A Forward-Backward View of Some Primal-Dual Optimization Methods in Image Recovery
Patrick Louis Combettes, Laurent Condat, Jean-Christophe Pesquet, Bang Cong Vu

To cite this version:
hal-01098038

HAL Id: hal-01098038
https://hal.science/hal-01098038
Submitted on 22 Dec 2014

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.
ABSTRACT

A wide array of image recovery problems can be abstracted into the problem of minimizing a sum of composite convex functions in a Hilbert space. To solve such problems, primal-dual proximal approaches have been developed which provide efficient solutions to large-scale optimization problems. The objective of this paper is to show that a number of existing algorithms can be derived from a general form of the forward-backward algorithm applied in a suitable product space. Our approach also allows us to develop useful extensions of existing algorithms by introducing a variable metric. An illustration to image restoration is provided.

Index Terms— convex optimization, duality, parallel computing, proximal algorithm, variational methods, image recovery.

1. INTRODUCTION

Many image recovery problems can be formulated in Hilbert spaces $\mathcal{H}$ and $(\mathcal{G}_i)_{1 \leq i \leq m}$ as structured optimization problems of the form

$$\min_{x \in \mathcal{H}} \sum_{i=1}^{m} g_i(L_ix), \quad (1)$$

where, for every $i \in \{1, \ldots, m\}$, $g_i$ is a proper lower semi-continuous convex function from $\mathcal{G}_i$ to $[-\infty, +\infty]$ and $L_i$ is a bounded linear operator from $\mathcal{H}$ to $\mathcal{G}_i$. For example, the functions $(g_i \circ L_i)_{1 \leq i \leq m}$ may model data fidelity terms, smooth or nonsmooth measures of regularity, or hard constraints on the solution. In recent years, many algorithms have been developed to solve such a problem by taking advantage of recent advances in convex optimization, especially in the development of proximal tools (see [12, 29] and the references therein). In image processing, however, solving such a problem still poses a number of conceptual and numerical challenges. First of all, one often looks for methods which have the ability to split the problem by activating each of the functions through elementary processing steps which can be computed in parallel. This makes it possible to reduce the complexity of the original problem and to benefit from existing parallel computing architectures. Secondly, it is often useful to design algorithms which can exploit, in a flexible manner, the structure of the problem. In particular, some of the functions may be Lipschitz differentiable in which case they should be exploited through their gradient rather than through their proximity operator, which is usually harder to implement (examples of proximity operators with closed-form expression can be found in [6, 12]). In some problems, the functions $(g_i)_{1 \leq i \leq m}$ can be expressed as the infimal convolution of simpler functions (see [9] and the references therein). Last but not least, in image recovery, the operators $(L_i)_{1 \leq i \leq m}$ may be of very large size so that their inverses are costly (e.g., in reconstruction problems). Finding algorithms which do not require to perform inversions of these operators is thus of paramount importance.

Note that all the existing convex optimization algorithms do not have these desirable properties. For example, the Alternating Direction Method of Multipliers (ADMM) [18, 17, 20] requires a stringent assumption of invertibility of the involved linear operator. Parallel versions of ADMM [28] and related Parallel Proximal Algorithm (PPXA) [11, 25] usually need to compute a linear inversion to be performed at each iteration. Also, early primal-dual algorithms [4, 5, 7, 10, 16, 21] did not make it possible to handle smooth functions through their gradients. Only recently, have primal-dual methods been proposed with this feature. Such work was initiated in [13] in the line of [4] and subsequent developments can be found in [2, 3, 8, 9, 15, 27, 30]. As will be seen in the present paper, another advantage of these approaches is that they can be coupled with variable metric strategies which can potentially accelerate their convergence.

In Section 2, we provide some background on convex analysis and monotone operator theory. In Section 3, we introduce a general form of the forward-backward algorithm which uses a variable metric. This algorithm is employed in Section 4 to develop a versatile family of primal-dual proximal methods. Several particular instances of this framework are discussed. Finally, we provide illustrating numerical results in Section 5.

2. NOTATION AND BACKGROUND

Monotone operator theory [1] provides a both insightful and elegant framework for dealing with convex optimization problems and developing new solution algorithms that cannot be devised using purely variational tools. We summarize a number of related concepts that will be needed.

Throughout, $\mathcal{H}$, $\mathcal{G}$, and $(\mathcal{G}_i)_{1 \leq i \leq m}$ are real Hilbert spaces. We denote the scalar product of a Hilbert space by $\langle \cdot | \cdot \rangle$ and the associated norm by $\| \cdot \|$. The symbol $\rightharpoonup$ denotes weak convergence, and $\text{Id}$ denotes the identity operator. We denote by $\mathcal{B}(\mathcal{H}, \mathcal{G})$ the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$, we set $\mathcal{S}(\mathcal{H}) = \ldots$
The class of lower semicontinuous convex functions \( f \) and its subdifferential of \( \text{gra} \ A \). Let \( U \subset H \) be a Hilbert space, \( \text{dom} \ f \subset H \), \( f \) be convex and differentiable with a Lipschitzian gradient. For every \( i \in \{1, \ldots, m\} \), let \( r_i \in H_i \) be a strictly positive integer, let \( g_i \in \Gamma_0(G_i) \), let \( \ell_i \in \Gamma_0(G_i) \) be strongly convex, and suppose that \( 0 \neq L_i \in B(H_i, G_i) \). Suppose that

\[
\begin{align*}
\text{Problem 4.1} \quad & \text{Let } z \in H, \text{ let } m \text{ be a strictly positive integer, let } f \in \Gamma_0(H), \text{ and let } h : H \to \mathbb{R} \text{ be convex and differentiable with a Lipschitzian gradient. For every } i \in \{1, \ldots, m\}, \text{ let } r_i \in H_i, \text{ let } g_i \in \Gamma_0(G_i), \text{ let } \ell_i \in \Gamma_0(G_i) \text{ be strongly convex}, \text{ and suppose that } 0 \neq L_i \in B(H_i, G_i). \text{ Suppose that } \\
& z \in \mathbb{R} \left( \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \circ \partial \ell_i)(L_i \cdot r_i) + \nabla h \right). \tag{11}
\end{align*}
\]

Consider the problem

\[
\begin{align*}
\text{Problem 4.1} \quad & \text{Let } z \in H, \text{ let } m \text{ be a strictly positive integer, let } f \in \Gamma_0(H), \text{ and let } h : H \to \mathbb{R} \text{ be convex and differentiable with a Lipschitzian gradient. For every } i \in \{1, \ldots, m\}, \text{ let } r_i \in H_i, \text{ let } g_i \in \Gamma_0(G_i), \text{ let } \ell_i \in \Gamma_0(G_i) \text{ be strongly convex}, \text{ and suppose that } 0 \neq L_i \in B(H_i, G_i). \text{ Suppose that } \\
& z \in \mathbb{R} \left( \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \circ \partial \ell_i)(L_i \cdot r_i) + \nabla h \right). \tag{11}
\end{align*}
\]

and the dual problem

\[
\begin{align*}
\text{Problem 4.1} \quad & \text{Let } z \in H, \text{ let } m \text{ be a strictly positive integer, let } f \in \Gamma_0(H), \text{ and let } h : H \to \mathbb{R} \text{ be convex and differentiable with a Lipschitzian gradient. For every } i \in \{1, \ldots, m\}, \text{ let } r_i \in H_i, \text{ let } g_i \in \Gamma_0(G_i), \text{ let } \ell_i \in \Gamma_0(G_i) \text{ be strongly convex}, \text{ and suppose that } 0 \neq L_i \in B(H_i, G_i). \text{ Suppose that } \\
& z \in \mathbb{R} \left( \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \circ \partial \ell_i)(L_i \cdot r_i) + \nabla h \right). \tag{11}
\end{align*}
\]

Finally, \( \ell_1^*(\mathbb{N}) \) denotes the set of summable sequences in \([0, +\infty[\).
Note that in the special case when \( \ell_i = \iota(0) \), \( g_i \square \ell_i \) reduces to \( g_i \) in (12).

Let us now examine how Problem 4.1 can be reformulated from the standpoint of monotone operators. To this end, let us define \( g \in \Gamma_0(\mathcal{G}) \), \( \ell \in \Gamma_0(\mathcal{G}) \) and \( L \in \mathcal{B}(\mathcal{H}, \mathcal{G}) \) by

\[
g : v \mapsto \sum_{i=1}^m g_i(v_i), \quad \ell : v \mapsto \sum_{i=1}^m \ell_i(v_i)
\]

and

\[
L : x \mapsto (L_1x, \ldots, L_mx).
\]  

(14)

Let us now introduce the product space \( \mathcal{K} = \mathcal{H} \oplus \mathcal{G} \) and the operators

\[
A : \mathcal{K} \to 2^\mathcal{K}
\]

\[
(x, v) \mapsto (\partial f(x) - z + L^*v) \times (-Lx + \partial g^*(v) + r)
\]

(15)

and

\[
B : \mathcal{K} \to \mathcal{K}
\]

\[
(x, v) \mapsto (\nabla h(x), \nabla \ell^*(v)).
\]  

(16)

The operator \( A \) can be shown to be maximally monotone, whereas \( B \) is cocoercive. A key observation in this context is that, if there exists \((\tau, \sigma) \in \mathcal{K}\) such that \((\tau, \sigma) \in \arg\{A + B\}\), then \((\tau, \sigma)\) is a pair of primal-dual solutions to Problem 4.1 [13]. This connection with the construction for a zero of \( A + B \) makes it possible to apply a forward-backward algorithm as discussed in Section 3, by using a linear operator \( V_n \in \mathcal{B}(\mathcal{K}, \mathcal{K}) \) to change the metric at each iteration \( n \). Depending on the form of this operator various algorithms can be obtained.

4.2. A first class of primal-dual algorithms

Let \( \alpha \in [0, +\infty[ \), let \((U_n)_{n\in\mathbb{N}} \) be a sequence in \( \mathcal{P}_\alpha(\mathcal{H}) \) such that \((\forall n \in \mathbb{N}) \quad U_{n+1} \supseteq U_n \). For every \( i \in \{1, \ldots, m\} \), let \((U_{i,n})_{n\in\mathbb{N}} \) be a sequence in \( \mathcal{P}_\alpha(\mathcal{G}_i) \) such that \((\forall n \in \mathbb{N}) \quad U_{i,n+1} \supseteq U_{i,n} \). A first possible choice for \((V_n)_{n\in\mathbb{N}} \) is given by

\[
(V_n)_{n\in\mathbb{N}} \text{ is given by } (\forall n \in \mathbb{N}) \quad V^{-1}_n : (x, v) \mapsto (U^{-1}_{n-1} x - L^*v, -Lx + \bar{U}^{-1}_n v)
\]

where

\[
\bar{U}_n : \mathcal{G} \to \mathcal{G} : (v_1, \ldots, v_m) \mapsto (U_{1,n}v_1, \ldots, U_{m,n}v_m)
\]

(17)

The following result constitutes a direct extension of [14, Example 6.4]:

**Proposition 4.2** Let \( x_0 \in \mathcal{H} \), and let \((a_n)_{n\in\mathbb{N}} \) and \((c_n)_{n\in\mathbb{N}} \) be absolutely summable sequences in \( \mathcal{H} \). For every \( i \in \{1, \ldots, m\} \), let \( v_{i,0} \in \mathcal{G}_i \), let \((b_{i,n})_{n\in\mathbb{N}} \) and \((d_{i,n})_{n\in\mathbb{N}} \) be absolutely summable sequences in \( \mathcal{G}_i \). For every \( n \in \mathbb{N} \), let \( \mu_n \in [0, +\infty[ \) be a Lipschitz constant of \( U_i^{1/2} \circ \nabla h \circ U_i^{1/2} \) and, for every \( i \in \{1, \ldots, m\} \), let \( \nu_{i,n} \in [0, +\infty[ \) be a Lipschitz constant of \( U_i^{1/2} \circ \nabla \ell^*_i \circ U_i^{1/2} \). Let \( (\lambda_n)_{n\in\mathbb{N}} \) be a sequence in \( [0, 1] \) such that \( \inf_{n\in\mathbb{N}} \lambda_n > 0 \). For every \( n \in \mathbb{N} \), set

\[
\delta_n = \left( \sum_{i=1}^m \left\| U_{i,n} L_i \right\| \right)^{-1/2} - 1,
\]

and suppose that

\[
\inf_{n\in\mathbb{N}} (1 + \delta_n) \max\{\mu_n, \nu_{1,n}, \ldots, \nu_{m,n}\} > \frac{1}{2}
\]  

Then \((x_n)_{n\in\mathbb{N}} \) converges weakly to a solution to (12), for every \( n \in \{1, \ldots, m\} \). \((v_n)_{n\in\mathbb{N}} \) converges weakly to some \( \bar{v} \in \mathcal{G} \), and \((\tau, \sigma) \) is a solution to (13).

In the special case when \( U_n \equiv \iota(0) \), for every \( i \in \{1, \ldots, m\} \), \( \lambda_n \equiv 1 \). For every \( n \in \mathbb{N} \), let \((\nu_{i,n})_{n\in\mathbb{N}} \) be diagonal matrices, and no error term is taken into account. Algorithm (21) then, for every \( n \in \mathbb{N} \), \( \lambda_n \equiv 1 \). For every \( n \in \mathbb{N} \), \((\nu_{i,n})_{n\in\mathbb{N}} \) are diagonal matrices, \( h = 0 \), and \((\forall i \in \{1, \ldots, m\}) \lambda_i = \iota(0) \) appears also to be closely related to the adaptive method in [19].

4.3. A second class of primal-dual algorithms

Let \( \alpha \in [0, +\infty[ \), let \((U_n)_{n\in\mathbb{N}} \) be a sequence in \( \mathcal{P}_\alpha(\mathcal{H}) \) such that \((\forall n \in \mathbb{N}) \quad U_{n+1} \supseteq U_n \). For every \( i \in \{1, \ldots, m\} \), let \((U_{i,n})_{n\in\mathbb{N}} \) be a sequence in \( \mathcal{P}_\alpha(\mathcal{G}_i) \) such that \((\forall n \in \mathbb{N}) \quad U_{i,n+1} \supseteq U_{i,n} \). A second possible choice for \((V_n)_{n\in\mathbb{N}} \) is given by the following diagonal form:

\[
(V_n)_{n\in\mathbb{N}} \text{ is given by } (\forall n \in \mathbb{N}) \quad V_n^{-1} : (x, v) \mapsto (U_n^{-1} x, (\bar{U}_n^{-1} - LU_n L^*) v)
\]

where \( \bar{U}_n \) is given by (18).

The following result can then be deduced from Theorem 3.1. Its proof is skipped due to the lack of space.

**Proposition 4.3** Let \( x_0 \in \mathcal{H} \), and let \((c_n)_{n\in\mathbb{N}} \) be an absolutely summable sequence in \( \mathcal{H} \). For every \( i \in \{1, \ldots, m\} \), let \( v_{i,0} \in \mathcal{G}_i \), let \((b_{i,n})_{n\in\mathbb{N}} \) and \((d_{i,n})_{n\in\mathbb{N}} \) be absolutely summable sequences in \( \mathcal{G}_i \). For every \( n \in \mathbb{N} \), let \( \mu_n \in [0, +\infty[ \) be a Lipschitz constant of \( U_i^{1/2} \circ \nabla h \circ U_i^{1/2} \) and, for every \( i \in \{1, \ldots, m\} \), let \( \nu_{i,n} \in [0, +\infty[ \) be a Lipschitz constant of \( U_i^{1/2} \circ \nabla \ell^*_i \circ U_i^{1/2} \). Let \( (\lambda_n)_{n\in\mathbb{N}} \) be a sequence in \( [0, 1] \) such that \( \inf_{n\in\mathbb{N}} \lambda_n > 0 \). For every \( n \in \mathbb{N} \), set

\[
\zeta_n = 1 - \frac{\sup_{i=1}^m \left\| U_{i,n} L_i \right\|}{\sqrt[1/2]{\lambda_n}}
\]

and suppose that

\[
\inf_{n\in\mathbb{N}} \left( \zeta_n \mu_n, \nu_{1,n}, \ldots, \nu_{m,n} \right) > \frac{1}{2}
\]  

(24)
Two observed images \( w_2 \) (SNR = 5.87 dB) (b), blurred image \( w_2 \) (SNR = 16.63 dB) (c), and restored image \( \hat{x} \) (SNR = 21.61 dB) (d).

\[
\begin{align*}
\text{Set} & \quad \text{For } n = 0, 1, \ldots , \quad s_n = x_n - U_n (\nabla h(x_n) + c_n - z) \\
& \quad \text{For } i = 1, \ldots , n, \quad q_{i,n} = \text{prox}_{\frac{\mu}{2}} (v_{i,n} + U_i x_n - \nabla \ell_i (v_{i,n})) \\
& \quad \quad -d_{i,n} - r_i \big) + b_{i,n} \\
& \quad \quad = v_{i,n+1} + \lambda_n (q_{i,n} - v_{i,n}) \\
& \quad \quad p_n = s_n - U_n \sum_{i=1}^n \ell_i q_{i,n} \\
& \quad x_{n+1} = x_n + \lambda_n (p_n - x_n). \\
\end{align*}
\]

Assume that \( f = 0 \). Then \((x_n)_{n \in \mathbb{N}}\) converges weakly to a solution to (12), for every \( i \in \{1, \ldots , m\} \), \((v_{i,n})_{n \in \mathbb{N}}\) converges weakly to some \( \Psi \in G_i \), and \((\Psi_1, \ldots , \Psi_m)\) is a solution to (13).

The algorithm proposed in [23, 8] is a special case of the previous one, in the absence of errors, when \( m = 1 \), and \( G_1 \) is finite dimensional spaces, \( \ell_1 = \ell_{(0)} \), \( U_n \equiv \tau \text{Id} \) with \( \tau \in [0, +\infty] \), \( U_{1,n} \equiv \sigma \text{Id} \) with \( \sigma \in [0, +\infty] \), and no relaxation \((\lambda_n \equiv 1)\) or a constant one \((\lambda_n \equiv \kappa < 1)\) is performed.

5. APPLICATION TO IMAGE RESTORATION

We illustrate the flexibility of the proposed primal-dual algorithms on an image recovery example. Two observed images \( w_1 \) and \( w_2 \) of the same scene \( \Psi \in \mathbb{R}^N \) \((N = 256^2)\) available (see Fig. 1(a)-(e)). The first one is corrupted with a noise with a variance \( \theta_1^2 = 576 \), while the second one has been degraded by a linear operator \( H \in \mathbb{R}^{N \times N} \) \((7 \times 7 \text{ uniform blur})\) and a noise with variance \( \theta_2^2 = 25 \). The noise components are mutually statistically independent, additive, zero-mean, white, and Gaussian distributed. Note that this kind of multivariate restoration problem is encountered in some push-broom satellite imaging systems.

An estimate \( \tilde{x} \) of \( \Psi \) is computed as a solution to (12) where \( m = 2, z = 0, r_1 = 0, r_2 = 0 \).

\[
\begin{align*}
\epsilon_1 = \ell_{(0)} \\
g_1 = \ell_{[0, 255] \times} \quad g_2 = \kappa \| \cdot \|_{1,2}, \\
& f = 0, \quad \ell_1 = \ell_2 = \ell_{(0)} \quad \text{(28)}
\end{align*}
\]

Fig. 2. Normalized norm of the error on the iterate vs computation time (in seconds) for Experiment 1 (blue, dash dot line) and Experiment 2 (red, continuous line).

while the second one has been degraded by a linear operator \( H \in \mathbb{R}^{N \times N} \) \((7 \times 7 \text{ uniform blur})\) and a noise with variance \( \theta_2^2 = 25 \). The noise components are mutually statistically independent, additive, zero-mean, white, and Gaussian distributed. Note that this kind of multivariate restoration problem is encountered in some push-broom satellite imaging systems.

An estimate \( \tilde{x} \) of \( \Psi \) is computed as a solution to (12) where \( m = 2, z = 0, r_1 = 0, r_2 = 0 \).

\[
\begin{align*}
\epsilon_1 = \ell_{(0)} \\
g_1 = \ell_{[0, 255] \times} \quad g_2 = \kappa \| \cdot \|_{1,2}, \\
& f = 0, \quad \ell_1 = \ell_2 = \ell_{(0)} \quad \text{(28)}
\end{align*}
\]

The restored image is displayed in Fig. 1(d). Fig. 2 shows the convergence profile of the algorithm. We plot the evolution of the normalized Euclidean distance (in log scale) between the iterates and \( \tilde{x} \) in terms of computational time (Matlab R2011b codes running on a single-core Intel i7-2620M CPU@2.7 GHz with 8 GB of RAM). An approximation of \( \tilde{x} \) obtained after 5000 iterations is used. This result illustrates the fact that an appropriate choice of the metric may be beneficial in terms of speed of convergence.
6. REFERENCES


