Linear inverse problems with noise: primal and primal-dual splitting
François-Xavier Dupé, Jalal M. Fadili, Jean-Luc Starck

To cite this version:
François-Xavier Dupé, Jalal M. Fadili, Jean-Luc Starck. Linear inverse problems with noise: primal and primal-dual splitting. 2011. hal-00575610

HAL Id: hal-00575610
https://hal.archives-ouvertes.fr/hal-00575610
Submitted on 10 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Linear inverse problems with noise: primal and primal-dual splitting

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

Jalal M. Fadili GREYC
ENSICAEN-Université de
Caen
14050 Caen, France
jfadili@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 for solving linear inverse problems when the observations are corrupted by noise. A proper data fidelity term (log-likelihood) is introduced to reflect the statistics of the noise (e.g. Gaussian, Poisson). On the other hand, as a prior, the images to restore are assumed to be positive and sparsely represented in a dictionary of waveforms. Piecing together the data fidelity and the prior terms, the solution to the inverse problem is cast as the minimization of a non-smooth convex functional. We establish the well-posedness of the optimization problem, characterize the corresponding minimizers, and 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
A lot of works have already been dedicated to linear inverse problems with Gaussian noise (see [4] 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 pioneer work for Gaussian noise by [4], many other methods have appeared for managing linear inverse problems 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, [4] 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, [4] presented a deconvolution algorithm with TV and sparsity regularization, and [4] a denoising algorithm for multiplicative noise. This scheme however necessitates to solve a least-squares problem which can be done explicitly only in some cases.

In this paper, we propose a framework for solving linear inverse problems when the observations are corrupted by noise. In order to form the data fidelity term, we take the exact likelihood. As a prior, the images to restore are assumed to be positive and sparsely represented in a dictionary of atoms. The solution to the inverse problem is cast as the minimization of a 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 $\| \cdot \|$ the norm associated with the inner product in $\mathcal{H}$, and $I$ the identity operator on $\mathcal{H}$. $\| \cdot \|_{p, p} \geq 1$ is the $\ell_p$ norm. $x$ and $a$ are respectively reordered vectors of image samples and transform coefficients. We denote by $\text{ri} \mathcal{C}$ 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 $\text{dom} f = \{ x \in \mathcal{H} \mid f(x) < +\infty \} \neq \emptyset$. $\Gamma_0(\mathcal{H})$ is the class of all proper lower semi-continuous (lsc) convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$. We denote by $\| M \| = \max_{x \neq 0} \frac{| M x |}{\| x \|}$ the spectral norm of the linear operator $M$, and $\ker( M ) := \{ x \in \mathcal{H} : M x = 0, x \neq 0 \}$ its kernel.

Let $x \in \mathcal{H}$ be an $\sqrt{n} \times \sqrt{n}$ image. $x$ can be written as
the superposition of elementary atoms $\varphi_\gamma$ parameterized by $\gamma \in \mathcal{I}$ such that $x = \sum_{\gamma \in \mathcal{I}} \alpha_\gamma \varphi_\gamma = \Phi \alpha$, $|\mathcal{I}| = L$, $L \geq n$. We denote by $\Phi : \mathcal{H}^L \to \mathcal{H}$ the dictionary (typically a frame of $\mathcal{H}$), whose columns are the atoms all normalized to a unit $\ell_2$-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 $H : \mathcal{K} \to \mathcal{H}$, and contaminated by a noise $\varepsilon$ through a composition operator $\odot$ (e.g. addition),

$$y \sim Hx \odot \varepsilon .$$

The linear inverse problem at hand is to reconstruct 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 $\Phi$ as measured by a sparsity-promoting penalty $\Psi$ supposed throughout to be convex but non-smooth, e.g. the $\ell_1$ norm.

2.1 Gaussian noise case

For Gaussian noise, we consider the following formation model,

$$y = Hx + \varepsilon ,$$

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

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

$$f_{\text{Gaussian}} : \eta \in \mathbb{R}^n \mapsto \| \eta - y \|^2 / (2\sigma^2) .$$

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

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

2.2 Poisson noise case

The observed image is then a discrete collection of counts $y = (y[i])_{i \in \mathbb{G}}$ 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 $(Hx)_i$. Formally, this writes in a vector form as

$$y \sim \mathcal{P}(Hx) .$$

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

$$p(y|x) = \prod_i \frac{(Hx)[i]^{y[i]} \exp(-Hx)[i])}{y[i]!} .$$

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

$$f_{\text{Poisson}} : \eta \in \mathbb{R}^n \mapsto \sum_{i=1}^{n} f_p(\eta[i]),$$

where

- if $y[i] > 0$, $f_p(\eta[i]) = \left\{ \begin{array}{ll} -y[i] \log(\eta[i]) + \eta[i] & \text{if } \eta[i] > 0, \\ +\infty & \text{otherwise,} \end{array} \right.$
- if $y[i] = 0$, $f_p(\eta[i]) = \left\{ \begin{array}{ll} \eta[i] & \text{if } \eta[i] \in [0, +\infty), \\ +\infty & \text{otherwise.} \end{array} \right.$

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

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

2.3 Multiplicative noise

We consider the case without linear operator and as in $\mathcal{I}$ with a $M$-look full developed speckle noise,

$$y = xe, \quad \varepsilon \sim \Gamma(M, 1/M) .$$

In order to simplify the problem, the logarithm of the observation is considered, $\log(y) = \log(x) + \log(\varepsilon) = z + \omega$. And in $\mathcal{I}$, the authors proof that the anti log-likelihood yields,

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

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

Proposition 3. $f_{\text{Multi}}$ 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$^4$:

$$\arg\min_{\alpha \in \mathbb{H}^L} J(\alpha), \quad J : \alpha \mapsto f_1 \circ H \circ \Phi(\alpha) + \gamma \Psi(\alpha) + \zeta \circ \Phi(\alpha) .$$

The data fidelity term $f_1$ reflect the noise statistics, the penalty function $\Psi : \alpha \mapsto \sum_{i=0}^{L} \psi_i(\alpha[i])$ is positive, additive, and chosen to enforce sparsity, $\gamma > 0$ is a regularization parameter and $\zeta$ is 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 $\mathcal{I}$ for others examples).

$^4$Our framework and algorithms extend to an analysis-type prior just as well.
From the objective in \( f_{1, \gamma, \delta} \), we get the following.

**Proposition 4.**

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

(ii) Suppose that \( f_1 \) is strictly convex, then \( f_1 \circ \mathbf{H} \circ \Phi \) remains strictly convex if \( \Phi \) is an orthobasis and \( \ker(\mathbf{H}) = \emptyset \).

(iii) Suppose that \( (0, +\infty) \cap \mathbf{H} ([0, +\infty)) \neq \emptyset \). Then \( J \in \Gamma_0(\mathbf{H}) \).

**2.5 Well-posedness of \( f_{1, \gamma, \delta} \)**

Let \( \mathcal{M} \) be the set of minimizers of problem \( f_{1, \gamma, \delta} \). Suppose that \( \Psi \) is coercive. Thus \( J \) is coercive. Therefore, the following holds:

**Proposition 5.**

(i) Existence: \( f_{1, \gamma, \delta} \) has at least one solution, i.e. \( \mathcal{M} \neq \emptyset \).

(ii) Uniqueness: \( f_{1, \gamma, \delta} \) has a unique solution if \( \Psi \) 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 \( f_{1, \gamma, \delta} \). At the heart of the splitting framework is the notion of proximity operator.

**Definition 6.** Let \( F \in \Gamma_0(\mathbf{H}) \). Then, for every \( x \in \mathbf{H}, \) the function \( y \mapsto F(y) + \| y - x \|^2/2 \) achieves its infimum at a unique point denoted by \( \text{prox}_F x \). The operator \( \text{prox}_F: \mathbf{H} \to \mathbf{H} \) 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.** Let \( F_k \in \Gamma_0(\mathbf{H}), \ k \in \{1, \cdots, K\} \) and let \( G : (x_k)_{1 \leq k \leq K} \mapsto \sum_k F_k(x_k) \). Then \( \text{prox}_G = (\text{prox}_{F_k})_{1 \leq k \leq K} \).

For Gaussian noise, we can easily prove that with \( f_1 \) as defined in \( 4 \), we have

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

\[
\text{prox}_{\beta f_{\text{Gaussian}}} x = \frac{\beta y + \sigma^2 x}{\beta + \sigma^2}.
\]

The following result can be proved easily by solving the proximal optimization problem in Definition 4 with \( f_1 \) as defined in \( 4 \), see also \( 6 \).

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

\[
\text{prox}_{\beta f_{\text{Poisson}}} x = \left( x[i] - \beta + \sqrt{(x[i] - \beta)^2 + 4\beta y[i]} \right) / 2 \cdot \mathbf{1}_{x \in \mathbf{G}}.
\]

As with multiplicative noise \( f_{\text{Multi}} \) involves the exponential, we need the W-Lambert function \( \mathbf{L} \) in order to derive a closed form of the proximity operator.

**Lemma 10.** Let \( y \) be the observations, the proximity operator associated to \( f_{\text{Multi}} \) is,

\[
\text{prox}_{\beta f_{\text{Multi}}} x = x - \beta M - W(-\beta M \exp(x - \log(y) - \beta M))
\]

where \( W \) is the W-Lambert function.

We now turn to \( \text{prox}_{\beta \Phi} \) which is given by Lemma 4 and the following result:

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

\[
\alpha[i] = \begin{cases} 0 & \text{if } |\beta[i]| \leq \delta \psi_i'(0) \\ 
\beta[i] - \delta \psi_i'(\alpha[i]) & \text{if } |\beta[i]| > \delta \psi_i'(0) 
\end{cases}
\]

Among the most popular penalty functions \( \psi_i \) 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 \( \text{ST} \) 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 \( \Gamma_0(\mathbf{H}) \), verifying domain qualification conditions:

\[
\arg\min_{x \in \mathbf{H}} \left( F(x) = \sum_{k=1}^K F_k(x) \right).
\]

Proximal splitting methods for solving \( (13) \) are iterative algorithms which may evaluate the individual proximity operators \( \text{prox}_{F_k} \), 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 \( (13) \) to the case \( K = 2 \) in the product space \( \mathbf{H}^K \). 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 \( 4 \) is very similar in spirit to Spingarn’s method, with some refinements.

**3.2.2 Application to noisy inverse problems**

Problem \( f_{\text{Poisson}} \) is amenable to the form \( (13) \), by wisely introducing auxiliary variables. As \( f_{\text{Multi}} \) involves two linear operators (\( \Phi \) and \( \mathbf{H} \)), we need two of them, that we
define as $x_1 = \Phi \alpha$ and $x_2 = Hx_1$. The idea is to get rid of the composition of $\Phi$ and $H$. Let the two linear operators $L_1 = [1 \ 0 \ -\Phi]$ and $L_2 = [-H \ 1 \ 0]$. Then, the optimization problem $\min_{x_1, x_2} f_1(x_2) + \iota_c(x_1) + \gamma \Psi(\alpha(x_2)) \min_{x_1, x_2} G(x_1, x_2, \alpha)$ can be equivalently written:

\[
\begin{align*}
\min_{(x_1, x_2, \alpha) \in H \times K \times H^*} & f_1(x_2) + \iota_c(x_1) + \gamma \Psi(\alpha(x_2)) \\
& + \iota_{\ker L_1}(x_1, x_2, \alpha) \quad + \iota_{\ker L_2}(x_1, x_2, \alpha) .
\end{align*}
\]

Notice that in our case $K = 3$ by virtue of separability of the proximity operator of $G$ in $x_1, x_2$ and $\alpha$; see Lemma 1.

Algorithm 1: Primal scheme for solving $\min_{x_1, x_2} f_1(x_2)$

Parameters: The observed image $y$, the dictionary $\Phi$, number of iterations $N_{\text{iter}}$, $\mu > 0$ and regularization parameter $\gamma > 0$.

Initialization:

\[ \forall i \in \{1, 2, 3\}, \quad P_{\text{iter}, i}(0, 0, 0)^T \cdot z_0 = (0, 0, 0)^T. \]

Main iteration:

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

- Data fidelity (Lemmas 11 and 10):\[ \xi_{t(1)}[1] = \text{prox}_{f_1}(P_{\text{iter}, 1}(1)). \]
- Sparsity-penalty (Lemma 11):\[ \xi_{t(1)}[2] = \text{prox}_{\gamma / 3}(P_{\text{iter}, 1}[2]). \]
- Positivity constraint: $\xi_{t(1)}[3] = P_{\text{C}}(P_{\text{iter}, 1}[3]).$
- Auxiliary constraints with $L_1$ and $L_2$: (Lemma 13):\[ \xi_{t(2)} = P_{\text{ker} L_1}(P_{\text{iter}, 2}), \quad \xi_{t(3)} = P_{\text{ker} L_2}(P_{\text{iter}, 3}). \]

Average the proximity operators:

\[ \xi_t = (\xi_{t(1)} + \xi_{t(2)} + \xi_{t(3)}/3. \]

- Choose $\theta_i \in [0, 2]$.\[ \theta_i(\xi_{t(1)} - \xi_t - \xi_{t(2)})/3. \]

- Update the components:

\[ \forall i \in \{1, 2, 3\}, \quad R_{t(1), i}(t) = P_{\text{iter}, i}(t, 0, 0) + \theta_i(2\xi_t - z_i - \xi_{t(1)}). \]

- Update the coefficients estimate:

\[ z_{t+1} = z_t + \theta_i(\xi_t - z_t). \]

End main iteration

Output: Reconstructed image $x^* = z_{N_{\text{iter}}, 3}.$

The proximity operators of $F$ and $\Psi$ are easily accessible through Lemmas 11 and 10. The projector onto $C$ is trivial for most of the case (e.g. positive orthant, closed interval). It remains now to compute the projector on $\ker L_i$, $i = 1, 2$, which by well-known linear algebra arguments, is obtained from the projector on the image of $L_i$.

Lemma 12. The proximity operator associated to $\ker L_i$ is

\[ P_{\text{ker} L_i} = I - L_i^T (L_i \circ L_i)^{-1} L_i . \] (16)

The inverse in the expression of $P_{\text{ker} L_i}$ is $(I + \Phi \circ \Phi^T)^{-1}$ can be computed efficiently when $\Phi$ is a tight frame. Similarly, for $L_2$, the inverse writes $(I + H \circ H^*)^{-1}$, and its computation can be done in the domain where $H$ 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 Theorem 3.4.

Proposition 13. Let $(z_t)_{t \in \mathbb{N}}$ be a sequence generated by Algorithm 1. Suppose that Proposition 4-(iiii) is verified, and $\sum_{t \in \mathbb{N}} \theta_t = +\infty$. Then $(z_t)_{t \in \mathbb{N}}$ converges to a (non-strict) global minimizer of $\min_{x_1, x_2} f_1(x_2)$.

3.2.3 Splitting on the dual: Primal-dual algorithm

Our problem $\min_{x_1, x_2} f_1(x_2)$ can also be rewritten in the form,

\[ \min_{x_1, x_2} F \circ K(\alpha) + \gamma \Psi(\alpha) \]

where now $K = (H \circ \Phi)$. 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 pre-conditioned version of ADMM, to minimize objectives of the form $(17)$ was proposed in 12. Transposed to our setting, this scheme gives the steps summarized in Algorithm 2.

Adapting the arguments of 9, convergence of the sequence $(\alpha_t)_{t \in \mathbb{N}}$ generated by Algorithm 2 is ensured.

Proposition 14. Suppose that Proposition 4-(iiii) holds. Let $\zeta = (\|F\|_2^2 (1 + \|H\|_2^2))$, choose $\tau > 0$ and $\sigma < 1$, and let $(\alpha_t)_{t \in \mathbb{N}}$ as defined by Algorithm 2. Then, $(\alpha_t)_{t \in \mathbb{N}}$ converges to a (non-strict) global minimizer of $\min_{x_1, x_2} f_1(x_2)$ at the rate $O(1/t)$ on the restricted duality gap.

3.3 Discussion

Algorithms 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 $\Phi$ and $H$, 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 $H$. 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 $\min_{x_1, x_2} f_1(x_2)$, both algorithms still apply by introducing wisely chosen auxiliary variables.

4. EXPERIMENTAL RESULTS

4.1 Deconvolution under Poisson noise

Our algorithms were applied to deconvolution. In all experiments, $\Psi$ was the $l_1$-norm. Table 1 summarizes the mean absolute error (MAE) and the execution times for this scheme gives the steps summarized in Algorithm 2.

Recently, a primal-dual scheme, which turns to be a pre-conditioned version of ADMM, to minimize objectives of the form $(17)$ was proposed in 12. Transposed to our setting, this scheme gives the steps summarized in Algorithm 2.

Adapting the arguments of 9, convergence of the sequence $(\alpha_t)_{t \in \mathbb{N}}$ generated by Algorithm 2 is ensured.

Proposition 14. Suppose that Proposition 4-(iiii) holds. Let $\zeta = (\|F\|_2^2 (1 + \|H\|_2^2))$, choose $\tau > 0$ and $\sigma < 1$, and let $(\alpha_t)_{t \in \mathbb{N}}$ as defined by Algorithm 2. Then, $(\alpha_t)_{t \in \mathbb{N}}$ converges to a (non-strict) global minimizer of $\min_{x_1, x_2} f_1(x_2)$ at the rate $O(1/t)$ on the restricted duality gap.

3.3 Discussion

Algorithms 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 $\Phi$ and $H$, 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 $H$. 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 $\min_{x_1, x_2} f_1(x_2)$, both algorithms still apply by introducing wisely chosen auxiliary variables.
Algorithm 2: Primal-dual scheme for solving (P_{f_1, γ, ω}).

Parameters: The observed image \( y \), the dictionary \( D \), number of iterations \( N_{\text{iter}} \), proximal steps \( σ > 0 \) and \( τ > 0 \), and regularization parameter \( γ > 0 \).

Initialization:
\[ α_0 = 0, \xi_0 = 0, η_0 = 0. \]

Main iteration:
For \( t = 0 \) to \( N_{\text{iter}} - 1 \),
- Data fidelity (Lemmas 5 and 10):
  \[ ξ_{t+1} = (I - σ \operatorname{prox}_{f_1/σ})(ξ_t/σ + H^T Φ(α_t)). \]
- Positivity constraint: \( η_{t+1} = (I - σP_δ)(η_t/σ + Φ(α_t)). \)
- Sparsity-penalty (Lemma 11):
  \[ α_{t+1} = \operatorname{prox}_{γ^2/2}\|\cdot\|_2(α_t - τΦ^T(Φ^T ξ_{t+1} + η_{t+1})). \]
- Update the coefficients estimate:
  \[ α_{t+1} = 2α_{t+1} - α_t. \]

End main iteration
Output: Recovered image \( \hat{x} = Φα_{N_{\text{iter}}} \).

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

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>Times (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL-MRS</td>
<td>63.6</td>
<td>43.0</td>
</tr>
<tr>
<td>RL-TV</td>
<td>63.6</td>
<td>46.5</td>
</tr>
<tr>
<td>Stable G</td>
<td>63.6</td>
<td>311.6</td>
</tr>
<tr>
<td>PIDAL-FS</td>
<td>63.6</td>
<td>342.6</td>
</tr>
</tbody>
</table>

Figure 1: Objective function for deconvolution under Poisson noise in function if iterations (left) and times (right).

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 curvelets transform and the mask was created 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 the image. Fig. 3 displays the objective as a function of the iteration number and times (in s). As we can clearly see that Algorithm 2 converges faster than Algorithm 1.

4.3 Denoising with Multiplicative noise

As we work on the logarithm the problem (see 2.3, 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. 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
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


