Linear inverse problems with various noise models and mixed regularizations

François-Xavier Dupé
AIM UMR CNRS - CEA
91191 Gif-sur-Yvette, France
francois-
xavier.dupe@cea.fr

Jalal M. Fadili
GREYC-ENSICAEN-Univ.
Caen
14050 Caen, France
Jalal.Fadili@greyc.ensicaen.fr

Jean-Luc Starck
AIM UMR CNRS - CEA
91191 Gif-sur-Yvette, France
jean-luc.starck@cea.fr

ABSTRACT

In this paper, we propose two algorithms to solve a large class of linear inverse problems when the observations are corrupted by various types of noises. A proper data fidelity term (log-likelihood) is introduced to reflect the statistics of the noise (e.g. Gaussian, Poisson, multiplicative, etc.) independently of the degradation operator. On the other hand, the regularization is constructed through different terms reflecting a priori knowledge on the images. Piecing together the data fidelity and the prior terms, the solution to the inverse problem is cast as the minimization of a composite non-smooth convex objective functional. We establish the well-posedness of the optimization problem, characterize the corresponding minimizers for different kind of noises. Then we solve it by means of primal and primal-dual proximal splitting algorithms originating from the field of non-smooth convex optimization theory. Experimental results on deconvolution, inpainting and denoising with some comparison to prior methods are also reported.

Keywords

Inverse Problems, Poisson noise, Gaussian noise, Multiplicative noise, Duality, Proximity operator, Sparsity.

1. INTRODUCTION

Previous work A lot of works have already been dedicated to linear inverse problems with Gaussian noise (see [16] for a comprehensive review), while linear inverse problems in presence of other kind of noise such as Poisson noise have attracted less interest, presumably because noises properties are more complicated to handle. Such inverse problems have however important applications in imaging such as restoration (e.g. deconvolution in medical and astronomical imaging), or reconstruction (e.g. computerized tomography).

Since the work for Gaussian noise by [9], many other methods have appeared for managing linear inverse problem with sparsity regularization. But they limited to the Gaussian case. In the context of Poisson linear inverse problems using sparsity-promoting regularization, a few recent algorithms have been proposed. For example, [11] stabilize the noise and proposed a family of nested schemes relying upon proximal splitting algorithms (Forward-Backward and Douglas-Rachford) to solve the corresponding optimization problem. The work of [4] is in the same vein. These methods may be extended to other kind of noise. However, nested algorithms are time-consuming since they necessitate to sub-iterate. Using the augmented Lagrangian method with the alternating method of multipliers algorithm (ADMM), which is nothing but the Douglas-Rachford splitting applied to the Fenchel-Rockafellar dual problem, [13] presented a deconvolution algorithm with TV and sparsity regularization, and [1] a denoising algorithm for multiplicative noise. This scheme however necessitates to solve a least-square problem which can be done explicitly only in some cases.

Contributions In this paper, we propose a framework for solving linear inverse problems when the observations are corrupted by various types of noise. In order to form the data fidelity term, we take the exact likelihood associated to the noise model. As a prior, the images are assumed to comply with several regularity properties or/and constraints reflecting knowledge about the original image. The solution to the inverse problem is cast as the minimization of a composite non-smooth convex functional, for which we prove well-posedness of the optimization problem, characterize the corresponding minimizers, and solve them by means of primal and primal-dual proximal splitting algorithms originating from the realm of non-smooth convex optimization theory. Convergence of the algorithms is also shown. Experimental results and comparison to other algorithms on deconvolution are finally conducted.

Notation and terminology Let $\mathcal{H}$ a real Hilbert space, here a finite dimensional vector subspace of $\mathbb{R}^n$. We denote by $\|.\|$ the norm associated with the inner product in $\mathcal{H}$, and $\mathbf{I}$ is the identity operator on $\mathcal{H}$. $\|.\|_p, p \geq 1$ is the ℓ_p norm. x and α are respectively reordered vectors of image samples and transform coefficients. We denote by $\mathrm{ri}\,\mathcal{C}$ the relative interior of a convex set $\mathcal{C}$. A real-valued function f is coercive, if $\lim_{\|x\|\to+\infty} f(x) = +\infty$, and is proper if its domain is non-empty $\mathrm{dom}\,f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \emptyset$. $\Gamma_0(\mathcal{H})$ is the class of all proper lower semicontinuous (lsc) convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$. We denote by $\|\mathbf{M}\| = \max_{x\neq 0} \frac{\|\mathbf{M}x\|}{\|x\|}$ the spectral norm of the linear operator $\mathbf{M}$, and $\ker(\mathbf{M}) := \{x \in \mathcal{H} : \mathbf{M}x = 0, x \neq 0\}$ its

kernel.

Let $x \in \mathcal{H}$ be a n-pixels image. x can be written as the superposition of elementary atoms φ_γ parameterized by $\gamma \in \mathcal{I}$ such that $x = \sum_{\gamma \in \mathcal{I}} \alpha_\gamma \varphi_\gamma = \boldsymbol{\Phi}\alpha$, $|\mathcal{I}| = L$, $L \geqslant n$. We denote by $\boldsymbol{\Phi} : \mathcal{H}' \to \mathcal{H}$ the dictionary (typically a frame of $\mathcal{H}$), whose columns are the atoms all normalized to a unit ℓ_2-norm

2. PROBLEM STATEMENT

Consider the image formation model where an input image of n pixels x is indirectly observed through the action of a bounded linear operator $\mathbf{H} : \mathcal{H} \to \mathcal{K}$, with $\mathcal{K}$ a real Hilbert space (usually a subspace of $\mathbb{R}^m$, $m > 0$), and contaminated by a noise ε through a composition operator $\odot$ (e.g. addition),

$$y \sim \mathbf{H}x \odot \varepsilon . \tag{1}$$

The linear inverse problem at hand is to reconstruct x from the observed image y.

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 statistics of the noise. As a prior, the image is supposed to be economically (sparsely) represented in a pre-chosen dictionary $\boldsymbol{\Phi}$ as measured by a sparsity-promoting penalty Ψ supposed throughout to be convex but non-smooth, e.g. the ℓ_1 norm.

2.1 Gaussian noise case

For Gaussian noise, we consider the following formation model,

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

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

>From the probability density function, the negative log-likelihood writes:

$$f_{\text{Gaussian}} : \eta \in \mathcal{H} \mapsto \|\eta - y\|_2^2 / (2\sigma^2) . \tag{3}$$

>From this function, we can directly derive the following result,

Proposition 1. f_{Gaussian} is a proper, strictly convex and lsc function.

2.2 Poisson noise case

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 $(\mathbf{H}x)_i$. Formally, this writes in a vector form as

$$y \sim \mathcal{P}(\mathbf{H}x) . \tag{4}$$

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

$$p(y|x) = \prod_i \frac{((\mathbf{H}x)[i])^{y[i]} \exp\left(-(\mathbf{H}x)[i]\right)}{y[i]!} . \tag{5}$$

Taking the negative log-likelihood, we arrive at the following data fidelity term:

$$f_{\text{Poisson}} : \eta \in \mathcal{H} \mapsto \sum_{i=1}^{n} f_{\text{P}}(\eta[i]), \tag{6}$$

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

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

Using classical results from convex theory, we can show that,

Proposition 2. f_{Poisson} is a proper, convex and lsc function. f_{Poisson} is strictly convex if and only if $\forall i \in \{1, \ldots, n\}$, $y[i] \neq 0$.

2.3 Multiplicative noise

We consider here the same context of multiplicative as in [1]. With multiplicative noise, a usual approach to improve the signal to noise ratio consists in averaging independent observation of the same resolution. When considering SAR/SAS system, this method is called *multilook*, i.e. M-look in the case of the averaging of M images. For fully developed speckle, the averaged images are Gamma distributed,

$$y = x\varepsilon, \quad \varepsilon \sim \Gamma(M, 1/M) . \tag{7}$$

In order to simplify the problem, the logarithm of the observation is considered,

$$\log(y) = \log(x) + \log(\varepsilon) = z + \omega . \tag{8}$$

And in [1], the authors proof that the anti log-likelihood yields,

$$f_{\text{Multi}} : \eta \in \mathcal{H} \mapsto M \sum_{i=1}^{n} (z[i] + \exp(\log(y[i]) - z[i])) . \tag{9}$$

Using classical results from convex theory, we can directly derive,

Proposition 3. f_{Multi} is a proper, strictly convex and lsc function.

2.4 Optimization problem

Our aim is then to solve the following optimization problems, under a synthesis-type sparsity prior[1] where $\mathcal{H}'$ is the Hilbert space defined by the dictionary,

$$\operatorname*{argmin}_{\alpha \in \mathcal{H}'} J(\alpha),$$
$$J : \alpha \mapsto f_1 \circ \mathbf{H} \circ \boldsymbol{\Phi}(\alpha) + \sum_{i=1}^{K} R_i(\alpha) . \tag{$\text{P}_{f_1, \gamma, \psi}$}$$

The data fidelity term f_1 reflect the noise statistics and the $R_i, 1 \leqslant i \leqslant K$ the K prior terms. In this paper, we restrict to the case $K = 2$, with $K_1 = \Psi$, the penalty function $\Psi : \alpha \mapsto \sum_{i=0}^{L} \psi_i(\alpha[i])$ which is positive, additive, and chosen

---

[1]Our framework and algorithms extend to an analysis-type prior just as well.

to enforce sparsity, $\gamma > 0$ is a regularization parameter and $K_2 = \iota_{\mathcal{C}} \circ \boldsymbol{\Phi}$ the indicator function of the convex set $\mathcal{C}$ (e.g. the positive orthant for Poissonian data).

For the rest of the paper, we assume that f_1 is a proper, convex and lsc function, i.e. $f_1 \in \Gamma_0(\mathcal{H})$. This is true for many kind of noises including Poisson, Gaussian, Laplacian...(see [3] for others examples).

>From the objective in $(\mathrm{P}_{f_1, \gamma, \psi})$, we get the following,

Proposition 4.
(i) f_1 is a convex functions and so are $f_1 \circ \mathbf{H}$ and $f_1 \circ \mathbf{H} \circ \boldsymbol{\Phi}$.
(ii) Suppose that f_1 is strictly convex, then $f_1 \circ \mathbf{H} \circ \boldsymbol{\Phi}$ remains strictly convex if $\boldsymbol{\Phi}$ is an orthobasis and $\ker(\mathbf{H}) = \emptyset$.
(iii) Suppose that $(0, +\infty) \cap \mathbf{H}([0, +\infty)) \neq \emptyset$. Then $J \in \Gamma_0(\mathcal{H})$.

2.5 Well-posedness of $(\mathrm{P}_{f_1, \gamma, \psi})$
Let $\mathcal{M}$ be the set of minimizers of problem $(\mathrm{P}_{f_1, \gamma, \psi})$. Suppose that Ψ is coercive. Thus J is coercive. Therefore, the following holds:

Proposition 5.
(i) Existence: $(\mathrm{P}_{f_1, \gamma, \psi})$ has at least one solution, i.e. $\mathcal{M} \neq \emptyset$.
(ii) Uniqueness: $(\mathrm{P}_{f_1, \gamma, \psi})$ has a unique solution if Ψ is strictly convex, or under (ii) of Proposition 4.

3. ITERATIVE MINIMIZATION ALGORITHMS
3.1 Proximal calculus
We are now ready to describe the proximal splitting algorithms to solve $(\mathrm{P}_{f_1, \gamma, \psi})$. At the heart of the splitting framework is the notion of proximity operator.

Definition 6 ([14]). Let $F \in \Gamma_0(\mathcal{H})$. Then, for every $x \in \mathcal{H}$, the function $y \mapsto F(y) + \|x - y\|_2^2 / 2$ achieves its infimum at a unique point denoted by $\mathrm{prox}_F x$. The operator $\mathrm{prox}_F : \mathcal{H} \rightarrow \mathcal{H}$ thus defined is the proximity operator of F.

Then, the proximity operator of the indicator function of a convex set is merely its orthogonal projector. One important property of this operator is the separability property:

Lemma 7 ([7]).
Let $F_k \in \Gamma_0(\mathcal{H})$, $k \in \{1, \cdots, K\}$ and let $G : (x_k)_{1 \leqslant k \leqslant K} \mapsto \sum_k F_k(x_k)$. Then $\mathrm{prox}_G = (\mathrm{prox}_{F_k})_{1 \leqslant k \leqslant K}$.

For Gaussian noise, we can easily prove that with f_1 as defined in (3),

Lemma 8. Let y be the observation, the proximity operator associated to f_{Gaussian} (i.e. the Gaussian anti log-likelihood) is,

$$\mathrm{prox}_{\beta f_{\mathrm{Gaussian}}} x = \frac{\beta y + \sigma^2 x}{\beta + \sigma^2} \ . \tag{10}$$

The following result can be proved easily by solving the proximal optimization problem in Definition 6 with f_1 as defined in (6), see also [5].

Lemma 9. Let y be the count map (i.e. the observations), the proximity operator associated to f_{Poisson} (i.e. the Poisson anti log-likelihood) is,

$$\mathrm{prox}_{\beta f_{\mathrm{Poisson}}} x = \left(\frac{x[i] - \beta + \sqrt{(x[i] - \beta)^2 + 4\beta y[i]}}{2} \right)_{1 \leqslant i \leqslant n} . \tag{11}$$

As with multiplicative noise f_{Multi} involves the exponential, we need the W-Lambert function [8] in order to derive a closed form of the proximity operator,

Lemma 10. Let y be the observations, the proximity operator associated to f_{Multi} is,

$$\mathrm{prox}_{\beta f_{\mathrm{Multi}}} x = \big(x[i] - \beta M - \mathrm{W}\left(-\beta M \exp(x[i] - \log(y[i]) - \beta M)\right) \big)_{1 \leqslant i \leqslant n} , \tag{12}$$

where W is the W-Lambert function, i.e. the function such that $\mathrm{W}(a) \exp(\mathrm{W}(a)) = a$, $a \in \mathbb{R}$.

We now turn to $\mathrm{prox}_{\gamma \Psi}$ which is given by Lemma 7 and the following result:

Theorem 11 ([12]). Suppose that $\forall i$: (i) ψ_i is convex even-symmetric, non-negative and non-decreasing on $\mathbb{R}^+$, and $\psi_i(0) = 0$; (ii) ψ_i is twice differentiable on $\mathbb{R} \setminus \{0\}$; (iii) ψ_i is continuous on $\mathbb{R}$, and admits a positive right derivative at zero $\psi'_{i+}(0) = \lim_{h \rightarrow 0^+} \frac{\psi_i(h)}{h} > 0$. Then, the proximity operator $\mathrm{prox}_{\delta \psi_i}(\beta) = \hat{\alpha}(\beta)$ has exactly one continuous solution decoupled in each coordinate $\beta[i]$:

$$\hat{\alpha}[i] = \begin{cases} 0 & \text{if } |\beta[i]| \leqslant \delta \psi'_{i+}(0) \\ \beta_i - \delta \psi'_i(\hat{\alpha}[i]) & \text{if } |\beta[i]| > \delta \psi'_{i+}(0) \end{cases} \tag{13}$$

Among the most popular penalty functions ψ_i satisfying the above requirements, we have $\psi_i(\alpha[i]) = |\alpha[i]|$, $\forall i$, in which case the associated proximity operator is soft-thresholding, denoted ST in the sequel.

3.2 Splitting on the primal problem
3.2.1 Splitting for sums of convex functions
Suppose that the objective to be minimized can be expressed as the sum of K functions in $\Gamma_0(\mathcal{H})$, verifying domain qualification conditions:

$$\underset{x \in \mathcal{H}}{\mathrm{argmin}} \left(F(x) = \sum_{k=1}^K F_k(x) \right) . \tag{14}$$

Proximal splitting methods for solving (14) are iterative algorithms which may evaluate the individual proximity operators prox_{F_k}, supposed to have an explicit convenient structure, but never proximity operators of sums of the F_k.

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 (18) 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, which performs independent proximal steps on each F_k, and then computes the next iterate by essentially averaging the individual proximity operators. The scheme described in [6] is very similar in spirit to Spingarn's method, with some refinements.

3.2.2 Application to noisy inverse problems

Problem $(\mathrm{P}_{f_1,\gamma,\psi})$ is amenable to the form (14), by wisely introducing auxiliary variables. As $(\mathrm{P}_{f_1,\gamma,\psi})$ involves two linear operators ($\boldsymbol{\Phi}$ and $\mathbf{H}$), we need two of them, that we define as $x_1 = \boldsymbol{\Phi}\alpha$ and $x_2 = \mathbf{H}x_1$. The idea is to get rid of the composition of $\boldsymbol{\Phi}$ and $\mathbf{H}$. Let the two linear operators $\mathbf{L}_1 = [\mathbf{I} \quad 0 \quad -\boldsymbol{\Phi}]$ and $\mathbf{L}_2 = [-\mathbf{H} \quad \mathbf{I} \quad 0]$. Then, the optimization problem $(\mathrm{P}_{f_1,\gamma,\psi})$ can be equivalently written:

$$\operatorname*{argmin}_{(x_1,x_2,\alpha)\in\mathcal{H}\times\mathcal{K}\times\mathcal{H}'} \underbrace{f_1(x_2) + \imath_{\mathcal{C}}(x_1) + \gamma\Psi(\alpha)}_{G(x_1,x_2,\alpha)} + \tag{15}$$

$$\imath_{\ker\mathbf{L}_1}(x_1,x_2,\alpha) + \imath_{\ker\mathbf{L}_2}(x_1,x_2,\alpha) , \tag{16}$$

where $\imath_{\ker\mathbf{L}_i}$ are the indicatrice function of the kernel space of the operator $\mathbf{L}_i$, in others words these two indicatrice functions will enforce the equality constraints $x_1 = \boldsymbol{\Phi}\alpha$ and $x_2 = \mathbf{H}x_1$. Notice that in our case $K = 3$ by virtue of separability of the proximity operator of G in x_1, x_2 and α; see Lemma 7.

Algorithm 1: Primal scheme for solving $(\mathrm{P}_{f_1,\gamma,\psi})$.

Parameters: The observed image y, the dictionary $\boldsymbol{\Phi}$, number of iterations N_{iter}, $\mu > 0$ and regularization parameter $\gamma > 0$.
Initialization:
$\forall i \in \{1,2,3\}, \quad p_{(0,i)} = (0,0,0)^{\mathrm{T}}.\ z_0 = (0,0,0)^{\mathrm{T}}.$
Main iteration:
For $t = 0$ **to** $N_{\mathrm{iter}} - 1$,

- Data fidelity (Lemmas 8, 9 and 10):
  $\overline{\xi_{(t,1)}[1]} = \mathrm{prox}_{\mu f_1/3}(p_{(t,1)}[1]).$
- Sparsity-penalty (Lemma 11):
  $\overline{\xi_{(t,1)}[2]} = \mathrm{prox}_{\mu\gamma\Psi/3}(p_{(t,1)}[2]).$
- Positivity constraint: $\xi_{(t,1)}[3] = \mathcal{P}_{\mathcal{C}}(p_{(t,1)}[3]).$
- Auxiliary constraints with $\mathbf{L}_1$ and $\mathbf{L}_2$: (Lemma 12):
  $\overline{\xi_{(t,2)}} = \mathcal{P}_{\ker\mathbf{L}_1}(p_{(t,2)}), \xi_{(t,3)} = \mathcal{P}_{\ker\mathbf{L}_2}(p_{(t,3)}).$
- Average the proximity operators:
  $\xi_t = (\xi_{(t,1)} + \xi_{(t,2)} + \xi_{(t,3)})/3.$
- Choose $\theta_t \in\ ]0,2[.$
- Update the components:
  $\forall i \in \{1,2,3\}, \quad p_{(t+1,i)} = p_{(t,i)} + \theta_t(2\xi_t - z_t - \xi_{(t,i)}).$
- Update the coefficients estimate:
  $z_{t+1} = z_t + \theta_t(\xi_t - z_t).$

End main iteration
Output: Reconstructed image $x^{\star} = z_{N_{\mathrm{iter}}}[0].$

The proximity operators of f_1 and Ψ are easily accessible through Lemmas 8, 9, 10 and 11. The projector onto $\mathcal{C}$ is trivial for most of the case (e.g. positive orthant, closed

interval). It remains now to compute the projector on $\ker\mathbf{L}_i$, $i = 1,2$, which by well-known linear algebra arguments, is obtained from the projector onto the image of $\mathbf{L}_i^*$.

Lemma 12. *The proximity operator associated to $\imath_{\ker\mathbf{L}_i}$ is*

$$\mathcal{P}_{\ker\mathbf{L}_i} = \mathbf{I} - \mathbf{L}_i^*(\mathbf{L}_i \circ \mathbf{L}_i^*)^{-1}\mathbf{L}_i . \tag{17}$$

The inverse in the expression of $\mathcal{P}_{\ker\mathbf{L}_1}$ is $(\mathbf{I} + \boldsymbol{\Phi} \circ \boldsymbol{\Phi}^{\mathrm{T}})^{-1}$ can be computed efficiently when $\boldsymbol{\Phi}$ is a tight frame. Similarly, for $\mathbf{L}_2$, the inverse writes $(\mathbf{I} + \mathbf{H} \circ \mathbf{H}^*)^{-1}$, and its computation can be done in the domain where $\mathbf{H}$ is diagonal; e.g. Fourier for convolution or pixel domain for mask.

Finally, the main steps of our primal scheme are summarized in Algorithm 1. Its convergence is a corollary of [6][Theorem 3.4].

Proposition 13. *Let $(z_t)_{t\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Suppose that Proposition 4-(iii) is verified, and $\sum_{t\in\mathbb{N}} \theta_t(2 - \theta_t) = +\infty$. Then $(z_t)_{t\in\mathbb{N}}$ converges to a (non-strict) global minimizer of $(\mathrm{P}_{f_1,\psi})$.*

3.2.3 Splitting on the dual: Primal-dual algorithm

Our problem $(\mathrm{P}_{f_1,\gamma,\psi})$ can also be rewritten in the form,

$$\operatorname*{argmin}_{\alpha\in\mathcal{H}'} F \circ \mathbf{K}(\alpha) + \gamma\Psi(\alpha) \tag{18}$$

where now $\mathbf{K} = \begin{pmatrix} \mathbf{H} \circ \boldsymbol{\Phi} \\ \boldsymbol{\Phi} \end{pmatrix}$ and $F : (x_1, x_2) \mapsto f_1(x_1) + \imath_{\mathcal{C}}(x_2)$. Again, one may notice that the proximity operator of F can be directly computed using the separability in x_1 and x_2.

Recently, a primal-dual scheme, which turns to be a preconditioned version of ADMM, to minimize objectives of the form (18) was proposed in [2]. Transposed to our setting, this scheme gives the steps summarized in Algorithm 2.

Adapting the arguments of [2], convergence of the sequence $(\alpha_t)_{t\in\mathbb{N}}$ generated by Algorithm 2 is ensured.

Proposition 14. *Suppose that Proposition 4-(iii) holds. Let $\zeta = \|\boldsymbol{\Phi}\|^2 (1 + \|\mathbf{H}\|^2)$, choose $\tau > 0$ and σ such that $\sigma\tau\zeta < 1$, and let $(\alpha_t)_{t\in\mathbb{R}}$ as defined by Algorithm 2. Then, $(\alpha)_{t\in\mathbb{N}}$ converges to a (non-strict) global minimizer $(\mathrm{P}_{f_1,\gamma,\psi})$ at the rate $O(1/t)$ on the restricted duality gap.*

3.3 Discussion

Algorithm 1 and 2 share some similarities, but exhibit also important differences. For instance, the primal-dual algorithm enjoys a convergence rate that is not known for the primal algorithm. Furthermore, the latter necessitates two operator inversions that can only be done efficiently for some $\boldsymbol{\Phi}$ and $\mathbf{H}$, while the former involves only application of these linear operators and their adjoints. Consequently, Algorithm 2 can virtually handle any inverse problem with a bounded linear $\mathbf{H}$. In case where the inverses can be done efficiently, e.g. deconvolution with a tight frame, both algorithms have comparable computational burden. In general,

if other regularizations/constraints are imposed on the solution, in the form of additional proper lsc convex terms that would appear in $(\mathrm{P}_{f_1,\gamma,\psi})$, both algorithms still apply by introducing wisely chosen auxiliary variables.

Algorithm 2: Primal-dual scheme for solving $(\mathrm{P}_{f_1,\gamma,\psi})$.

Parameters: The observed image y, the dictionary $\mathbf{\Phi}$, number of iterations N_{iter}, proximal steps $\sigma > 0$ and $\tau > 0$, and regularization parameter $\gamma > 0$.

Initialization:
$\alpha_0 = \bar{\alpha}_0 = 0$ $\xi_0 = \eta_0 = 0$.

Main iteration:
For $t = 0$ **to** $N_{\mathrm{iter}} - 1$,

- Data fidelity (Lemmas 8, 9 and 10):
  $\overline{\xi_{t+1}} = (\mathbf{I} - \sigma\,\mathrm{prox}_{f_1/\sigma})(\xi_t/\sigma + \mathbf{H} \circ \mathbf{\Phi}\bar{\alpha}_t)$.
- Positivity constraint: $\eta_{t+1} = (\mathbf{I} - \sigma\mathcal{P}_\mathcal{C})(\eta_t/\sigma + \mathbf{\Phi}\bar{\alpha}_t)$.
- Sparsity-penalty (Lemma 11):
  $\overline{\alpha_{t+1}} = \mathrm{prox}_{\tau\gamma\Psi}\left(\alpha_t - \tau\mathbf{\Phi}^{\mathrm{T}}\left(\mathbf{H}^*\xi_{t+1} + \eta_{t+1}\right)\right)$.
- Update the coefficients estimate: $\bar{\alpha}_{t+1} = 2\alpha_{t+1} - \alpha_t$

End main iteration
Output: Reconstructed image $x^\star = \mathbf{\Phi}\alpha_{N_{\mathrm{iter}}}$.

# 4. EXPERIMENTAL RESULTS
4.1 Deconvolution under Poisson noise
Our algorithms were applied to deconvolution. In all experiments, Ψ was the ℓ_1-norm. Table 1 summarizes the mean absolute error (MAE) and the execution times for an astronomical image, where the dictionary consisted of the wavelet transform (here the translation-invariant discret wavelet transform which form a tight frame and so is a redundant transform) and the PSF was that of the Hubble telescope. Our algorithms were compared to state-of-the-art alternatives in the literature. In summary, flexibility of our framework and the fact that Poisson noise was handled properly, demonstrate the capabilities of our approach, and allow our algorithms to compare very favorably with other competitors. The computational burden of our approaches is also among the lowest, typically faster than the PIDAL algorithm [13]. Fig. 1 displays the objective as a function of the iteration number and times (in s). We can clearly see that Algorithm 2 converges faster than Algorithm 1.

	RL-MRS [15]	RL-TV [10]	StabG [11]	PIDAL-FS [13]
MAE	63.5	52.8	43	43.6
Times	230s	4.3s	311s	342s
	Alg. 1	Alg. 2		
MAE	46	43.6		
Times	183s	154s		

Table 1: MAE and execution times for the deconvolution of the sky image.

## 4.2 Inpainting with Gaussian noise
We also applied our algorithms to inpainting with Gaussian noise. In all experiments, Ψ was the ℓ_1-norm. Fig 2 summarizes the results with the PSNR and the execution times for the Cameraman, where the dictionary consisted of the same wavelet transform as used for deconvolution and the mask was create from a random process (here with about 34% of missing pixels). Notice that both algorithms leads to the same solution which gives a good reconstruction of
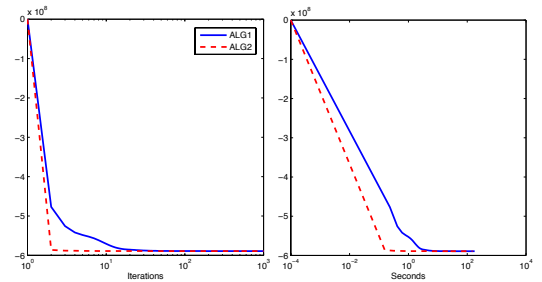
Figure 1: Objective function for deconvolution under Poisson noise in function of iterations (left) and times (right).

the image. Fig. 3 displays the objective as a function of the iteration number and times (in s). Again, we can clearly see that Algorithm 2 converges faster than Algorithm 1.


Original   Masked and noisy (PSNR = 11.1)

Alg. 1 (PSNR = 25.8)   Alg. 2 (PSNR = 25.8)
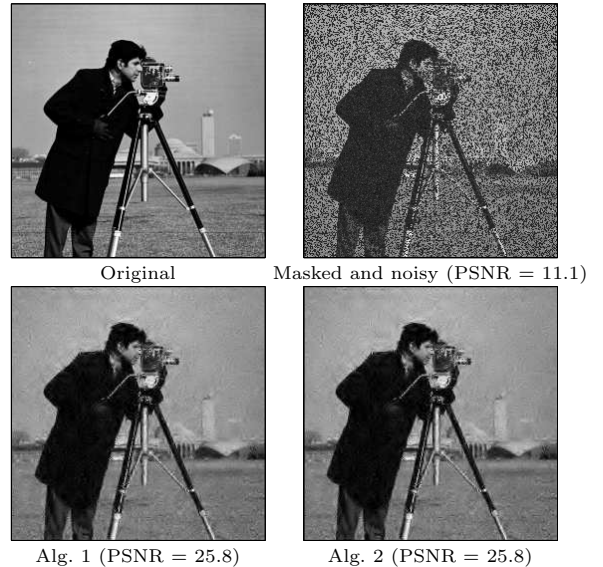
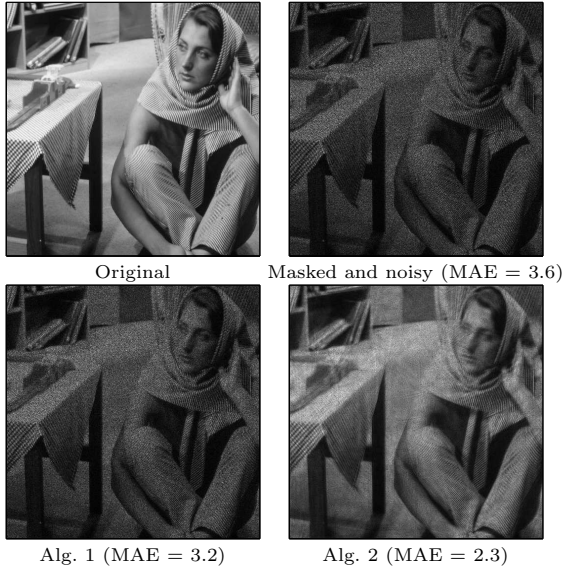**Figure 2: Inpainting results for the Cameraman using our two algorithms.**

4.3 Denoising with Multiplicative noise
As imply by (8), the final estimate for each algorithm is given by taking the exponential of the result. In all experiments, Ψ was the ℓ_1-norm. The Barbara image was set to a maximal intensity of 30 and the minimal to a non-zero value in order to avoid issues with the logarithm. The noise was added using $M = 10$ which leads to a medium level of noise. Fig 4 summarizes the results with the MAE and the execution times for Barbara, where the dictionary consisted of the curvelets transform. The difference between the visual results of the two algorithms can be explain be the fact we did not stop at the convergence, but instead stop after one thousand iterations. Our methods give correct reconstruction of the image. Fig. 5 displays the objective as a function of the iteration number and times (in s). Again, we can clearly see that Algorithm 2 converges faster than Algorithm 1.

5. CONCLUSION
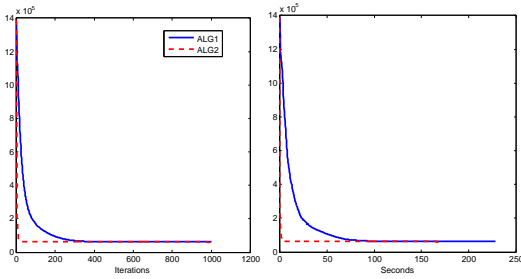
Figure 3: Objective function for inpainting with Gaussian noise in function of iterations (left) and times (right).

Figure 5: Objective function for denoising with multiplicative noise in function of iterations (left) and times (right).

Original   Masked and noisy (MAE = 3.6)
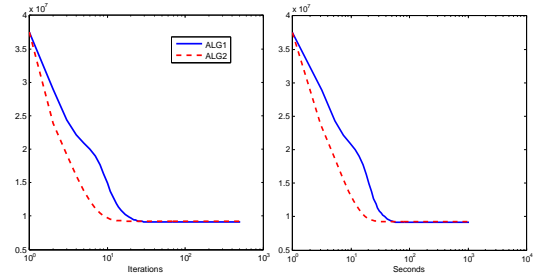
Alg. 1 (MAE = 3.2)   Alg. 2 (MAE = 2.3)

Figure 4: Denoising results for Barbara using our two algorithms.

In this paper, we proposed two provably convergent algorithms for solving the linear inverse problems with a sparsity prior. The primal-dual proximal splitting algorithm seems to perform better in terms of convergence speed than the primal one. Moreover, its computational burden is lower than most comparable of state-of-art methods. Inverse problems with multiplicative noise does not enter currently in this framework, we will consider its adaptation to such problems in future work.

6. REFERENCES

[1] J. Bioucas-Dias and M. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(7):1720–1730, 2010.

[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Technical report, CMAP, Ecole Polytechnique, 2010.

[3] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inv. Prob.*, 23:1495–1518, 2007.

[4] C. Chaux, J.-C. Pesquet, and N. Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 2(2):730–762, 2009.

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

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

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

[8] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996. 10.1007/BF02124750.

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

[10] N. Dey, L. Blanc-Féraud, J. Zerubia, C. Zimmer, J.-C. Olivo-Marin, and Z. Kam. A deconvolution method for confocal microscopy with total variation regularization. In *IEEE ISBI*, pages 1223–1226, Arlington, USA, 15-18 April 2004.

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

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

[13] M. Figueiredo and J. Bioucas-Dias. Restoration of poissonian images using alternating direction optimization. *IEEE Transactions on Image Processing*, 19(12), 2010.

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

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

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