



HAL
open science

A generic first-order primal-dual method for convex optimization involving Lipschitzian, proximable and linear composite terms

Laurent Condat

► **To cite this version:**

Laurent Condat. A generic first-order primal-dual method for convex optimization involving Lipschitzian, proximable and linear composite terms. 2011. hal-00609728v2

HAL Id: hal-00609728

<https://hal.science/hal-00609728v2>

Preprint submitted on 22 Jul 2011 (v2), last revised 4 Aug 2013 (v5)

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.

A generic first-order primal-dual method for convex optimization involving Lipschitzian, proximable and linear composite terms

Laurent Condat*

July 22, 2011

Abstract

We propose a new first-order algorithm for solving jointly the primal and dual formulations of large-scale convex minimization problems involving the sum of a smooth function with Lipschitzian gradient, nonsmooth proximable functions and linear composite functions. The gradient and the linear operators present in the formulation are called explicitly, while the other functions are processed individually via their proximity operators. This work brings together and notably extends several classical splitting schemes like the forward-backward [1] and Douglas-Rachford [2] methods, as well as recent primal-dual methods designed for linear composite terms [3–5].

1 Introduction – Problem Formulation

Let \mathcal{X} and \mathcal{Y} be two real Hilbert spaces with inner products and norms denoted respectively by $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$. We denote by $\Gamma_0(\mathcal{H})$ the class of proper, lower semi-continuous, convex functions from a Hilbert space \mathcal{H} to $\mathbb{R} \cup \{+\infty\}$ and by J^* the Fenchel-Rockafellar conjugate of $J \in \Gamma_0(\mathcal{H})$, defined by $J^*(s) = \sup_{s' \in \mathcal{H}} \langle s, s' \rangle - J(s')$. We define Moreau's proximity operator of $J \in \Gamma_0(\mathcal{H})$ by $\text{prox}_J(s) = \text{argmin}_{s' \in \mathcal{H}} J(s') + \frac{1}{2} \|s - s'\|^2$. We also define the subdifferential of $J \in \Gamma_0(\mathcal{H})$ as the set-valued operator $\partial J : s \mapsto \{s' \in \mathcal{H} : \forall s'' \in \mathcal{H}, \langle s'' - s, s' \rangle + J(s) \leq J(s'')\}$. If J is differentiable at s , then $\partial J(s) = \{\nabla J(s)\}$. For background in convex analysis, we refer the readers to textbooks, e.g. [6].

A wide range of problems in areas such as partial differential equations, mechanics, economics, signal and image processing, or operations research, can be reduced to solving minimization problems. In this article, we consider the generic convex optimization problem

$$\text{Find } \hat{x} \in \text{argmin}_{x \in \mathcal{X}} F(x) + G(x) + H(Lx), \quad (1)$$

where

*L. Condat is with the Image Team of the GREYC laboratory, a joint CNRS-UCBN-ENSICAEN research unit in Caen, France. Contact: see <http://www.greyc.ensicaen.fr/~lcondat/>.

- $F : \mathcal{X} \rightarrow \mathbb{R}$ is convex, Fréchet-differentiable on \mathcal{X} and its gradient ∇F is β -Lipschitz continuous for some $\beta \in [0, +\infty]$; that is,

$$\|\nabla F(x) - \nabla F(x')\| \leq \beta \|x - x'\| \text{ for every } (x, x') \in \mathcal{X}^2. \quad (2)$$

- $G \in \Gamma_0(\mathcal{X})$ and $H \in \Gamma_0(\mathcal{Y})$ are “simple”, in the sense that their proximity operators have a closed-form representation, or at least can be solved efficiently with high precision. We recall that $\text{prox}_{\sigma H^*}$ can be easily computed from $\text{prox}_{H/\sigma}$ if necessary, thanks to Moreau’s identity $\text{prox}_{\sigma H^*}(y) = y - \sigma \text{prox}_{H/\sigma}(y/\sigma)$.

- $L : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator with adjoint denoted by L^* and induced norm

$$\|L\| = \sup \{\|Lx\| : x \in \mathcal{X}, \|x\| \leq 1\} < +\infty. \quad (3)$$

- The set of minimizers of (1) is assumed to be nonempty.

Under suitable qualification conditions, the corresponding *dual* formulation of the *primal* problem (1) is [6]

$$\text{Find } \hat{y} \in \underset{y \in \mathcal{Y}}{\text{argmin}} (F + G)^*(-L^*y) + H^*(y) \quad (4)$$

$$\Leftrightarrow \text{Find } \hat{y} \in \underset{y \in \mathcal{Y}}{\text{argmin}} \inf_{x' \in \mathcal{X}} F^*(-L^*y - x') + G^*(x') + H^*(y). \quad (5)$$

The primal and dual variables \hat{x} and \hat{y} are solutions to the primal-dual formulation of (1), which is the saddle-point problem [7]

$$\text{Find } (\hat{x}, \hat{y}) \in \underset{x \in \mathcal{X}}{\text{argmin}} \max_{y \in \mathcal{Y}} F(x) + G(x) - H^*(y) + \langle Lx, y \rangle. \quad (6)$$

Moreover, the classical Kuhn-Tucker theory [8] asserts that if $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$ are solutions to the variational inclusions

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G(\hat{x}) + L^* \hat{y} + \nabla F(\hat{x}) \\ -L\hat{x} + \partial H^*(\hat{y}) \end{pmatrix}, \quad (7)$$

then \hat{x} and \hat{y} are solutions to (1) and (4), respectively, as well as (6); see [9, eqns (4.10)–(4.18)] for a proof. In the following, we assume that the solution set of the inclusions (7) is nonempty. This is the case if the solution set of (1) is nonempty and

$$0 \in \text{sri}\{Lx - y : x \in \text{dom}(G), y \in \text{dom}(H)\}, \quad (8)$$

where $\text{dom}(J) = \{s \in \mathcal{H} : J(s) < +\infty\}$ and $\text{sri}(\Omega)$ is the strong relative interior of the convex subset Ω of \mathcal{Y} ; i.e., the set of points $y \in \mathcal{Y}$ such that the cone generated by $-y + \Omega$ is a closed vector subspace of \mathcal{Y} ; see [9, Remark 4.3] for a proof.

The advantage in solving (7) instead of the inclusion $0 \in \nabla F(\hat{x}) + \partial G(\hat{x}) + L^* \partial H(L\hat{x})$ associated to (1) is twofold: 1) the composite function $H \circ L$ has been split; 2) we obtain not only the primal solution \hat{x} but also the dual solution \hat{y} , and the proposed algorithm actually uses their intertwined properties to update the primal and dual variables alternatively and efficiently.

In this article, we propose a new algorithm to solve (7). Compared to the method presented in [9], which can handle this inclusion problem, our approach relies on a different splitting of (7). In brief, we show that the algorithm we propose has the same structure as the classical forward-backward splitting method [1], when expressed in terms of nonexpansive operators in $\mathcal{X} \times \mathcal{Y}$ equipped with an inner product which is *not* the one of the direct sum $\mathcal{X} \oplus \mathcal{Y}$.

We may observe that there is "room" in the dual inclusion of (7) for an additional term $\nabla K^*(y_n)$, which yields a more symmetric treatment of the primal and dual problems. The obtained variational inclusions characterize the following primal problem, which includes an infimal convolution [9]:

$$\text{Find } \hat{x} \in \operatorname{argmin}_{x \in \mathcal{X}} \inf_{y' \in \mathcal{Y}} F(x) + G(x) + H(Lx - y') + K(y') \quad (9)$$

with the same hypotheses as earlier and an additional function $K \in \Gamma_0(\mathcal{Y})$ such that K^* is Fréchet-differentiable on \mathcal{Y} with β' -Lipschitz gradient for some $\beta' > 0$; this is equivalent for K to be $1/\beta'$ -strongly convex; that is, $K(y) - \frac{1}{2\beta'}\|y\|^2$ is convex. We leave the study of this more general framework for future work.

The article is organized as follows. In Sect. 2, we present a new algorithm to solve (7) and we show in Sect. 3 that it converges under reasonable assumptions. Then, in Sect. 4, we present parallel variants of the algorithm adapted to minimization problems with more than two proximable terms.

2 Proposed Algorithm

The proposed algorithm is as follows:

Algorithm 1.

Choose the parameters $\tau > 0$, $\sigma > 0$, $\rho_n \in]0, 1]$ and the initial estimate $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, then iterate, for every $n \geq 0$,

$$\left| \begin{array}{l} 1. \quad \tilde{x}_{n+1} = \operatorname{prox}_{\tau G}(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau L^* y_n) + e_{G,n} \\ 2. \quad \tilde{y}_{n+1} = \operatorname{prox}_{\sigma H^*}(y_n + \sigma L(2\tilde{x}_{n+1} - x_n)) + e_{H,n} \\ 3. \quad (x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n) \end{array} \right. \quad (10)$$

where the error terms $e_{F,n}$, $e_{G,n}$, $e_{H,n}$ model the inexact computation of the operators ∇F , $\operatorname{prox}_{\tau G}$, $\operatorname{prox}_{\sigma H^*}$, respectively.

2.1 Relationship to Existing Optimization Methods

The proposed algorithm is able to solve the general problem (1) iteratively without inner loops and by activating separately $\operatorname{prox}_{\tau G}$, $\operatorname{prox}_{\sigma H^*}$, L and L^* without any other implicit (inverse) operator. In particular, no inverse operator of the form $(I + \alpha L^* L)^{-1}$ is required, where I denotes the identity operator. To our knowledge, the only existing method of the literature having this feature is the very recent proposal of Combettes and Pesquet [9]. Their algorithm requires two calls to ∇F , L , L^*

per iteration, against only one with our algorithm. Whether this is a sign of faster convergence of our algorithm remains to be shown in practical applications.

Some authors have studied the use of nested algorithms to solve (1) (in the case $L = I$) for practical imaging problems [10–12]. This approach consists in embedding an iterative algorithm as an inner loop inside each iteration of another iterative method. However, the method is applicable if the number of inner iterations is kept small, a scenario where convergence is not proven. Note that Algorithm 1 bears similarities with a nested algorithm in which only one subiteration of the forward-backward method, to approximate prox_{F+G} , would be embedded in the primal-dual algorithm of [3], or the Douglas-Rachford method if $L = I$. This method has been found to give good results for an imaging application in [11], where it was used empirically without convergence proof.

We also note that the proposed algorithm may turn out to be a particular case of the general framework proposed in [13], although this is not clear at first glance and remains to be studied.

We now show that in some particular cases, our algorithm reverts to classical splitting methods of the literature.

2.1.1 Case $F = 0$

If the smooth term F is absent of the problem, the proposed algorithm exactly reverts to the primal-dual algorithm presented in [3] in the finite-dimensional setting and $\rho_n = 1$, and also proposed in another form in [4, 5]. The convergence of this algorithm has been proved in [14] for the exact computation case (no error terms) with $\tau\sigma\|L\|^2 < 1$ and constant parameter $\rho_n = \rho \in]0, 2[$ (note the smaller range $]0, 1[$ accessible if $F \neq 0$). Accelerations of the convergence in the sense of the partial primal-dual gap, based on variable parameters τ_n, σ_n , are discussed in [3].

When $F = 0$, the primal-dual method of [15] and the method in [16] can be used as well. They yield algorithms different than ours.

Note that these methods cannot be used to solve the problem (1) if $F \neq 0$, because they involve the proximity operator of $F + G$, which is usually intractable. Even in the simple case where G is the quadratic function $\frac{\lambda}{2}\|Mx - b\|^2$ for a bounded linear operator M , the proximity operator of G requires to apply the operator $(I + \lambda M^* M)^{-1}$, which may be feasible (e.g. using FFTs for some inverse problems in imaging) but complicated to implement (especially if particular care is paid to the treatment at the boundaries for multi-dimensional problems) and slow. By contrast, considering $\frac{\lambda}{2}\|Mx - b\|^2$ as the function F with Lipschitz continuous gradient in our framework yields an algorithm with simple calls to M and M^* . An alternative consists in incorporating $\frac{\lambda}{2}\|Mx - b\|^2$ into the term $H(Lx)$ using a product space technique, see [3, eq. (74)]. Such parallel strategies are discussed in Sect. 4.

If $F = 0$ and $L = I$, our algorithm reverts to the classical Douglas-Rachford splitting method [2], as discussed in [3]. However, this equivalence requires to set $\sigma = 1/\tau$, which goes beyond the application conditions of our Theorem 1. If $\sigma = 1/\tau$, weak convergence of both the primal and dual variables is not guaranteed.

2.1.2 Case $H(Lx) = 0$

The degenerate case $H(Lx) = 0$ is not so interesting, because the primal and dual problems are then uncoupled. However, if we set $L = 0$ or $H = 0$ and we focus on the computations of the primal variable x only, we obtain the iteration

$$x_{n+1} = \rho_n \text{prox}_{\tau_n G}(x_n - \tau_n(\nabla F(x_n) + e_{F,n})) + e_{G,n} + (1 - \rho_n)x_n, \quad (11)$$

which is exactly the classical forward-backward splitting method [1]. It is known to converge if the error terms are absolutely summable and for every $n \in \mathbb{N}$, $\rho_n \in [\epsilon, 1]$ for some $\epsilon > 0$ and $\tau_n \in [\delta, \frac{2}{\beta} - \delta]$ for some $\delta > 0$.

We note that if we set $G = 0$ and $L = I$ in our framework, we obtain another algorithm to minimize the sum of a function F with Lipschitz continuous gradient and a proximable function H , along with the dual problem.

3 Convergence Proof

We now state the main result of this article:

Theorem 1. *Let us assume that, in Algorithm 1, $\rho_n \in [\epsilon, 1]$ for every $n \in \mathbb{N}$ and some $\epsilon \in]0, 1]$, and that the parameters $\tau > 0$ and $\sigma > 0$ are such that*

$$\boxed{\frac{1}{\tau} - \sigma \|L\|^2 > \frac{\beta}{2}}, \quad (12)$$

where the Lipschitz constant β is defined in (2). We also assume that the errors are absolutely summable: $\sum_{n \in \mathbb{N}} \|e_{F,n}\| < +\infty$, $\sum_{n \in \mathbb{N}} \|e_{G,n}\| < +\infty$, $\sum_{n \in \mathbb{N}} \|e_{H,n}\| < +\infty$. Then, the sequence (x_n, y_n) computed by Algorithm 1 converges weakly to a pair (\hat{x}, \hat{y}) solution to (7).

The sequel of this section is devoted to the proof of this result.

We first define the vector space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ and the bounded linear operator on \mathcal{Z}

$$P: \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\tau}I & -L^* \\ -L & \frac{1}{\sigma}I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (13)$$

We now define the inner product $\langle \cdot, \cdot \rangle_I$ and norm $\|\cdot\|_I = \langle \cdot, \cdot \rangle_I^{1/2}$ in \mathcal{Z} as

$$\langle z, z' \rangle_I = \langle x, x' \rangle + \langle y, y' \rangle, \text{ for every } z = (x, y), z' = (x', y') \in \mathcal{Z}. \quad (14)$$

We denote by \mathcal{Z}_I the Hilbert space defined by \mathcal{Z} equipped with this inner product. Then, P is self-adjoint and, from (12), positive-definite in \mathcal{Z}_I . Hence, we can define another inner product $\langle \cdot, \cdot \rangle_P$ and norm $\|\cdot\|_P = \langle \cdot, \cdot \rangle_P^{1/2}$ in \mathcal{Z} as

$$\langle z, z' \rangle_P = \langle z, Pz' \rangle_I, \text{ for every } (z, z') \in \mathcal{Z}^2. \quad (15)$$

We denote by \mathcal{Z}_P the Hilbert space defined by \mathcal{Z} equipped with this inner product. The crux of the proof of Theorem 1 will be to prove the weak convergence in \mathcal{Z}_P and

not in \mathcal{X}_I ; that is, $\langle z_n - \hat{z}, Pz' \rangle_I \rightarrow 0$ as $n \rightarrow +\infty$, for every $z' \in \mathcal{Z}$, where $z_n = (x_n, y_n)$ is generated by Algorithm 1 and $\hat{z} = (\hat{x}, \hat{y})$. Obviously, weak convergence in \mathcal{X}_p is equivalent to weak convergence in \mathcal{X}_I .

We now recall some classical definitions and properties of monotone operator theory and convex optimization. Let \mathcal{H} be a real Hilbert space and $T: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a set-valued operator. We denote by $\text{ran}(T) = \{v \in \mathcal{H} : \exists u \in \mathcal{H}, v \in Tu\}$ the range of T , by $\text{gra}(T) = \{(u, v) \in \mathcal{H}^2 : v \in Tu\}$ its graph, and by T^{-1} its inverse; that is, the set-valued operator with graph $\{(v, u) \in \mathcal{H}^2 : v \in Tu\}$. T is said monotone if $\forall (u, u') \in \mathcal{H}^2, \forall (v, v') \in Tu \times Tu', \langle u - u', v - v' \rangle \geq 0$ and maximal monotone if there exists no monotone operator T' such that $\text{gra}(T) \subset \text{gra}(T') \neq \text{gra}(T)$. $T: \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive if it is 1-Lipschitz continuous, see (2), and firmly nonexpansive if $2T - I$ is nonexpansive. The resolvent $(I + T)^{-1}$ of a maximal monotone operator is single-valued on \mathcal{H} and firmly nonexpansive. The subdifferential ∂J of $J \in \Gamma_0(\mathcal{H})$ is maximal monotone and $(I + \partial J)^{-1} = \text{prox}_J$.

Lemma 1. (Convergence of the forward-backward iterative scheme) [17, Corollary 6.5]. Let $T_1: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ and $T_2: \mathcal{H} \rightarrow \mathcal{H}$ be maximal monotone operators such that αT_2 is firmly nonexpansive for some $\alpha > \frac{1}{2}$ and $0 \in \text{ran}(T_1 + T_2)$. We consider the forward-backward algorithm: Fix $s_0 \in \mathcal{H}$ and, for every $n \in \mathbb{N}$, set

$$s_{n+1} = \rho_n \left((I + T_1)^{-1} (s_n - T_2(s_n) - e_{2,n}) + e_{1,n} \right) + (1 - \rho_n) s_n, \quad (16)$$

where $\rho_n \in [\epsilon, 1]$ for some $\epsilon \in]0, 1]$ and $(e_{1,n}, e_{2,n}) \in \mathcal{H}^2$ are such that $\sum_{n \in \mathbb{N}} \|e_{1,n}\| < +\infty$ and $\sum_{n \in \mathbb{N}} \|e_{2,n}\| < +\infty$. Then s_n converges weakly to $\hat{s} \in \mathcal{H}$ such that $0 \in (T_1 + T_2)(\hat{s})$.

Lemma 2. (Baillon-Haddad theorem) [18, Corollaire 10]. Let $J: \mathcal{H} \rightarrow \mathbb{R}$ be convex, Fréchet-differentiable on \mathcal{H} and such that $\alpha \nabla J$ is nonexpansive for some $\alpha > 0$. Then, $\alpha \nabla J$ is firmly nonexpansive.

We are now equipped to prove Theorem 1. Let us first consider the error-free case $e_{F,n} = e_{G,n} = e_{H,n} = 0$. We introduce the notation $z = (x, y)$. Then, for every n , the following inclusion is satisfied:

$$-\underbrace{\begin{pmatrix} \nabla F(x_n) \\ 0 \end{pmatrix}}_{B(z_n)} \in \underbrace{\begin{pmatrix} \partial G(\tilde{x}_{n+1}) + L^* \tilde{y}_{n+1} \\ -L \tilde{x}_{n+1} + \partial H^*(\tilde{y}_{n+1}) \end{pmatrix}}_{A(\tilde{z}_{n+1})} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & -L^* \\ -L & \frac{1}{\sigma} I \end{pmatrix}}_P \underbrace{\begin{pmatrix} \tilde{x}_{n+1} - x_n \\ \tilde{y}_{n+1} - y_n \end{pmatrix}}_{(\tilde{z}_{n+1} - z_n)}, \quad (17)$$

or equivalently

$$\tilde{z}_{n+1} = (I + P^{-1} \circ A)^{-1} \circ (I - P^{-1} \circ B)(z_n). \quad (18)$$

Considering now the over-relaxation step and the error terms, we obtain

$$z_{n+1} = \rho_n \left((I + P^{-1} \circ A)^{-1} (z_n - P^{-1} \circ B(z_n) - e_{2,n}) + e_{1,n} \right) + (1 - \rho_n) z_n. \quad (19)$$

with $e_{1,n} = (e_{G,n}, e_{H,n})$ and $e_{2,n} = P^{-1}(e_{F,n}, -2Le_{G,n})$. (20)

It is now obvious that the proposed algorithm has the structure of the classical forward-backward splitting scheme. The idea of expressing an iteration of Algorithm 1 under the form (17) was inspired by the work [14], where the primal-dual method of [3] (the case $F = 0$ in our setting) was reformulated as a proximal point algorithm.

Thus, to apply Lemma 1 with $\mathcal{H} = \mathcal{Z}_P$, $T_1 = P^{-1} \circ A$ and $T_2 = P^{-1} \circ B$, it remains to show that

(i) T_1 and T_2 are maximal monotone in \mathcal{Z}_P . This is an immediate consequence of the maximal monotony of A and B in \mathcal{Z}_I .

(ii) $0 \in \text{ran}(T_1 + T_2)$. This is an immediate consequence of the hypothesis $0 \in \text{ran}(A + B)$.

(iii) The error sequences $e_{1,n}$ and $e_{2,n}$ in (20) are absolutely summable for the $\|\cdot\|_P$ -norm. This is immediate, because $e_{F,n}$, $e_{G,n}$, $e_{H,n}$ are absolutely summable and P , P^{-1} , L are bounded.

(iv) αT_2 is firmly nonexpansive in \mathcal{Z}_P for some $\alpha > \frac{1}{2}$. Let us prove this property. For every $z = (x, y)$, $z' = (x', y') \in \mathcal{Z}$, we have

$$\|T_2(z) - T_2(z')\|_P^2 = \langle P^{-1} \circ B(z) - P^{-1} \circ B(z'), B(z) - B(z') \rangle_I \quad (21)$$

$$= \left\langle \frac{1}{\sigma} \left(\frac{1}{\sigma\tau} I - L^* L \right)^{-1} (\nabla F(x) - \nabla F(x')), \nabla F(x) - \nabla F(x') \right\rangle \quad (22)$$

$$\leq \frac{1}{\sigma} \left(\frac{1}{\sigma\tau} - \|L\|^2 \right)^{-1} \|\nabla F(x) - \nabla F(x')\|^2 \quad (23)$$

$$\leq \frac{\beta^2}{\sigma} \left(\frac{1}{\sigma\tau} - \|L\|^2 \right)^{-1} \|x - x'\|^2 = \frac{\beta}{\alpha} \|x - x'\|^2 \quad (24)$$

where we define $\alpha = \frac{1}{\beta} \left(\frac{1}{\tau} - \sigma \|L\|^2 \right)$ and the linear operator $Q = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$ of \mathcal{Z} .

$P - \beta\alpha Q$ is positive semi-definite, so that

$$\beta\alpha \|x - x'\|^2 = \beta\alpha \langle (z - z'), Q(z - z') \rangle_I \leq \langle (z - z'), P(z - z') \rangle_I = \|z - z'\|_P^2. \quad (25)$$

Putting together (24) and (25), we get

$$\alpha \|T_2(z) - T_2(z')\|_P \leq \|z - z'\|_P, \quad (26)$$

so that αT_2 is nonexpansive in \mathcal{Z}_P , with $\alpha > \frac{1}{2}$ by hypothesis (12). Finally, let us define on \mathcal{Z} the function $J : (x, y) \mapsto F(x)$. Then, in \mathcal{Z}_P , $\nabla J = T_2$. Therefore, from Lemma 2, αT_2 is firmly nonexpansive in \mathcal{Z}_P . \square

4 Variants

We first remark that Algorithm 1 is not symmetric with respect to x and y , since the computation of \tilde{y}_{n+1} uses the over-relaxed version $2\tilde{x}_{n+1} - x_n$ of x_{n+1} , while the computation of \tilde{x}_{n+1} uses y_n . There is no reason a priori to privilege Algorithm 1 over the following Algorithm 2, where the roles of x and y are switched. The convergence conditions are identical for the two algorithms.

Algorithm 2.

Choose the parameters $\sigma > 0$, $\tau > 0$, $\rho_n \in]0, 1]$ and the initial estimate $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, then iterate, for every $n \geq 0$,

$$\begin{cases} 1. & \tilde{y}_{n+1} = \text{prox}_{\sigma H^*}(y_n + \sigma Lx_n) + e_{H,n} \\ 2. & \tilde{x}_{n+1} = \text{prox}_{\tau G}(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau L^*(2\tilde{y}_{n+1} - y_n)) + e_{G,n} \\ 3. & (x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n) \end{cases} \quad (27)$$

We now investigate the problem with $m \geq 2$ composite functions

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\text{argmin}} F(x) + G(x) + \sum_{i=1}^m H_i(L_i x). \quad (28)$$

with the same assumptions on F and G as in the problem (1), m functions $H_i \in \Gamma_0(\mathcal{Y}_i)$ defined on real Hilbert spaces \mathcal{Y}_i and m bounded linear functions $L_i : \mathcal{X} \rightarrow \mathcal{Y}_i$. The set of minimizers of (28) is assumed nonempty. At the same time, we consider the dual problem

$$\text{Find } (\hat{y}_1, \dots, \hat{y}_m) \in \underset{y_1 \in \mathcal{Y}_1, \dots, y_m \in \mathcal{Y}_m}{\text{argmin}} (F + G)^*(-\sum_{i=1}^m L_i^* y_i) + \sum_{i=1}^m H_i^*(y_i) \quad (29)$$

$$\Leftrightarrow \text{Find } (\hat{y}_1, \dots, \hat{y}_m) \in \underset{y_1 \in \mathcal{Y}_1, \dots, y_m \in \mathcal{Y}_m}{\text{argmin}} \inf_{x' \in \mathcal{X}} F^*(-\sum_{i=1}^m L_i^* y_i - x') + G^*(x') + \sum_{i=1}^m H_i^*(y_i). \quad (30)$$

Although these primal and dual problems are more general than (1) and (4), respectively, they can be recast as particular cases of them using product spaces. The remainder of this section is devoted to the study of the parallel variants of Algorithms 1 and 2 that we obtain along this line.

We introduce the bold notation $\mathbf{y} = (y_1, \dots, y_m)$ for an element of the Hilbert space $\mathcal{Y} = \mathcal{Y}_1 \oplus \dots \oplus \mathcal{Y}_m$. We define the function $\mathbf{H} \in \Gamma_0(\mathcal{Y})$ by $\mathbf{H}(\mathbf{y}) = \sum_{i=1}^m H_i(y_i)$ and the linear function $\mathbf{L} : \mathcal{X} \rightarrow \mathcal{Y}$ by $\mathbf{L}x = (L_1 x, \dots, L_m x)$.

We have the following properties:

$$\mathbf{H}^*(\mathbf{y}) = \sum_{i=1}^m H_i^*(y_i), \quad (31)$$

$$\mathbf{L}^* \mathbf{y} = \sum_{i=1}^m L_i^* y_i, \quad (32)$$

$$\text{prox}_{\sigma \mathbf{H}^*}(\mathbf{y}) = (\text{prox}_{\sigma H_1}(y_1), \dots, \text{prox}_{\sigma H_m}(y_m)). \quad (33)$$

Thus, we can rewrite (28) and (29) as

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\text{argmin}} F(x) + G(x) + \mathbf{H}(\mathbf{L}x), \quad (34)$$

$$\text{Find } \hat{\mathbf{y}} \in \underset{\mathbf{y} \in \mathcal{Y}}{\text{argmin}} (F + G)^*(-\mathbf{L}^* \mathbf{y}) + \mathbf{H}^*(\mathbf{y}), \quad (35)$$

which exactly take the form of (1) and (4). Hence, we can rewrite Algorithms 1 and 2 by doing the appropriate substitutions and we obtain the two following algorithms,

respectively:

Algorithm 3.

Choose the parameters $\tau > 0, \sigma > 0, \rho_n \in]0, 1]$ and the initial estimate $(x_0, y_{1,0}, \dots, y_{m,0}) \in \mathcal{X} \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$, then iterate, for every $n \geq 0$,

$$\left\{ \begin{array}{l} 1. \quad \tilde{x}_{n+1} = \text{prox}_{\tau G}(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau \sum_{i=1}^m L_i^* y_{i,n}) + e_{G,n} \\ 2. \quad x_{n+1} = \rho_n \tilde{x}_{n+1} + (1 - \rho_n)x_n \\ 3. \quad \forall i = 1, \dots, m, \tilde{y}_{i,n+1} = \text{prox}_{\sigma H_i^*}(y_{i,n} + \sigma L_i(2\tilde{x}_{n+1} - x_n)) + e_{H_i,n} \\ 4. \quad \forall i = 1, \dots, m, y_{i,n+1} = \rho_n \tilde{y}_{i,n+1} + (1 - \rho_n)y_{i,n} \end{array} \right. \quad (36)$$

Algorithm 4.

Choose the parameters $\sigma > 0, \tau > 0, \rho_n \in]0, 1]$ and the initial estimate $(x_0, y_{1,0}, \dots, y_{m,0}) \in \mathcal{X} \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$, then iterate, for every $n \geq 0$,

$$\left\{ \begin{array}{l} 1. \quad \forall i = 1, \dots, m, \tilde{y}_{i,n+1} = \text{prox}_{\sigma H_i^*}(y_{i,n} + \sigma L_i x_n) + e_{H_i,n} \\ 2. \quad \forall i = 1, \dots, m, y_{i,n+1} = \rho_n \tilde{y}_{i,n+1} + (1 - \rho_n)y_{i,n} \\ 3. \quad \tilde{x}_{n+1} = \text{prox}_{\tau G}(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau \sum_{i=1}^m L_i^*(2\tilde{y}_{n+1} - y_n)) + e_{G,n} \\ 4. \quad x_{n+1} = \rho_n \tilde{x}_{n+1} + (1 - \rho_n)x_n \end{array} \right. \quad (37)$$

Similarly, using the fact that $\|L\|^2 = \sum_{i=1}^m \|L_i\|^2$, we obtain the following spin-off of Theorem 1 as follows:

Theorem 2. *Let us assume that, in Algorithm 3 or 4, $\rho_n \in [\epsilon, 1]$ for every $n \in \mathbb{N}$ and some $\epsilon \in]0, 1]$, and that the parameters $\tau > 0$ and $\sigma > 0$ are such that*

$$\boxed{\frac{1}{\tau} - \sigma \sum_{i=1}^m \|L_i\|^2 > \frac{\beta}{2}}, \quad (38)$$

where the Lipschitz constant β is defined in (2). We also assume that the errors are absolutely summable: $\sum_{n \in \mathbb{N}} \|e_{F,n}\| < +\infty, \sum_{n \in \mathbb{N}} \|e_{G,n}\| < +\infty, \sum_{n \in \mathbb{N}} \|e_{H_i,n}\| < +\infty$ and that the following qualification condition is met:

$$(0, \dots, 0) \in \text{sri}\{L_i x - y_i : x \in \text{dom}(G), \forall i = 1, \dots, m, y_i \in \text{dom}(H_i)\}. \quad (39)$$

Then, the sequence (x_n, y_n) computed by Algorithm 3 or 4 converges weakly to a pair (\hat{x}, \hat{y}) solution to (28) and (29).

We note that if one of the function in (28) is $\frac{\lambda}{2}\|Mx - b\|^2$ for some linear operator M , assigning this term to F or to one of the $H_i(L_i x)$ yields different algorithms. Which one is the most efficient depends on the problem at hand and on the way the algorithms are implemented.

Acknowledgements

The author wants to thank Jalal Fadili for fruitful discussions that stimulated this work.

References

- [1] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [2] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
- [3] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [4] E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.
- [5] X. Zhang, M. Burger, and S. Osher, "A unified primal-dual algorithm framework based on Bregman iteration," *Journal of Scientific Computing*, pp. 1–27, 2010.
- [6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.
- [7] L. McLinden, "An extension of Fenchel's duality theorem to saddle functions and dual minimax problems," *Pacific J. Math.*, vol. 50, pp. 135–158, 1974.
- [8] R. T. Rockafellar, *Conjugate Duality and Optimization*. Philadelphia, PA: SIAM, 1974.
- [9] P. L. Combettes and J. C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum monotone operators, Tech. Rep. arXiv:1107.0081, 2011.
- [10] C. Chau, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 730–762, June 2009.
- [11] F.-X. Dupé, M. J. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Processing*, vol. 18, no. 2, pp. 310–321, 2009.
- [12] J. M. Fadili and G. Peyré, "Total variation projection with first order schemes," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 657–669, Mar. 2011.
- [13] R. D. C. Monteiro and B. F. Svaiter, "Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented Lagrangian method," Tech. Rep., 2010, preprint.
- [14] B. S. He and X. M. Yuan, "Convergence analysis of primal-dual algorithms for total variation image restoration," Tech. Rep., 2010, submitted to *SIAM Journal on Imaging Science*.

- [15] L. M. Briceño-Arias and P. L. Combettes, “A monotone+skew splitting model for composite monotone inclusions in duality,” *SIAM J. Optim.*, 2011, to appear.
- [16] G. Chen and M. Teboulle, “A proximal-based decomposition method for convex minimization problems,” *Math. Programming*, vol. 64, pp. 81–101, 1994.
- [17] P. Combettes, “Solving monotone inclusions via compositions of nonexpansive averaged operators,” *Optimization*, vol. 53, no. 5-6, pp. 475–504, 2004.
- [18] J.-B. Baillon and G. Haddad, “Quelques propriétés des opérateurs angle-bornés et n -cycliquement monotones,” *Israel J. Math.*, vol. 26, pp. 137–150, 1977.