Sparse Representation-based Image Deconvolution by Iterative Thresholding

M.J. Fadili [a,*]

[a] *Image Processing Group GREYC CNRS UMR 6072, 14050 Caen Cedex, France*

J.-L. Starck [b]

[b] *DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, 91191 Gif sur Yvette, France*

Abstract

Image deconvolution algorithms with overcomplete sparse representations and fast iterative thresholding methods are presented. The image to be recovered is assumed to be sparsely represented in a redundant dictionary of transforms. These transforms are chosen to offer a wider range of generating atoms; allowing more flexibility in image representation and adaptativity to its morphological content. The deconvolution inverse problem is formulated as the minimization of an energy functional with a sparsity-promoting regularization (e.g. ℓ_1 norm of the image representation coefficients). As opposed to quadratic programming solvers based on the interior point method, here, recent advances in fast solution algorithms of such problems, i.e. Stagewise Iterative Thresholding, are exploited to solve the optimization problem and provide fast and good image recovery results. Some theoretical aspects as well as computational and practical issues are investigated. Illustrations are provided for potential applicability of the method to astronomical data.

Key words: Deconvolution, Overcomplete sparse representations, Iterative thresholding, Hard thresholding.

1. Introduction

Deconvolution is a longstanding problem in many areas of signal and image processing (e.g. biomedical imaging, astronomy, remote-sensing, seismology, etc). Suppose that we observe the random process:

$$Y_s = F\left(X \circledast h\right)_s \odot \epsilon_s \tag{1}$$

where $\odot$ is any composition of two arguments (e.g. '+', '$\cdot$'), $s \in S$ is the location index, ϵ_s is the random noise generally assumed to be additive white gaussian noise (AWGN) but is not necessarily so (e.g. speckle, Poisson), $\circledast$ is the (circular) convolution operator with the space-invariant point spread function (PSF) h and F is a possibly non-linear transformation reflecting the nonlinearity of the sensor.

When the noise is AWGN, the sensor is linear, and in a finite-dimensional discrete setting ($\mathbf{H}$ is block-Toeplitz), the above degradation equation reduces to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} \tag{2}$$

Deconvolution is to recover $\mathbf{x}$ from $\mathbf{y}$ which is known to be an inverse ill-posed problem.

* Corresponding author.
Email addresses: Jalal.Fadili@greyc.ensicaen.fr (M.J. Fadili), jstarck@cea.fr (J.-L. Starck).

Brief overview

Research in astronomical image deconvolution has recently seen considerable work, partly triggered by the Hubble space telescope (HST) optical aberration problem at the beginning of its mission that motivated astronomers to improve current algorithms or develop new and more efficient ones. Since then, deconvolution of astronomical images has proven in some cases to be crucial for extracting scientific content.

There is an extensive statistical literature on deconvolution problems. The Wiener filter is the best filter (amongst the linear ones) for signal/image deconvolution [1]. It is equivalent to the maximum a posteriori (MAP) and posterior conditional mean bayesian estimators if both the signal and the noise are mutually independent gaussian. Stochastic approaches with ML estimators or within a bayesian framework have also been extensively studied (e.g. Markov random fields [2–5]). In order to confine candidate solutions to some particular functional spaces, several variational approaches have been developed in the literature to tackle the deconvolution problem either with linear (e.g. Tikhonov [6]) or nonlinear edge-preserving regularization (e.g. potential functions on the TV norm) [7–11]. If no prior information is available on the solution, except its positivity, maximum entropy method (MEM) is a possible course of action to derive the prior probability density function (pdf). The wavelet domain multiscale extension off the MEM has been proposed in [12,13].

Wavelet-based deconvolution methods have received considerable attention over the last decade. Donoho in [14] gave the first discussion of wavelet thresholding in linear inverse problems and introduced the wavelet-vaguelette decomposition; see also [15]. The approach was refined in [16] who proposed a mirror wavelet basis adapted to capture the singularity of the spectrum of the inverse of h. Authors in [17] advocated a hybrid approach coined ForWaRD. Deconvolution methods with sparsity promoting regularization over wavelet coefficients have been recently proposed, either by adopting a bayesian expectation-maximization framework [18–20], or by introducing surrogate functionals [21]. These algorithms were concerned with wavelet ortho-bases and generalization to frames was proposed in [22]. In [23–25], regularization from the multiresolution support was proposed. To the best of our knowledge, Starck et al. [26] were the first to introduce a combined deconvolution method using the wavelet and the curvelet transforms in order to benefit from the advantages of each. The interested reader may refer in [27] for an excellent overview.

This paper

We here propose some image deconvolution algorithms with overcomplete sparse representations and fast iterative thresholding algorithms. The image to be recovered is assumed to be sparsely represented in a redundant dictionary of transforms. Such representations differ from the more traditional basis representation because they offer a wider range of generating elements; potentially, this wider range allows more flexibility in image representation and adaptativity to its morphological content, and hence more effectiveness at tasks like image restoration. The deconvolution inverse problem is formulated as the minimization of an energy functional with a sparsity-promoting regularization. As opposed to quadratic programming solvers based on the interior point method, here, several Stagewise Iterative Thresholding algorithms are proposed to solve the optimization problem and provide fast and good image recovery results. Some of them have appealing finite-stopping properties and do not require the choice of auxiliary regularization hyperparameters. Theoretical aspects as well as computational and practical issues are discussed.

This paper is organized as follows. In Section 2, we will introduce the necessary ingredients of sparse representations and concepts from modern harmonic analysis. In Section 3, we state the optimization problem underlying the deconvolution problem, give the different algorithms to solve it, and investigate their properties. Section 4 is devoted to presenting and discussing experimental results. We conclude with a brief summary of our work and some perspectives on its likely future development.

2. Sparse image representation

2.1. *Sparse representations*

Suppose $\mathbf{x} \in \mathcal{H}$. An $\sqrt{n} \times \sqrt{n}$ image $\mathbf{x}$ can be written as the superposition of (a few) elementary functions $\phi_\gamma(s)$ (atoms) parameterized by $\gamma \in \Gamma$ such that (Γ is countable):

$$\mathbf{x}(s) = \sum_{\gamma \in \Gamma} \boldsymbol{\alpha}_\gamma \phi_\gamma(s), \text{ Card } \Gamma = L \tag{3}$$

Popular examples of Γ include: frequency (Fourier), scale-translation (wavelets), scale-translation-frequency (wavelet packets), translation-duration-frequency (cosine packets), scale-translation-angle (e.g. curvelets, bandlets, contourlets, wedgelets, etc.). The dictionary $\mathbf{\Phi}$ is the $n \times L$ matrix whose columns are the generating atoms $(\phi_\gamma)_{\gamma \in \Gamma}$, which are supposed to be normalized to a unit ℓ_2-norm. The forward transform is defined via a non-necessarily square full rank matrix $\mathbf{T} = \mathbf{\Phi}^T \in \mathbb{R}^{L \times n}$, with $L \geq N$. When $L > n$ the dictionary is said to be redundant or overcomplete.

Owing to recent advances in modern harmonic analysis, many redundant systems, like the undecimated wavelet transform, curvelet, contourlet, steerable or complex wavelet pyramids, were shown to be very effective in sparsely representing images. By sparsity, we mean that we are seeking a good representation of $\mathbf{x}$ with only very few non-zero coefficients, i.e. $\|\boldsymbol{\alpha}\|_0 \ll n$. In most practical situations, the dictionary is built by taking union of one or several (sufficiently incoherent) transforms, generally each corresponding to an orthogonal basis or a tight frame. In the case of a simple orthogonal basis, the inverse transform is trivially $\mathbf{\Phi} = \mathbf{T}^T$. Whereas assuming that $\mathbf{T}$ is a tight frame implies that the frame operator satisfies $\mathbf{T}^T\mathbf{T} = c\mathbf{I}$, where $c > 0$ is the tight frame constant. Hence, $\mathbf{T}^T = \mathbf{\Phi}$ is the Moore-Penrose pseudo-inverse transform (corresponding to the minimal dual synthesis frame), up to the constant c. In other words, the pseudo-inverse reconstruction operator $\mathbf{T}^+$ reduces to $c^{-1}\mathbf{\Phi}$.

2.2. *Choosing a dictionary*

Choosing an appropriate dictionary is a key step towards a good sparse representation, hence restoration. A core idea here is the concept of morphological diversity. When the transforms are amalgamated in one dictionary, they have to be chosen such that each leads to sparse representations over the parts of the images it is serving. Thus, for example (see Fig. 1), to represent efficiently isotropic structures in an image, a qualifying choice is the wavelet transform [28]. The curvelet system [29,30] is a very good candidate for representing piecewise smooth (C^2) images away from C^2 contours. The ridgelet transform [31,32] has been shown to be very effective for representing global lines in an image. The local DCT (LDCT) [28] and the Wave Atom [33] are well suited to represent warped locally stationary or oscillatory textures. These transforms are also computationally tractable particularly in large-scale applications, and never explicitly implementing $\mathbf{\Phi}$ and $\mathbf{T}$. The associated implicit fast analysis and synthesis operators have typical complexities of $O(n)$ (e.g. orthogonal or bi-orthogonal wavelet transform) or $O(n \log n)$ (e.g. ridgelet, curvelet, LDCT, wave atom transforms). Another desirable requirement that the merged transforms have to satisfy is that when a transform sparsely represents a part in the image, it yields non-sparse representations on the other content type. This is intimately related to the concept of mutual incoherence which plays a fundamental role in the solution of overcomplete (redundant) sparse linear systems. For a given dictionary, it is defined as the maximum in magnitude of the off-diagonal elements of its Gram matrix. See e.g. [34–38] for more details.

3. Sparse Iterative-Thresholding Deconvolution

Let's consider the following minimization problem:

$$(\mathsf{L}_{\psi,\sigma}): \ \min \ \Psi(\boldsymbol{\alpha}) \ \ s.t. \ \|\mathbf{y} - \mathbf{H}\mathbf{\Phi}\boldsymbol{\alpha}\|_{\mathcal{H}} \leq \sigma \tag{4}$$

This problem can also be written in the augmented Lagrangian form,

$$(\mathsf{P}_{\psi,\lambda}): \ \min \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{\Phi}\boldsymbol{\alpha}\|_{\mathcal{H}}^2 + \lambda \sum_{\gamma \in \Gamma} \psi(\boldsymbol{\alpha}_\gamma) \tag{5}$$

Note that $(\mathsf{P}_{\psi,\lambda})$ is equivalent to the MAP/MPLE with independent and identically distributed coefficients. ψ is commonly assumed non-negative, continuous, even-symmetric, and non-decreasing on $\mathbb{R}^+$, but is not necessarily smooth at point zero to produce sparse solutions.

For the ℓ_1 penalty, $(\mathsf{L}_{\ell_1,\sigma})$ can be cast as a convex program and solved using e.g. Iteratively-Reweighted Least-Squares, second order cone programming with interior point algorithms, or active set methods. $(\mathsf{P}_{\ell_1,\lambda})$ corresponds to the Basis Pursuit Denoising (BPDN) [39]. However, these solvers become rapidly time-consuming and intractable in most large-scale practical problems such as those arising in image processing.

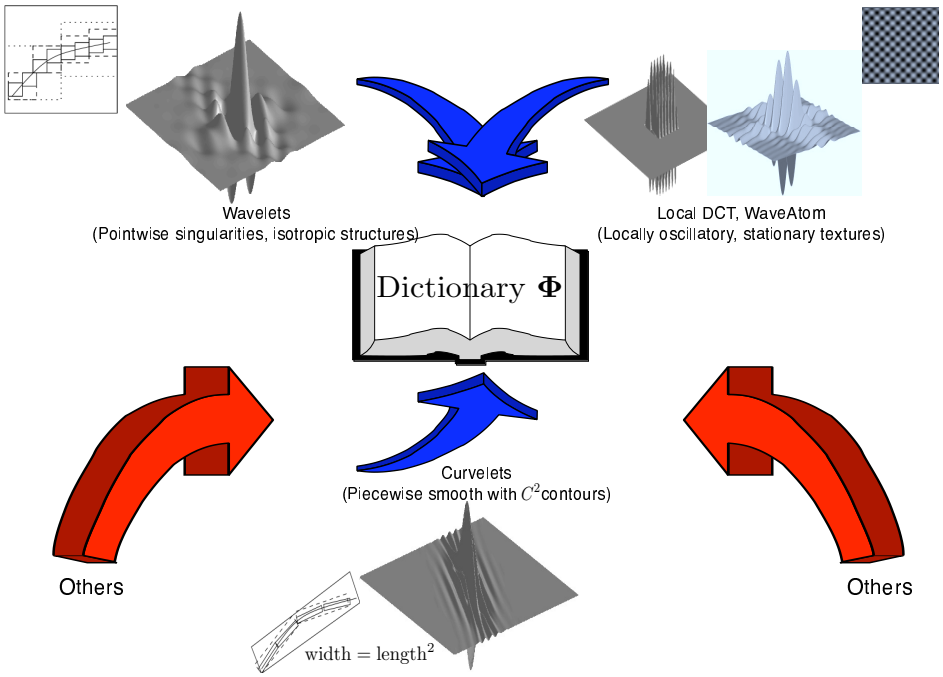

Fig. 1. Example of a typical dictionary in image processing applications.

3.1. *Proximal forward-backward splitting*

The following sequence of iterates solves $(\mathsf{P}_{\psi,\lambda})$:

$$\boldsymbol{\alpha}^{(t+1)} = \mathrm{prox}_{\mu_t,\lambda,\psi}\left(\boldsymbol{\alpha}^{(t)} + \mu_t \boldsymbol{\Phi}^* \mathbf{H}^*\left(\mathbf{y} - \mathbf{H}\boldsymbol{\Phi}\boldsymbol{\alpha}^{(t)}\right)\right) \tag{6}$$

with appropriate μ_t, $\mathrm{prox}_{\mu_t,\lambda,\psi}$ is the Moreau proximity operator associated with the function ψ (see [40] for an introduction to proximal calculus). This proximity operator can be expressed analytically for a wide class of sparsity penalties [41] (i.e. soft thresholding for the ℓ_1 penalty).

Theorem 1 *Suppose that ψ satisfies:*
H 1. *ψ is even-symmetric, non-negative and non-decreasing on $[0,+\infty)$, and $\psi(0)=0$.*
H 2. *ψ is twice differentiable on $\mathbb{R}\setminus\{0\}$ but not necessarily convex.*
H 3. *ψ is continuous on $\mathbb{R}$, it is not necessarily smooth at zero and admits a positive right derivative at zero $\psi'_+(0) = \lim_{h\to 0^+} \frac{\psi(h)}{h} > 0$.*
H 4. *The function $\alpha + \lambda\psi'(\alpha)$ is unimodal on $(0,+\infty)$.*
Then, the proximity operator $\mathrm{prox}_{\delta,\psi}\left(\mathbf{z}^{(t)}\right)$ has exactly one continuous solution decoupled in each coordinate z_l:

$$\hat{\boldsymbol{\alpha}}_l(z_l) = \begin{cases} 0 & \text{if } |z_l| \le \delta\psi'_+(0) \\ z_l - \delta\psi'(\hat{\boldsymbol{\alpha}}_l) & \text{if } |z_l| > \delta\psi'_+(0) \end{cases} \tag{7}$$

A proof can be found in [41].

3.1.1. *Convergence analysis*
In this section, we shall give the convergence properties of the iterative sequence in Eq.6. Convergence is examined in both the weak and strong topologies.

Theorem 2
Let the dictionary be a finite union of K frames, each frame with an upper-bound B_k. Suppose that ψ is coercive and proper convex (but not necessarily smooth). Moreover, suppose that $0 < \|\mathbf{H}\|^2 \le C$ and $0 < \mu_t < 2/\left(C\sum_k B_k\right)$. Then the following holds:

(i) Problem $(\mathsf{P}_{\psi,\lambda})$ *has at least one solution.*
(ii) Problem $(\mathsf{P}_{\psi,\lambda})$ *has exactly one solution if* ψ *is strictly convex.*
(iii) The instance of coefficient iterates $\{\boldsymbol{\alpha}^{(t)}, t \geq 0\}$ *is asymptotically regular and converges weakly to a solution of the deconvolution algorithm.*
(iv) If $\psi(\alpha) = |\alpha|^p,\ 1 \leq p \leq 2$, *the sequence of coefficient iterates converges strongly to a solution of the deconvolution problem.*

The proof follows from [41] and [40, Theorem 5.5] using the canonical basis of $\mathbb{R}^L$.

Remark 1
a. The above result generalizes formally to multi-frames the work of [18,21,20]. It also bears similarities with an independent work of [22].
b. If the frequency response of the PSF does not vanish, then $\mathbf{H\Phi}$ *is bounded below and the solution to* $(\mathsf{P}_{\psi,\lambda})$ *is unique.*
c. The sequence of iterates is asymptotically regular. This furnishes a simple stopping rule test, typically $\left\|\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)}\right\| < \delta$.
d. The case of a single basis:
- *In this case,* $\mathbf{\Phi}$ *is a bijective bounded linear operator. The solution in terms of coefficients is equivalent to the solution in terms of image.*
- *If one has additional constraints to be imposed on the solution such as forcing the image to lie in some closed convex subset of* $\mathcal{H}$ *(e.g. positive cone), the iteration in Eq. 6 can be easily adapted by incorporating the projector onto the constraint set.*

e. The case of a frames:
- *It is less straightforward (at least theoretically) to impose some side constraints on the solution such as positivity.*
- *When the dictionary is a union tight frames, the algorithm is fast only involving fast implicit analysis and synthesis operators. This will be discussed in Section 3.3.*

The proximal approach is simple, competitive and faster than interior-point-based solvers (see Table.3 for its computational complexity). It presents, however, several drawbacks:
– We can only ensure convergence behavior for convex penalties. But typically, soft thresholding, which is associated to the convex ℓ_1-penalty is known to yield biased estimates.
– The choice of the regularization hyperparameter λ is delicate and not intuitive. Finding its optimal value necessitates the use of model selection criteria such as GCV or stochastic sampling approaches (MCMC). These methods are computationally expensive and may produce smoothed images. This second point can be a real problem when we seek compact structures such as is the case in astronomical imaging.

3.2. *Stagewise Iterative Hard Thresholding*

In recent years, considerable attention has focused on the Sparse Solutions Problem (SSP) of linear (i) underdetermined systems; see e.g. [39,35,38,42], and (ii) overdetermined systems (statistical community [43–45]). But, some of the most powerful theoretical results are associated with some of the most computationally intensive methods (e.g. greedy such as OMP, or moderately greedy LARS/Lasso, BPDN). Recent wave of interest in fast solutions to SSP for large-scale problems [46–48]. These approaches can be seen as variants of stagewise forward regression with slightly different forward progress steps (adaptation of LARS/Lasso/homotopy).

Let's denote $\mathbf{A} = \mathbf{H\Phi}$ [1]. Consider the so-called Lasso minimization problem as stated by [43]:

$$(\mathsf{Q}_q): \ \min \ \|\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}\|_{\mathcal{H}}^2 \ \ s.t. \ \|\boldsymbol{\alpha}\|_1 \leq q \tag{8}$$

which is a least-squares fit subject to an ℓ_1 budget constraint. Note that problems $(\mathsf{P}_{\ell_1,\lambda})$ and (Q_q) are equivalent under an appropriate correspondence of parameters λ and q. Indeed, $(\mathsf{P}_{\ell_1,\lambda})$ traces out a solution path for $\lambda \in [0, +\infty)$ with $\hat{\boldsymbol{\alpha}}_{\lambda\to+\infty} = 0$, and $\hat{\boldsymbol{\alpha}}$ getting closer to the solution of $(\mathsf{L}_{\ell_1,0})$ as $\lambda \to 0^+$. (Q_q) identifies a similar path with $\hat{\boldsymbol{\alpha}}_{q=0} = 0$ and $\hat{\boldsymbol{\alpha}}_{q\to+\infty}$ tending to $\hat{\boldsymbol{\alpha}}_{\lambda\to0^+}$. See discussion in [48].

In [43–45], the Lasso, LARS and Homotopy methods were proposed to solve (Q_q) for all $q \geq 0$ (or $(\mathsf{P}_{\ell_1,\lambda})$ for all $\lambda \in [0, \left\|\mathbf{A}^T\mathbf{y}\right\|_\infty]$. The key observation is that the homotopy [2] method (and also LARS/Lasso) follows

[1] In the following, we will assume that the columns of $\mathbf{A}$ are normalized to a unit ℓ_2-norm
[2] Homotopy because the objective undergoes a homotopy from ℓ_2 to ℓ_1 as q decreases.

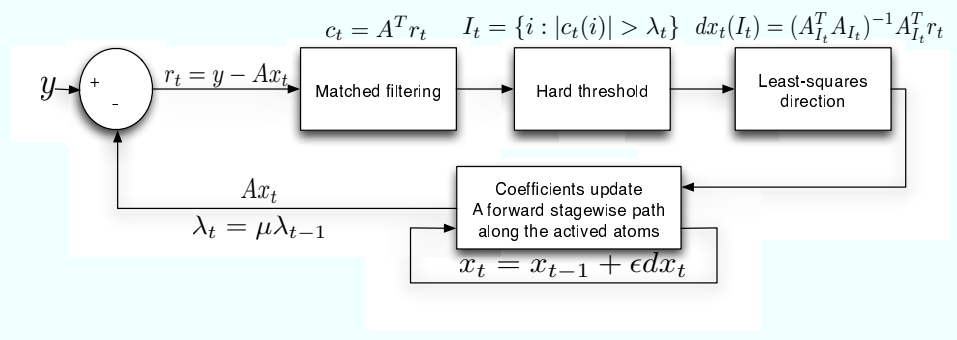

Fig. 2. IT-LARS algorithm.

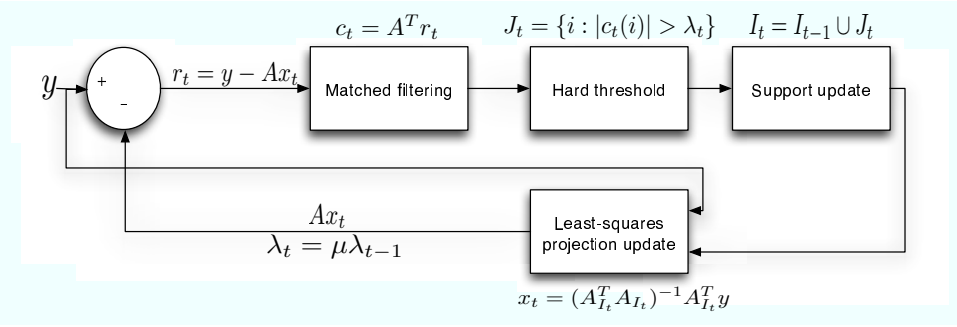

Fig. 3. StOMP algorithm.

a polygonal solution path for all $\lambda \in [0, \left\|\mathbf{A}^T\mathbf{y}\right\|_\infty]$ [45], with changes occurring only at some critical values of λ. Based on these observations, authors in [45] presented the homotopy algorithm, which follows the solution path by jumping from vertex to vertex of this polygonal path. It starts at at $\boldsymbol{\alpha} = 0$ and $\lambda = \left\|\mathbf{A}^T\mathbf{y}\right\|_\infty$, and successively obtains the solutions at a special problem-adapted sequence λ_t. However, the homotopy method computes these solutions by updating the active set considering one coordinate at a time as a candidate to enter or leave the active set. The same behavior is valid for the LARS as the latter is a variant of homotopy where the coefficients are not allowed to leave the active set once activated; see the recent work of [48] who shed light on the intimate connections between LARS, Homotopy and greedy algorithms such as OMP for solving underdetermined SSP. Consequently, such moderately greedy stagewise procedures, although faster than interior point-based solvers [48], can still be computationally demanding for large-scale problems. Therefore, we here adapt the same ideas and propose two stagewise iterative-thresholding (IT) algorithms where the sequence $(\lambda_t)_{t\geq 0}$ is allowed to be strictly decreasing [49,47].

In Fig.2, we summarize the main steps of a stagewise variant of the LARS algorithm (hence the name IT-LARS). The t-th iteration applies matched filtering to the current residual, getting a vector of residual correlations c_t, which is supposed to contain a small fraction of significant nonzero coefficients. The next step performs hard thresholding with threshold λ_t to find the significant nonzeros. Thresholding yields a set of coordinates $I_t = \{i : |c_t(i)| > \lambda_t\}$. The IT-LARS computes an update direction $d\boldsymbol{\alpha}_t$ by projecting the residual r_t onto subspace spanned by the columns of $\mathbf{A}$ belonging to I_t. Then, an update of the solution is computed by augmenting $\boldsymbol{\alpha}_{t-1}$ in the direction $d\boldsymbol{\alpha}_t$ with a step size ϵ (typically chosen equal to the threshold decrease step). The algorithm terminates when the residual $\|r_t\|_2 / \sqrt{n}$ becomes less than the noise standard deviation σ, or when the threshold λ is less than say $k\sigma$ (typically $k = 3$ to 4).

This algorithm is a fast stagewise approximation to LARS. It provides a good approximation to the solution of (Q_q). Furthermore, it is very fast for bases and tight frames (see detailed discussion in the next subsection).

In Fig.2, the steps of the StOMP algorithm are summarized. The StOMP differs from the IT-LARS in the way the estimate update is computed. At each iteration, the StOMP performs a hard thresholding with threshold λ_t to find the significant nonzeros. And a new support estimate is computed by merging the newly selected coordinates with the support estimate at the previous iteration. The next estimate update is computed by least-squares projection of the observation $\mathbf{y}$ in the linear subspace spanned by the columns of $\mathbf{A}$ belonging to the current merged support. As for IT-LARS, termination can be checked using the same stopping criteria on the residual or the threshold.

3.3. *Computational complexity*

The operators $\mathbf{H}$ (resp. $\mathbf{H}^*$) and $\mathbf{\Phi}$ (resp. $\mathbf{\Phi}^T$) are never represented explicitly in matrix form. Rather, there are implemented as fast implicit operators taking a vector $\mathbf{x}$, and returning $\mathbf{Hx}$ (resp $\mathbf{H}^*\mathbf{x}$) and $\mathbf{\Phi x}$ (resp. $\mathbf{\Phi}^T\mathbf{x}$). Multiplication by $\mathbf{H}$ and $\mathbf{H}^*$ costs two FFTs, that is $O(n \log n)$ operations [3] . The complexity of multiplications by $\mathbf{\Phi}$ and $\mathbf{\Phi}^T$ was discussed in Section 2 for some popular transforms.

The computation bottleneck of IT-LARS (resp. StOMP) lies in calculating the least-squares projection step to get $d\boldsymbol{\alpha}_t$ (resp. $\boldsymbol{\alpha}_t$). Following [47], this is implemented via a conjugate gradient (CG) solver, with each iteration involving multiplications by $\mathbf{H}$ (FFT), $\mathbf{\Phi}$ (fast synthesis operator), $\mathbf{H}^*$ (FFT) and $\mathbf{\Phi}^T$ (fast analysis operator). The same reasoning leads to the same computational complexity for the StOMP algorithm. This is summarized in Table 3.

Algorithm	CC
Proximal	$O(2N_{\text{iter}}(\underbrace{V}_{\text{Transform}} + \underbrace{n \log n}_{\text{FFT}})$
IT	$O\left(S(l+2)(V + n \log n)\right)$
StOMP	$O((S(l+2)(V + n \log n))$

Table 1
Computational complexity (CC) of each iterative deconvolution algorithm. N_{iter} is the pre-specified number of iterations of the proximal algorithm. l is the number of iterations for the CG solver for LS projection. S is the number of stages (at most k for k-sparse solutions). V is the complexity of the transform(s) in the dictionary: typically $O(n)$ or $O(n \log n)$ for most popular transforms.

4. Results

To illustrate potential applicability of the three iterative-thresholding algorithms presented above to astronomical image deconvolution, a series of experiments with both simulated and real datasets were conducted. For all experiments: the proximal method was run with $\psi(\alpha) = |\alpha|$ with 100 iterations, and λ was optimized manually to get the highest PSNR. The IT-LARS and StOMP algorithms were run with a threshold decreasing step $\mu = 0.75$. These algorithms terminated when either the residual ℓ_2-norm was less than σ or the threshold λ_t less than 3σ.

A simulated image of spike sources is depicted in Fig.4, upper left. The simulated PSF and the blurred noisy image are shown respectively in the lower left and upper middle quadrant. Restored images are shown respectively in Fig.4 upper right, lower middle and right. As the sources are spiky, a qualifying candidate dictionary is the Dirac basis. Because of the ℓ_1-norm penalty, the proximal approach clearly leads to oversmooth estimates, which degrades the final PSNR. This effect disappears with the iterative hard-thresholding algorithms (a gain of up to 16dB in PSNR). This comes at no additional computational cost and the actual execution times of the different algorithms were comparable (few seconds on this image). Note also that there was no need to choose or adjust the regularization parameter for IT-LARS and StOMP solvers.

A second experiment with a simulated Hubble Space Telescope Wide Field Camera image of a distant cluster of galaxies is shown in Fig.4, upper left. The aberrated and noisy image is shown in the upper middle panel (with the same simulated PSF as in Fig.4) . Deconvolved images are shown respectively in Fig.4 upper right, lower middle and right. The sources being isotropic, a natural choice of the dictionary is the wavelet transform. Again, the proximal approach yields biased estimates. The increase in PSNR when using iterative hard-thresholding algorithms instead of the proximal approach is up to 9dB. The running times of the different algorithms were also comparable (tens of seconds on this image).

Finally, the algorithms were applied to a real image of Jupiter acquired during the calibration stage of the Deep Impact project. Fig.4 shows the observed aberrated and noisy image, top right. The approximate measured PSF is depicted in Fig.4 top right. Restored images with the IT-LARS and the proximal algorithm are portrayed in Fig.4 bottom left and right. Based on the expected morphological content of the image to be restored, the dictionary consisted of the curvelet transform. The details seem to be sharper on the image restored using IT-LARS which is in total agreement with the simulations.

[3] Note that one may expect $O(2n \log n)$. But we supposed our image to contain n pixels.

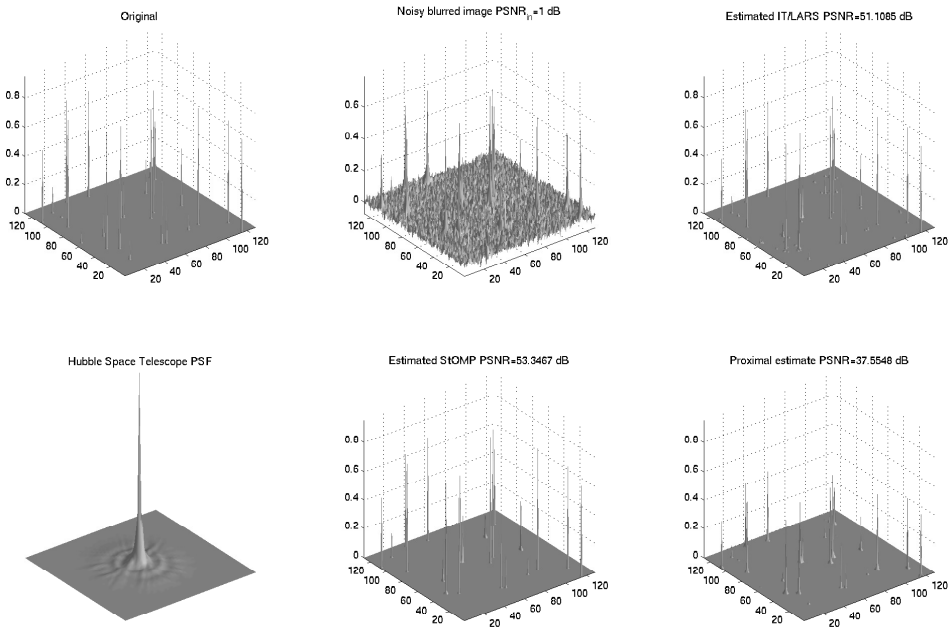

Fig. 4. Simulated image of spike sources. From left to right and top to bottom: original image, noisy blurred image, restored with IT-LARS PSNR = $51.1dB$, simulated PSF, restored with StOMP PSNR = $53.34dB$, restored with the proximal method PSNR = $37.55dB$. The dictionary contained the Dirac basis.

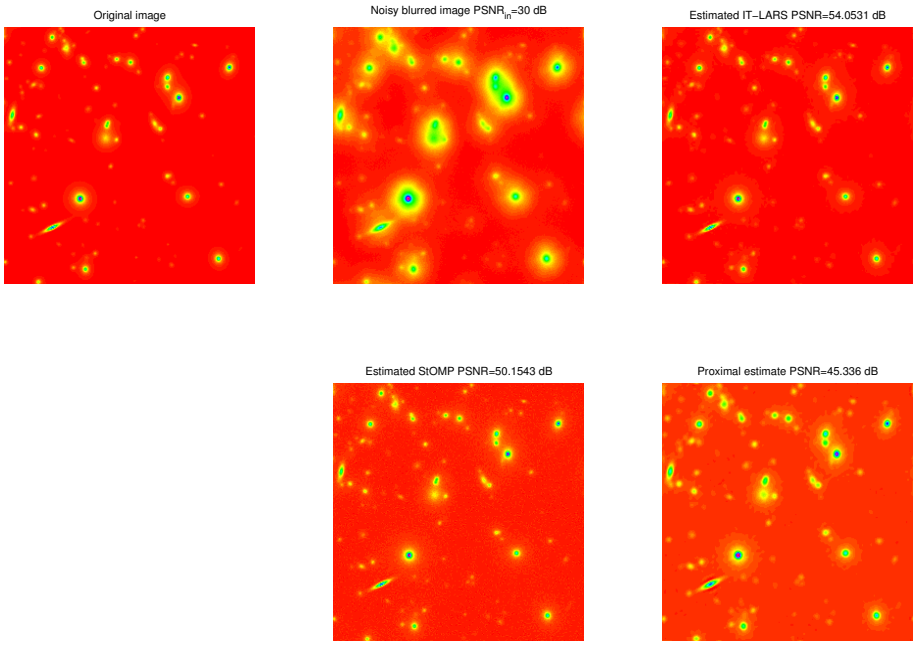

Fig. 5. Simulated Hubble Space Telescope Wide Field Camera image of a distant cluster of galaxies. From left to right and top to bottom: original image, noisy aberrated image, restored with IT-LARS PSNR = $54.05dB$, restored with StOMP PSNR = $50.15dB$, restored with the proximal method PSNR = $45.33dB$. As the image contained isotropic structures, the dictionary was composed of the wavelet transform.

5. Conclusion

In this paper, image deconvolution algorithms with overcomplete sparse representations and fast iterative thresholding solvers were proposed. The proximal approach uses soft thresholding and yields biased (over-smooth) estimates. The iterative hard thresholding-based alternatives, namely IT-LARS and StOMP do not suffer from this drawback. Furthermore, they exhibit an appealing auto-stopping behavior and relieve the user from the delicate choice of the regularization parameter.

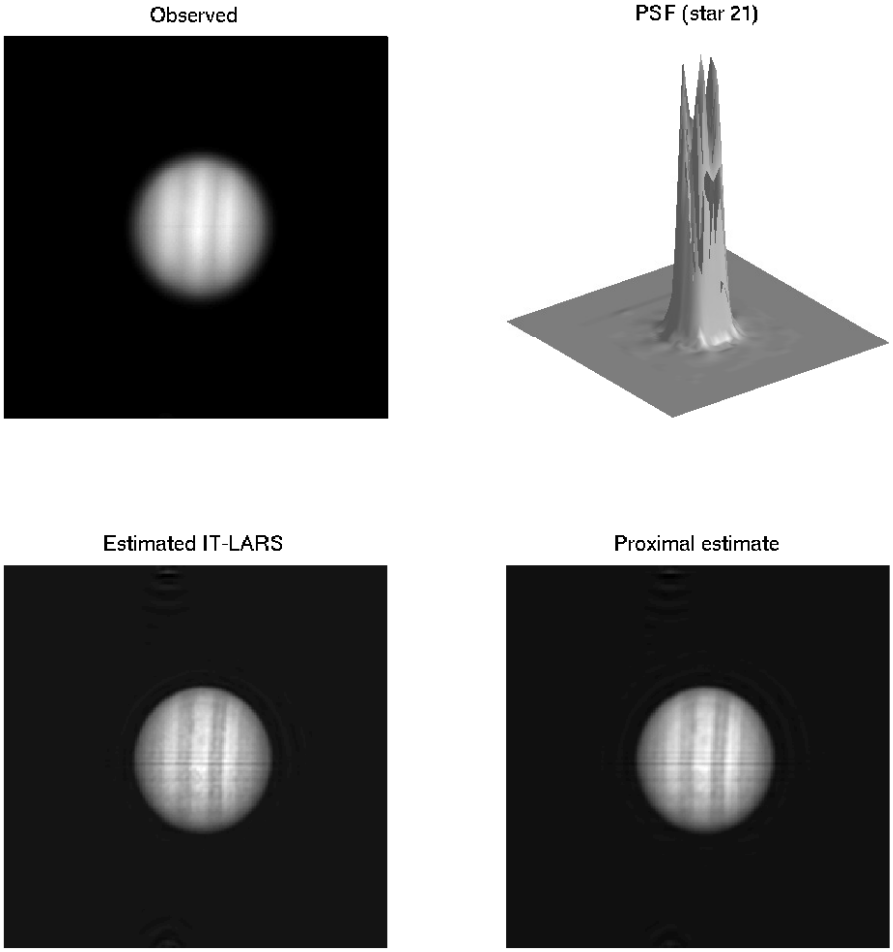

Fig. 6. Deconvolution of a real image of Jupiter acquired during calibration of the Deep Impact project. From left to right and top to bottom: observed noisy aberrated image, measured PSF, restored with IT-LARS, restored with the proximal method. The dictionary consisted of the curvelet transform.

However, several aspects remain to investigate. For instance, the theoretical properties of stagewise iterative hard thresholding for solving linear inverse problems, and deconvolution problems in particular. Extension to deconvolution with other noise models beyond the traditional AWGN case (e.g. Poisson without variance stabilization) is also an important perspective. We believe these points will be our next milestones.

References

[1] A. Katsaggelos. *Digital Image Processing*. Springer-Verlag, 1993.

[2] F. Jeng and J.Woods. Compound gauss-markov random elds for image estimation. *ITSP*, 39:683–697, 1991.

[3] D. Geman, G. Reynolds, and C. Yang. Stochastic algorithms for restricted image spaces and experiments in deblurring. In A. Jain R. Chellappa, editor, *Markov Random Fields Theory and Applications*, pages 39–68. Academic Press, 1993.

[4] R. Molina, A. Katsaggelos, J. Mateos, and J. Abad. Compound gauss markov random fields for astronomical image restoration. *VA*, 40:539, 1996.

[5] A. Jalobeanu. *Modèles, estimation bayśienne et algorithmes pour la dćonvolution d'images satellitaires et aériennes*. PhD thesis, Université de Nice Sophia Antipolis, 2001.

[6] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *English translation of Dokl Akad Nauk SSSR*, 151:501–504, 1963.

[7] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *PAMI*, 14:367–383, 1992.

[8] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[9] L. Blanc-Féraud and M. Barlaud. Edge-preseving restoration of astrophysical images. *VA*, 40(4), 1996.

[10] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *ITIP*, 6(2):298–311, 1997.

[11] Gilles Aubert and Luminita Vese. A variational method in image recovery. *SIAM Journal on Numerical Analysis*, 34(5):1948–1979, 1997.

[12] E. Pantin and J.-L. Starck. Deconvolution of astronomical images using the multiscale maximum entropy method. *Astronomy and Astrophysics, Supplement Series*, 315:575–585, 1996.

[13] J.-L. Starck, F. Murtagh, P. Querre, and F. Bonnarel. Entropy and astronomical data analysis: Perspectives from multiresolution analysis. *Astronomy and Astrophysics*, 368:730–746, 2001.
[14] D.L. Donoho. Nonlinear solution of inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.
[15] I.M. Johnstone. Wavelet deconvolution in a periodic setting. *JR Statist. Soc. B*, 66(3):547–573, 2004.
[16] J. Kalifa, S. Mallat, and B. Rougé. Image deconvolution in mirror wavelet bases. In *IEEE ICIP*, volume 1, pages 565–569, 1998.
[17] R. Neelamani, H. Choi, and R. G. Baraniuk. ForWard: fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, 52(2):418–433, 2004.
[18] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.
[19] M. Figueiredo and R. Nowak. A bound optimization approach to wavelet-based image deconvolution. In *IEEE ICIP*, 2005.
[20] J. Bioucas-Dias. Bayesian wavelet-based image deconvolution: a gem algorithm exploiting a class of heavy-tailed priors. *IEEE Transactions on Image Processing*, 15(4):937–951, 2006.
[21] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57:1413–1541, 2004.
[22] Gerd Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Applied and Computational Harmonic Analysis*, 22(1):43–60, 2007.
[23] J.-L. Starck and F. Murtagh. Image restoration with noise suppression using the wavelet transform. *Astronomy and Astrophysics*, 288:343–348, 1994.
[24] J.-L. Starck, A. Bijaoui, and F. Murtagh. Multiresolution support applied to image filtering and deconvolution. *CVGIP: Graphical Models and Image Processing*, 57:420–431, 1995.
[25] F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics, Supplement Series*, 112:179–189, 1995.
[26] J.-L. Starck, M.K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83(10):2279–2283, 2003.
[27] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer-Verlag, 2nd edition, 2006.
[28] S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, 1998.
[29] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L.L. Schumaker, editors, *Curve and Surface Fitting: Saint-Malo 1999*, Nashville, TN, 1999. Vanderbilt University Press.
[30] E.J. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. Technical report, CalTech, Applied and Computational Mathematics, 2005. To appear in SIAM Multiscale Model. Simul.
[31] E.J. Candès and D.L. Donoho. Ridgelets: the key to high dimensional intermittency? *Philosophical Transactions of the Royal Society of London A*, 357:2495–2509, 1999.
[32] J. L. Starck, E. Candes, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.
[33] L. Demanet and L. Ying. Wave atoms and sparsity of oscillatory patterns. Technical report, CalTech, Applied and Computational Mathematics, 2005. submitted to ACHA.
[34] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.
[35] D.L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52:6–18, 2006.
[36] A.M. Bruckstein and M. Elad. A generalized uncertainty principle and sparse representation in pairs of $\mathbb{R}^N$ bases. *IEEE Transactions on Information Theory*, 48:2558–2567, 2002.
[37] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.
[38] T.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(11):2231–2242, 2004.
[39] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.
[40] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.
[41] M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 2006. Submitted.
[42] E.J. Candès and J. Romberg. Practical signal recovery from random projections. Technical report, CalTech, Applied and Computational Mathematics, 2005.
[43] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.
[44] B. Efron, T. Hastie, I. M. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.
[45] M. R. Osborne, B. Presnell, and B. A. Turlach. A new approach to variable selection in least squares problems. *IMA J. Numerical Analysis*, 20:389–403, 2000.
[46] J. Bobin, J.-L. Starck, M.J. Fadili, Y. Moudden, and D.L. Donoho. Morphological component analysis : New results. *ITIP*, 2006. Submitted.
[47] D.L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report, Stanford University, Department of Statistics, 2006. Submitted.
[48] D.L. Donoho and Y. Tsaig. Fast solution of ℓ_1-norm minimization problems when the solution may be sparse. Technical report, Stanford University, Department of Statistics, 2006. Submitted.
[49] J.-L. Starck, M. Elad, and D.L. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132, 2004.