Invited Paper

Tomographic reconstruction with non-linear diagonal estimators

Jérôme Kalifa[a], Andrew Laine[a] and Peter D. Esser[b]

[a]Department of Biomedical Engineering, Columbia University, New York, NY
[b]Department of Radiology, Columbia-Presbyterian Medical Center, New York, NY

ABSTRACT

In tomographic reconstruction, the inversion of the Radon transform in the presence of noise is numerically unstable. Reconstruction estimators are studied where the regularization is performed by a thresholding in a wavelet or wavelet packet decomposition. These estimators are efficient and their optimality (minimax sense for bounded variation images) can be established when the decomposition provides a near-diagonalization of the inverse Radon transform operator and a compact representation of the object to be recovered. Several new estimators are investigated in different decompositions. First numerical results already exhibit a strong metrical and perceptual improvement over current reconstruction methods. These estimators are implemented with fast non-iterative algorithms, and are expected to outperform Filtered Back-Projection and iterative procedures for PET, SPECT and X-ray CT devices.

Keywords: Global tomographic reconstruction, wavelets and wavelet packets, bounded variation images, minimax optimality.

1. INTRODUCTION

1.1. Tomographic reconstruction

We are interested in the recovery of an image f from its tomographic projections Y, also called sinograms, and defined as:

$$Y = \mathcal{R}f + W \tag{1}$$

where $f[n_1, n_2] \in \mathbb{C}^{N^2}$ is the observed image, W is an additive noise, usually modeled as Gaussian or Poisson, and $\mathcal{R}$ is the discrete Radon transform. The discrete Radon transform is derived from its continuous version $\mathcal{R}_c$, which is equivalent to the X-ray transform in two dimensions and is defined as [1]

$$(\mathcal{R}_c f_c)(t, \alpha) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_c(x_1, x_2)\delta(x_1 \cos\alpha + x_2 \sin\alpha - t)dx_1 dx_2. \tag{2}$$

where $f_c(x_1, x_2) \in \mathbf{L}^2(\mathbb{R}^2)$, δ is the Dirac mass, $\alpha \in [0, 2\pi)$, and $t \in \mathbb{R}$. In the discrete Radon transform, a line integral along $x_1 \cos\alpha + x_2 \sin\alpha = t$ can be approximated by a summation of the pixel values inside the strip $t - 1/2 \leq n_1 \cos\alpha + n_2 \sin\alpha < t + 1/2$.

Tomographic reconstruction is ubiquitous in medical imaging. Imaging devices such as X-ray CTs, Positron Emission Tomography (PET), Single Positron Emission Computerized Tomography (SPECT) measure the density or the metabolic activity of a section of the patient's body (ie, roughly speaking, produce sinograms Y), and an estimation of the image f representing the observed section is derived by a tomographic reconstruction procedure from the sinograms.

A direct inversion of the Radon transform is computed with an amplification of high frequency components of Y in the direction of t, followed by a spatial interpolation. A fundamental difficulty of tomographic reconstruction comes from the fact that the Radon transform is a smoothing transform, and inverting the Radon transform in presence of additive noise is an *ill-posed* inverse problem,[2] because $\mathcal{R}_c^{-1}$ is not a bounded linear operator; numerically speaking, a direct computation of $\mathcal{R}^{-1}f$ would be contaminated by a huge additive noise $Z = \mathcal{R}^{-1}W$.

1.2. Regularization Procedures

Filtered Back-Projection (FBP) and its derivatives are the most popular regularization methods for tomographic reconstruction. These are linear filtering techniques in the Fourier space. In these procedures, the amplification of high frequencies to compute the inverse Radon transform is attenuated to damp out high frequency components in which the amplitude of the noise $Z = \mathcal{R}^{-1}W$ would be too high. However, FBP suffers from serious performance limitations, due to the fact that the vectors of the Fourier basis provide a good representation (diagonalization) of the Radon operator, but are not adapted to represent spatially inhomogeneous data such as medical images. In other words, the Fourier basis is suboptimal for the denoising and restoration of piece-wise regular signals because it does not provide a compact representation of this type of objects.

To improve the performance of tomographic reconstruction procedures, researchers have studied iterative statistical model-based techniques, designed to implement Expectation-Maximization (EM) and Maximum A Posteriori (MAP) estimators. [3,4] In some cases, these approaches can provide a significant improvement over FBP. However, these estimators suffer from the following drawbacks:

- Computation time. Almost all the corresponding algorithms are too computer-intensive and not usable yet for clinical applications, with the exception of OS-EM[5] (an accelerated implementation of an EM estimator). In MAP methods, useful priors usually give local maxima, and the computational cost of relaxation methods is prohibitive.

- Theoretical understanding and justification. EM estimators lack theoretical foundations to understand and characterize the estimation error. Some MAP estimators are in some cases better understood, yet no optimality for a realistic model has been established.

- Convergence. EM estimators are ill-conditioned, in the sense that the corresponding iterative algorithms have to be stopped after a limited number of iterations. Beyond this critical number, the noise is magnified, and EM and OS-EM converge to a non-ML (Maximum Likelihood) solution.

This study aims at building a family of non-linear estimators, for which the estimation error is understood and the optimality is established for a realistic prior model of the data, which can be implemented with fast non-iterative algorithms, and which provide better numerical results, both metrically and perceptually. Such algorithms should also be flexible enough to be compatible with other specific needs, such as local tomographic reconstruction, limited angle tomography, and post-processing algorithms.

1.3. Non-linear Diagonal Estimators

A general strategy was recently advocated by Kalifa and Mallat[6] to solve linear inverse problems:

1. "Diagonalization" of the problem. Design a decomposition in which the Gram inverse operator -here, $\mathcal{R}^{-1*}\mathcal{R}^{-1}$- of the operator to be inverted is nearly diagonal, and in which the object to be estimated is compactly represented.
2. Design diagonal operators (typically thresholding rules) in this representation to estimate the decomposition coefficients of the signal to be recovered.

The Wavelet-Vaguelette Decomposition of Donoho,[7] which has been studied numerically by Kolaczyk[8] for tomographic reconstruction and refined by Lee and Lucier,[9] and the curvelet-based estimator of Candès and Donoho,[10] are particular examples. Section 2 briefly recalls the underlying formalism and the justifications of this approach. The asymptotic optimality is characterized on a set modeling the prior information on the signals, which often happens, especially for natural images, to be a more realistic and simpler model than probabilistic priors. Section 3 implements this strategy to build simple minimax optimal estimators for bounded variation images in 1-D and 2-D orthogonal wavelet bases. Section 4 investigates the benefits of wavelet packets, as well as directional selectivity on images using better 2-D wavelet transforms. The corresponding estimators have similar minimax optimality properties than the estimators in orthogonal wavelet bases of section 3, but can provide significant numerical enhancements. Section 5 shows a numerical application on a simulation of PET acquisition of the Shepp-Logan phantom. Preliminary results already show a strong metrical and perceptual improvement over Filtered Back-Projection.

Notation: if β_1 and β_2 depend upon the parameters of the problem, such as the signal size N or any other parameter, we write $\beta_1 \sim \beta_2$ and say that they are equivalent, if there exists two constant $A, B > 0$ such that for all values of these parameters $A\,\beta_1 \leq \beta_2 \leq B\,\beta_1$.

2. DIAGONAL ESTIMATORS FOR LINEAR INVERSE PROBLEMS

The results summarized in this section are detailed in.[6] We consider the estimation problem in (1), which is also equivalent to the denoising problem

$$X = f + Z \tag{3}$$

where $X = \mathcal{R}^{-1}Y$ and $Z = \mathcal{R}^{-1}W$. The noise W is considered Gaussian of variance σ^2 to simplify the explanations. Z is still Gaussian because the inverse Radon transform is a linear operator. Its covariance operator is $K = \sigma^2 \mathcal{R}^{-1*}\mathcal{R}^{-1}$.

2.1. Thresholding Estimators

A thresholding estimator $\tilde{F}$ of f in a basis $\mathcal{B} = \{g_m\}_m$ is computed with

$$\tilde{F} = \sum_m \rho_{T_m} \left(\langle X, g_m \rangle \right) g_m \ , \tag{4}$$

when the denoising is performed after the application of the inverse operator $\mathcal{R}^{-1}$, or with

$$\tilde{F} = \mathcal{R}^{-1} \left(\sum_m \rho_{T_m} \left(\langle Y, g_m \rangle \right) g_m \right), \tag{5}$$

when the denoising is performed before the application of $\mathcal{R}^{-1}$. ρ_{T_m} is a thresholding operator. T_m are the threshold values, and are different in (4) and (5). Typical thresholding rules include hard thresholding

$$\rho_{T_m}(x) = \left\{ \begin{array}{ll} x & \text{if } |x| > T_m \\ 0 & \text{if } |x| \leq T_m \end{array} \right. , \tag{6}$$

and soft thresholding

$$\rho_{T_m}(x) = \left\{ \begin{array}{ll} x - T_m & \text{if } x \geq T_m \\ x + T_m & \text{if } x \leq -T_m \\ 0 & \text{if } |x| \leq T_m \end{array} \right. . \tag{7}$$

If the noise Z in (3) was white, the choice of the basis $\mathcal{B}$ would only depend on the prior information on the object f. Indeed, Donoho and Johnstone[11] proved that, with a proper choice of threshold values T_m, thresholding estimators in an orthonormal basis are nearly optimal for white noise removal if the basis provides a sparse signal representation, which means that the basis concentrates the energy of the signal on a few coefficients. In our situation, the choice of $\mathcal{B}$ also depends on the noise Z.

2.2. Near-Diagonalization of the Estimation Problem

The assumption underlying diagonal estimators in a basis is that each coefficient in this decomposition can be estimated independently. As a consequence, such estimators are efficient if the coefficients of the noise and of the object to be recovered are indeed nearly independent in the basis $\mathcal{B}$. This means that $\mathcal{B}$ must provide a near-diagonalization of the covariance operator K of the noise Z and of the prior information on the object f.

The near-diagonalization of the covariance operator K of Z is measured by preconditioning K with its diagonal. Let K_d be the diagonal operator in the basis $\mathcal{B}$, whose diagonal values are equal to the diagonal values of K, noted σ_m^2:

$$\sigma_m^2 = \langle K g_m, g_m \rangle.$$

The square root $K_d^{1/2}$ is the diagonal matrix whose coefficients are σ_m. The diagonal preconditioning factor of K^{-1} in the basis $\mathcal{B}$ is defined as

$$\lambda_{\mathcal{B}} = \| K_d^{1/2} K^{-1} K_d^{1/2} \|_s \ . \tag{8}$$

It satisfies $\lambda_{\mathcal{B}} \geq 1$. We have $\lambda_{\mathcal{B}} = 1$ 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.

The best method to measure the near-diagonalization of the prior information on f depends on its nature. In a classical Bayes estimation framework, f is modeled as a realization of a random process F, whose probability distribution is known

a priori. However, complex signals such as natural images are highly non-Gaussian, and there is no probabilistic model that incorporates their spatial inhomogeneity. The prior information often defines a set Θ to which signals are guaranteed to belong, without specifying their probability distribution. For example, many images have some form of piecewise regularity, which can be characterized by bounded Besov norms [12] or a bounded total variation, which specify a prior set Θ. With this type of prior model, the near-diagonalization of the prior information can be measured with orthosymmetric sets. Any signal f can be decomposed as $f = \sum_m f_\mathcal{B}[m]g_m$. A set Θ is *orthosymmetric* in $\mathcal{B}$ if for any $f \in \Theta$ and for any $a[m]$ with $|a[m]| \leq 1$ then

$$\sum_m a[m] \, f_\mathcal{B}[m] \, g_m \in \Theta \, .$$

This means that the set Θ is elongated along the directions of the vectors g_m of $\mathcal{B}$.

2.3. Minimax estimation

When the prior model on the signals is a set Θ, the estimation error for an estimator $\tilde{F} = DX$ is characterized by the maximum risk over Θ:

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

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

$$r^n(\Theta) = \inf_{D \in \mathcal{O}_n} r(D, \Theta).$$

In practice, one must find a decision operator D that is simple to implement and such that $r(D, \Theta)$ is close to the minimax risk $r^n(\Theta)$.

Minimax estimation aims at studying the robustness of estimators; In a Bayesian framework, one has to be careful to design a Bayes prior which is not too optimistic, otherwise its performance will be overestimated for some natural signals. This task is difficult for complex signals such as natural images, and yields Bayes priors which are usually close to the "worst" signals of the set Θ used as a model for minimax estimation. This means that, for complex signals, a robust Bayesian prior will be roughly similar to a minimax prior, while more complex to design and with a less convenient formalism.

2.4. Minimax optimality

Let $\mathcal{B}$ be a decomposition in which the covariance K of the noise Z is nearly diagonal with a preconditioning factor $\lambda_\mathcal{B}$ defined in (8), and in which the set Θ is orthosymmetric. Let D be a thresholding estimator in $\mathcal{B}$, as defined in section 2.1. The corresponding maximum thresholding risk is $r^t(\Theta) = r(D, \Theta) = \sup_{f \in \Theta} \mathsf{E}\{\|DX - f\|^2\}$, and we clearly have $r^t(\Theta) \geq r^n(\Theta)$. Let us choose $T_m = \sigma_m \sqrt{2 \log_e N}$ for the threshold values in (6) and (7) when the estimator is given by equation (4), and let $T_m = \sigma \sqrt{2 \log_e N}$ when the estimator is given by equation (5). One can show [6] that

$$r^t(\Theta) \sim r^n(\Theta)\lambda_\mathcal{B} \log_e N. \tag{9}$$

This shows that the thresholding risk $r^t(\Theta)$ remains of the same order as the minimax risk $r^n(\Theta)$ up to a $\lambda_\mathcal{B} \log_e N$ factor. In some cases, the factor $\log_e N$ can even be reduced to a constant independent of N. $\lambda_\mathcal{B}$ is independent of N is $\mathcal{R}^{-1*}\mathcal{R}^{-1}$ is nearly diagonal in $\mathcal{B}$.

We now apply these results to the tomographic reconstruction of images.

3. APPLICATION TO THE INVERSION OF THE RADON TRANSFORM

This section shows that orthogonal 1-D and 2-D wavelet transforms are simple examples of decompositions which satisfy the conditions of section 2 for the inversion of the Radon transform. The prior set Θ of images is defined with a bounded variation norm.

The thresholding estimation for tomographic reconstruction can be performed in the spatial domain, on the signal $X[n_1, n_2]$ of equation (3), which is obtained after the application of the inverse Radon transform. It can also be performed in the Radon domain on the sinograms Y of equation (1), before applying the inverse Radon transform.

3.1. Bounded Variation Images

The set Θ used as a prior model on the object f is defined with a bounded variation norm. Bounded variation images may include sharp transitions such as discontinuities. Large classes of images including medical images, in fact images with no fractal textures, have a bounded total variation. For square discrete image of N^2 pixels, the total variation is defined as

$$\|f\|_V = \frac{1}{N} \sum_{n_1, n_2 = 0}^{N-1} \left(\left| f[n_1, n_2] - f[n_1 - 1, n_2] \right|^2 + \left| f[n_1, n_2] - f[n_1, n_2 - 1] \right|^2 \right)^{1/2}.$$

We say that an image has a bounded variation if $\|f\|_V$ is bounded by a constant independent of the resolution N. Let Θ be the set of images that have a total variation bounded by C

$$\Theta = \left\{ f \ : \ \|f\|_V \leq C \right\}.$$

The co-area formula[13] proves that the total variation has a simple interpretation as the average length of level sets in the image.

3.2. Wavelet bases

1-D Wavelet Basis A discrete orthonormal wavelet basis of $\mathbb{C}^N$ is constructed with a dilated discrete wavelet $\psi_j[n]$ at each scale $1 < 2^j \leq N$. The wavelet ψ_j is translated: $\psi_{j,m}[n] = \psi_j \left[\frac{n - 2^j m}{2^j} \right]$, and at the largest scale $\psi_{N,0}[n] = 1$. One can construct such discrete wavelets so that the resulting family

$$\mathcal{B}_1 = \left\{ \psi_{j,m} \right\}_{1 < j \leq \log_2(N), \, 0 \leq m < 2^j} \tag{10}$$

is an orthonormal basis of $\mathbb{C}^N$. Moreover, the N wavelet coefficients in this basis can be calculated with $O(N)$ operations, with a fast filter bank algorithm.[14]

2-D Wavelet Basis A separable wavelet basis of $\mathbb{C}^{N^2}$ is constructed with the one-dimensional wavelets ψ_j defined in (10). At each scale 2^j there are three wavelets $\psi_j^\alpha[n_1, n_2]$ for $\alpha = 1, 2, 3$. These wavelets are uniformly translated to define $\psi_{j,m_1,m_2}^\alpha[n_1, n_2] = \psi_j^\alpha \left[\frac{n_1 - 2^j m_1}{2^j}, \frac{n_2 - 2^j m_2}{2^j} \right]$, and $\psi_N^0[n_1, n_2] = 1/N$. The separable wavelet family

$$\mathcal{B}_2 = \left\{ \psi_N^0 \, , \, \psi_{j,m_1,m_2}^1 \, , \, \psi_{j,m_1,m_2}^2 \, , \, \psi_{j,m_1,m_2}^3 \right\}_{1 < j \leq \log_2(N), \, 0 \leq m_1, m_2 < 2^j}$$

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.[14]

Orthosymmetry in Wavelet Bases The set Θ of bounded variation images is not orthosymmetric in the 2-D wavelet basis. However, it can be embedded in two close orthosymmetric sets Θ_1 and Θ_2 such that $\Theta_1 \subset \Theta \subset \Theta_2$. For a given estimator $\tilde{F} = DX$, the estimation risks satisfy $r(D, \Theta_1) \leq r(D, \Theta) \leq r(D, \Theta_2)$. If one can show that $r(D, \Theta_1) \sim r(D, \Theta_2)$, it follows immediately that $r(D, \Theta)$ is of the same order. The set Θ of bounded variation images can be embedded in two sets defined with discrete Besov norms[12] in which the risk equivalence is satisfied for the estimators under study.

The representation of bounded variation images using the 1-D wavelet basis is slightly more difficult to characterize, but follows a similar idea. Images are segmented in different directions, and each direction is studied independently. Once again, the set Θ is embedded in two close sets in which the orthosymmetry can be established for the different directions.

3.3. Near-Diagonalization of the Operator

For a given α, let $\mathcal{R}_\alpha f(t) = \mathcal{R}f(t, \alpha)$, and let $Y_\alpha = \mathcal{R}_\alpha f + W_\alpha$ be the noisy tomographic projection of f at the angle α. The 1-D orthogonal wavelet decomposition in $\mathcal{B}_1$ is used to estimate each projections Y_α; here, the operator to be inverted is $\mathcal{R}_\alpha$, and the operation is repeated for each value of α.

The 2-D wavelet basis $\mathcal{B}_2$ is used to perform the whole estimation on the sinograms Y or on the back-projected data $\mathcal{R}^{-1}Y$ in a single pass, without separating the estimation in each direction α. In that case, $\mathcal{R}$ is the operator to be inverted.

The inverse Radon transform is a particular example of a class of linear homogeneous operators, often referred as Calderón-Zygmund operators, for which the vectors of the wavelet basis are almost eigenfunctions. For a detailed presentation of the mathematical relations between wavelet bases and Calderón-Zygmund operators, we refer to.[15] Theorem 3.1 shows that this property is satisfied in both the 1-D case where $\mathcal{B}_1$ is used for the inversion of $\mathcal{R}_\alpha$ and in the 2-D case, where $\mathcal{B}_2$ is used for the inversion or $\mathcal{R}$.

Let K_1 and K_2 be respectively the operators $\sigma^2 \mathcal{R}_\alpha^{-1*} \mathcal{R}_\alpha^{-1}$ and $\sigma^2 \mathcal{R}^{-1*} \mathcal{R}^{-1}$. Let $K_{1,d}$ and $K_{2,d}$ be respectively the diagonal operators in the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, whose diagonal values are respectively equal to the diagonal values of K_1 and K_2 in $\mathcal{B}_1$ and $\mathcal{B}_2$.

THEOREM 3.1. *There exists λ_1 and λ_2 such that for all $N > 0$, K_1 and K_2 satisfy*

- *In the 1-D wavelet basis $\mathcal{B}_1$*

$$\|K_{1,d}^{1/2} K_1^{-1} K_{1,d}^{1/2}\|_s \leq \lambda_1.$$

- *In the 2-D wavelet basis $\mathcal{B}_2$*

$$\|K_{2,d}^{1/2} K_2^{-1} K_{2,d}^{1/2}\|_s \leq \lambda_2.$$

This theorem shows that the diagonal preconditioning factors λ_1 and λ_2, which correspond to the factor $\lambda_\mathcal{B}$ in (9) do not grow with N.

3.4. Minimax optimality of wavelet estimators

Different orthogonal wavelet estimators The thresholding estimation in the 2-D orthogonal wavelet basis $\mathcal{B}_2$ can be performed before or after the application of $\mathcal{R}^{-1}$. The latter is given by equation (5), where $\mathcal{B} = \mathcal{B}_2$, and it is equivalent to the Wavelet-Vaguelette Decomposition (WVD) estimator of Donoho.[7] Let $\tilde{F}_2^{iw} = D_2^{iw} Y$ be this estimator.

The estimator is given by equation (4) when it is performed in the Radon domain $Im(\mathcal{R})$. It is equivalent to the Vaguelette-Wavelet Decomposition (VWD) as studied by Abramovich and Silverman.[16] Let $\tilde{F}_2^{wi} = D_2^{wi} Y$ be this estimator.

The thresholding estimation in the 1-D basis $\mathcal{B}_1$ is simple to compute when performed in the Radon domain. The estimation is computed independently in each direction α according to

$$\tilde{F}_\alpha = \mathcal{R}_\alpha^{-1} \left(\sum_{j,m} \rho_{T_{j,m}} \left(\langle Y_\alpha, \psi_{j,m} \rangle \right) \psi_{j,m} \right). \tag{11}$$

The 1-D wavelet denoising must be implemented before the spatial interpolation for the inverse Radon transform, but can be implemented before or after the amplification of high frequencies. This defines two different but very close estimators. We write formally that the estimation is always performed before the inversion of the Radon transform and that the corresponding estimator is unique, because the properties of the 1-D wavelet denoising before or after the amplification of high frequencies are extremely similar. Let $\tilde{F}_1 = D_1 Y$ be this estimator. Note that the same formal simplification is made for the estimator $\tilde{F}_2^{wi}$ in the 2-D wavelet basis, which can be implemented before or after the amplification of high frequencies.

Relation with the ridgelet transform The application of the 1-D wavelet transform in the basis $\mathcal{B}_1$ in the Radon domain $Im(\mathcal{R})$ is related to the ridgelet transform.[17] The ridgelet transform is a decomposition in a collection of ridge functions. Roughly speaking, the ridgelet transform can be implemented by computing the 1-D wavelet transforms of the Radon transform $\mathcal{R}_\alpha$ in each direction α. Ridgelets are designed to provide an optimal representation of straight contours in images. As a consequence, one can show that the thresholding estimator using a 1-D wavelet decomposition in the Radon domain is not only optimal to recover bounded variation images, but also to recover straight edges in images.

Thresholding estimation risk Let $r^t{}_2^{iw}(\Theta) = r(D_2^{iw}, \Theta)$, let $r^t{}_2^{wi}(\Theta) = r(D_2^{wi}, \Theta)$, and let $r^t{}_1(\Theta) = r(D_1, \Theta)$ be respectively the three estimation risks of the thresholding estimators $\tilde{F}_2^{iw}, \tilde{F}_2^{wi}, \tilde{F}_1$.

Theorem 3.2 shows that the asymptotic behavior of the estimation risk is identical for the three thresholding estimators $\tilde{F}_2^{iw}, \tilde{F}_2^{wi}, \tilde{F}_1$ and that these estimators are quasi-minimax optimal. Recall that $r^n(\Theta)$ is the minimax estimation risk defined in section 2.3. Under mild conditions, one can show that

THEOREM 3.2.

$$\sigma^{\frac{4}{3}} C N^{\frac{3}{2}} \sim r^n(\Theta) \leq r^{t\,iw}_2(\Theta) \sim r^{t\,wi}_2(\Theta) \sim r^t_1(\Theta) \sim \sigma^{\frac{4}{3}} C N^{\frac{3}{2}} \log_e N \tag{12}$$

The thresholding risks are of the same order as the minimax risk up to a $\log_e N$ factor, which can in some cases be reduced to a constant independent of N.

One can also characterize the linear minimax risk $r^l(\Theta)$, which is the minimax risk when restricting to linear estimators. The linear minimax risk gives the best possible performance than can be achieved by Filtered Back-Projection and other linear procedures in the Fourier space. $r^l(\Theta)$ satisfies, under mild conditions

$$r^l(\Theta) \sim \sigma^{\frac{2}{3}} N^{\frac{5}{3}} C^{\frac{4}{3}}. \tag{13}$$

Equations (12) and (13) show that the minimax and thresholding risks are smaller than the minimax linear risk. A linear estimator tries to reduce the noise without smoothing too much the singularities. A non negligible residual noise will remain in the smooth parts of the signal.

4. REFINING THE CHOICE OF DECOMPOSITION

Section 3 provides theoretical foundations which justify the use of wavelet thresholding estimators for tomographic reconstruction. While the asymptotic minimax optimality has been established for the estimators under study, their numerical performances can be enhanced. This section studies different refinement in the choice of the decompositions in which the thresholding estimation is computed. These enhanced estimators still satisfy the same asymptotic minimax optimality properties.

4.1. Translation-invariant decompositions

Improved estimators can be obtained with the translation invariant algorithm of Coifman and Donoho, [18] by averaging estimators for translated versions of the degraded signal. Such an estimator can be calculated by thresholding the non-decimated coefficients computed with the "algorithme à trous", which can implement a fast wavelet or wavelet packet transform, without subsampling.[19] This fast translation-invariant filter bank algorithm requires $O(N)$ operations to compute the estimation for a signal of N samples.

4.2. Benefits of Wavelet Packets

Wavelet packet bases are constructed with the conjugate mirror filters used for wavelet transforms. Wavelet packet bases divide the Fourier domain in separate intervals of various sizes,[20] and can enhance the near-diagonalization of the Radon transform operator. In equation (8), the preconditioning factor $\lambda_{\mathcal{B}}$ can be made closer to 1 in wavelet packets bases than in conventional wavelet bases. On the other hand, wavelet packet bases provide in general less compact representations of natural piece-wise regular signals and images.

The Fourier basis achieves a strict diagonalization of the Radon transform operator, while the wavelet basis provides a good representation of inhomogeneous signals such as medical images. It is natural to search for the best basis which achieves an optimal trade-off between the spatial representation of the data and the representation of the Radon transform operator. Suppose we have a dictionary $\{\mathcal{B}^\gamma\}_\gamma$ of orthogonal wavelet packet bases. The best basis algorithm of Coifman and Wickerhauser [20] can be used to determine the best basis $\mathcal{B}^{\gamma_1} = \{g_m^{\gamma_1}\}_m$ in this dictionary. The additive criterion to be minimized must be an approximation of the final estimation error. If one has a prior model on the decomposition coefficients of the data, which is the case in tomography since phantoms are available, the criterion can be chosen as the ideal attenuation or projection risk. [11,21]

This approach can be used with 1-D or 2-D bases, and for any of the three wavelet estimators studied in section 2.3. In each case, the resulting estimation risks will be equal or lower than their wavelet counterparts $r^{t\,iw}_2(\Theta)$, $r^{t\,wi}_2(\Theta)$, $r^t_1(\Theta)$ of section 3.4, while sharing the same asymptotic quasi-minimax optimality properties. The benefits of using thresholding estimators in wavelet packet bases rather than in wavelet bases depend on the nature of the data f, and especially on its spectral distribution and on the importance of its oscillatory elements, and on the level and nature of the noise W.

Remark: Wavelet packet bases can achieve an arbitrarily fine dyadic segmentation of the Fourier domain. The amplification of frequencies to invert the smoothing effect of the Radon transform can be accurately approximated by an amplification of the wavelet packet coefficients, without having to perform a Fourier transform and its inverse. Usually, the vectors of the wavelet packet basis $\mathcal{B}^{\gamma_2}$ in which the amplification is performed need to have a more compact support in the Fourier domain

than the best wavelet packet basis $\mathcal{B}^{\gamma_1}$ for the estimation. Because of the tree structure of wavelet packet decompositions, it is not necessary to reconstruct the signal to compute directly the wavelet packet coefficients in the best basis $\mathcal{B}^{\gamma_1}$ from the amplified coefficients in $\mathcal{B}^{\gamma_2}$. This procedure is simpler and can also save computation time in some cases, as compared to the computation of a Fourier transform and reconstruction.

4.3. Better Wavelet Transforms

The *dyadic wavelet transform*,[22] which is implemented with the translation-invariant "algorithme à trous", provides an orientation selectivity in 2-D.

In two dimensions, orientation plays a crucial role to correlate the wavelets with the signal structures. In a 2-D dyadic wavelet decomposition composed of two wavelets, all orientations are recovered with an appropriate linear combination of the two wavelets. The dyadic wavelet transform is a decomposition in a family of wavelets which are dilated and translated from two functions $\psi^{d,1}$ and $\psi^{d,2}$ respectively oscillating in the horizontal and vertical directions.

Let $\psi^{\vec{a}}(x_1, x_2) = a_1 \psi^{d,1}(x_1, x_2) + a_2 \psi^{d,2}(x_1, x_2)$ be the wavelet whose orientation is $\vec{a}$. The vector $\vec{a}_{opt}$ such that, for any image f

$$\left| \langle f, \psi^{\vec{a}_{opt}} \rangle \right|^2 = \sup_{\vec{a}} \left| \langle f, \psi^{\vec{a}} \rangle \right|^2 = a_1 \left| \langle f, \psi^{d,1} \rangle \right|^2 + a_2 \left| \langle f, \psi^{d,2} \rangle \right|^2 \tag{14}$$

is given by : $\vec{a}_{opt} = \left(\dfrac{\langle f, \psi^{d,1} \rangle}{M}, \dfrac{\langle f, \psi^{d,2} \rangle}{M} \right)$, and we have

$$M^2 = \left| \langle f, \psi^{d,1} \rangle \right|^2 + \left| \langle f, \psi^{d,2} \rangle \right|^2.$$

This means that, instead of applying the thresholding estimation algorithm to the wavelet decomposition in the two directions, we must threshold the modulus. This is equivalent to selecting first a direction in which the partial derivative is maximum at the scale 2^j, and thresholding the amplitude of the partial derivative in this direction. This can be viewed as an adaptive choice of the wavelet direction in order to maximize the correlation with the signal.

The *complex wavelet transform* of Kingsbury[23] is designed to bring similar benefits. The fast complex wavelet decomposition provides shift invariance and a good directional selectivity, while not as accurate as the modulus of a dyadic wavelet transform.

To take further advantage of the geometrical prior information on images, and of the fact that images incorporate many curved edges, Candès and Donoho have investigated the benefits of a *curvelet* decomposition,[10] which is designed to recover curved edges. The theoretical properties of this approach are appealing, however numerical applications do not yet fulfill the expectations.

5. NUMERICAL EXAMPLES

Figure 1 shows a simulation of acquisition on the classical Shepp-Logan phantom, using the characteristics of a standard PET scanner, with 256 angular positions and 192 detectors. Figure 1(a) is a reconstructed image from sinograms without noise. This image will be used as our reference image. Figure 1(b) is a reconstructed image from sinograms degraded with Poisson noise, using the Filtered Back-Projection algorithm. The Signal-to-Noise Ratio (SNR) of this image as compared to the reference image is $27.2db$.

Section 3 and 4 show that many non-linear minimax thresholding estimators can be implemented. One can choose between 1-D or 2-D decompositions, regularization in the Radon domain or after the inversion of the Radon transform, wavelet or wavelet packets, different 2-D wavelet decompositions, etc... The choice of threshold values and thresholding operator is also important. Note that in some cases, the combination of regularization algorithms in both the Radon domain and the spatial domain can also provide a supplemental numerical enhancement on the results. The wavelet/wavelet packet representation in the Radon domain has different numerical properties in terms of signal/noise discrimination than the wavelet/wavelet packet representation in space. Combining the best of these different representations can be made by choosing the right combination of threshold values and thresholding procedures.

Figure 1(c) shows a reconstructed image using hard thresholding in a 1-D wavelet decomposition in the Radon domain, as suggested in section 3.4, starting from the same noisy sinograms as in Figure 1(b). The resulting SNR is $33.1db$. Figure 1(d) is a reconstructed image with a hard thresholding operator in a 2-D wavelet packet basis in the spatial domain, after the inversion of the Radon transform. The resulting SNR is $32.3db$. These two simple estimation algorithms already provide a significant metrical and perceptual improvement over an optimized Filtered Back-Projection method, while many possible enhancements have not been implemented yet. The increase in SNR is between $5db$ and $6db$ as compared to the image reconstructed with Filtered Back-Projection.

6. NEXT DEVELOPMENTS

Non-linear thresholding estimators are expected to provide fast reconstruction algorithms which will outperform current FBP-methods and iterative procedures in X-ray CTs, PET and SPECT devices. An extensive work has to be conducted to apply and evaluate these estimators on medical data from different devices . Promising results have already been obtained on simulations of PET and SPECT data.

It is important to mention that the wavelet representations are very useful for different applications in tomography. Holschnei-der,[24] Walnut,[25] Peyrin et al.[26] have established relations between wavelet transforms and the Radon transform. Since then, many publications have specifically studied the benefits of wavelets for local tomographic reconstruction, [27–34] and for limited angle tomography.[35,36] Also, the dyadic wavelet transform is an ideal tool for edge detection and segmentation [22] which are common problems in medical imaging. One can expect a wavelet-based representation for global tomographic reconstruction to be a convenient and flexible framework for other specific problems arising from tomographic reconstruction, and for post-processing algorithms. This issue calls for deeper investigations.

REFERENCES

1. S. Deans, *The Radon transform and some of its applications*, John Wiley and Sons, 1983.
2. M. Bertero, *Advances in Electronics and Electron Physics*, ch. ”Linear inverse and ill-posed problems”. Academic Press, 1989.
3. L. Shepp and Y. Vardi, “Maximum likelihood reconstruction for emission tomography,” *Transactions on Medical Imaging* , pp. 113–122, 1982.
4. G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, New York:Wiley, 1997.
5. H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Transactions on Medical Imaging* , 1994.
6. J. Kalifa and S. Mallat, “Thresholding estimators for inverse problems and deconvolution,” 1999. Submitted for publication, available at http://www.cmap.polytechnique.fr/~kalifa/.
7. D. Donoho, “Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions,” *J. of Appl. and Comput. Harmonic Analysis* **2**(2), pp. 101–126, 1995.
8. E. Kolaczyk, “A wavelet shrinkage approach to tomographic image reconstruction,” *J. Amer. Statist. Assoc.* **91**, pp. 1079–1090, 1996.
9. N. Lee and B. Lucier, “Wavelet methods for inverting the Radon transform with noisy data,” *IEEE Trans. on IP, to appear* , 2000.
10. E. Candes and D. Donoho, “Recovering edges in ill-posed linear inverse problems: Optimality of curvelet frame,” tech. rep., Dept. of Statistics, Stanford University, 2000.
11. D. Donoho and I. Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” *Biometrika* **81**, pp. 425–455, December 1994.
12. M. Frazier and B. Jawerth, “Decomposition of Besov spaces,” *Indiana Univ. Math. J.* **34**(4), pp. 777–789, 1985.
13. W. Ziemer, *Weakly differentiable functions : Sobolev spaces and functions of bounded variations*, vol. 120, Springer-Verlag, 1989.
14. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.
15. Y. Meyer and R. Coifman, *Wavelets - Calderón-Zygmund and multilinear operators*, Cambridge Studies in Advanced Mathematics, 1997.

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

17. E. Candes and D. Donoho, "Ridgelets: the key to high-dimensional intermittency?," in *Wavelets: the key to intermittent information?*, No. 357, pp. 2495–2509, Phil. Trans. R. Soc. Lond. A., Sept. 1999.

18. R. Coifman and D. Donoho, "Translation invariant de-noising," Tech. Rep. 475, Dept. of Statistics, Stanford University, May 1995.

19. M. J. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithms," *IEEE Trans. Signal Proc.* **40**, pp. 2464–2482, October 1992.

20. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Info. Theory* **38**, pp. 713–718, March 1992.

21. J. Kalifa, S. Mallat, and B. Rougé, "Restauration d'images par paquets d'ondelettes," in *GRETSI conf. de trait. du signal et des images*, (Grenoble, France), 1997.

22. S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Patt. Recog. and Mach. Intell.* **14**, pp. 710–732, July 1992.

23. N. G. Kingsbury, "Image processing with complex wavelets," in *Wavelets: the key to intermittent information?*, Phil. Trans. R. Soc. Lond. A., 1999.

24. M. Holschneider, "Inverse Radon transforms through inverse wavelet transforms," *Inverse problems* **7**, pp. 853–861, 1991.

25. D. Walnut, *Probabilistic and Stochastic Methods in Analysis with Applications, J.S Byrnes et al., eds*, ch. Applications of Gabor and wavelet expansions to the Radon transform, pp. 187–205. Kluwer Academic Publishers, Boston, 1992.

26. F. Peyrin, M. Zaim, and R.Goutte, "Construction of wavelet decompositions for tomographic images," *J. Math. Imaging Vision (special issue on wavelet theory and applications)* **3**, pp. 105–121, 1993.

27. C. Berestein and D. Walnut, *Wavelets in Medicine and Biology, A. Aldroubi and M. Unser, eds.*, ch. Wavelets and Local Tomography. CRC Press, 1996.

28. M. Bhatia, W. Karl, and A. .Willsky, "Tomographic reconstruction and estimation based on multiscale natural-pixel bases," *IEEE Trans. Image Proc.* **6**, pp. 463–478, 1997.

29. T. Olson, "Optimal time-frequency projections for localized tomography," *Annals of Biomedical Engineering* **23**, pp. 622–836, 1995.

30. F. Rashid-Farrokhi, K. Liu, C. Berenstein, and D. Walnut, "Wavelet-based multiresolution local tomography," *IEEE Trans. Image Proc.* **22**, pp. 1412–1430, 1997.

31. S. Zhao, G.Wang, and J.Hsieh, "Wavelet sampling and localization schemes for the Radon transform in two dimensions," *SIAM J. of Appl. Math.* **57**, 1749-1762 1997.

32. M. Bottema, B. Moran, and S. Suorova, "An application of wavelets in tomography," *Digital Signal Processing* **8**, pp. 244–254, 1998.

33. W. Maldych, "Tomography, approximate reconstructions, and continuous wavelet transforms," *J. of Appl. and Comput. Harmonic Analysis* **7**, pp. 54–100, 1999.

34. S. Zhao, "Wavelet filtering for filtered backprojection in computer tomography," *J. of Appl. and Comput. Harmonic Analysis* **6**, pp. 346–373, 1999.

35. T. Olson, "Limited angle tomography via multiresolution analysis and oversampling," in *IEEE-SP Symposium on time-frequency and time-scale analysis*, N. j. IEEE Press, Piscataway, ed., pp. 215–218, 1992.

36. B. Sahiner and A. Yagle, *Time-frequency and wavelets in biomedical signal processing*, ch. Inversion of the Radon transform under wavelet constraints, pp. 473–498. IEEE Press, 1997. Metin Akay, editor.
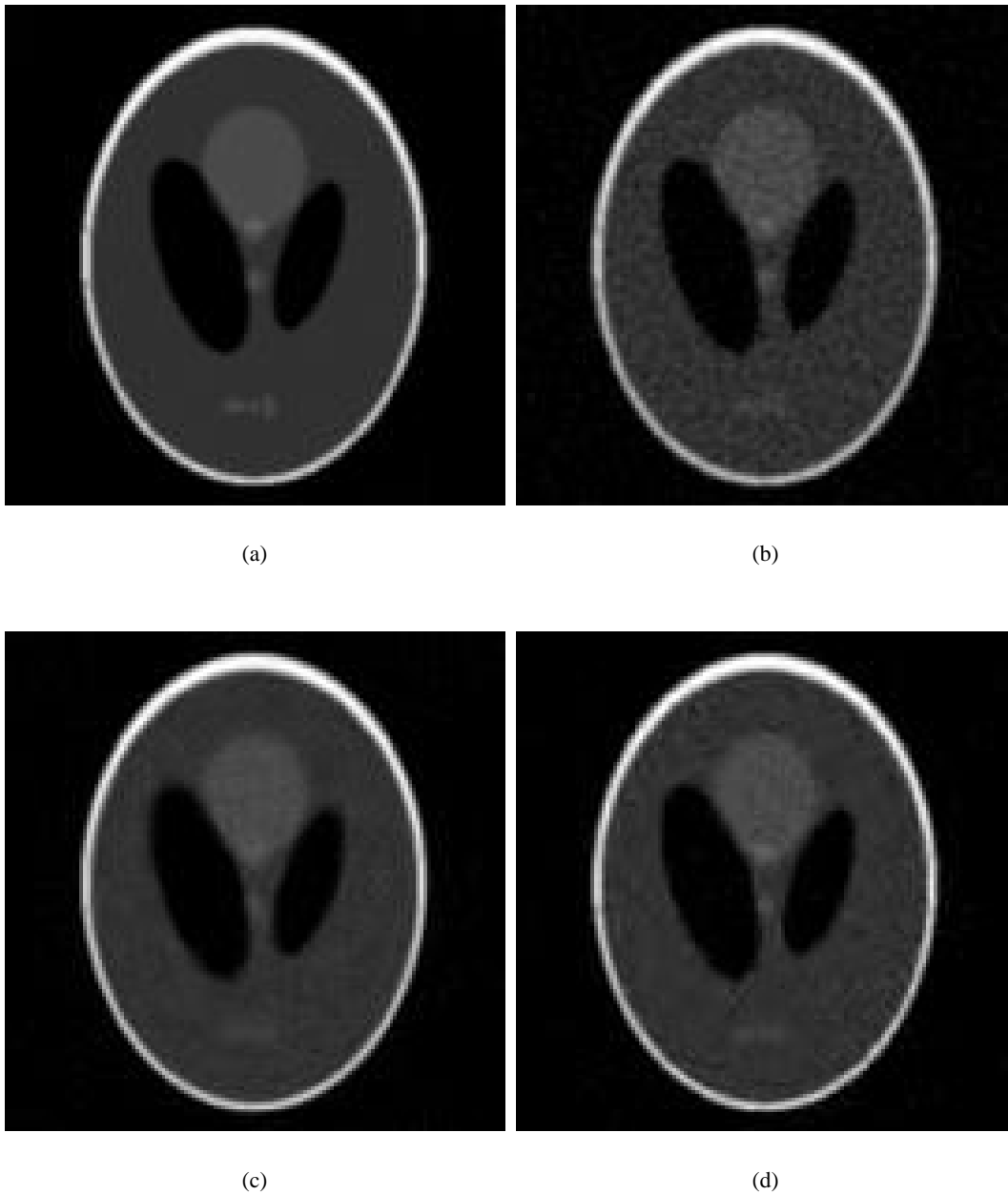
(a)

(b)

(c)

(d)

Figure 1. (a) Reference image. (b) Reconstructed image with a Filtered Back-Projection. (SNR=$27.2db$) (c) Reconstructed image with a 1-D wavelet thresholding in the Radon domain (SNR=$33.1db$) (d) Reconstructed image with a wavelet packet thresholding after the inversion of the Radon transform (SNR=$32.4db$).