_m \rangle|^2).$$

In practice, f is not know so one cannot compute the *oracle decision* (9), but it provides a lower bound for the risk of the thresholding estimator (5), calculated in (8)

$$r_t(f) \geq r_d(f) = \sum_{m=0}^{N-1} \min(\sigma_m^2, |\langle f, g_m \rangle|^2).$$

It results that $r_t(\Theta) \geq \sup_{f \in \Theta} r_d(f) = r_d(\Theta)$.

To guarantee that the thresholding risk $r_t(f)$ is not too much above the lower bound $r_d(f)$, Donoho and Johnstone [10] as well as Johnstone and Silverman [15] choose a threshold T_m proportional to the noise standard deviation

$$T_m = \sigma_m \sqrt{2 \log_e N^d}. \tag{10}$$

For a deconvolution, σ_m may become much larger than the maximum amplitude that a signal coefficient can reach, in which case we should systematically set the corresponding coefficient to 0. For signals in Θ, the maximum coefficient amplitude in the direction of g_m is known a priori: $s_\mathcal{B}[m] = \sup_{f \in \Theta} |\langle f, g_m \rangle|$. The threshold T_m is thus modified accordingly

$$T_m = \begin{cases} \sigma_m \sqrt{2 \log_e N^d}, & \text{if } \sigma_m \sqrt{2 \log_e N^d} \leq s_\mathcal{B}[m] \\ \infty, & \text{otherwise} \end{cases}. \tag{11}$$

Setting $T_m = \infty$ guarantees that $\rho_{T_m}(X_\mathcal{B}[m]) = 0$ if $\sigma_m \sqrt{2 \log_e N^d} > s_\mathcal{B}[m]$. With this threshold definition, if $W[n]$ and hence $Z[n]$ are Gaussian random vectors then using the results of Donoho and Johnstone [10], we prove in [12] that

$$r_d(\Theta) \leq r_t(\Theta) \leq 4 (\log_e N^d + 1) r_d(\Theta). \tag{12}$$

This guarantees that $r_t(\Theta)$ remains of the order of the oracle decision risk $r_d(\Theta)$.

Now that the thresholds are chosen, it remains to optimize the basis. Since (12) shows that the thresholding risk is of the same order as $r_d(\Theta)$, we optimize the basis $\mathcal{B}$ by minimizing

$$r_d(f) = \sum_{m=0}^{N-1} \min(\sigma_m^2, |\langle f, g_m \rangle|^2) \quad \text{for } f \in \Theta. \tag{13}$$

Since $\mathcal{B}$ is orthonormal, we have an energy conservation equation

$$\|f\|^2 = \sum_{m=0}^{N-1} |\langle f, g_m \rangle|^2$$

which applied to the noise gives

$$\mathbf{E}\{\|Z\|^2\} = \sum_{m=0}^{N-1} \mathbf{E}\{|\langle Z, g_m \rangle|^2\} = \sum_{m=0}^{N-1} \sigma_m^2.$$

However, depending upon $\mathcal{B}$ the energy of $\|f\|^2$ and $\mathbf{E}\{\|Z\|^2\}$ may be spread over few coefficients of large amplitudes or diluted over many basis coefficients of small amplitudes. To minimize (13), we need to concentrate the huge energy $\mathbf{E}\{\|Z\|^2\}$ over very few vectors g_m which produce coefficients σ_m much larger than $|\langle f, g_m \rangle|$, and among the remaining vectors g_p concentrate the energy $\|f\|^2$ over few large coefficients $|\langle f, g_p \rangle|$ that are above the noise level σ_p.

B. Basis Choice for Deconvolution

We further study the choice of basis to optimize a thresholding estimator for a deconvolution, and give a condition to adapt the basis to the noise power spectrum. Applied to wavelet bases, this condition specifies a limited class of "mild" deconvolution problems where a wavelet thresholding estimator is nearly optimal.

The Karhunen-Loève energy compaction theorem [16] proves that the basis that diagonalizes the covariance of Z is the basis that best concentrates the expected energy of a random vector Z over any specified number of basis vectors. Since $Z = W \star u^{-1}$ is circular stationary, the Karhunen-Loève basis that diagonalizes its covariance is the discrete Fourier basis. The variance of the noise coefficients of Z in the Fourier basis is given by the power spectrum for each frequency index k

$$\sigma_k^2 = P_Z[k] = \begin{cases} \frac{P_W[k]}{|\hat{u}[k]|^2}, & \text{if } \hat{u}[k] \neq 0 \\ 0, & \text{if } \hat{u}[k] = 0 \end{cases}.$$

The Fourier basis is therefore optimal to concentrate the noise energy, but if the signal f includes sharp transitions as it is often the case in images, then the Fourier basis does not concentrate efficiently the energy of f over few coefficients. It is therefore necessary to choose a different basis $\mathcal{B} = \{g_m\}_{1 \leq m \leq N^d}$, which still concentrates efficiently the noise energy. This means that the noise variances in $\mathcal{B}$ should remain of the same order as the power spectrum values. In the direction of g_m, the variance of Z is related to its power spectrum and to the Fourier transform $\hat{g}_m[k]$ by

$$\sigma_m^2 = \frac{1}{N^d} \sum_k |\hat{g}_m[k]|^2 P_Z[k]. \tag{14}$$

If S_m is the frequency support of $\hat{g}_m[k]$, since $N^{-d} \sum_k |\hat{g}_m[k]|^2 = \|g_m\|^2 = 1$ we derive from (14) that

$$\min_{k \in S_m} P_Z[k] \leq \sigma_m^2 \leq \max_{k \in S_m} P_Z[k]. \tag{15}$$

The inequalities (15) may also remain valid for a set S_m smaller than the frequency support of $\hat{g}_m[k]$ if $\hat{g}_m[k]$ has a fast decay outside this frequency set so that the sum in (14) is dominated by the values of $P_Z[k]$ in S_m. We shall then say that $\hat{g}_m$ has an energy *essentially concentrated* in S_m. If there exists a constant λ such that

$$\max_{k \in S_m} P_Z[k] \leq \lambda^2 \min_{k \in S_m} P_Z[k] \quad \text{for } 1 \leq m \leq N^d \tag{16}$$

then (15) implies that the noise variance σ_m^2 is of the same order as the power spectrum values for $k \in S_m$, up to the factor λ^2. If W is a white noise of variance σ^2 then $P_W[k] = \sigma^2$ so $P_Z[k] = \sigma^2 |\hat{u}^{-1}[k]|^2$. The condition (16) is then equivalent to

$$\max_{k \in S_m} |\hat{u}[k]| \leq \lambda \min_{k \in S_m} |\hat{u}[k]| \quad \text{for } 1 \leq m \leq N^d. \tag{17}$$

Wavelet bases concentrate efficiently the energy of piecewise regular signals over few coefficients and are thus a priori good candidates to restore such signals. The condition (16) shows we must also guarantee that $P_Z[k]$ has a small relative variation of the frequency support of each wavelet. For one-dimensional

signals, a discrete wavelet orthonormal basis is constructed by translating N periodic discrete wavelets $\psi_j[n]$ for each scale $2^j \leq 1$

$$\psi_{j,q}[n] = \psi_j[n - N2^j q] \quad \text{for } 0 \leq q < 2^{-j}.$$

The support of $\psi_j[n]$ is proportional to 2^j whereas its Fourier transform $\widehat{\psi}_j[k]$ has an energy essentially concentrated in the frequency band $|k| \in [2^{-j-1}, 2^{-j}]$, as illustrated in Fig. 1. At the largest scale $2^j = 1$, the wavelet is constant $\psi_{1,0}[n] = 1$. With an appropriate design of the discrete wavelets $\psi_j[n]$[16], the resulting family $\{\psi_{j,q}\}_{L < j \leq 1, \, 0 \leq q < 2^{-j}}$ is an orthonormal basis of $\mathbb{C}^N$, with $L = -\log_2 N$.

Since $|\widehat{\psi}_{j,q}[k]| = |\widehat{\psi}_j[k]|$ which is essentially nonzero for $|k| \in [2^{-j-1}, 2^{-j}]$, the condition (16) can be rewritten

$$\max_{|k| \in [2^{-j-1}, 2^{-j}]} P_Z[k] \leq \lambda^2 \min_{|k| \in [2^{-j-1}, 2^{-j}]} P_Z[k] \quad \text{for } 2^j \leq 1. \tag{18}$$

To simplify notations, in the following we write $\alpha[k] \sim \beta[k]$ if there exists two constants A and B of the order of 1 such that

$$\forall k \ , \quad A\,\alpha[k] \leq \beta[k] \leq B\,\alpha[k].$$

For a deconvolution problem with a white noise W, the condition (16) is equivalent to (17) which can be rewritten

$$\max_{|k| \in [2^{-j-1}, 2^{-j}]} |\widehat{u}[k]| \leq \lambda \min_{|k| \in [2^{-j-1}, 2^{-j}]} |\widehat{u}[k]| \quad \text{for } 2^j \leq 1. \tag{19}$$

If $|\widehat{u}[k]|$ has a rational growth which means that there exists a real exponent γ (positive or negative) such that

$$\forall 1 \leq |k| \leq \frac{N}{2}, \ \ |\widehat{u}[k]| \sim |k|^\gamma \tag{20}$$

then (19) is satisfied for $\lambda \sim 2^{|\gamma|}$. A wavelet basis thus concentrates efficiently the noise energy.

Donoho [11] and Johnstone and Silverman [15] have studied deconvolution problems where the transfer function satisfies (20), in which case wavelet bases are well adapted to perform the thresholding estimation. For example, computing the discrete derivative of a signal in presence of white noise can be interpreted as a deconvolution problem, where $u[n]$ is an integrator that inverts the finite difference operator. This integrator satisfies (20) for $\gamma = -1$, and Donoho [11] proved the minimax optimality of the resulting wavelet thresholding estimator over different classes Θ of piecewise regular signals. Tomographic image reconstructions can also be interpreted as a deconvolution problem where u satisfies a rational growth condition and where wavelet bases thus provide efficient thresholding estimators [11].

If the transfer function $\widehat{u}[k]$ vanishes then the deconvolution is much more unstable than for filters satisfying (20), because $|\widehat{u}^{-1}[k]|$ has a very fast relative variation in the frequency neighborhood of its zeros. Next section shows that in this case, wavelet bases are not well adapted to perform the deconvolution and must be replaced by bases having a better frequency resolution.
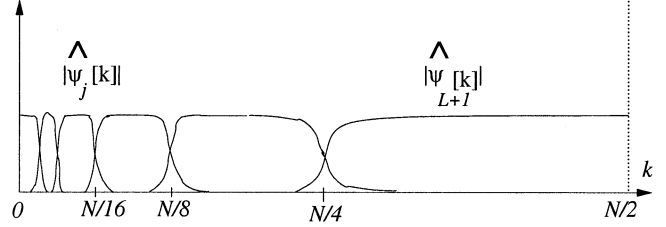

Fig. 1. A wavelet ψ_j has a Fourier transform $\widehat{\psi}_j[k]$ whose energy is essentially concentrated in the interval $[2^{-j-1}, 2^{-j}]$, for $-\log_2 N < j < 0$.

III. HYPERBOLIC DECONVOLUTIONS

Digital measuring devices include a low-pass filter which limits the frequency band of the analog input signal to an interval $[-\omega_0, \omega_0]$ and sample uniformly the output at intervals $\Delta = \pi/\omega_0$, to avoid any aliasing. However, system imperfections often produce a transfer function $\widehat{u}(\omega)$ that differs significantly from a perfect low-pass filter equal to 1 on $[-\omega_0, \omega_0]$ and 0 outside. To reduce the aliasing, $\widehat{u}(\omega)$ must vanish at $\pm\omega_0$ but $\widehat{u}(\omega)$ is generally continuous at $\pm\omega_0$ and decreases to zero like $(|\omega| - |\omega_0|)^p$ for some $p \geq 1$. This progressive decay to zero attenuates the highest signal frequencies in the neighborhood of ω_0, and thus produces a blur relatively to the ideal signal that would be obtained with a perfect low-pass filter. On the resulting sampled signal, for a sampling distance π/ω_0 renormalized to 1, this blur is equivalent to a convolution with a discrete filter $u[n]$ whose transfer function $\widehat{u}[k]$ has a zero of order $p \geq 1$ at the highest frequency $k = \pm N/2$, which we write

$$|\widehat{u}[k]| \sim \left|\frac{2k}{N} - 1\right|^p. \tag{21}$$

To restore the highest frequencies, it is equivalent to estimating the ideal digital signal $f[n]$ from

$$Y[n] = f \star u[n] + W[n]$$

where $W[n]$ is the additive noise of the measuring device. To simplify the explanations, in the following, we shall suppose that W is a white noise of variance σ^2. Hence the power spectrum $P_Z[k]$ of the deconvolved noise $Z = W \star u^{-1}$ has a hyperbolic growth when the frequency k is in the neighborhood of $\pm N/2$

$$P_Z[k] = P_W[k]\,|\widehat{u}^{-1}[k]|^2 = \sigma^2\,|\widehat{u}^{-1}[k]|^2 \sim \sigma^2 \left|\frac{2k}{N} - 1\right|^{-2p}. \tag{22}$$

This deconvolution problem is called a *hyperbolic deconvolution* of order p.

Section III-A shows that a wavelet basis does not have enough frequency resolution to obtain an efficient estimator for hyperbolic deconvolutions. A mirror wavelet basis is obtained by modifying the wavelet basis at high frequencies in order to adapt it to hyperbolic deconvolutions. The implementation and numerical performance of thresholding estimators in mirror wavelet bases is described in Section III-B.

A. Mirror Wavelets

Section II-B explains that wavelet bases are a priori a good candidate for the deconvolution of piecewise regular signals

because they efficiently approximate such signals with few nonzero coefficients, but they must also be well adapted to the noise spectrum. The resulting condition (19) implies that

$$\lambda_j = \frac{\max_{|k|\in[2^{-j-1}, 2^{-j}]} |\widehat{u}[k]|}{\min_{|k|\in[2^{-j-1}, 2^{-j}]} |\widehat{u}[k]|}$$

should remain of the order of 1 at all scales $2^j \leq 1$. For the low-pass filter (21), this is valid as long as $2^j < 2N^{-1}$. However, at the finest scale $2^j = 2^{L+1} = 2 N^{-1}$, the transfer function $|\widehat{u}[k]|$ varies by a huge factor on the highest frequency interval $[N/4, N/2-1]$, with $\lambda_{L+1} \sim (N/2)^p$. It implies that the finest scale wavelets $\psi_{L+1,q}$ do not have enough frequency resolution to efficiently concentrate the noise energy as a Fourier basis would do.

We now explain how to modify the wavelet basis to construct a basis that is well adapted to concentrate both the signal and the noise energy. At scales $2^j < 2N^{-1}$, the fact that λ_j is of the order of 1 indicates that the wavelets $\psi_{j,q}$ have enough frequency resolution, so we keep them as is. Over the highest frequency interval $[N/4, N/2]$, it is however necessary to replace the finest scale wavelets $\{\psi_{L+1,q}[n]\}_{1\leq q\leq N/2}$ by new vectors having a better frequency resolution. Let $\mathbf{W}_{L+1}$ be the space generated by $\{\psi_{L+1,q}\}_{0\leq q<N/2}$. The new vectors must also define an orthonormal basis of $\mathbf{W}_{L+1}$ and hence must have Fourier transforms that cover the highest frequency interval $[N/4, N/2]$. To guarantee that $|\widehat{u}[k]| \sim |(2k/N) - 1|^p$ does not vary by a factor larger than 2^p over the frequency support of each vector, this support should typically be included in an interval of the form $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$ for $0 \geq j > L+1$. But we also want to construct new vectors that are still well adapted to approximate efficiently piecewise regular signals f. This means producing few large amplitude coefficients at the location of sharp transitions. The new vectors must therefore have a spatial support as small as possible and hence a frequency support as large as possible. A solution is to build a new basis of $\mathbf{W}_{L+1}$ composed of vectors having a Fourier transform whose energy is spread as much as possible over each interval $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$ for $0 \geq j > L+1$, as illustrated in Fig. 2.

Mirror wavelets are defined from the original wavelets by

$$\widetilde{\psi}_{j,q}[n] = (-1)^{n-1}\psi_{j,q}[1-n]. \qquad (23)$$

Their Fourier transforms satisfy $|\widehat{\widetilde{\psi}}_{j,m}[k]| = |\widehat{\psi}_{j,q}[N/2 - k]|$, and thus have an energy mostly concentrated in $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$ as we wanted. We shall prove that the family of translated mirror wavelets

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

is an orthonormal basis of $\mathbf{W}_{L+1}$ by relating these mirror wavelets to the finest scale wavelets $\psi_{L+1,m}$, with a wavelet packet decomposition [17]. As a result, adding the largest scale wavelets to this family

$$\mathcal{B} = \left\{ \psi_{j,q},\, \widetilde{\psi}_{j,q} \right\}_{0\leq q<2^{-j},\, L+2\leq j\leq 1}$$

defines an orthonormal basis of $\mathbb{C}^N$, called a mirror *wavelet basis*.
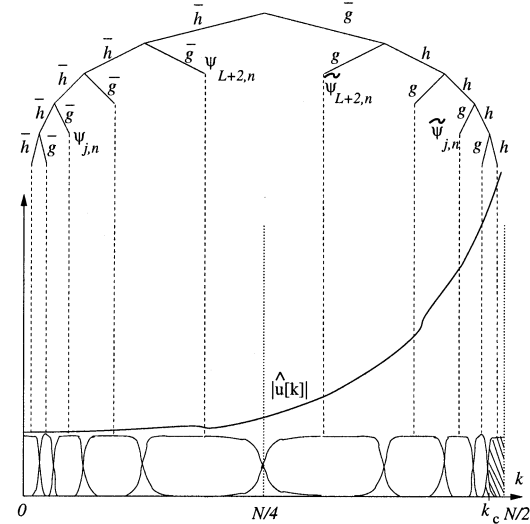

Fig. 2. A mirror wavelet basis is computed with a wavelet packet filter bank tree, where each branch corresponds to a convolution with the indicated filter followed by a subsampling by a factor 2. The resulting wavelets $\psi_{j,m}$ and mirror wavelets $\widetilde{\psi}_{j,m}$ have a Fourier transform shown below. The inverse filter $|\widehat{u}^{-1}[k]|$ has a hyperbolic growth but varies by a relatively small factor on the frequency support of each mirror wavelet. There exists a cut-off frequency k_c above which the variance of the deconvolved noise is too large and the thresholding sets all coefficients to zero.

Let us first recall that discrete orthonormal wavelets $\psi_j[n]$ are obtained with a cascade of circular convolutions of two finite impulse response conjugate mirror filters $h[n]$ and $g[n]$[16]. Their discrete Fourier transforms are given by

$$\widehat{\psi}_j[k] = \widehat{g}[2^{j-L-1}k] \prod_{p=0}^{j-L-2} \widehat{h}[2^p k] \quad \text{with } 2^{-L} = N. \qquad (24)$$

For a periodic signal $f[n]$, the discrete wavelet coefficients $d_j[q] = \langle f, \psi_{j,q}\rangle$ are computed with a fast filter bank algorithm that requires $O(N)$ operations. We initially set $a_L[n] = f[n]$ and define $\bar{g}[n] = g[-n]$ and $\bar{h}[n] = h[-n]$. Then, for any $L \leq j < 0$

$$a_{j+1}[q] = a_j \star \bar{h}[2q] \quad \text{and} \quad d_{j+1}[q] = a_j \star \bar{g}[2q] \qquad (25)$$

and $d_1[0] = a_0[0]$. The wavelet coefficients f are calculated with (25) at all scales $1 \geq 2^j \geq N^{-1}$ with $O(N)$ operations. The signal $a_j[n]$ provides an "approximation" of f at a scale 2^j. Let $\phi_j[n]$ be the signal whose Fourier transform is

$$\widehat{\phi}_j[k] = \prod_{p=0}^{j-L-1} \widehat{h}[2^p k]. \qquad (26)$$

One can verify that $a_j[q] = \langle f[n], \phi_j[n - 2^{j-L}q]\rangle$.

The $N/2$ finest scale wavelet coefficients $d_{L+1}[q]$ carry the highest frequencies of $f[n]$ essentially concentrated in the interval $[N/4, N/2]$. Mirror wavelet coefficients are obtained by applying a fast mirror wavelet transform to $d_{L+1}[q]$. This transform associates to an arbitrary initial signal $\widetilde{a}_{L+1}[q]$ a family of $L+1$ discrete signals $\{\widetilde{d}_j\}_{0\leq j\leq L+2}$ defined recursively by

$$\widetilde{d}_{j+1}[q] = \widetilde{a}_j \star g[2q] \quad \text{and} \quad \widetilde{a}_{j+1}[q] = \widetilde{a}_j \star h[2q]. \qquad (27)$$

If we set $\widetilde{a}_{L+1}[q] = d_{L+1}[q]$, since $g[n] = (-1)^{1-n} h[1 - n]$, one can derive from (25), (27) and (23) that the resulting coefficients are the mirror wavelet coefficients of f

$$\widetilde{d}_j[q] = \langle f, \widetilde{\psi}_{j,q} \rangle.$$

Mirror wavelets are thus obtained from the finest scale orthonormal wavelets with a cascade of filtering and subsampling with conjugate mirror filters, and are thus a particular instance of orthonormal wavelet packet basis [17]. General results on conjugate mirror filters and wavelet packets guarantee that this filter bank algorithm defines an orthonormal basis of the space $\mathbf{W}_{L+1}$ generated by the finest scale wavelets. Fig. 2 illustrates the filter bank algorithm that computes all wavelet and mirror wavelet coefficients with $O(N)$ operations.

The inverse mirror mirror wavelet transform is computed with the inverse wavelet packet algorithm. Let us denote

$$\check{x}[n] = \begin{cases} x\left[\frac{n}{2}\right], & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}.$$ (28)

For any $L + 1 \leq j < 0$, the coefficients $\widetilde{a}_j[q]$ are computed with

$$\widetilde{a}_j[q] = \check{\widetilde{a}}_{j+1} \star \bar{g}[q] + \check{\widetilde{d}}_{j+1} \star \bar{h}[q].$$ (29)

This reconstructs the finest scale wavelet coefficients $d_{L+1}[q] = \widetilde{a}_{L+1}[q]$. Then the signal f is reconstructed with the fast inverse wavelet transform algorithm

$$a_j[q] = \check{a}_{j+1} \star h[q] + \check{d}_{j+1} \star g[q] \text{ for } L \leq j < 0.$$ (30)

This inverse transform thus reconstructs $f[n] = a_0[n]$ with $O(N)$ operations.

B. Thresholding in a Mirror Wavelet Basis

This section describes the implementation of a thresholding estimator in a mirror wavelet basis and gives numerical results. Following the general definition (5), the thresholding estimator decomposes the deconvolved signal $X = Y \star u^{-1}$ in a mirror wavelet basis and thresholds its coefficients

$$\widetilde{F} = D_t X = \sum_{j=L+2}^{1} \sum_{q=0}^{2^{-j}-1} \rho_{T_j} (\langle X, \psi_{j,q} \rangle) \psi_{j,m}$$

$$+ \sum_{j=L+2}^{1} \sum_{q=0}^{2^{-j}-1} \rho_{\widetilde{T}_j} \left(\langle X, \widetilde{\psi}_{j,q} \rangle \right) \widetilde{\psi}_{j,m}.$$ (31)

The thresholds are proportional to the noise variance, as long as this one is not above the largest possible signal coefficient. The variance $\sigma_j^2 = \mathbf{E}\{|\langle Z, \psi_{j,q} \rangle|^2\}$ does not depend upon the position index q and following (11) we set

$$T_j = \sigma_j \sqrt{2 \log_e N}.$$ (32)

Since $|\widehat{u}[k]| \sim |(2k/N) - 1|^p$, one can verify with (14) that σ_j^2 remains of the order of the variance σ^2 of the white noise W. In most applications, σ^2 is not too large so that the resulting thresholds T_j are indeed below the maximum amplitude of wavelet coefficients.

On the opposite, the variance $\widetilde{\sigma}_j^2 = \mathbf{E}\{|\langle Z, \widetilde{\psi}_{j,q} \rangle|^2\}$ of mirror wavelet coefficients becomes very large when the scale 2^j in-

creases. Since the support of $\widehat{\widetilde{\psi}_{j,q}}[k]$ is essentially concentrated in the interval $[N/2 - 2^{-j}, N/2 - 2^{-j-1}]$, we derive from (14) and (22) that

$$\widetilde{\sigma}_j^2 \sim P_Z\left[\frac{N}{2} - 2^{-j}\right] \sim \sigma^2 \left|N 2^j\right|^{2p}.$$

Let $s_j = \sup_{f \in \Theta} |\langle f, \widetilde{\psi}_{j,m} \rangle|$ be the maximum amplitude of mirror wavelet coefficients for signals in a set Θ. According to (11) the thresholds are defined by

$$\widetilde{T}_j = \begin{cases} \widetilde{\sigma}_j \sqrt{2 \log_e N}, & \text{if } \widetilde{\sigma}_j \sqrt{2 \log_e N} \leq s_j \\ \infty, & \text{otherwise} \end{cases}.$$ (33)

Since $\widetilde{\sigma}_j$ increases quickly when the scale increases, when 2^j is above a critical scale 2^c we have $\widetilde{T}_j = \infty$ which sets to zero all mirror wavelet coefficients at this scale. The corresponding mirror wavelets have their frequency support above a cut-off frequency k_c, as shown in Fig. 2.

The risk produced by the mirror wavelet thresholding estimator (31) can be reduced with several procedures which currently lack of theoretical foundations, but bring significant improvements.

- The hard thresholding function ρ_T defined in (6) can be replaced by a soft thresholding function introduced by Donoho and Johnstone [10]:

$$\rho_T(x) = \begin{cases} x - T, & \text{if } x > T \\ x + T, & \text{if } x < -T \\ 0, & \text{if } |x| \leq T \end{cases}.$$ (34)

  This guarantees with a high probability that the thresholded coefficients have an amplitude below the amplitude of the original signal coefficients. Hence, the estimated signal $\widetilde{F}$ is at least as regular as the original one.

- Choosing thresholds T_j and $\widetilde{T}_j$ having an amplitude smaller than the theoretical values (32) and (33) can reduce the estimation risk. These thresholds are thus multiplied by a factor $\beta < 1$, and for a soft thresholding we typically have $\beta \sqrt{2 \log_e N} = 2$.

- The time origin is set arbitrarily, yet it defines the location of wavelets and mirror wavelets that are centered at points $2^j q$. To remove a stationary noise, Coifman and Donoho [18] showed that a thresholding estimator is improved with a translation invariant procedure which avoids this grid artifact. Let D_t be the thresholding operator in the mirror wavelet basis. The N periodic signal $X[n]$ is translated $X_l[n] = X[n - l]$ for $0 \leq l < N$. We compute a thresholding estimation $D_t X_l$ of each translated signal and perform an averaging after an inverse translation

$$\widetilde{F} = \frac{1}{N} \sum_{l=0}^{N-1} D_t X_l[n + l].$$ (35)

  In wavelet and wavelet packet bases, which are partially translation invariant, Coifman and Donoho [18] give a fast translation-invariant filter bank algorithm which requires $O(N \log N)$ operations to compute $\widetilde{F}$.

- The error $\widetilde{F} - f$ can be considered as a residual noise, after the mirror wavelet thresholding estimation. This noise has a power spectrum which is nearly flat at high frequencies,
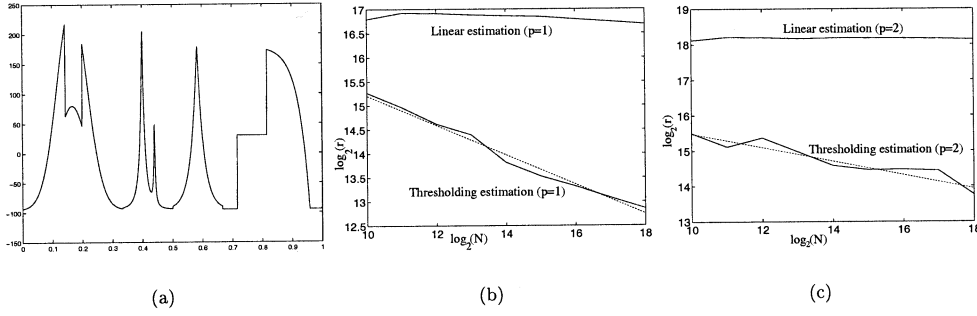
Fig. 3. (a): Piecewise regular signal. (b,c): For the signal (a) sampled at intervals $1/N$, if $N\sigma^2$ remains constant, the risk of an optimized linear deconvolution is nearly constant. For a thresholding deconvolution in a mirror wavelet basis, $\log_2 r_t(f)$ decays linearly as a function of $\log_2 N$, with a slope that depends on the degree p of the zeros of $\hat{u}[k]$ at $k = \pm N/2$.

as opposed to the original deconvolved noise $Z = X - f$. This residual noise can be further reduced by thresholding the coefficients of $\widetilde{F}$ over the finest scale wavelets $\psi_{L+1,q}$. The resulting thresholding estimator is calculated with a hard thresholding function ρ_{T_h}

$$\widetilde{F}_b = \sum_{j > L+1, m} \langle \widetilde{F}, \psi_{j,q} \rangle \psi_{j,q}$$

$$+ \sum_{q=0}^{N/2-1} \rho_{T_h}\left(\langle \widetilde{F}, \psi_{L+1,q}\rangle\right) \psi_{L+1,q}. \quad (36)$$

This removes some oscillations produced by mirror wavelets in the neighborhood of singularities. A soft thresholding as opposed to a hard thresholding would add an undesirable smoothing. The value of the hard threshold T_h is chosen to be proportional to the standard deviation σ_h of the finest scale wavelet coefficients $\left\{\langle \widetilde{F}, \psi_{L+1,q}\rangle\right\}_{0 \le q < N/2}$ calculated for $f = 0$.

In summary, the restoration algorithm is decomposed in the following steps:

```
1)Deconvolution: X = Y ⋆ ũ⁻¹.
2) For all 0 ≤ l < N, let X_l[n] = X[n − l]
  a) Decomposition of X_l in the mirror
wavelet basis B with the filter bank al-
gorithm of Fig. 2.
  b) Computation of D_t X_l by soft thresh-
olding the mirror wavelet coefficients of
X_l.
3) Computation of the translation-in-
variant estimator F̃ = (1/N)∑_{l=0}^{N−1} D_t X_l[n+l].
4)Hard thresholding of the finest scale
wavelet coefficients of F̃ with (36).
```

With a fast implementation of the translation invariant loop, the overall complexity of this algorithm is $O(N \log N)$ for a signal of size N.

Piecewise regular signals can be modeled as signals whose total variations are bounded by a constant C:

$$\Theta = \left\{ f[n] : \|f\|_{tv} = \sum_{n=0}^{N-1} |f[n] - f[n-1]| \le C \right\}.$$

If W is a Gaussian white noise, we prove in [12] that a thresholding in a mirror wavelet basis has a maximum risk that is asymptotically of the same order as the minimax risk over Θ, and much below the maximum risk obtained by any linear estimation [19].

To see the evolution of the risk as a function of the signal size, we define $f[n] = f(n/N)$ for $0 \le n < N$, where $f(x)$ is a bounded variation function defined over $[0, 1]$. Fig. 3(a) gives an example. Observe that modifying N does not change the total variation $\|f\|_{tv}$. The noise variance σ^2 is normalized so that $\mathbf{E}\{\|W\|^2\} = N\sigma^2$ remains constant. Fig. 3(b)-(c) gives the risk $r = \|F - f\|^2$ for two hyperbolic deconvolution problems, where the transfer functions $\hat{u}[k]$ have respectively $p = 1$ and $p = 2$ zeros at $k = \pm N/2$. The top curves in Fig. 3(b)-(c) give the values of the deconvolution risk obtained with linear estimators that are optimized in order to minimize the maximum risk over Θ. Since Θ is translation invariant and the noise is stationary, these estimators are also translation invariant and are therefore convolutions. We prove in [12] that this minimum linear risk remains nearly constant when N increases, if $N\sigma^2$ remains constant. On the contrary, it is proved in [12] that the risk r_t of a mirror wavelet thresholding is such that $\log_2(r_t)$ decreases when $\log_2 N$ increases, with a slope equal to $-1/(2p+1)$. The numerical results of Fig. 3(b)-(c) verify closely these theoretical predictions. Indeed, we observe that the risk of optimal linear estimators remain nearly constant whereas the risk of mirror wavelet thresholding estimators decrease with a slope of $-0.32 \approx -1/(2p+1)$ for $p = 1$, and a slope of $-0.18 \approx -1/(2p+1)$ for $p = 2$.

IV. IMAGE DEBLURRING

The diffraction of the optics of a camera produces a blur that is equivalent to a low-pass filter that attenuates the highest image frequencies. Moreover, the electronics of the photo-receptors add a noise, so we measure

$$Y[n_1, n_2] = f \star u[n_1, n_2] + W[n_1, n_2] \quad (37)$$

where $W[n_1, n_2]$ is typically a white noise of variance σ^2. Inverting this degradation is particularly important for satellite images, to optimize their resolution. 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 noises. The satellite

movement thus produces another blur that is combined with the diffraction of the optics [20]. Calibration procedures can compute the system transfer function u and the noise variance σ^2. The image Fig. 5(b) is a simulated satellite image provided by the French spatial agency (CNES). It is calculated from an airplane image shown in Fig. 5(a) with the impulse response calibrated on a new observatory satellite.

In the following, we shall concentrate on the particular case where the impulse response is a separable low-pass filter

$$u[n_1, n_2] = u_1[n_1]\, u_2[n_2] \tag{38}$$

and suppose that the discrete Fourier transforms of u_1 and u_2 have respectively a zero of order p_1 and p_2 at $\pm N/2$

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

This assumption is valid for many satellite systems.

For $k = (k_1, k_2)$, the pseudo-inverse filter is defined like in (2) and the deconvolved noise $Z[n_1, n_2] = W \star u^{-1}[n_1, n_2]$ has a power spectrum

$$\begin{aligned} P_Z[k_1, k_2] &= P_W[k_1, k_2]\, |\widehat{u}^{-1}[k_1, k_2]|^2 \\ &= \frac{\sigma^2}{|\widehat{u}_1[k_1]|^2\, |\widehat{u}_2[k_2]|^2} \\ &\sim \sigma^2 \left| \frac{2k_1}{N} - 1 \right|^{-2p_1} \left| \frac{2k_2}{N} - 1 \right|^{-2p_2}. \end{aligned} \tag{39}$$

Section IV-A applies the results of Section II-B to construct a mirror wavelet basis well adapted to perform such a hyperbolic deconvolution on images. Numerical implementations and examples are given in Section IV-B.

A. Separable Mirror Wavelet Bases

To minimize the risk of a thresholding estimator, Section II-B explains that the basis must be composed of vectors which concentrate efficiently the noise and the signal energy over few coefficients. Since the noise is stationary, the ideal basis is the two-dimensional discrete Fourier basis which diagonalizes its covariance. However, a Fourier basis does not provide an efficient representation of typical images that include edges. Wavelet bases are particularly efficient to construct sparse representations of images, hence their use in the JPEG-2000 compression standard. Like in one-dimension, a thresholding algorithm in a wavelet basis gives disastrous results for a hyperbolic deconvolution because the finest scale wavelets have a Fourier transform that are not sufficiently well localized. We thus replace these fine scale wavelets with separable mirror wavelets.

A separable wavelet basis of $\mathbb{C}^{N^2}$ is constructed with separable products of discrete periodic scaling signals ϕ_j defined in (26) and the corresponding discrete periodic wavelet ψ_j defined in (24). At each scale 2^j, there are three wavelets

$$\begin{aligned} \psi_j^1[n_1, n_2] &= \phi_j[n_1]\, \psi_j[n_2] \\ \psi_j^2[n_1, n_2] &= \psi_j[n_1]\, \phi_j[n_2] \\ \psi_j^3[n_1, n_2] &= \psi_j[n_1]\, \psi_j[n_2]. \end{aligned}$$

Let $\psi_{j,q_1,q_2}^\alpha[n_1, n_2] = \psi_j^\alpha[n_1 - 2^{j-L}q_1,\, n_2 - 2^{j-L}q_2]$ and $\psi_1^0[n_1, n_2] = 1/N$. The separable wavelet family

$$\left\{ \psi_1^0,\ \psi_{j,q_1,q_2}^1,\ \psi_{j,q_1,q_2}^2,\ \psi_{j,q_1,q_2}^3 \right\}_{L < j \leq 1,\, 0 \leq q_1, q_2 < 2^{-j}}$$

is an orthonormal basis of $\mathbb{C}^{N^2}$, with $L = -\log_2 N$.

Section II-B shows in (17) that a basis concentrates the noise energy nearly as well as the Fourier basis, if over the frequency support of each of its vectors the amplitude of the transfer function

$$|\widehat{u}[k_1, k_2]| \sim \left| \frac{2k_1}{N} - 1 \right|^{p_1} \left| \frac{2k_2}{N} - 1 \right|^{p_2}$$

has a relative variation bounded by a constant λ that is of the order of 1. At scales $2^j > 2^{L+1}$, the lower frequency wavelets

$$\mathcal{B}_0 = \left\{ \psi_1^0,\ \psi_{j,q_1,q_2}^1,\ \psi_{j,q_1,q_2}^2,\ \psi_{j,q_1,q_2}^3 \right\}_{L+1 < j \leq 1,\, 0 \leq q_1, q_2 < 2^{-j}} \tag{40}$$

have a Fourier transform mostly concentrated in the lower frequency square $[-N/4, N/4]^2$, where $|\widehat{u}[k_1, k_2]|$ remains nearly constant. On the opposite, the remaining $3N^2/4$ finer scale wavelets

$$\left\{ \begin{aligned} \psi_{L+1,q_1,q_2}^1[n_1, n_2] &= \phi_{L+1,q_1}[n_1]\, \psi_{L+1,q_2}[n_2] \\ \psi_{L+1,q_1,q_2}^2[n_1, n_2] &= \psi_{L+1,q_1}[n_1]\, \phi_{L+1,q_2}[n_2] \\ \psi_{L+1,q_1,q_2}^3[n_1, n_2] &= \psi_{L+1,q_1}[n_1]\, \psi_{L+1,q_2}[n_2] \end{aligned} \right\}_{0 \leq q_1, q_2 < N/2} \tag{41}$$

have a Fourier transform 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 $|\widehat{u}[k_1, k_2]|$ varies by huge factors. It is therefore necessary to replace these finest scale wavelets by other vectors that have a smaller frequency support inside which $|\widehat{u}[k_1, k_2]|$ varies by a factor of the order of 1. This was first observed by Rougé [20] who proposed to use wavelet packets for the deconvolution of images.

Since $\widehat{u}[k_1, k_2] = \widehat{u}_1[k_1]\, \widehat{u}_2[k_2]$, where $\widehat{u}_1[k]$ and $\widehat{u}_2[k]$ have a zero at the highest frequency $k = \pm N/2$, we use the results of Section III-A to construct a separable mirror wavelet basis which segments the horizontal and vertical frequency axes into intervals where these transfer functions vary by relatively small factors. Let $\mathbf{W}_{L+1}^2$ be the space generated by the $3N^2/4$ finest scale wavelets (41). To construct a mirror wavelet basis of $\mathbf{W}_{L+1}^2$ we shall use the one-dimensional mirror wavelets $\widetilde{\psi}_{j,q}[n]$ constructed in (23). Let us remind from Section III-A that for one-dimensional signals, the finest scale wavelets $\{\psi_{L+1,q}[n]\}_{1 \leq q \leq N/2}$ and the one-dimensional mirror wavelets $\{\widetilde{\psi}_{j,q}[n]\}_{L+2 \leq j \leq 1,\, 0 \leq q < 2^{-j}}$ are two orthonormal bases of the same space. We may thus perform an orthogonal change of basis by replacing the one-dimensional finest scale wavelets that appear in (41) by the corresponding family of mirror wavelets. To simplify notations, we shall write the one-dimensional low-frequency scaling signal $\phi_{L+1,q}[n] = \widetilde{\psi}_{L+1,q}[n]$. This orthogonal change of basis defines a new orthonormal
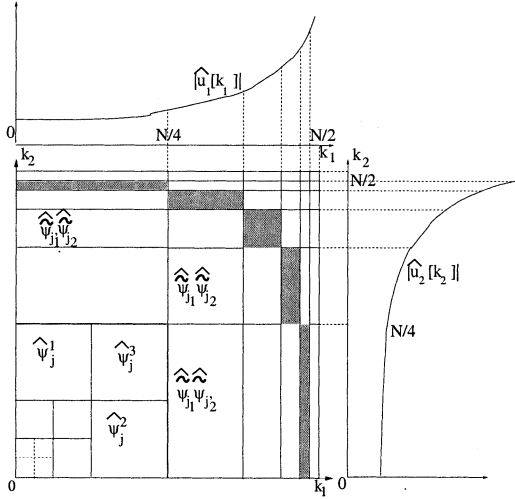
Fig. 4.   An inverse filter $|\widehat{u}^{-1}[k_1, k_2]|$ is the product of $|\widehat{u}^{-1}[k_1]|$ by $|\widehat{u}^{-1}[k_2]|$ which have a hyperbolic growth as shown by the curves on the top and on the left. The separable mirror wavelets (42) segments the frequency plane (k_1, k_2) into rectangles over which $|\widehat{u}^{-1}[k_1, k_2]|$ varies by a relatively small factor. The lower frequencies are covered by separable wavelets $\widehat{\psi}_j^\alpha[k_1, k_2]$, and the higher frequencies are covered by separable mirror wavelets $\widehat{\widetilde{\psi}}_{j_1}[k_1]\,\widehat{\widetilde{\psi}}_{j_2}[k_2]$. The grey squares correspond to critical frequencies beyond which the thresholding sets all coefficients to zero.

basis of $\mathbf{W}_{L+1}^2$ composed of $3N^2/4$ separable mirror wavelets defined by

$$\mathcal{B}_1 = \left\{ \widetilde{\psi}_{j_1, q_1}[n_1]\, \widetilde{\psi}_{j_2, q_2}[n_2] \right\}_{\substack{L+1 \le j_1, j_2 \le 1,\, 0 \le q_1 < 2^{-j_1},\, 0 \le q_2 < 2^{-j_2} \\ (j_1, j_2) \ne (L+1, L+1)}}.$$
(42)

The union $\mathcal{B} = \mathcal{B}_0 \cup \mathcal{B}_1$ of low frequency wavelets and high frequency mirror wavelets is therefore an orthonormal basis of $\mathbb{C}^{N^2}$. This two-dimensional mirror wavelet basis segments the Fourier plane into rectangles illustrated in Fig. 4. It is a particular instance of anisotropic wavelet packet basis as defined in [21], 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 which requires $O(N^2)$ operations for an image of N^2 pixels. Let us denote the wavelet and mirror coefficients

$$d_j^\alpha[q_1, q_1] = \langle f, \psi_{j, q_1, q_2}^\alpha \rangle \text{ and } \widetilde{d}_{j_1, j_2}[q_1, q_1] = \langle f, \widetilde{\psi}_{j_1, q_1} \widetilde{\psi}_{j_2, q_2} \rangle.$$
(43)

Let us first compute the two-dimensional wavelet transform. We write a separable product of filters $hg[n_1, n_2] = h[n_1]\, g[n_2]$. We initiate $a_L[n_1, n_2] = f[n_1, n_2]$ and for $L \le j < 0$ the fast wavelet transform computes

$$d_{j+1}^1[q_1, q_2] = a_j \star \overline{h}\overline{g}[2q_1, 2q_2]$$
$$d_{j+1}^2[q_1, q_2] = a_j \star \overline{g}\overline{h}[2q_1, 2q_2]$$
$$d_{j+1}^3[q_1, q_2] = a_j \star \overline{g}\overline{g}[2q_1, 2q_2]$$
(44)

and

$$a_{j+1}[q_1, q_2] = a_j \star \overline{h}\overline{h}[2q_1, 2q_2].$$

The mirror wavelet coefficients are calculated by applying the fast one-dimensional mirror wavelet transform along

(a)

(b)

(c)

(d)

Fig. 5.   (a) Original airplane image. (b) Simulation of a satellite image provided by the CNES ($PSNR = 27.6$ db). (c) Deconvolution with a translation invariant thresholding in a mirror wavelet basis ($PSNR = 30.6$ db). (d) Deconvolution calculated with a circular convolution optimized for satellite images ($PSNR = 29.8$ db).

the rows and columns of the images of finest scale wavelet coefficients d_{L+1}^α for $\alpha = 1, 2, 3$. The fast mirror wavelet transform is the filter bank algorithm (27) that associates to any one-dimensional signal $\widetilde{a}_{L+1}[q]$ of size $N/2$ a family of $L+1$

one-dimensional discrete signals $\{\widetilde{d}_j[q]\}_{0 \leq j \leq L+2}$ of sizes 2^{-j}. Applying the one-dimensional mirror wavelet transform along each column of the wavelet coefficient image $d^1_{L+1}[q_1, q_2]$ divides each column into $L + 1$ one-dimensional signals of sizes 2^{-j_2} for $0 \leq j_2 \leq L+2$. One can verify that aggregating together theses signals along the $N/2$ columns yields $L + 1$ images of mirror wavelet coefficients $\{\widetilde{d}_{L+1,j_2}\}_{0 \leq j_2 \leq L+2}$ of 2^{-j_2} rows and $N/2$ columns. Similarly, applying the one-dimensional mirror wavelet transform along each row of the wavelet coefficient image $d^2_{L+1}[q_1, q_2]$ yields $L + 1$ images of mirror wavelet coefficients $\{\widetilde{d}_{j_1,L+1}\}_{0 \leq j_1 \leq L+2}$ of 2^{-j_1} columns and $N/2$ rows. Let us now apply the one-dimensional mirror wavelet transform along each row of the wavelet coefficient image $d^3_{L+1}[q_1, q_2]$. It yields $L + 1$ images $\{\widetilde{d}^3_{j_1,L+1}\}_{0 \leq j_1 \leq L+2}$ of 2^{-j_1} columns and $N/2$ rows. For each j_1, applying the one-dimensional mirror wavelet transform along each column of the image $d^3_{j_1,L+1}[q_1, q_2]$ yields $L + 1$ images of mirror wavelet coefficients $\{\widetilde{d}_{j_1,j_2}\}_{0 \leq j_2 \leq L+2}$, of 2^{-j_1} columns and 2^{-j_2} rows. This algorithm computes all mirror wavelet coefficients with $O(N^2)$ operations.

The original image is reconstructed from its mirror wavelet coefficients by first computing the finest scale wavelet coefficients d^α_{L+1} for $\alpha = 1, 2, 3$ by inverting each one-dimensional mirror wavelet transform using the inverse algorithm defined in (29). Once all wavelet coefficients are calculated, the image $f[n_1, n_2]$ is recovered with the fast separable two-dimensional inverse wavelet transform [16], which also requires $O(N^2)$ operations.

B. Thresholding for Image Restoration

The thresholding estimation in the separable mirror wavelet basis is a direct extension of the one-dimensional mirror wavelet thresholding algorithm in Section III-B. It decomposes the deconvolved signal $X = Y \star u^{-1}$ in a separable mirror wavelet basis and thresholds its coefficients.

Over the $N^2/4$ lower frequency wavelets ψ^α_{j,q_1,q_2}, the noise Z has a variance that is independent on the position index (q_1, q_2) and which remains of the order of σ^2

$$\sigma^2_{j,\alpha} = \mathbf{E}\{|\langle Z, \psi^\alpha_{j,q_1,q_2} \rangle|^2\} \sim \sigma^2.$$

These variances are typically much below the maximum amplitude of the signal wavelet coefficients, so according to (11) the resulting thresholds are

$$T_{j,\alpha} = \sigma_{j,\alpha} \sqrt{2 \log_e N^2}. \tag{45}$$

For mirror wavelets, the noise variance is

$$\widetilde{\sigma}^2_{j_1,j_2} = \mathbf{E}\{|\langle Z, \widetilde{\psi}_{j_1,q_1} \widetilde{\psi}_{j_2,q_2} \rangle|^2\}.$$

For any set Θ of images, we compute $s_{j_1,j_2} = \sup_{f \in \Theta} |\langle f, \widetilde{\psi}_{j_1,q_1} \widetilde{\psi}_{j_2,q_2} \rangle|$. According to (11), the thresholds are

$$\widetilde{T}_{j_1,j_2} = \begin{cases} \widetilde{\sigma}_{j_1,j_2} \sqrt{2 \log_e N^2}, & \text{if } s_{j_1,j_2} \leq \widetilde{\sigma}_{j_1,j_2} \sqrt{2 \log_e N^2} \\ \infty, & \text{otherwise.} \end{cases} \tag{46}$$

Since $\widetilde{\sigma}_{j_1,j_2}$ increases when 2^{j_1} and 2^{j_2} increase, we shall have $\widetilde{T}_{j_1,j_2} = \infty$ when 2^{j_1} or 2^{j_2} are above critical scales. This defines critical frequencies illustrated by the gray area of Fig. 4,

above which the thresholding algorithm sets all coefficients to zero, removing the corresponding image highest frequencies.

Like in one dimension, the risk of the thresholding estimator is reduced with several procedures.

- The thresholding of mirror wavelet coefficients is performed with the soft thresholding function (34).
- The amplitudes of the thresholds (45) and (46) are reduced by multiplying them with a factor $\beta < 1$.
- A translation invariant estimation is implemented like in (35) by translating the deconvolved image $X_{l_1,l_2}[n_1, n_2] = X[n_1 - l_1, n_2 - l_2]$, computing the thresholding estimation, making an inverse translation and averaging the N^2 estimates for all translation parameters. A fast filter bank implementation of the translation invariant procedure [18] requires $O(N^2 \log N)$ operations.
- The mirror wavelet thresholding estimation $\widetilde{F}$ can include fine scale oscillations produced by mirror wavelets in the neighborhood of edges. Such oscillations are attenuated by hard thresholding on the wavelet coefficients $\langle \widetilde{F}, \psi^\alpha_{L+1,q_1,q_2} \rangle$ at the finest scale 2^{L+1}, for all positions (q_1, q_2) and $\alpha = 1, 2, 3$.

The thresholding algorithm in the separable mirror wavelet basis thus follows the same steps as the algorithm described in Section III-B for one dimensional signals. It requires $O(N^2 \log N)$ with the fast filter bank implementation of the translation invariant procedure [18].

Fig. 5(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 Fig. 5(b). It was calculated by the CNES with a low-pass filter calibrated on one of its satellite, which has 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 256. Fig. 5(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 Fig. 5(d), calculated with a circular convolution estimator, and optimized to minimize the risk over satellite images. The linear deconvolution sharpens the image but leaves a visible noise in the regular parts of the image. The thresholding algorithm removes completely the noise in these regions while improving the restoration of edges and oscillatory parts. This thresholding algorithm was benchmarked as superior to all competing algorithms by photo-interpreters of the French spatial agency (CNES) for the deconvolution of satellite images. Deconvolution procedures minimizing different energies including a regularization term [5], [6], [13], [22] as well as thresholding algorithms in wavelet bases [14] have been tested.

To check that the parameters of the mirror wavelet restoration algorithm were not optimized for a specific type of image or distortion operator, Fig. 6 shows the result for a different image. The image is degraded by a convolution with a different low-pass filter, and by the addition of a Gaussian white noise of variance $\sigma = 1$. Once again, edges and high frequency textures are restored. The PSNR of the degraded image is 26.6 dB whereas the PSNR of the restored image is 31.2 dB. These nu-

(a)

(b)

(c)

Fig. 6. (a) Original image. (b) Degraded image by a low-pass filter and the addition of a Gaussian white noise ($PSNR = 26.6$ db). (c) Restored image with a translation invariant thresholding in a mirror wavelet basis ($PSNR = 31.2$ db).

merical results are supported by a mathematical study in [12] that considers a set Θ of images having a bounded total variation. The images in Figs. 5(a) and 6(a) belong to such a set Θ. For a hyperbolic deconvolution with a Gaussian white noise W, we prove in [12] that a thresholding estimator in a separable mirror wavelet basis produces a maximum risk over Θ that is asymptotically equivalent to the minimax risk.

V. CONCLUSION

In presence of a stationary additive noise, a deconvolution problem is equivalent to the removal of a colored stationary noise whose power spectrum may be very large at certain frequencies. Thresholding estimators decompose the noisy signal in an orthonormal basis and threshold the resulting coefficients. This paper studies the optimization of the basis in order to min-

imize the estimation risk of a thresholding estimator. We show that the basis must concentrate the noise energy and concentrate the signal energy over few basis coefficients.

Many digital measuring devices produce discrete signals whose frequencies are attenuated by a low-pass filter that vanishes at the highest frequencies. In this case, wavelet bases do not have enough frequency resolution to restore the signal high frequencies. For one-dimensional signals and images, mirror wavelet bases are designed to optimize such a deconvolution, with an application to the restoration of satellite images.

REFERENCES

[1] M. Bertero, "Linear inverse and ill-posed problems," in *Advances in Electronics and Electron Physics*. New York: Academic, 1989.

[2] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Recognit. Machine Intell.*, vol. 14, pp. 367–383, Mar. 1992.

[3] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

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

[5] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, no. 2, pp. 167–188, 1997.

[6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298–311, Feb. 1997.

[7] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 2024–2036, 1989.

[8] J. Idier, "Convex half-quadratic criteria and interacting auxilary variables for image restoration," *IEEE Trans. Image Processing*, vol. 10, pp. 1001–1009, July 2001.

[9] A. K. Katsaggelos, Ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991.

[10] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, Dec. 1994.

[11] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions," *J. Appl. Comput. Harmon. Anal.*, vol. 2, no. 2, pp. 101–126, 1995.

[12] J. Kalifa and S. Mallat, "Thresholding estimators for linear inverse problems and deconvolutions," Ann. Statist., 2002, to be published.

[13] C. Heinrich and G. Demoment, "Minimization of strictly convex functions: An improved optimality test based on fenchel duality," *Inv. Probl.*, vol. 16, pp. 795–810, 2000.

[14] J. L. Starck and A. Bijaoui, "Filtering and deconvolution by the wavelet transform," *Signal Process.*, vol. 35, pp. 195–211, 1994.

[15] I. M. Johnstone and B. W. Silverman, *Wavelet Threshold Estimators for Data With Correlated Noise*. Stanford, CA: Dept. Statistics, Stanford Univ., 1995.

[16] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.

[17] R. R. Coifman *et al.*, "Wavelet analysis and signal processing," in *Wavelets and their Applications*, B. R. Ruskai *et al.*, Eds. Boston, MA: Jones and Barlett, 1992, pp. 153–178.

[18] R. R. Coifman and D. Donoho, "Translation invariant denoising," in *Wavelets and Statistics*. New York: Springer-Verlag, 1995, p. 125.

[19] J. Kalifa, S. Mallat, and B. Rougé, "Minimax solution of inverse problems and deconvolution," *Proc. SPIE*, pp. 42–57, July 1999.

[20] B. Rougé, "Théorie de la Chaine Image Optique et Restauration," Ph.D., Université Paris-Dauphine, Paris, France, 1997.

[21] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. New York: Peters, 1994.

[22] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE ICIP*, Nov 1994, pp. 31–35.

Jérôme Kalifa was born in Paris, France. He graduated from Université Dauphine, Paris, in 1995 and received the Ph.D. in applied mathematics from Ecole Polytechnique, Paris, in 1999.

He joined the Department of Biomedical Engineering at Columbia University, New York, in 1999, as a Postdoctoral Fellow. In 2001, he co-founded and joined Let It Wave, a startup company based in Paris and specialized in geometrical image processing. His research interests include harmonic analysis, inverse problems and nonparametric estimation in signal processing, as well as medical and spatial image processing.

Bernard Rougé is a Researcher with CMLA, ENS Cachan, Paris, France, and Centre National d'Etudes Spatiales (French Space Agency). He works on the acquisition chain of satellite images, which includes problems such as sampling, restoration and image matching. He leads a team of researchers and engineers working on the design of future optical instruments for European satellites.

Mr. Rougé is a member of GRETSI, ESIPCO, and PSIP.

Stéphane Mallat was born in Paris, France. He graduated from Ecole Polytechnique in 1984 and from Ecole Nationale Supérieure des Télécommunications, Paris, in 1985. He received the Ph.D. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1988.

In 1988, he joined the Computer Science Department of the Courant Institute of Mathematical Sciences at New York University, became Associate Professor in 1993, and is now a Research Professor. In the fall of 1994, he was a Visiting Professor in the Electrical Engineering Department at the Massachusetts Institute of Technology, Cambridge, and in the spring of 1994 in the Applied Mathematics Department at the University of Tel Aviv, Israel. Since 1995, he has been a Professor in the Applied Mathematics Department at Ecole Polytechnique, Paris. His research interest include computer vision, signal processing, and diverse applications of wavelet transforms. He is the author of the book *A Wavelet Tour of Signal Processing* (New York: Academic, 1998).

Dr. Mallat received the 1990 IEEE Signal Processing Society's paper award, the 1993 Alfred Sloan fellowship in Mathematics, the 1997 Outstanding Achievement Award from the SPIE Optical Engineering Society, and the 1997 Blaise Pascal Prize in applied mathematics, from the French Academy of Sciences.