Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors

F.-X. Dupé[a,b,*], M. J. Fadili[a], J.-L. Starck[b]

[a]GREYC UMR CNRS 6072, Université de Caen Basse-Normandie/ENSICAEN, 6 Bd Maréchal Juin, 14050 Caen, France
[b]Laboratoire AIM, UMR CEA-CNRS-Paris 7, Irfu, SEDI-SAP, Service d'Astrophysique, CEA Saclay, F-91191 GIF-Sur-YVETTE CEDEX, France

Abstract

In this paper, we propose a Bayesian MAP estimator for solving the deconvolution problems when the observations are corrupted by Poisson noise. Towards this goal, a proper data fidelity term (log-likelihood) is introduced to reflect the Poisson statistics of the noise. On the other hand, as a prior, the images to restore are assumed to be positive and sparsely represented in a dictionary of waveforms such as wavelets or curvelets. Both analysis and synthesis-type sparsity priors are considered. Piecing together the data fidelity and the prior terms, the deconvolution problem boils down to the minimization of non-smooth convex functionals (for each prior). We establish the well-posedness of each optimization problem, characterize the corresponding minimizers, and solve them by means of proximal splitting algorithms originating from the realm of non-smooth convex optimization theory. Experimental results are conducted to demonstrate the potential applicability of the proposed algorithms to astronomical imaging datasets.

Keywords: Deconvolution, Poisson noise, Proximal iteration, Iterative thresholding, Sparse representations.

1. Introduction

Deconvolution is a longstanding problem in many areas of signal and image processing (e.g. biomedical imaging [3, 31, 33], astronomy [3, 38], remote-sensing, to quote a few). For example, research in astronomical image deconvolution has seen considerable work, partly triggered by the Hubble space telescope (HST) optical aberration problem at the beginning of its mission. In biomedical imaging, researchers are also increasingly relying on deconvolution to improve the quality of images acquired using complex optical systems, like confocal microscopes [31]. Deconvolution may then prove crucial for exploiting images and extracting scientific content, and is more and more involved in everyday research.

*Corresponding author
Email addresses: francois-xavier.dupe@cea.fr (F.-X. Dupé), Jalal.Fadili@greyc.ensicaen.fr (M. J. Fadili), jean-luc.starck@cea.fr (J.-L. Starck)

An extensive literature exist on the deconvolution problem [1, 25, 41]. In presence of Poisson noise, several deconvolution methods have been proposed such as Tikhonov-Miller inverse filter and Richardson-Lucy (RL) algorithms; see [33, 38] for comprehensive reviews. The RL has been extensively used in many applications, but as it tends to amplify the noise after a few iterations, several extension have been proposed. For example, wavelet-regularized RL algorithm has been proposed by several authors [24, 36, 37]. The interested reader may refer to [3, 13] for a more complete review on deconvolution with Poisson noise.

In the context of deconvolution with either Gaussian or Poisson noise, sparsity-promoting regularization over orthobasis or frame dictionaries has been recently proposed [7, 10, 11, 13, 20, 21, 22, 32, 40].

For instance, in [13] the authors proposed to stabilize the Poisson noise using the Anscombe variance stabilizing transform [2] to bring the problem back to deconvolution with additive white Gaussian noise, but at the price of a non-linear data fidelity term. However, stabilization has a cost, and the performance of such an approach degrades in low intensity regimes. On the other hand, the authors in [32] use directly the data fidelity term related to Poisson noise (see Section 3) and exploit its additive structure together with some particularities of the composition of the convolution operator with tight frames in order to solve the problem using a proximal framework. However, their framework only applies to some special convolution kernels. Using the augmented Lagrangian method with the alternating method of multipliers algorithm, [21] presented a deconvolution algorithm with either a synthesis or an analysis prior. In fact, ADMM is nothing but the Douglas-Rachford splitting [15] applied to the Fenchel-Rockafellar dual objective.

Contributions. In this paper, we propose an image deconvolution algorithm for data blurred and contaminated by Poisson noise using analysis and synthesis-type sparsity priors. In order to form the data fidelity term, we take the exact Poisson likelihood. Putting together the data fidelity and the prior terms, the deconvolution problem is formulated as the minimization of a maximum a posteriori (MAP) objective functional involving three terms: the data fidelity term; a non-negativity constraint (as Poisson data are positive by definition); and a regularization term, in the form of a non-smooth penalty that enforces the sparsity of the sought after image over a —possibly redundant— dictionary of waveforms. We establish the well-posedness of our optimization problems and characterize the corresponding minimizers. We then solve them by means of proximal splitting algorithms originating from the field of non-smooth convex optimization theory. More precisely we take benefit from a generalization of Douglas-Rachford splitting to product spaces; see Section 4.2. In order to use this proximal splitting algorithm, the proximity operator of each individual term in the objective functional is established. Experimental results on several imaging datasets are carried out to demonstrate the potential applicability of the proposed algorithms and compare them with some competitors.

Notation and terminology. Let $\mathcal{H}$ a real Hilbert space, here a finite dimensional real vector space. We denote by $\|.\|$ the norm associated with the inner product $\langle ., . \rangle$ in $\mathcal{H}$, and $\mathbf{I}$ is the identity operator on $\mathcal{H}$. x and α are respectively reordered vectors of image samples and coefficients. We denote by $\operatorname{ri} \mathcal{C}$ the relative interior of a convex set $\mathcal{C}$.

A real-valued function f is coercive, if $\lim_{\|z\| \to +\infty} f(z) = +\infty$. The domain of f is defined by $\operatorname{dom} f = \{z \in \mathcal{H} \mid f(z) < +\infty\}$ and f is proper if $\operatorname{dom} f \neq \emptyset$. A real-valued

function f is lower semi-continuous (lsc) if $\liminf_{z \to z_0} f(z) \geqslant f(z_0)$. $\Gamma_0(\mathcal{H})$ is the class of all proper lsc convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$. The subdifferential of a function $f \in \Gamma_0(\mathcal{H})$ at $z \in \mathcal{H}$ is the set $\partial f(z) = \{u \in \mathcal{H} \mid \forall y \in \mathcal{H}, f(y) \geqslant f(z) + \langle u, y - z \rangle\}$. Its Legendre-Fenchel conjugate is f^*. We denote by $\imath_{\mathcal{C}}$ the indicator of the convex set $\mathcal{C}$: $\imath_{\mathcal{C}}(z) = \begin{cases} 0, & \text{if } z \in \mathcal{C}, \\ +\infty, & \text{otherwise.} \end{cases}$. See [26] for a more comprehensive account on convex analysis.

We denote by $\|\mathbf{M}\| = \max_{z \neq 0} \frac{\|\mathbf{M}z\|}{\|z\|}$ the spectral norm of $\mathbf{M}$.

2. Sparsity prior

Let $x \in \mathbb{R}^n$ be an $\sqrt{n} \times \sqrt{n}$ image. We denote by $\boldsymbol{\Phi}$ the dictionary, i.e. the $n \times L$ matrix whose columns are the generating waveforms (called atoms) $(\varphi_i)_{1 \leqslant i \leqslant L}$ all normalized to a unit ℓ_2-norm. The forward (analysis) transform is then defined by a non-necessarily square matrix $\boldsymbol{\Phi}^{\mathrm{T}} \in \mathbb{R}^{L \times n}$ with $L \geqslant n$. When $L > n$ the dictionary is said to be redundant or overcomplete. A dictionary matrix $\boldsymbol{\Phi}$ is said to be a frame with bounds c_1 and c_2, $0 < c_1 \leqslant c_2 < +\infty$, if $c_1 \|x\|^2 \leqslant \left\|\boldsymbol{\Phi}^{\mathrm{T}} x\right\|^2 \leqslant c_2 \|x\|^2$. A frame is tight when $c_1 = c_2 = c$, i.e. $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}} = c\mathbf{I}$.

In a synthesis prior model, the image x can be written as the superposition of the elementary atoms φ_i according to the following linear generative model $x = \sum_{i=1}^{L} \alpha[i]\varphi_i = \boldsymbol{\Phi}\alpha$. x is said to be sparse (resp. compressible)[1] in $\boldsymbol{\Phi}$ if it can be represented exactly (resp. approximately) as a superposition of a small fraction of the atoms in the dictionary $\boldsymbol{\Phi}$ compared to the ambient dimension n. In an analysis-type prior model, the transform coefficients $\boldsymbol{\Phi}^{\mathrm{T}} x$ of the image x are assumed to be sparse.

In the sequel, the dictionary $\boldsymbol{\Phi}$ corresponds to a frame and will be built by taking union of one or several transforms. For instance, for astronomical objects such as stars, the wavelet transform is a very good candidate [27]. For more anisotropic or elongated structures, ridgelets or curvelets would be a better choice [4].

3. Problem statement

Consider the image formation model where an input image of n pixels x is blurred by a point spread function (PSF) h and contaminated by Poisson noise. The observed image is then a discrete collection of counts $y = (y_i)_{1 \leqslant i \leqslant n}$ which are bounded, i.e. $y \in \ell_\infty$. Each count y_i is a realization of an independent Poisson random variable with a mean $(h \circledast x)_i$, where $\circledast$ is the circular convolution operator. Formally, this writes,

$$y_i \sim \mathcal{P}\left((h \circledast x)_i\right) . \tag{1}$$

This formula can be rewritten in a vector form as $y \sim \mathcal{P}(\mathbf{H}x)$, where $\mathbf{H}$ is a circular convolution matrix. The deconvolution problem at hand is to restore x from the observed count image y.

---

[1] With a slight abuse of terminology, in both cases we will use the term sparse.

A natural way to attack this problem would be to adopt a maximum a posteriori (MAP) bayesian framework with an appropriate likelihood function — the distribution of the observed data y given an original x — reflecting the Poisson statistics of the noise. As a prior, the image is supposed to be economically (sparsely) represented in a pre-chosen dictionary $\mathbf{\Phi}$ as measured by a sparsity-promoting penalty Ψ supposed throughout to be proper, lsc and convex but non-smooth, e.g. the ℓ_1 norm.

From the probability density function of a Poisson random variable, the likelihood writes:

$$p(y|x) = \prod_i \frac{((\mathbf{H}x)[i])^{y[i]} \exp\left(-(\mathbf{H}x)[i]\right)}{y[i]!} \ , \tag{2}$$

and the associated log-likelihood function is

$$\ell\ell(y|x) = \sum_i \left(y[i] \log((\mathbf{H}x)[i]) - (\mathbf{H}x)[i] - \log(y[i]!)\right) \ . \tag{3}$$

These formula are extended to the case $y = 0$, using the convention $0! = 1$. Taking negative log-likelihood, we arrive at the following data fidelity term:

$$f_1 \ : \eta \in \mathbb{R}^n \mapsto \sum_{i=1}^n f_{\text{poisson}}(\eta[i]), \tag{4}$$

$$\text{if } y[i] > 0, \quad f_{\text{poisson}}(\eta[i]) = \begin{cases} -y[i] \log(\eta[i]) + \eta[i] & \text{if } \eta[i] > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

$$\text{if } y[i] = 0, \quad f_{\text{poisson}}(\eta[i]) = \begin{cases} \eta[i] & \text{if } \eta[i] \in [0, +\infty), \\ +\infty & \text{otherwise.} \end{cases}$$

Our aim is then to solve the following optimization problems, with either an analysis-type prior,

$$\underset{x \in \mathbb{R}^n}{\text{argmin}} \ J(x),$$

$$J \ : \ x \mapsto f_1 \circ \mathbf{H}(x) + \underbrace{\gamma \Psi \circ \mathbf{\Phi}^{\mathrm{T}}(x)}_{f_2 \circ \mathbf{\Phi}^{\mathrm{T}}(x)} + \underbrace{\iota_{\mathcal{C}}(x)}_{f_3(x)} \ , \tag{$\mathrm{P}_{\gamma,\psi}$}$$

or a synthesis-type prior,

$$\underset{\alpha \in \mathbb{R}^L}{\text{argmin}} \ J'(\alpha),$$

$$J' \ : \ \alpha \mapsto f_1 \circ \mathbf{H} \circ \mathbf{\Phi}(\alpha) + \underbrace{\gamma \Psi(\alpha)}_{f_2(\alpha)} + \underbrace{\iota_{\mathcal{C}} \circ \mathbf{\Phi}(\alpha)}_{f_3 \circ \mathbf{\Phi}(\alpha)} \ . \tag{$\mathrm{P}'_{\gamma,\psi}$}$$

The penalty function Ψ is additive, i.e. $\Psi(\alpha) = \sum_{i=0}^L \psi(\alpha[i])$, $\gamma > 0$ is a regularization parameter and $\iota_{\mathcal{C}}$ is the indicator function of the closed convex set $\mathcal{C}$. In our case, $\mathcal{C}$ is the positive orthant since we are fitting Poisson intensities, which are positive by nature.

In ($\mathrm{P}_{\gamma,\psi}$), the solution image $x_{\mathbf{A}}^\star$ is directly targeted whose transform (analysis) coefficients $\mathbf{\Phi}^{\mathrm{T}} x_{\mathbf{A}}^\star$ are the sparsest. While in problem ($\mathrm{P}'_{\gamma,\psi}$), we are seeking a sparse set of coefficients $\alpha_S^\star$ and the solution signal or image is

synthesized from these representation coefficients and the dictionary $\mathbf{\Phi}$ as $x_{\mathsf{S}}^\star = \mathbf{\Phi}\alpha_{\mathsf{S}}^\star$.

Obviously, problems $(\mathrm{P}_{\gamma,\psi})$ and $(\mathrm{P}'_{\gamma,\psi})$ are not equivalent in general unless $\mathbf{\Phi}$ is an orthobasis. For overcomplete non-orthogonal dictionaries, the solutions to synthesis and analysis formulations are different. Indeed, in the synthesis prior the set of solutions $x_{\mathsf{S}}^\star$ is confined to the column space of the dictionary, whereas in the analysis formulation, the solutions $x_{\mathsf{A}}^\star$ are allowed to be arbitrary vectors in $\mathbb{R}^n$. Furthermore, for redundant dictionaries, there are much fewer unknowns in the analysis formulation, hence leading to a "simpler" optimization problem, although the proximity operators become more complicated because of composition even for simple functions. As opposed to the analysis approach, the synthesis approach has a constructive form providing an explicit description of the signals it represents, and as such, can benefit from higher redundancy to synthesize rich and complex signals. On the other hand, in the analysis approach, we can ask a signal or image to simultaneously agree with many columns of $\mathbf{\Phi}$. This might become impossible with a highly redundant dictionary.

It can be shown that for a Parseval tight frame (PTF) dictionary $\mathbf{\Phi}$, using the change of variable $\alpha = \mathbf{\Phi}^{\mathrm{T}}x$, an analysis-type prior formulation can be written as a linearly constrained synthesis-type prior formulation. The constraint ensures that α remains in the span of $\mathbf{\Phi}^{\mathrm{T}}$. Therefore, analysis-type prior forms can be seen as a subset of synthesis-type prior ones; see [17] for another geometrical argument. If a global minimizer of the synthesis-type prior satisfies the span analysis constraint, i.e. is a feasible analysis solution, then both problems are equivalent. Other equivalence conditions can be drawn if stronger assumptions (but useless to promote sparse solutions) are imposed on the penalty Ψ; see e.g. [17].

With the notable exception of [17], and a recent paper by [14] on compressed sensing (with PTF dictionaries), a little is known about the actual theoretical guarantees of analysis-type priors in general. From a practical point of view, there is no consensus either on the type of conclusions to be drawn from the experimental work reported in the literature. Some authors have reported results where the analysis prior is superior to its synthesis counterpart, e.g. [5, 17, 34]. Others have shown that they are more or less equivalent. In our work here, the synthesis prior turns out to be better on the Saturn image, presumably because the dictionary is very well suited to sparsely synthesize the image (see Section 6.3).

In short, these phenomena are still not very well understood, especially for general inverse problems operators and dictionaries $\mathbf{\Phi}$. More investigation is needed in this direction to extricate the deep distinctions and similarities between the two priors.

3.1. Properties of the objective functions

From $(\mathrm{P}_{\gamma,\psi})$ and $(\mathrm{P}'_{\gamma,\psi})$, we have the following,

Proposition 1.

(i) f_1 is a convex function and so are $f_1 \circ \mathbf{H}$ and $f_1 \circ \mathbf{H} \circ \mathbf{\Phi}$.

5

(ii) f_1 is strictly convex if $\forall i \in \{1, \ldots, n\}, y[i] \neq 0$. $f_1 \circ \mathbf{H}$ remains strictly convex if $\ker(\mathbf{H}) = \emptyset$, and so does $f_1 \circ \mathbf{H} \circ \mathbf{\Phi}$ if $\mathbf{\Phi}$ is an orthobasis and $\ker(\mathbf{H}) = \emptyset$.

(iii) The gradient of f_1 is

$$\nabla f_1(\eta) = (g(\eta[i]))_{1 \leqslant i \leqslant n}, \tag{5}$$

$$\text{if } y[i] > 0, \quad g(\eta[i]) = \begin{cases} 1 - \frac{y[i]}{\eta[i]} & \text{if } \eta[i] > 0, \\ +\infty & \text{else.} \end{cases}$$

$$\text{if } y[i] = 0, \quad g(\eta[i]) = \begin{cases} 1 & \text{if } \eta[i] \geqslant 0, \\ +\infty & \text{else.} \end{cases}$$

(iv) Suppose that $(0, +\infty) \cap \mathbf{H}([0, +\infty)) \neq \emptyset$ and Ψ has full domain. Then both $J \in \Gamma_0(\mathbb{R}^n)$ and $J' \in \Gamma_0(\mathbb{R}^L)$.

Proof. The proof of (i) and (ii) follow from well-known properties of convex functions. (iii) is obtained by straightforward calculations. (iv): J and J' are the sums of lsc convex functions which entails their convexity and lower semicontinuity. $\mathbf{\Phi}$ is a frame, hence surjective. Thus under the assumptions of (iv), the domains of both J and J' are non-empty, i.e. proper functions. $\square$

3.2. Well-posedeness of $(\mathrm{P}_{\gamma, \psi})$ and $(\mathrm{P}'_{\gamma, \psi})$

Let $\mathcal{M}$ and $\mathcal{M}'$ be respectively the sets of minimizers of problems $(\mathrm{P}_{\gamma, \psi})$ and $(\mathrm{P}'_{\gamma, \psi})$. Suppose that $\exists i \in \{1, \ldots, n\}$ such that $y[i] \neq 0$, otherwise the minimizer would be trivially 0. Thus J and J' are coercive ($\mathbf{\Phi}^\mathrm{T}$ is injective). Therefore, the following holds:

Proposition 2.

1. *Existence: each of $(\mathrm{P}_{\gamma, \psi})$ and $(\mathrm{P}'_{\gamma, \psi})$ has at least one solution, i.e. $\mathcal{M} \neq \emptyset$ and $\mathcal{M}' \neq \emptyset$.*

2. *Uniqueness: $(\mathrm{P}_{\gamma, \psi})$ and $(\mathrm{P}'_{\gamma, \psi})$ have unique solutions if Ψ is strictly convex, or under (ii) of Proposition 1.*

4. Iterative minimization algorithm

We are now ready to describe the proximal splitting algorithm to solve $(\mathrm{P}_{\gamma, \psi})$ and $(\mathrm{P}'_{\gamma, \psi})$. At the heart of this splitting framework is the notion of proximity operator which we are about to introduce.

4.1. Proximity operator

The proximity operator is a generalization of the projection onto a closed convex set, introduced by Moreau [28, 29, 30],

Definition 3 (Moreau[28]). *Let $f \in \Gamma_0(\mathcal{H})$. Then, for every $x \in \mathcal{H}$, the function $y \mapsto f(y) + \|x - y\|^2 / 2$ achieves its infimum at a unique point denoted by $\mathrm{prox}_f x$. The operator $\mathrm{prox}_f : \mathcal{H} \to \mathcal{H}$ thus defined is the proximity operator of f.*

The proximity operator enjoys many properties, among these, one can easily show that $\forall x, p \in \mathcal{H}$

$$p = \operatorname{prox}_f x \iff x - p \in \partial f(p) \iff \langle y - p, x - p \rangle + f(p) \leqslant f(y) \ \forall y \in \mathcal{H} \ , \quad (6)$$

which means that prox_f is the resolvent of the subdifferential of f [15]. Recall that the resolvent of the subdifferential ∂f is the single-valued operator $J_{\partial f} : \mathcal{H} \to \mathcal{H}$ such that $\forall x \in \mathcal{H}, x - J_{\partial f}(x) \in \partial f(J_{\partial f}) \iff J_{\partial f} = (\mathbf{I} + \partial f)^{-1}$.

4.2. A splitting algorithm for sums of convex functions

Suppose that the objective function can be expressed as the sum of K convex, lsc and proper functions,

$$\operatorname*{argmin}_{x \in \mathcal{H}} \left(f(x) = \sum_{i=1}^{K} f_i(x) \right) . \quad (7)$$

Proximal splitting methods for solving (7) are iterative algorithms which may evaluate the individual proximity operators $\operatorname{prox}_{f_i}$ (perhaps approximately), but never proximity operators of sums of the f_i, and a fortiori that of f. The key idea relies on the formulation of (7) which is such that the proximity operator of each individual function has some (relatively) convenient structure, while those of the sums do not in general. This turns out to be true in our case for problems $(\mathrm{P}_{\gamma,\psi})$ and $(\mathrm{P}'_{\gamma,\psi})$.

Splitting algorithms have an extensive literature since the 1970's, where the case $K = 2$ predominates. Usually, splitting algorithms handling $K > 2$ have either explicitly or implicitly relied on reduction of (7) to the case $K = 2$ in the product space $\mathcal{H}^K$. For instance, applying the Douglas-Rachford splitting to the reduced form produces Spingarn's method [35], which performs independent proximal steps on each f_i, and then computes the next iterate by essentially averaging the individual proximity operators. The method proposed in [9] is very similar in spirit to Spingarn's method, where moreover, it allows for inexact computation of the proximity operators.

For every $i \in \{1, \ldots, K\}$, let $(a_{t,i})_{t \in \mathbb{N}}$ be a sequence in $\mathcal{H}$. Let $(x_t)_{t \in \mathbb{N}}$ be the sequence as constructed by Algorithm 1. The following result due to [9] establishes the convergence of $(x_t)_{t \in \mathbb{N}}$. We reproduce it here for the sake of completeness.

Theorem 4 ([9]). *Suppose that f is coercive and let $(x_t)_{t \in \mathbb{N}}$ be a sequence generated by Algorithm 1 under the following assumptions,*

1. *f is a proper function;*

2. *$(0, \ldots, 0) \in \operatorname{ri} \{(x - x_1, \ldots, x - x_K) \mid x \in \mathcal{H}, x_1 \in \operatorname{dom} f_1, \ldots, x_K \in \operatorname{dom} f_K\}$;*

3. *$\sum_{t \in \mathbb{N}} \theta_t (2 - \theta_t) = +\infty$;*

4. *$\forall i \in \{1, \ldots, K\} \quad \sum_{t \in \mathbb{N}} \theta_t \left\| a_{(t,i)} \right\| < +\infty.$*

Then $(x_t)_{t \in \mathbb{N}}$ converges to a (non-strict) global minimizer.

Notice that the sequences $(a_{(t,i)})_{t \in \mathbb{N}}$ allow for some errors in the evaluation of the different proximity operators, with the proviso that these errors remain summable. This is useful when the closed forms of the proximity operators are difficult to construct and

Algorithm 1

Task: Solve the convex optimization problem (7).
Initialization:
Choose $\mu \in (0, +\infty)$, $(p_{(0,i)})_{1 \leqslant i \leqslant K} \in \mathcal{H}^K$, $(\omega_i)_{1 \leqslant i \leqslant K} \in (0,1]^K$ satisfying $\sum_{i=1}^K \omega_i = 1$ and $x_0 = \sum_{i=1}^K \omega_i p_{(0,i)}$.
Main iteration:
For $t = 0$ to $N_{\text{ext}} - 1$,

$$\forall i \in \{1, \ldots, K\} \quad \xi_{(t,i)} = \text{prox}_{\mu f_i/\omega_i} p_{(t,i)} + a_{(t,i)} \ ,$$

$$\xi_t = \sum_{i=1}^K \omega_i \xi_{(t,i)} \ ,$$

Choose $\theta_t \in]0, 2[\ ,$

$$\forall i \in \{1, \ldots, K\} \quad p_{(t+1,i)} = p_{(t,i)} + \theta_t \left(2\xi_t - x_t - \xi_{(t,i)} \right) \ ,$$

$$x_{t+1} = x_t + \theta_t (\xi_t - x_t) \ .$$

End main iteration
Output: A solution of (7): $x_{N_{\text{ext}}}$.

rather, must be computed via an inner iteration. This will turn to be the case for some terms involved in the two objective functions of $(P_{\gamma,\psi})$ and $(P'_{\gamma,\psi})$. More precisely, the difficulty at stake is how to deal with the composition with a bounded linear operator, here the circular convolution operator and/or the dictionary. This is what we are about to handle.

4.3. Proximity operator of a pre-composition with a linear operator

The following result will manifest its importance as a building block to solve the two problems $(P_{\gamma,\psi})$ and $(P'_{\gamma,\psi})$. Indeed, it expresses the proximity operator of a function $f \in \Gamma_0(\mathcal{H})$ composed with an affine operator $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto \mathbf{F}x - y$, $y \in \mathbb{R}^m$ where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is a bounded linear operator.

Theorem 5.
Let $\mathbf{F}$ be a linear bounded operator such that the domain qualification condition $\text{ri}(\text{dom}(f) \cap \text{Im}(\mathbf{A})) \neq \emptyset$ holds. Then $f \circ \mathbf{A} \in \Gamma_0(\mathbb{R}^m)$ and

(i) If $\mathbf{F}$ is a tight frame. Then,

$$\text{prox}_{f \circ \mathbf{A}}(x) = y + c^{-1}\mathbf{F}^{\text{T}}(\text{prox}_{cf} - \mathbf{I})\mathbf{A}(x). \tag{8}$$

(ii) If $\mathbf{F}$ is a general frame. Let $\tau_t \in (0, 2/c_2)$. Let $(u_t)_{t \in \mathbb{N}}$ be sequence of iterates provided by Algorithm 2. Then, $(u_t)_{t \in \mathbb{N}}$ converges to $\bar{u}$ and $(p_t)_{t \in \mathbb{N}}$ converges to $\text{prox}_{f \circ \mathbf{A}} x = x - \mathbf{F}^{\text{T}}\bar{u}$. More precisely, these two sequences converge linearly and the best rate is attained for $\tau_t \equiv 2/(c_1 + c_2)$:

$$\left\| p_t - \text{prox}_{f \circ \mathbf{A}}(x) \right\| \leqslant \sqrt{\frac{c_2}{c_1}} \left(\frac{c_2 - c_1}{c_2 + c_1} \right)^t \left\| p_0 - \text{prox}_{f \circ \mathbf{A}}(x) \right\| \ . \tag{9}$$

Algorithm 2

Task: Forward-backward algorithm to compute $\text{prox}_{f \circ \mathbf{A}}(x)$.

Parameters: The function f, the linear bounded operator $\mathbf{A}$, number of iterations N_{int} and step-size $\tau_t \in (0, 2/c_2)$.

Initialization: Choose $u_0 \in \mathbb{R}^m$, $p_0 = x - \mathbf{F}^{\text{T}} u_0$.

Main iteration:

For $t = 0$ **to** $N_{\text{int}} - 1$,

$$
\begin{aligned}
u_{t+1} &= \tau_t (\mathbf{I} - \text{prox}_{f/\tau_t})(u_t/\tau_t + \mathbf{A} p_t), \\
p_{t+1} &= x - \mathbf{F}^{\text{T}} u_{t+1}.
\end{aligned}
\tag{10}
$$

End main iteration

Output: The proximity operator of $f \circ \mathbf{A}$ at x : $p_{N_{\text{int}}}$.

(iii) *In all other cases, apply Algorithm 2 with $\tau_t \in (0, 2/c_2)$. Then, $(u_t)_{t \in \mathbb{N}}$ converges to $\bar{u}$, and $(p_t)_{t \in \mathbb{N}}$ converges to $\text{prox}_{f \circ \mathbf{A}} x = x - \mathbf{F}^{\text{T}} \bar{u}$ at the rate $\mathcal{O}(1/t)$.*

See the appendix for a concise proof.

When $\mathbf{F}$ is a frame, the convergence speed of the one-step forward-backward scheme [23, 39] depends clearly on the redundancy of the frame. The higher the redundancy, the slower the convergence, i.e. the number of necessary iterations to obtain an ε-accurate solution is $\mathcal{O}\left(\frac{c_1}{c_2} \log \varepsilon^{-1}\right)$. For the general case (iii) where $\mathbf{F}$ is not a frame, the algorithm necessitates as large as $\mathcal{O}(1/\varepsilon)$ iterations to reach an ε-accuracy on the iterates.

Assessing the convergence rate as we have just done is important when this algorithm will be used in a nested scheme as a sub-iteration to compute the proximity operator of the composition with a linear operator. Indeed, Theorem 4 and the discussion thereafter, clearly states the stability of Algorithm 1 to errors in the individual proximity operators of the functions $F_i, \ i = 1, \cdots, K$, provided that the errors remain summable. In a nutshell, this means that if Algorithm 2 is used within Algorithm 1, a sufficient number of inner iterations N_{int} must be chosen to ensure error summability. This number of inner iterations obviously depends on the convergence speed of Algorithm 2, hence on the structure of F.

4.4. Proximity operator of a sparsity-promoting penalty

To implement the above iteration, we need to express $\text{prox}_{\gamma \psi}$. The following result gives us the formula for a wide class of penalties ψ :

Lemma 6. *Suppose that ψ satisfies, (i) ψ is convex even-symmetric , non-negative and non-decreasing on $[0, +\infty)$, and $\psi(0) = 0$. (ii) ψ is twice differentiable on $\mathbb{R} \setminus \{0\}$. (iii) ψ is continuous on $\mathbb{R}$, it is not necessarily smooth at zero and admits a positive right derivative at zero $\psi'_+(0) = \lim_{h \to 0^+} \frac{\psi(h)}{h} > 0$. Then, the proximity operator of $\gamma \Psi(\alpha)$,*

$\mathrm{prox}_{\gamma\Psi}(\alpha)$ *has exactly one continuous solution decoupled in each coordinate* α_i :

$$\mathrm{prox}_{\gamma\psi}(\alpha_i) = \begin{cases} 0 & \text{if } |\alpha_i| \leqslant \gamma\psi'_+(0) \\ \alpha_i - \gamma\psi'(\bar{\alpha}_i) & \text{if } |\alpha_i| > \gamma\psi'_+(0) \end{cases} \tag{11}$$

A proof of this lemma can be found in [19]. A similar result is also proposed in [8]. Among the most popular penalty functions ψ satisfying the above requirements, we have $\psi(\alpha_i) = |\alpha_i|$, in which case the associated proximity operator is the popular soft-thresholding.

4.5. Proximity operator of the data fidelity f_1

The following result can be proved easily by solving the proximal optimization problem in Definition 3 with f_1 as defined in (4), see also [7].

Lemma 7. *Let y be the count map (i.e. the observations), the proximity operator associated to* f_1 *(i.e. the Poisson anti log-likelihood) is,*

$$\mathrm{prox}_{\beta f_1} x = \left(\frac{x[i] - \beta + \sqrt{(x[i] - \beta)^2 + 4\beta y[i]}}{2} \right)_{1 \leqslant i \leqslant n} . \tag{12}$$

5. Sparse iterative deconvolution

The game to play now is to plug the previous results to compute the proximity operators involved in problems $(\mathrm{P}_{\gamma,\psi})$ and $(\mathrm{P}'_{\gamma,\psi})$ in Algorithm 1.

5.1. Analysis-type prior

For problem $(\mathrm{P}_{\gamma,\psi})$, the different proximity operator are computed as follows:

(1) $\mathrm{prox}_{f_1 \circ \mathbf{H}}$ is obtained owing to Theorem 5(iii) and Lemma 7;

(2) $\mathrm{prox}_{f_2 \circ \mathbf{\Phi}^\mathrm{T}} = \mathrm{prox}_{\gamma\Psi \circ \mathbf{\Phi}^\mathrm{T}}$ is given by Theorem 5(iii) and Lemma 6;

(3) $\mathrm{prox}_{f_3} = \mathrm{prox}_{\iota_\mathcal{C}}$ is directly is the euclidean projection onto the closed convex set $\mathcal{C}$.

5.2. Synthesis-type prior

As far as problem $(\mathrm{P}'_{\gamma,\psi})$ is concerned, here is how the proximity operators are computed:

(1) $\mathrm{prox}_{f_1 \circ \mathbf{H} \circ \mathbf{\Phi}}$ is computed using Theorem 5 and Lemma 7: use Theorem 5(i) and Theorem 5(iii) if $\mathbf{\Phi}$ is a tight frame, or Theorem 5(iii) otherwise;

(2) $\mathrm{prox}_{f_2} = \mathrm{prox}_{\gamma\Psi}$ is given using Lemma 6;

(3) $\mathrm{prox}_{f_3 \circ \mathbf{\Phi}} = \mathrm{prox}_{\iota_\mathcal{C} \circ \mathbf{\Phi}}$ is given by Theorem 5(i) or (ii) and the projector onto the closed convex set $\mathcal{C}$: use Theorem 5(i) if $\mathbf{\Phi}$ is a tight frame, or Theorem 5(ii) if it is a general frame;

For example, let f_2 be the ℓ_1-norm (i.e. $\psi = |.|$) and $f_3 = \iota_\mathcal{C} \circ \mathbf{\Phi}$, where $\mathcal{C}$ is the positive orthant and $\mathbf{\Phi}$ is a tight frame, then $(\mathrm{P}'_{\gamma,\psi})$ is solved by Algorithm 3.

Algorithm 3

Task: Image deconvolution with Poisson noise, solve $(\mathrm{P}'_{\gamma,\psi})$.

Parameters: The observed image counts y, the dictionary $\boldsymbol{\Phi}$, number of main iterations N_{ext}, number of sub-iterations N_{int} for the proximal operator of the data fidelity term, $\mu > 0$ and regularization parameter γ.

Initialization:

$\forall i \in \{0,1,2\}, \quad p_{(0,i)} = \boldsymbol{\Phi}^{\mathrm{T}} y$.

$\alpha_0 = \boldsymbol{\Phi}^{\mathrm{T}} y$.

Main iteration:

For $t = 0$ **to** $N_{\mathrm{ext}} - 1$,

- Data fidelity (Lemma 7 and Theorem 5(i)-(iii)): $\xi_{(t,0)} = \mathrm{prox}_{\mu f_1 \circ \mathbf{H} \circ \boldsymbol{\Phi}/3}\, p_{(t,0)}$.

- Sparsity-penalty (Lemma 6): $\xi_{(t,1)} = \mathrm{prox}_{\mu\gamma\Psi/3}\, p_{(t,1)} = \mathrm{ST}_{\mu\gamma/3}(p_{(t,1)})$.

- Positivity constraint (Theorem 5(i)): $\xi_{(t,2)} = \mathrm{prox}_{\iota_\mathcal{C} \circ \boldsymbol{\Phi}/3}(p_{(t,3)})$.

- Average the proximity operators: $\xi_t = (\xi_{(t,0)} + \xi_{(t,1)} + \xi_{(t,2)})/3$.

- Choose $\theta_t \in {]0,2[}$.

- Update the components: $\forall i \in \{0,1,2\}, \quad p_{(t+1,i)} = p_{(t,i)} + \theta_t (2\xi_t - \alpha_t - \xi_{(t,i)})$.

- Update the coefficients estimate: $\alpha_{t+1} = \alpha_t + \theta_t(\xi_t - \alpha_t)$

End main iteration

Output: Deconvolved image $x^\star = \boldsymbol{\Phi}\alpha_{N_{\mathrm{ext}}}$.

5.3. Computational complexity and implementation details

The bulk of computation of our deconvolution algorithm is invested in applying $\boldsymbol{\Phi}$ (resp. $\mathbf{H}$) and its adjoint $\boldsymbol{\Phi}^{\mathrm{T}}$ (resp. $\mathbf{H}^{\mathrm{T}}$). These operators are never constructed explicitly, rather they are implemented as fast implicit operators taking a vector x, and returning $\boldsymbol{\Phi}x$ (resp. $\boldsymbol{\Phi}^{\mathrm{T}}x$) and $\mathbf{H}x$ (resp. $\mathbf{H}^{\mathrm{T}}x$). Multiplication by $\mathbf{H}$ or $\mathbf{H}^{\mathrm{T}}$ costs two FFTs, that is $2n \log n$ operations (n denotes the number of pixels). The complexity of $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^{\mathrm{T}}$ depends on the transforms in the dictionary: for example, the orthogonal wavelet transform costs $\mathcal{O}(n)$ operations, the translation-invariant discrete wavelet transform (TI-DWT) costs $\mathcal{O}(n \log n)$, the curvelet transform costs $\mathcal{O}(n \log n)$, etc. Let $V_{\boldsymbol{\Phi}}$ denote the complexity of applying the analysis or synthesis operator. Define N_{ext} and N_{int} as the maximal number of outer and inner iterations respectively in Algorithm 1 and Algorithm 2, and recall that L is the number of coefficients. The computational complexity of the algorithms for solving $(\mathrm{P}_{\gamma,\psi})$ and $(\mathrm{P}'_{\gamma,\psi})$ are summarized as follows:

11

Problem	Computational complexity
	$\boldsymbol{\Phi}$ orthobasis or tight frame
$(\mathrm{P}_{\gamma,\psi})$	$N_{\mathrm{ext}}\left(N_{\mathrm{int}}\left(4\mathcal{O}(n\log n)+2V_{\boldsymbol{\Phi}}\right)+\mathcal{O}(n)\right)$
$(\mathrm{P}'_{\gamma,\psi})$	$N_{\mathrm{ext}}\left(4V_{\boldsymbol{\Phi}}+N_{\mathrm{int}}\left(4n\log n\right)+\mathcal{O}(L)\right)$
	$\boldsymbol{\Phi}$ general frame or linear bounded operator
$(\mathrm{P}_{\gamma,\psi})$	$N_{\mathrm{ext}}\left(N_{\mathrm{int}}\left(4\mathcal{O}(n\log n)+2V_{\boldsymbol{\Phi}}\right)+\mathcal{O}(n)\right)$
$(\mathrm{P}'_{\gamma,\psi})$	$N_{\mathrm{ext}}\left(N_{\mathrm{int}}\left(4V_{\boldsymbol{\Phi}}+4n\log n+\mathcal{O}(L)\right)\right)$

6. Results

6.1. Experimental protocol

We applied the proposed algorithms on a simulated image of the Hubble Space Telescope Wide Field Camera of a distant cluster of galaxies [38] and an image of the planet Saturn. In the sequel, our algorithms are respectively dubbed Prox-FA (for the one with the analysis prior) and Prox-FS (for that associated to the synthesis prior). For both of them, the number of interior iterations N_{int} was set to 10 which was enough to ensure convergence (i.e. the error terms in the computation of the proximity operators were under control). The obtained results were compared with others state-of-the art algorithms: the naive approach treating the noise as if it were additive white Gaussian (NaiveG [42]), the Anscombe variance stabilizing approach (StabG [13]), and the algorithms proposed in [21] with the synthesis (PIDAL-FS) and the analysis (PIDAL-FA) priors. For all compared algorithms, the sparsity-promoting penalty was the ℓ_1-norm whose proximity operator is well-known soft-thresholding operator. For fair comparison, all algorithms were stopped when the relative change between two consecutive iterates was below a given tolerance $\delta > 0$, i.e.

$$\|x_{t+1}-x_t\|_2 \, / \, \|x_t\|_2 \le \delta \ . \tag{13}$$

The quantitative score of restoration quality was the the mean absolute error (MAE). The MAE was chosen as it is well suited to Poisson noise [13].

As usual in regularized inverse problems, the choice of the regularization parameter γ is crucial. When the reference image is available, its choice can be tweaked manually so as to maximize the restoration quality by minimizing some quality score; e.g. MAE. But of course, this is a tedious task and cumbersome for real data. To circumvent these difficulties, in [13], an objective model selection driven choice based on the GCV was proposed. An approximate closed-form of the GCV was derived (see for [13] details),

$$\mathrm{GCV}(\gamma) = \frac{\left\|2\sqrt{y+\frac{3}{8}}-2\sqrt{\mathbf{H}x^{\star}+\frac{3}{8}}\right\|^2}{(n-df)^2} \ , \tag{14}$$

$$\text{with } df \approx \mathrm{card}\left\{i=1,\ldots,L \ \big| \ |\alpha_i^{\star}| \ge \gamma\right\} \ ,$$

where $x^{\star}$ is the restored image provided by the estimator. It was shown in [13] that the GCV-based choice of γ, although turned out to be close to the value that actually minimizes the MAE and the MSE, was slightly upper-biased in practice. The GCV then was used as a first good guess that might need refinement.

	RL-MRS [38]	RL-TV [12]	FTITPR [44]	NaiveG [42]	StabG [13]
MAE	63.5	52.8	60.9	39.8	43

	PIDAL-FA [21]	PIDAL-FS [21]	Prox-FA	Prox-FS	
MAE	40	43.6	37.8	35.6	

Table 1: MAE for the deconvolution of the sky image.

6.2. Simulated sky image

Figure 1(a) displays a simulated image of the sky. The maximal intensity (mean photon count) in this image is ~18000. The blurred version by h (using the Hubble's PSF before being repaired [38]) and the noisy ones are depicted in Figure 1(b) and (c). For this image, the dictionary $\mathbf{\Phi}$ contained a tight frame of translation-invariant wavelets (TI-DWT). For this image, we also include in the comparative study the deconvolution obtained with Richardson-Lucy regularized with the wavelet multiresolution support (RL-MRS) as proposed in [36]. The latter was specifically designed for this kind of objects. We also compare to the total-variation regularized Richardson-Lucy algorithm (RL-TV) proposed in [12], and the fast translation invariant tree-pruning reconstruction combined with an EM algorithm (FTITPR [44]). For each compared method, the regularization parameter was tuned manually to achieve its best performance level (i.e. minimize MAE), and the convergence tolerance δ was set to 10^{-5}.

For this image, the best algorithm is clearly RL-MRS where most of the small faint structures as well as bright objects are well deconvolved. However, small artifacts from the wavelet transform are also present. The StabG algorithm does a good job and preserves most of the weak objects which are barely visible in the degraded noisy image. At this intensity regime, the NaiveG algorithm yielded satisfactory results, which are comparable to ours. FTITPR correctly restores most of the important structures with a clean background, but many faint objects are lost. RL-TV leads to a deconvolution similar to ours for bright objects, but the background is dominated by spurious artifacts. The PIDAL approaches give quite similar results, PIDAL-FA leads here to a smoother estimate than PIDAL-FS. In the other hand PIDAL-FS seems to preserve more sources, as it can be seen for example on the bottom left, where one can notice objects that do not appear in the results of PIDAL-FA. In the same way, our algorithms yield results comparable to their respective parts of PIDAL as expected since they solve the same optimization problems. Prox-FS shows one of the best deconvolution results with a clean background and good restoration of most objects, although a few small details are oversmooth (see on bottom left).

These qualitative results are confirmed by a quantitative score of the restoration quality in terms of the MAE, see Table 1. Notice that the hight MAE value of RL-MRS, which might seem surprising given its good visual result, might be explained by the artifacts that locally destroy the photometry.

6.3. Intensity level influence

To assess the impact of the intensity level (average photon counts), the different algorithms in the comparative study were applied to an image of the planet Saturn at several intensity levels. In this experiment, the PSF was a 7×7 moving average. The regularization parameter γ was chosen using the GCV criterion (14). Given that this

image is piecewise smooth away from smooth curvilinear structures (e.g. the rings), the curvelet dictionary will be a very good candidate to sparsify it [4].

Intensity Level	NaiveG	StabG	PIDAL-FA	PIDAL-FS	Prox-FA	Prox-FS
5	0.70 (33.49)	1.14 (54.54)	0.34 (16.04)	0.24 (11.78)	0.42 (20.09)	0.25 (11.96)
30	1.27 (10.13)	6.62 (52.78)	4.22 (33.62)	1.41 (11.23)	2.35 (18.74)	2.09 (16.66)
100	3.24 (7.75)	26.73 (63.94)	23.40 (55.97)	12.46 (29.81)	7.88 (18.85)	7.31 (17.49)
255	7.48 (7.02)	80.45 (75.46)	73.83 (69.26)	49.36 (46.30)	21.02 (19.72)	18.78 (17.62)

Table 2: Average MAE values and relative MAE (in parentheses and in percent) for the Saturn image as a function of the intensity level with the naive Gaussian (NaiveG) [43], stabilized gaussian (StabG) [13], PIDAL-FA and PIDAL-FS [21], and our two algorithms Prox-FA and Prox-FS.

At each intensity value, ten noisy and blurred replications were generated and the MAE was computed for each deconvolution algorithm. The average and relative (i.e. divided by the mean intensity) MAE values over the ten realizations are summarized in Table 2. As expected, for low intensity levels, the best results are obtained for the methods using data fidelity terms that account properly for Poisson noise. For higher levels, the performance of NaiveG gets better as the Poisson distribution tends to normal with increasing intensity. The difference in performance, especially for the high intensity regime, between our approaches and the PIDAL ones seem to originate from the optimization algorithms used (and their convergence then).

For this image, it seems that the synthesis prior is slightly better than the analysis one. This might seem contradictory with results reported in the Gaussian noise case by a few authors where the analysis prior was observed to be superior to its synthesis counterpart [5, 17, 34]. But one should avoid to infer strong conclusions from such limited body of experimental work. For instance, the noise is different in our setting, and the deep phenomena underlying differences and similarities between analysis and synthesis priors are still not very well understood and more investigation is needed in this direction both on the theoretical and practical sides.

The visual results obtained for a maximal intensity of 30 are portrayed in Figure 2. For all the methods, the curvelet transform is able to capture most of the curved structures inside the image and then leads to a good restoration of them. The original image contains some tiny satellites at the bottom right that appear as localized spikes. Prox-FA was able to recover one of them, though it is hardly visible in the restored image. Prox-FS however recovers several structures that look like background artifacts including the satellites. A wise solution would be to use another transform in the dictionary beside curvelets, which is well-adapted to point-like structures (e.g. wavelets or even Diracs).

7. Conclusion

In this paper, a novel sparsity-based fast iterative thresholding deconvolution algorithm that takes account of the presence of Poisson noise was presented. The Poisson noise was handled properly using its associated likelihood function to construct the data fidelity term. A careful theoretical study of the optimization problems and convergence guarantees of the iterative algorithms were provided. Several experimental experiments have shown the capabilities of our approach, which compares favorably with some state-of-the-art algorithms.

The present work may be extended along several lines. For example, our approach uses a proximal algorithm which is a generalization of the Douglas-Rachford splitting applied to the primal problem, but to the best of our knowledge, its convergence rate (even on the objective) is not known. Furthermore, in this paper, the regularization parameter was chosen based on the GCV formula proposed, see (14). But this formula was derived using some heuristics and was valid for the stabilized version of the deconvolution problem. This objective and automatic selection of the regularization parameter remains an open issue that is worth investigating in the future.

Appendix A. Proof of Theorem 5

As $\mathbf{A}$ is an affine operator, it preserves convexity. Thus, $f \circ \mathbf{A}$ is convex. Moreover, $\mathbf{A}$ is continuous and f is lsc, and so is $f \circ \mathbf{A}$. By the domain qualification condition, we then have $f \circ \mathbf{A} \in \Gamma_0(\mathbb{R}^m)$. Note that the domain assumption in the theorem is not needed for frames by surjectivity of $\mathbf{F}$.

By Fenchel-Rockafellar duality [16, Section III.4], we have

$$\bar{p} = \mathrm{prox}_{\mathrm{prox}_{f \circ \mathbf{A}}} \quad \Longleftrightarrow \quad \bar{p} = \underset{p \in \mathbb{R}^n}{\mathrm{arginf}} \frac{1}{2} \|p - x\|^2 + f \circ \mathbf{A}(p) \qquad (\mathrm{A.1})$$

$$\Longleftrightarrow \quad \bar{u} = \underset{u \in \mathbb{R}^m}{\mathrm{argmin}} \frac{1}{2} \left\| x - \mathbf{F}^{\mathrm{T}} u \right\|^2 + \langle u, y \rangle + f^*(u) \ , \qquad (\mathrm{A.2})$$

and the primal solution $\bar{p}$ is recovered from the dual solution $\bar{u}$ as

$$\bar{p} = x - \mathbf{F}^{\mathrm{T}} \bar{u} \ .$$

(i) $\mathbf{F}$ is a tight frame with constant $c > 0$. Applying the minimality condition to (A.2), using the fact that $\mathbf{F}\mathbf{F}^{\mathrm{T}} = c\mathbf{I}$ and Moreau Identity, we obtain

$$\begin{aligned}
(\mathbf{F}x - y) - c\bar{u} &\in & \partial f^*(\bar{u}) \\
c^{-1}(\mathbf{F}x - y) - \bar{u} &\in & \partial(c^{-1}f^*)(\bar{u}) \\
\bar{u} &=& \mathrm{prox}_{c^{-1}f^*} \left(c^{-1}(\mathbf{F}x - y) \right) \\
\bar{u} &=& c^{-1} \left(\mathbf{I} - \mathrm{prox}_{cf} \right) (\mathbf{F}x - y) \ , \qquad (\mathrm{A.3})
\end{aligned}$$

which leads to (8).

(ii) $\mathbf{F}$ is a general frame operator. From (A.2), $\frac{1}{2} \left\| x - \mathbf{F}^{\mathrm{T}} . \right\|^2 + \langle ., y \rangle$ is continuous with c_2-Lipschitz continuous gradient. Therefore, applying the forward-backward splitting to the minimization problem (A.2), we obtain (10). From strong convexity of the dual problem (A.2), the dual minimizer is unique.

To get the linear convergence rate, the proof is very classical and is an extension of the projected gradient descent one that can be found in classical monographs such as [6, Theorem 8.6.2].

(iii) In all other cases, the forward-backward splitting applied to (A.2) converges provided that $0 < \inf_t \tau_t \leq \sup_t \tau_t < 2/\|\mathbf{F}\|^2 = 2/c_2$. The proof of the convergence rate is technical and follows the same lines as [18, Theorem 1].

References

[1] Andrews, H. C., Hunt, B. R., 1977. Digital Image Restoration. Prentice-Hall.

[2] Anscombe, F. J., 1948. The Transformation of Poisson, Binomial and Negative-Binomial Data. Biometrika 35, 246–254.

[3] Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G., 2009. Image deblurring with Poisson data: from cells to galaxies. Inverse Problems 25 (123006).

[4] Candès, E. J., Donoho, D. L., 1999. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In: Cohen, A., Rabut, C., Schumaker, L. (Eds.), Curve and Surface Fitting: Saint-Malo 1999. Vanderbilt University Press, Nashville, TN.

[5] Candès, E. J., Wakin, M. B., Boyd, S. P., 2008. Enhancing sparsity by reweighted L1 minimization. Journal of Fourier Analysis and Applications 14 (5), 877–905.

[6] Ciarlet, P., 1982. Introduction à l'Analyse Numérique Matricielle et à l'Optimisation. Paris.

[7] Combettes, P. L., Pesquet, J.-., 2007. A Douglas-Rachford splittting approach to nonsmooth convex variational signal recovery. IEEE Journal of Selected Topics in Signal Processing 1 (4), 564–574.

[8] Combettes, P. L., Pesquet, J.-C., November 2007. Proximal thresholding algorithm for minimization over orthonormal bases. SIAM Journal on Optimization 18 (4), 1351–1376.

[9] Combettes, P. L., Pesquet, J.-C., 2008. A proximal decomposition method for solving convex variational inverse problems. Inverse Problems 24 (6).

[10] Combettes, P. L., Wajs, V. R., 2005. Signal recovery by proximal forward-backward splitting. SIAM Multiscale Model. Simul. 4 (4), 1168–1200.

[11] Daubechies, I., Defrise, M., Mol, C. D., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraints. Comm. Pure Appl. Math. 112, 1413–1541.

[12] Dey, N., Blanc-Féraud, L., Zerubia, J., Zimmer, C., Olivo-Marin, J.-C., Kam, Z., 2004. A deconvolution method for confocal microscopy with total variation regularization. In: Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, USA, 15-18 April 2004. IEEE, pp. 1223–1226.

[13] Dupé, F.-X., Fadili, M.-J., Starck, J.-L., 2009. A proximal iteration for deconvolving poisson noisy images using sparse representations. IEEE Transactions on Image Processing 18 (2), 310–321.

[14] E. J. Candès, Y. Eldar, D. N., Randall, P., 2011. Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic AnalysisTo appear.

[15] Eckstein, J., Bertsekas, D. P., 1992. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Programming 55, 293–318.

[16] Ekeland, I., Temam, R., 1974. Analyse convexe et problèmes variationnels. Dunod.

[17] Elad, M., Milanfar, P., Rubinstein, R., 2007. Analysis versus synthesis in signal priors. Inverse Problems 23 (3), 947–968.

[18] Fadili, M., Peyré, G., 2011. Total variation projection with first order schemes. IEEE Transactions on Image Processing 20 (3), 657–669.

[19] Fadili, M., Starck, J.-L., Murtagh, F., 2007. Inpainting and zooming using sparse representations. The Computer Journal 52 (1), 64–79.

[20] Fadili, M. J., Starck, J.-L., 2006. Sparse representation-based image deconvolution by iterative thresholding. In: ADA IV. Elsevier, France.

[21] Figueiredo, M., Bioucas-Dias, J., 2010. Restoration of poissonian images using alternating direction optimization. IEEE Transactions on Image Processing, Submitted.
URL \url{http://arxiv.org/abs/1001.2244}

[22] Figueiredo, M., Nowak, R., 2003. An em algorithm for wavelet-based image restoration. IEEE Transactions on Image Processing 12, 906–916.

[23] Gabay, D., 1983. Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (Eds.), Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems. North-Holland, Amsterdam.

[24] Jammal, G., Bijaoui, A., 2004. Dequant: a flexible multiresolution restoration framework. Signal Processing 84, 1049–1069.

[25] Jansson, P., 1997. Deconvolution of Images and Spectra. Academic Press, New York.

[26] Lemaréchal, C., Hiriart-Urruty, J.-B., 1996. Convex Analysis and Minimization Algorithms I, 2nd Edition. Springer.

[27] Mallat, S. G., 1998. A Wavelet Tour of Signal Processing, 2nd Edition. Academic Press.

[28] Moreau, J.-J., 1962. Fonctions convexes duales et points proximaux dans un espace hilbertien. CRAS Sér. A Math. 255, 2897–2899.

[29] Moreau, J.-J., 1963. Propriétés des applications "prox". Comptes Rendus de l'Académie des Sciences Série A Mathématiques 256, 1069–1071.

[30] Moreau, J.-J., 1965. Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France 93, 273–299.

[31] Pawley, J., 2005. Handbook of Confocal Microscopy. Plenum Press.

[32] Pustelnik, N., Chaux, C., Pesquet, J.-C., 2011. Parallel proximal algorithm for image restoration using hybrid regularization. IEEE Trans. on image processingTo appear.

[33] Sarder, P., Nehorai, A., 2006. Deconvolution Method for 3-D Fluorescence Microscopy Images. IEEE Sig. Pro. Mag. 23, 32–45.

[34] Selesnick, I., Figueiredo, M., 2009. Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In: SPIE conference on Signal and Image Processing: Wavelet Applications in Signal and Image Processing XIII.

[35] Spingarn, J. E., 1983. Partial inverse of a monotone operator. Appl. Math. Optim. 10 (3), 247–265.

[36] Starck, J.-L., Bijaoui, A., Murtagh, F., 1995. Multiresolution support applied to image filtering and deconvolution. CVGIP: Graphical Models and Image Processing 57, 420–431.

[37] Starck, J.-L., Murtagh, F., 1994. Image restoration with noise suppression using the wavelet transform. Astronomy and Astrophysics 288, 343–348.

[38] Starck, J.-L., Murtagh, F., 2006. Astronomical Image and Data Analysis. Springer.

[39] Starck, J.-L., Murtagh, F., Fadili, M., 2010. Sparse Signal and Image Processing: Wavelets, Curvelets and Morphological Diversity. Cambridge University Press, Cambridge, UK, in press.

[40] Starck, J.-L., Nguyen, M., Murtagh, F., 2003. Wavelets and curvelets for image deconvolution: a combined approach. Signal Processing 83, 2279–2283.

[41] Stark, H., 1987. Image Recovery: Theory and Application. Elsevier Science & Technology Books.

[42] Vonesch, C., Unser, M., 2007. Fast wavelet-regularized image deconvolution. In: Proceedings of the 2007 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington, DC, USA, April 12-16, 2007. pp. 608–611.

[43] Vonesch, C., Unser, M., 2008. A fast thresholded Landweber algorithm for general wavelet bases: Application to 3D deconvolution microscopy. In: Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, May 14-17, 2008. pp. 1351–1354.

[44] Willett, R., Nowak, R., 2004. Fast multiresolution photon-limited image reconstruction. In: IEEE ISBI.
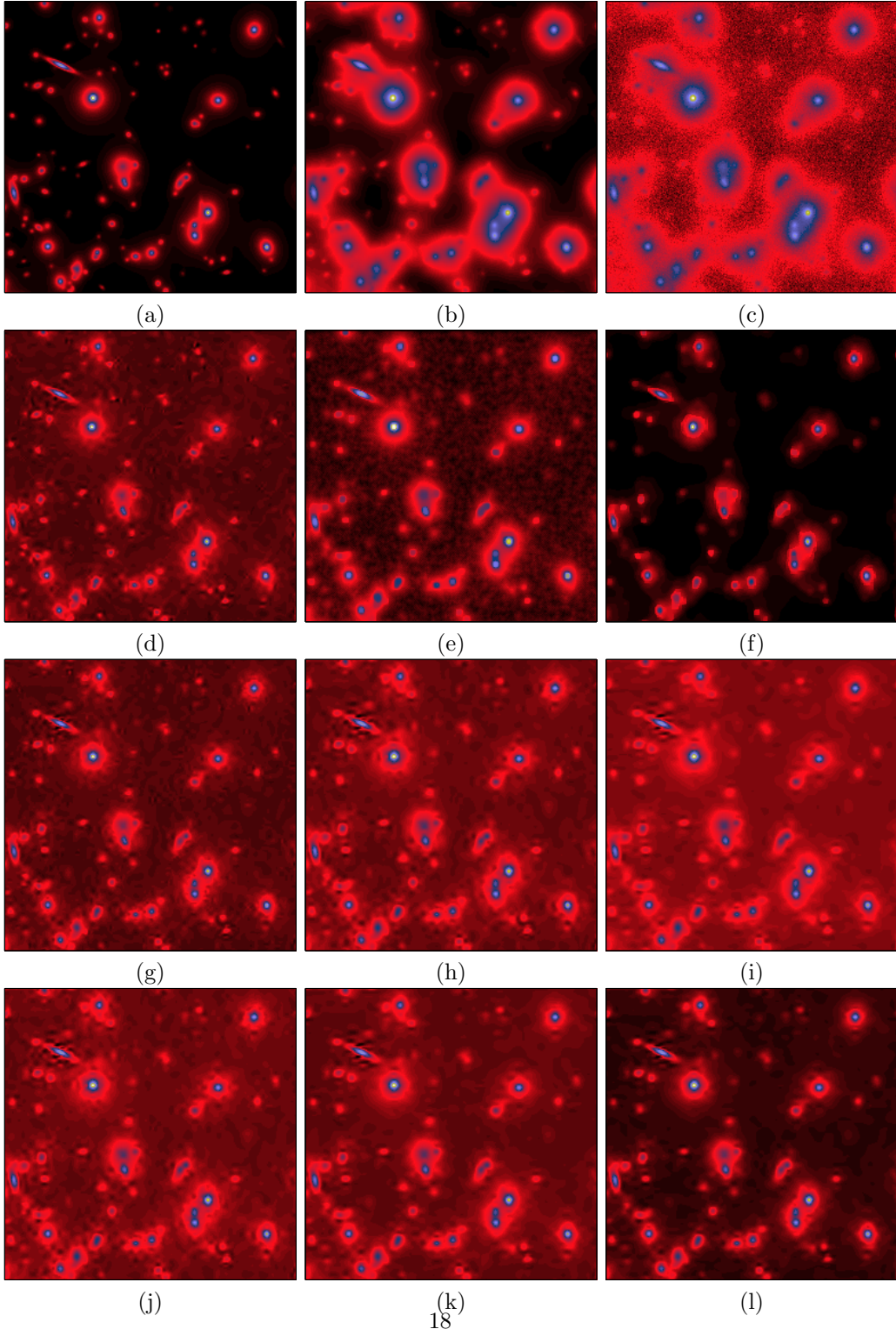
Figure 1: Deconvolution of the sky simulation. (a) Original, (b) Blurred, (c) Blurred and noisy, (d) RL-MRS [38], (e) RL-TV [12], (f) FTITPR [44], (g) NaiveG [42], (h) StabG [13], (i) PIDAL-FA [21], (j) PIDAL-FS [21], (k) Prox-FA, (l) Prox-FS.
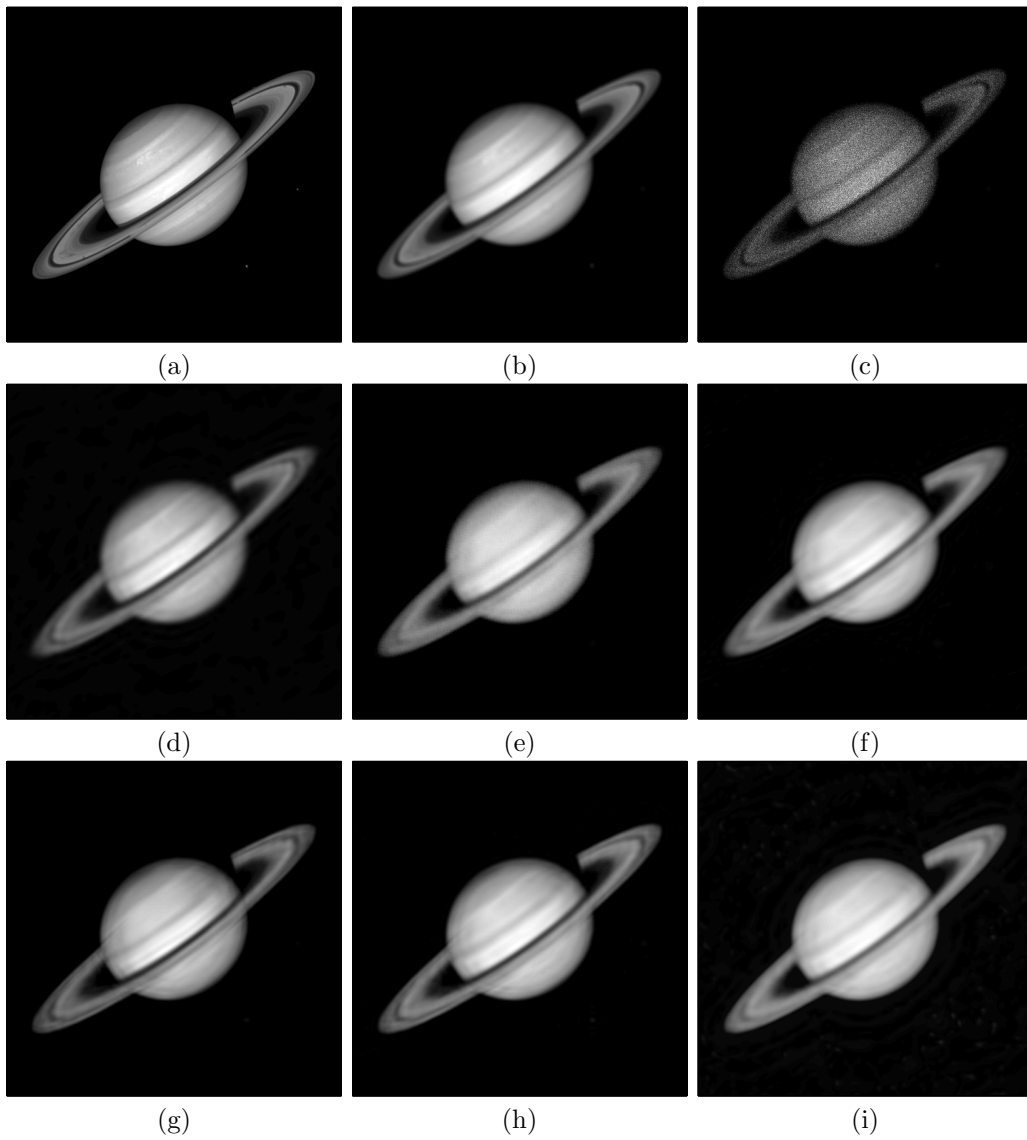
Figure 2: Deconvolution of Saturn image. (a) original, (b) Convolved, (c) Convolved and Noisy, (d) NaiveG [43], (e) StabG [13], (f) PIDAL-FA [21], (g) PIDAL-FS [21], (h) Prox-FA, (i) Prox-FS.