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

Laurent Condat*

December 16, 2011

Abstract

We propose a new first-order splitting 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, a nonsmooth proximable function, and linear composite functions. This is a full splitting approach in the sense that the gradient and the linear operators involved are called explicitly without any inversion, while the nonsmooth functions are processed individually via their proximity operators. This work brings together and notably extends several classical splitting schemes like the forward-backward [1–3] and Douglas-Rachford [1,3–5] methods, as well as a recent primal-dual method designed for problems with linear composite terms [6].

Keywords. Convex and nonsmooth optimization, operator splitting, primal-dual algorithm, forward-backward method, Douglas-Rachford method, monotone inclusion, proximal method, Fenchel-Rockafellar duality, saddle-point problem, monotone operator.

1 Introduction

Nonlinear and nonsmooth convex optimization problems are widely present in many disciplines, including signal and image processing, operations research, machine learning, game theory, economics, and mechanics. In many cases, the problem consists in finding a minimizer of the sum of compositions of a convex function with a linear operator, where the involved functions may be differentiable or not, and the variables may live in very high-dimensional spaces. The first-order *proximal splitting* algorithms are dedicated to the resolution of such problems. They proceed by *splitting* [7] in that the original problem is decomposed into an iterative sequence of much easier subproblems, which involves the functions individually. A

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

smooth (or differentiable) function is processed by evaluation of its gradient operator, while a nonsmooth function is processed via its Moreau proximity operator [8]; that is why such methods are called *proximal*. The use of proximal splitting methods is spreading rapidly in signal and image processing, especially to solve large-scale problems formulated with sparsity-promoting ℓ_1 penalties; we refer to [9] for a recent overview with a list of relevant references, and to [10–12] for examples of applications. In return, this quest of practitioners for efficient minimization methods has caused a renewed interest among mathematicians around splitting methods in monotone and nonexpansive operator theory, as can be judged from the numerous recent contributions, e.g. [5,6,13–22]. The most classical operator splitting methods to minimize the sum of two convex functions are the forward-backward method, proposed in [1, 2] and further developed in [3, 23–26], and the Douglas-Rachford method [1, 3–5]. In this work, we propose a new proximal splitting method for the generic template problem of minimizing the sum of *three* convex terms: a smooth function, a proximable function, and the composition of a proximable function with a linear operator. The method brings together and encompasses as particular cases the forward-backward and Douglas-Rachford methods, as well as a recent method for minimizing the sum of a proximable function and a linear composite term [6]. It is fully split in that the gradient, proximity, and linear operators are called individually; in particular, there is no implicit operation like an inner loop or a call to the inverse of a linear operator. Moreover, our method is *primal-dual*, since it provides, in addition to the (primal) solution, a solution to the dual convex optimization problem, in the frame of the now classical Fenchel-Rockafellar duality theory [27–30]. Hence, an equivalent interpretation of the method is that it finds saddle-points of convex-concave functions with bilinear coupling, here the Lagrangian associated to the primal and dual problems [31,32].

The paper is organized as follows. In Sect. 2, we present the convex optimization problem we are interested in, along with the associated dual and primal-dual formulations, and the corresponding variational inclusion. In Sect. 3, we present the iterative algorithm and provide conditions on the parameters under which convergence to a primal and a dual solutions is ensured. We also discuss the links with other splitting methods of the literature. The Sect. 4 is devoted to the proofs of convergence. Finally, in Sect. 5, we present some parallelized variants of the algorithm adapted to minimization problems with more than two proximable terms.

2 Problem Formulation

First, we introduce some definitions and notations. Let $\mathcal{H}$ be a real Hilbert space, with its inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$. We denote by $\Gamma_0(\mathcal{H})$ the set of proper, lower semi-continuous, convex functions from $\mathcal{H}$ to $\mathbb{R}\cup\{+\infty\}$. Let J belong to $\Gamma_0(\mathcal{H})$. Its domain is $\mathrm{dom}(J) = \{s \in \mathcal{H} \,:\, J(s) < +\infty\}$, its Fenchel-Rockafellar conjugate $J^* \in \Gamma_0(\mathcal{H})$ is defined by $J^*(s) = \sup_{s'\in\mathcal{H}}\langle s, s'\rangle - J(s')$ and its proximity operator by $\mathrm{prox}_J(s) = \arg\min_{s'\in\mathcal{H}} J(s') + \frac{1}{2}\|s - s'\|^2$. We define the subdifferential of J as the set-valued operator $\partial J : u \in \mathcal{H} \mapsto \{v \in \mathcal{H} \,:\, \forall u' \in \mathcal{H}, \langle u' - u, v\rangle + J(u) \leq J(u')\}$. If J is differentiable at s, then $\partial J(s) = \{\nabla J(s)\}$. The strong relative interior of a convex subset Ω of $\mathcal{H}$ is denoted by $\operatorname{sri}(\Omega)$. Finally, we denote by I the identity operator. For background in convex analysis, we refer the readers to textbooks, like [30].

Throughout the paper, $\mathcal{X}$ and $\mathcal{Y}$ are two real Hilbert spaces. We consider the *primal* optimization problem

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\arg\min}\ F(x) + G(x) + H(Lx), \tag{1}$$

where

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

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

  We note that the case $\beta = 0$ corresponds to ∇F being constant, which is the case for instance if the term F is to be ignored in the problem (1), and therefore set to zero.

- $G \in \Gamma_0(\mathcal{X})$ 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.

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

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

- The set of minimizers of (1) is supposed nonempty.

The corresponding *dual* formulation of the *primal* problem (1) is [30, Chapters 15 and 19]

$$\text{Find } \hat{y} \in \underset{y \in \mathcal{Y}}{\arg\min}\ (F + G)^*(-L^* y) + H^*(y), \tag{4}$$

where we note that $(F+G)^*(-L^* y) = \min_{x' \in \mathcal{X}} F^*(-L^* y - x') + G^*(x')$ [30, Proposition 15.2]. Without further assumption, the set of solutions to (4) may be empty.

We also consider the *primal-dual* problem associated to (1) and (4), which consists in finding a saddle point of the Lagrangian [30, Section 19.2]:

$$\text{Find } (\hat{x}, \hat{y}) \in \arg \min_{x \in \mathcal{X}} \max_{y \in \operatorname{dom}(H^*)} F(x) + G(x) - H^*(y) + \langle Lx, y \rangle. \tag{5}$$

We now introduce the monotone variational inclusion problem under investigation in this paper:

$$\text{Find } (\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y} \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G(\hat{x}) + L^* \hat{y} + \nabla F(\hat{x}) \\ -L\hat{x} + \partial H^*(\hat{y}) \end{pmatrix}. \tag{6}$$

The classical Karush-Kuhn-Tucker theory asserts that if the pair $(\hat{x}, \hat{y})$ is a solution to (6), then $\hat{x}$ is a solution to (1), $\hat{y}$ is a solution to (4) [30, Theorem 19.1], and $(\hat{x}, \hat{y})$

is a solution to (5) [30, Proposition 19.18]. The converse does not hold in general and the set of solutions to (6) may be empty. However, if the following qualification condition holds:

$$0 \in \operatorname{sri}\big(L(\operatorname{dom}(G)) - \operatorname{dom}(H)\big), \tag{7}$$

then the set of solutions to (4) is nonempty [30, Theorem 15.23] and for every primal solution $\hat{x}$ to (1) and dual solution $\hat{y}$ to (4) then $(\hat{x}, \hat{y})$ is a solution to (6) [30, Theorem 19.1].

Thus, in the following, we assume that the set of solutions to the inclusions (6) is nonempty, keeping in mind that (7) is a sufficient condition for this to hold.

The advantage in solving (6) 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 alternately and efficiently.

We may observe that there is "room" in the dual inclusion of (6) for an additional term $\nabla K^*(\hat{y})$, which yields a more symmetric formulation of the primal and dual problems. The obtained variational inclusions characterize the following primal problem, which includes an infimal convolution:

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\arg\min} \inf_{y' \in \mathcal{Y}} F(x) + G(x) + H(Lx - y') + K(y'), \tag{8}$$

where the additional function $K \in \Gamma_0(\mathcal{Y})$ is such that K^* is differentiable on $\mathcal{Y}$ with β'-Lipschitz gradient for some $\beta' \geq 0$. Also, since the proofs in Sect. 4 are derived in the general framework of monotone and nonexpansive operators, it would be straightforward to adapt the study to solve more general monotone inclusions, where, in (6), the subgradients would be replaced by arbitrary maximally monotone operators, the gradients by cocoercive operators and the proximity operators by resolvants. The intention of the author was to keep the study accessible to the practitioner interested in the optimization problem (1). We refer the reader interested in a more general formulation to the subsequent[1] work [22].

3 Proposed Algorithm

The proposed algorithm to solve (6) is the following:

Algorithm 3.1. Choose the stepsizes $\tau > 0$, $\sigma > 0$, the sequence $(\rho_n)_{n \in \mathbb{N}}$ in $]0, +\infty[$, and the initial estimate $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, then iterate, for every $n \geq 0$,

$$\left|\begin{array}{ll} 1. & \tilde{x}_{n+1} = \operatorname{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau L^* y_n\big) + e_{G,n} \\ 2. & \tilde{y}_{n+1} = \operatorname{prox}_{\sigma H^*}\big(y_n + \sigma L(2\tilde{x}_{n+1} - x_n)\big) + e_{H,n} \\ 3. & (x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n) \end{array}\right. \tag{9}$$

[1] This manuscript is a revised version of [33], which was made available online in July 2011. The same algorithm was proposed several months later in [22]—where [33] is cited—in the more general framework of monotone inclusions.

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

We recall that $\text{prox}_{\sigma H^*}$ can be easily computed from $\text{prox}_{H/\sigma}$ if necessary, thanks to Moreau's identity $\text{prox}_{\sigma H^*}(y) = y - \sigma\, \text{prox}_{H/\sigma}(y/\sigma)$.

Now, we claim at once the convergence results for Algorithm 3.1. The proofs are derived in Sect. 4. We refer to Remark 4.8 for a variant of Algorithm 3.1 where the primal and dual update steps are switched, with same convergence properties.

Theorem 3.2. *Let $\tau > 0$, $\sigma > 0$ and the sequences $(\rho_n)_{n\in\mathbb{N}}$, $(e_{F,n})_{n\in\mathbb{N}}$, $(e_{G,n})_{n\in\mathbb{N}}$, $(e_{H,n})_{n\in\mathbb{N}}$, be the parameters of Algorithm 3.1. Let β be the Lipschitz constant defined in* (2). *Suppose that $\beta > 0$ and that the following hold:*

(i) $\dfrac{1}{\tau} - \sigma\|L\|^2 \geq \dfrac{\beta}{2}$,

(ii) $\forall n \in \mathbb{N},\ \rho_n \in]0, \delta[$, *where we set*

$$\delta = \max\left(\frac{4\kappa}{2\kappa+1}, \min\left(\frac{3}{2}, \frac{1}{2}+\kappa\right)\right) \in [1,2[\quad \textit{and} \tag{10}$$

$$\kappa = \frac{1}{\beta}\left(\frac{1}{\tau} - \sigma\|L\|^2\right) \in [\tfrac{1}{2}, +\infty[, \tag{11}$$

(iii) $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$,

(iv) $\sum_{n\in\mathbb{N}} \rho_n\|e_{F,n}\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \rho_n\|e_{G,n}\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \rho_n\|e_{H,n}\| < +\infty$.

Then, there exists a pair $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ solution to (6) *such that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ computed by Algorithm 3.1 converge weakly to $\hat{x}$ and $\hat{y}$, respectively.*

Theorem 3.3. *Let $\tau > 0$, $\sigma > 0$ and the sequences $(\rho_n)_{n\in\mathbb{N}}$, $(e_{F,n})_{n\in\mathbb{N}}$, $(e_{G,n})_{n\in\mathbb{N}}$, $(e_{H,n})_{n\in\mathbb{N}}$, be the parameters of Algorithm 3.1. Let β be the Lipschitz constant defined in* (2) *and suppose that $\beta = 0$, so that F is an affine functional. Considering F as an affine functional or setting $F = 0$ and aggregating an affine functional with the term G yield exactly the same calculations in Algorithm 3.1. Thus, we assume without restriction that $F = 0$ and that the error terms $e_{F,n}$ are all zero. Suppose that the following hold:*

(i) $\sigma\tau\|L\|^2 < 1$,

(ii) $\forall n \in \mathbb{N},\ \rho_n \in]0, 2[$,

(iii) $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$,

(iv) $\sum_{n\in\mathbb{N}} \rho_n\|e_{G,n}\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \rho_n\|e_{H,n}\| < +\infty$.

Then, there exists a pair $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ solution to (6) *such that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ computed by Algorithm 3.1 converge weakly to $\hat{x}$ and $\hat{y}$, respectively.*

3.1 Relationships to Existing Optimization Methods

The proposed Algorithm 3.1 is able to solve the general problem (1) iteratively, using calls to the operators ∇F, $\mathrm{prox}_{\tau G}$, $\mathrm{prox}_{\sigma H^*}$, L and L^*, without any other implicit (inverse) operator nor inner loop. In particular, no inverse operator of the form $(I + \alpha L^* L)^{-1}$ is required. To our knowledge, the only existing method of the literature having this feature is the very recent one of Combettes and Pesquet [20], based on a different splitting of (6). 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.

We also note that a general and abstract framework has been proposed in [13], of which the problem (1) could be considered as a particular instance. The algorithms obtained along this line of research remain to be studied.

Some authors have studied the use of nested algorithms to solve (1) for practical imaging problems [34–36]. 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.

Now, in some particular cases, Algorithm 3.1 reverts to classical splitting methods of the literature.

3.1.1 The Case $F = 0$

If the smooth term F is absent of the problem, Algorithm 3.1 exactly reverts to the primal-dual algorithm of Chambolle and Pock [6]. These authors proved the convergence in the finite-dimensional case, assuming that $\tau\sigma\|L\|^2 < 1$ and $\rho_n = 1$, and without error terms. The convergence has been proven in a different way in [16] with a constant relaxation parameter $\rho_n = \rho \in]0,2[$ and the same other hypotheses. Thus, Theorem 3.3 extends the range of applicability of this primal-dual algorithm. We note that this method has been proposed in other forms in [14, 15].

When $F = 0$, the primal-dual method in [18] and the method in [37] can be used to solve (1), as well. They yield algorithms different from Algorithm 3.1.

However, 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 the Fast Fourier Transform if M has shift-invariance properties) but slow and complicated to implement (especially if particular care is paid to the treatment at the boundaries for multi-dimensional problems). By contrast, considering $\frac{\lambda}{2}\|Mx - b\|^2$ as the function F 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 [6, eq. (74)]. Such parallel strategies are discussed in Sect. 5.

If $F = 0$ and $L = I$, let us discuss the relationship with the classical Douglas-Rachford splitting method [1, 3–5]. For simplicity, we assume that, for every $n \in \mathbb{N}$,

$\rho_n = 1$ and $e_{F,n} = e_{G,n} = e_{H,n} = 0$. Also, we introduce the auxiliary variable $s_n = x_n - \frac{1}{\sigma} y_n$. Then, we can rewrite Algorithm 3.1 as

Algorithm 3.4. Choose the parameters $\tau > 0$, $\sigma > 0$ and the initial estimate $(x_0, s_0) \in \mathcal{X}^2$, then iterate, for every $n \geq 0$,

$$\left| \begin{array}{ll} 1. & x_{n+1} = \text{prox}_{\tau G}\big((1-\tau\sigma)x_n + \tau\sigma s_n\big) \\ 2. & s_{n+1} = s_n - x_{n+1} + \text{prox}_{\frac{1}{\sigma}H}(2x_{n+1} - s_n) \end{array} \right. \tag{12}$$

If additionally we set $\sigma = 1/\tau$, Algorithm 3.4 exactly reverts to the Douglas-Rachford method. In that case, it is known that s_n converges weakly to a limit $\hat{s}$ and it was shown recently [5] that x_n converges weakly to a primal solution $\hat{x} = \text{prox}_{\tau G}(\hat{s})$ and $y_n = \sigma(x_n - s_n)$ converges weakly to a dual solution $\hat{y}$. We note that this limit case $\sigma\tau = 1$ is not covered by Theorem 3.3.

Interestingly, Algorithm 3.4 in the case $\sigma\tau < 1$ was studied in [17] and presented as a Douglas-Rachford method with inertia, because of the linear combination of x_n and s_n used to compute x_{n+1}. However, when reintroducing relaxation in the algorithm, there is a slight difference between the algorithm proposed in [17] and Algorithm 3.1.

3.1.2 The 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 only, we obtain the iteration

$$x_{n+1} = \rho_n \text{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n})\big) + e_{G,n} + (1-\rho_n)x_n, \tag{13}$$

which is exactly the forward-backward splitting method [3, 24, 26]. It is known to converge weakly to a primal solution $\hat{x}$ if $0 < \tau < \frac{2}{\beta}$, $(\rho_n)_{n\in\mathbb{N}}$ is a sequence in $[\epsilon, 1]$ for some $\epsilon > 0$, and the error terms are absolutely summable.

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.

4 Proofs of Convergence

In this section, we prove Theorems 3.2 and 3.3.

First, we recall some definitions and properties of operator theory. We refer the readers to [30] for more details. In the following, $\mathcal{H}$ is a real Hilbert space. Let $M : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. We denote by $\text{ran}(M) = \{v \in \mathcal{H} \ : \ \exists u \in \mathcal{H}, v \in Mu\}$ the range of M, by $\text{gra}(M) = \{(u, v) \in \mathcal{H}^2 \ : \ v \in Mu\}$ its graph, and by M^{-1} its inverse; that is, the set-valued operator with graph $\{(v, u) \in \mathcal{H}^2 \ : \ v \in Mu\}$. We define $\text{zer}(M) = \{u \in \mathcal{H} \ : \ 0 \in Mu\}$. M is said to be monotone if $\forall (u, u') \in \mathcal{H}^2, \forall (v, v') \in Mu \times Mu'$, $\langle u - u', v - v'\rangle \geq 0$ and maximally monotone if there exists no monotone operator M' such that $\text{gra}(M) \subset \text{gra}(M') \neq \text{gra}(M)$.

Let $T:\mathcal{H}\to\mathcal{H}$ be an operator. We define $\operatorname{fix}(T)=\{x\in\mathcal{H} : Tx=x\}$. T is said to be nonexpansive if it is 1-Lipschitz continuous on $\mathcal{H}$, see (2), and firmly nonexpansive if $2T-I$ is nonexpansive. Let $\alpha\in]0,1]$. T is said to be α-averaged if there exists a nonexpansive operator T' such that $T=\alpha T'+(1-\alpha)I$. We denote by $\mathscr{A}(\mathcal{H},\alpha)$ the set of α-averaged operators on $\mathcal{H}$. Clearly, $\mathscr{A}(\mathcal{H},1)$ is the set of nonexpansive operators and $\mathscr{A}(\mathcal{H},\frac{1}{2})$ is the set of firmly nonexpansive operators.

The resolvent $(I+M)^{-1}$ of a maximally monotone operator $M:\mathcal{H}\to 2^{\mathcal{H}}$ is defined and single-valued on $\mathcal{H}$ and firmly nonexpansive. The subdifferential ∂J of $J\in\Gamma_0(\mathcal{H})$ is maximally monotone and $(I+\partial J)^{-1}=\operatorname{prox}_J$.

Lemma 4.1 (Krasnosel'skii Mann iteration) [3, Lemma 5.1]. *Let $T:\mathcal{H}\to\mathcal{H}$ be a nonexpansive operator, let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ and let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $\operatorname{fix}(T)\neq\emptyset$, $\sum_{n\in\mathbb{N}}\rho_n(1-\rho_n)=+\infty$, and $\sum_{n\in\mathbb{N}}\rho_n\|e_n\|<+\infty$. Let $s_0\in\mathcal{H}$ and set, for every $n\in\mathbb{N}$,*

$$s_{n+1}=s_n+\rho_n\big(T(s_n)+e_n-s_n\big). \tag{14}$$

Then, $(s_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{s}\in\operatorname{fix}(T)$.

Lemma 4.2 (Proximal point algorithm). *Let $M:\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma\in]0,+\infty[$, let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ and let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $\operatorname{zer}(M)\neq\emptyset$, $\sum_{n\in\mathbb{N}}\rho_n(2-\rho_n)=+\infty$, and $\sum_{n\in\mathbb{N}}\rho_n\|e_n\|<+\infty$. Let $s_0\in\mathcal{H}$ and set, for every $n\in\mathbb{N}$,*

$$s_{n+1}=s_n+\rho_n\big((I+\gamma M)^{-1}(s_n)+e_n-s_n\big). \tag{15}$$

Then, $(s_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{s}\in\operatorname{zer}(M)$.

Proof. Set $T=(I+\gamma M)^{-1}$, $T'=2T-I$, and, $\forall n\in\mathbb{N}$, $\rho'_n=\rho_n/2$ and $e'_n=2e_n$. Then, $\sum_{n\in\mathbb{N}}\rho'_n(1-\rho'_n)=+\infty$, $\sum_{n\in\mathbb{N}}\rho'_n\|e'_n\|<+\infty$, and we can rewrite (19) as

$$s_{n+1}=s_n+\rho'_n\big(T'(s_n)+e'_n-s_n\big). \tag{16}$$

Moreover, $T\in\mathscr{A}(\mathcal{H},\frac{1}{2})$ [30, Corollary 23.8], so that $T'\in\mathscr{A}(\mathcal{H},1)$, and $\operatorname{fix}(T')=\operatorname{fix}(T)=\operatorname{zer}(M)$ [30, Proposition 23.38]. Thus, we obtain the desired result by applying Lemma 4.1 to the iteration (16). □

Lemma 4.3 (Composition of averaged operators). *Let $\alpha\in]0,1]$, $T_1\in\mathscr{A}(\mathcal{H},\frac{1}{2})$ and $T_2\in\mathscr{A}(\mathcal{H},\alpha)$. Then $T_1\circ T_2\in\mathscr{A}(\mathcal{H},\alpha')$ with*

$$\alpha'=\min\left(\frac{1+\alpha}{2},\max\left(\frac{2}{3},\frac{2\alpha}{\alpha+1}\right)\right). \tag{17}$$

Proof. If $\alpha\in]0,1[$, then it is shown in [3, Lemma 2.2(iii)] that $T_1\circ T_2\in\mathscr{A}(\mathcal{H},\alpha_1)$ with $\alpha_1=\max(\frac{2}{3},\frac{2\alpha}{\alpha+1})$. Clearly, this property also holds if $\alpha=1$. Furthermore, there exists $T'_1\in\mathscr{A}(\mathcal{H},1)$ and $T'_2\in\mathscr{A}(\mathcal{H},1)$ such that $T_1=(T'_1+I)/2$ and $T_2=\alpha T'_2+(1-\alpha)I$. Then, $T_1\circ T_2=\frac{1}{2}T'_1\circ T_2+\frac{\alpha}{2}T'_2+\frac{1-\alpha}{2}I\in\mathscr{A}(\mathcal{H},\alpha_2)$ with $\alpha_2=\frac{1+\alpha}{2}$. Hence, $T_1\circ T_2\in\mathscr{A}(\mathcal{H},\alpha_1)\cap\mathscr{A}(\mathcal{H},\alpha_2)=\mathscr{A}(\mathcal{H},\min(\alpha_1,\alpha_2))$. □

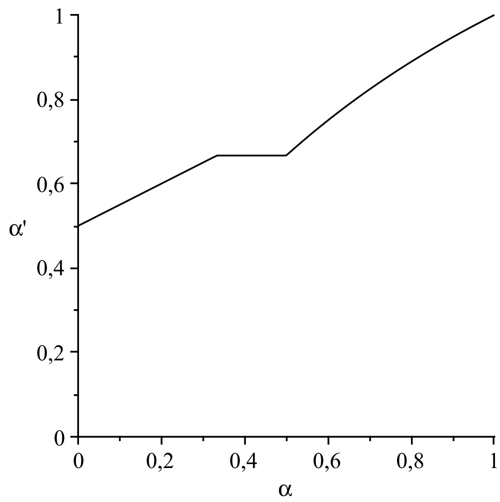

Figure 1: α' as a function of α in Lemma 4.3, according to (17).

Lemma 4.4 (Forward-backward iteration). *Let $M_1 : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\kappa \in]0,+\infty[$, let $M_2 : \mathcal{H} \to \mathcal{H}$ be κ-cocoercive; that is, $\kappa M_2 \in \mathscr{A}(\mathcal{H}, \frac{1}{2})$. Let $\gamma \in]0,2\kappa]$, and set*

$$\delta = \max\left(\frac{4\kappa}{2\kappa+\gamma}, \min\left(\frac{3}{2}, \frac{1}{2} + \frac{\kappa}{\gamma}\right)\right) \in [1,2[. \tag{18}$$

Furthermore, let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence in $]0,\delta[$ such that $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$ and let $(e_{1,n})_{n\in\mathbb{N}}$ and $(e_{2,n})_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \rho_n \|e_{1,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}} \rho_n \|e_{2,n}\| < +\infty$. Suppose that $\operatorname{zer}(M_1 + M_2) \neq \varnothing$, let $s_0 \in \mathcal{H}$, and set, for every $n \in \mathbb{N}$,

$$s_{n+1} = \rho_n\Big((I+\gamma M_1)^{-1}\big(s_n - \gamma(M_2(s_n) + e_{2,n})\big) + e_{1,n}\Big) + (1-\rho_n)s_n. \tag{19}$$

Then, $(s_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{s} \in \operatorname{zer}(M_1 + M_2)$.

Proof. Set $T_1 = (I+\gamma M_1)^{-1}$, $T_2 = I - \gamma M_2$, and $T = T_1 \circ T_2$. Then $T_1 \in \mathscr{A}(\mathcal{H}, \frac{1}{2})$ [30, Corollary 23.8] and $T_2 \in \mathscr{A}(\mathcal{H}, \frac{\gamma}{2\kappa})$ [30, Proposition 4.33]. Hence, Lemma 4.3 implies that $T \in \mathscr{A}(\mathcal{H}, 1/\delta)$. Now, set $T' = \delta T + (1-\delta)I$ and, $\forall n \in \mathbb{N}$, $\rho'_n = \rho_n/\delta$ and $e'_n = \delta\big(T_1(T_2(s_n) + e_{2,n}) + e_{1,n} - T_1(T_2(s_n))\big)$. Then, we can rewrite (19) as (16). Moreover, T' is nonexpansive, $\operatorname{fix}(T') = \operatorname{fix}(T) = \operatorname{zer}(M_1 + M_2)$ [30, Proposition 25.1(iv)] and $\sum_{n\in\mathbb{N}} \rho'_n(1-\rho'_n) = +\infty$. From the nonexpansivity of T_1, we obtain, $\forall n \in \mathbb{N}$,

$$\|e'_n\| \le \delta\|e_{1,n}\| + \delta\big\|T_1\big(T_2(s_n) + e_{2,n}\big) - T_1\big(T_2(s_n)\big)\big\| \le \delta\|e_{1,n}\| + \delta\|e_{2,n}\|. \tag{20}$$

Hence, $\sum_{n\in\mathbb{N}} \rho'_n \|e'_n\| < +\infty$. Thus, the result follows from Lemma 4.1. $\square$

Lemma 4.5 (Baillon-Haddad theorem) [30, Corollary 18.16]. *Let $J : \mathscr{H} \to \mathbb{R}$ be convex, differentiable on $\mathscr{H}$ and such that $\kappa\nabla J$ is nonexpansive for some $\kappa \in]0,+\infty[$. Then, ∇J is κ-cocoercive; that is, $\kappa\nabla J$ is firmly nonexpansive.*

Proof of Theorem 3.2. In essence, we show that Algorithm 3.1 has the structure of a forward-backward iteration, when expressed in terms of nonexpansive operators on $\mathscr{X} \times \mathscr{Y}$ equipped with a particular inner product. Let us define the vector space $\mathcal{Z} = \mathscr{X} \times \mathscr{Y}$ and the bounded linear operator on $\mathcal{Z}$,

$$P : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\tau} I & -L^* \\ -L & \frac{1}{\sigma} I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{21}$$

Then, let the inner product $\langle \cdot, \cdot \rangle_I$ in $\mathcal{Z}$ be defined as

$$\langle z, z' \rangle_I = \langle x, x' \rangle + \langle y, y' \rangle, \text{ for every } z = (x, y), z' = (x', y') \in \mathcal{Z}. \tag{22}$$

By endowing $\mathcal{Z}$ with this inner product, we obtain the Hilbert space denoted by $\mathcal{Z}_I$. Then, in $\mathcal{Z}_I$, P is bounded, self-adjoint, and, from the assumptions $\beta > 0$ and (i), positive definite. 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

$$\langle z, z' \rangle_P = \langle z, P z' \rangle_I, \text{ for every } (z, z') \in \mathcal{Z}^2. \tag{23}$$

By endowing $\mathcal{Z}$ with this inner product, we obtain the Hilbert space denoted by $\mathcal{Z}_P$. Thereafter, $\forall n \in \mathbb{N}$, we denote by $z_n = (x_n, y_n)$ the iterates of Algorithm 3.1. Then, our aim is to prove the existence of $\hat{z} = (\hat{x}, \hat{y}) \in \mathcal{Z}$ such that $(z_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{z}$ in $\mathcal{Z}_P$; that is, for every $z' \in \mathcal{Z}$, $\langle z_n - \hat{z}, P z' \rangle_I \to 0$ as $n \to +\infty$. Since P is bounded from below in $\mathcal{Z}_I$, P^{-1} is well defined and bounded in $\mathcal{Z}_I$, so that the notions of weak convergence in $\mathcal{Z}_P$ and in $\mathcal{Z}_I$ are equivalent.

Now, let us consider the error-free case $e_{F,n} = e_{G,n} = e_{H,n} = 0$. For every $n \in \mathbb{N}$, the following inclusion is satisfied by $\tilde{z}_{n+1} = (\tilde{x}_{n+1}, \tilde{y}_{n+1})$ computed by Algorithm 3.1:

$$-\underbrace{\begin{pmatrix} \nabla F(x_n) \\ 0 \end{pmatrix}}_{B(z_n)} \in \underbrace{\begin{pmatrix} \partial G(\tilde{x}_{n+1}) + L^* \tilde{y}_{n+1} \\ -L\tilde{x}_{n+1} + \partial H^*(\tilde{y}_{n+1}) \end{pmatrix}}_{A(\tilde{z}_{n+1})} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & -L^* \\ -L & \frac{1}{\sigma} I \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} \tilde{x}_{n+1} - x_n \\ \tilde{y}_{n+1} - y_n \end{pmatrix}}_{(\tilde{z}_{n+1} - z_n)}, \tag{24}$$

or equivalently,

$$\tilde{z}_{n+1} = (I + P^{-1} \circ A)^{-1} \circ (I - P^{-1} \circ B)(z_n). \tag{25}$$

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

$$z_{n+1} = \rho_n \Big((I + P^{-1} \circ A)^{-1} \big(z_n - P^{-1} \circ B(z_n) - e_{2,n} \big) + e_{1,n} \Big) + (1 - \rho_n) z_n. \tag{26}$$

$$\text{with } e_{1,n} = (e_{G,n}, e_{H,n}) \text{ and } e_{2,n} = P^{-1}(e_{F,n}, -2L e_{G,n}), \tag{27}$$

and we recognize the structure of the forward-backward iteration (19). Thus, it is sufficient to check the conditions of application of Lemma 4.4 for the convergence result of Theorem 3.2 to follow. To that end, we set $\mathscr{H} = \mathcal{Z}_P$, $M_1 = P^{-1} \circ A$ and $M_2 = P^{-1} \circ B$. Let δ and κ be defined in (10) and (11).

• The operator $(x,y)\mapsto \partial G(x)\times\partial H^*(y)$ is maximally monotone in $\mathcal{Z}_I$, by Theorem 20.40, Corollary 16.24, Propositions 20.22 and 20.23 of [30]. Moreover, the skew operator $(x,y)\mapsto(L^*y,-Lx)$ is maximally monotone in $\mathcal{Z}_I$ [30, Example 20.30] and has full domain. Hence, A is maximally monotone [30, Corollary 24.4(i)]. Thus, M_1 is monotone in $\mathcal{Z}_P$ and from the injectivity of P, M_1 is maximally monotone in $\mathcal{Z}_P$.
• Let us show the cocoercivity of M_2. For every $z=(x,y)$, $z'=(x',y')\in\mathcal{Z}$, we have

$$
\begin{aligned}
\|M_2(z)-M_2(z')\|_P^2 &= \left\langle P^{-1}\circ B(z)-P^{-1}\circ B(z'), B(z)-B(z')\right\rangle_I &(28)\\
&= \left\langle \tfrac{1}{\sigma}\left(\tfrac{1}{\sigma\tau}I-L^*L\right)^{-1}\left(\nabla F(x)-\nabla F(x')\right), \nabla F(x)-\nabla F(x')\right\rangle &(29)\\
&\le \tfrac{1}{\sigma}\left(\tfrac{1}{\sigma\tau}-\|L\|^2\right)^{-1}\|\nabla F(x)-\nabla F(x')\|^2 &(30)\\
&\le \tfrac{\beta^2}{\sigma}\left(\tfrac{1}{\sigma\tau}-\|L\|^2\right)^{-1}\|x-x'\|^2=\tfrac{\beta}{\kappa}\|x-x'\|^2. &(31)
\end{aligned}
$$

We define the linear operator $Q:(x,y)\mapsto(x,0)$ of $\mathcal{Z}$. Then, $P-\beta\kappa Q$ is positive semidefinite in $\mathcal{Z}_I$, so that

$$\beta\kappa\|x-x'\|^2=\beta\kappa\langle(z-z'),Q(z-z')\rangle_I\le\langle(z-z'),P(z-z')\rangle_I=\|z-z'\|_P^2. \qquad (32)$$

Putting together (31) and (32), we get

$$\kappa\|M_2(z)-M_2(z')\|_P\le\|z-z'\|_P, \qquad (33)$$

so that κM_2 is nonexpansive in $\mathcal{Z}_P$. Let us define on $\mathcal{Z}$ the function $J:(x,y)\mapsto F(x)$. Then, in $\mathcal{Z}_P$, $\nabla J=M_2$. Therefore, from Lemma 4.5, κM_2 is firmly nonexpansive in $\mathcal{Z}_P$.
• We set $\gamma=1$. Since $\kappa\ge\frac{1}{2}$, $\gamma\in\,]0,2\kappa]$.
• Since P^{-1} and L are bounded and the norms $\|\cdot\|_I$ and $\|\cdot\|_P$ are equivalent, we get from the assumption (iv) that $\sum_{n\in\mathbb{N}}\rho_n\|e_{1,n}\|_P<+\infty$ and $\sum_{n\in\mathbb{N}}\rho_n\|e_{2,n}\|_P<+\infty$.
• The set of solutions to (6) is $\mathrm{zer}(A+B)=\mathrm{zer}(M_1+M_2)$ and it is nonempty by assumption. □

Proof of Theorem 3.3. We follow the same line of proof as for Theorem 3.2. Now, $\beta=0$, but from the assumption (i), P is still bounded from below and defines a valid scalar product on $\mathcal{Z}$. $F=0$ so that $B=0$ and (26) becomes

$$z_{n+1}=\rho_n\big((I+P^{-1}\circ A)^{-1}(z_n)+e_{1,n}\big)+(1-\rho_n)z_n, \qquad (34)$$

and we recognize the structure of the proximal point algorithm (15) with $\gamma=1$, as shown in [16]. Thus, the convergence result of Theorem 3.3 follows from Lemma 4.2, whose conditions of application have been checked in the proof of Theorem 3.2. □

Remark 4.6 on Lemma 4.4. To our knowledge, the range of parameters under which convergence of the stationary forward-backward iteration is ensured, as given in Lemma 4.4, is wider than the one found in the literature. If $\gamma<2\kappa$, then $\delta>1$ and we can proceed without relaxation by setting $\rho_n=1$. However, if $\gamma=2\kappa$, then $\delta=1$ and we have to choose under-relaxation factors $\rho_n < 1$.

Algorithm 4.7. Choose the parameters $\sigma > 0$, $\tau > 0$, the sequence $(\rho_n)_{n\in\mathbb{N}}$ in $]0,+\infty[$, and the initial estimate $(x_0, y_0) \in \mathscr{X} \times \mathscr{Y}$, then iterate, for every $n \geq 0$,

$$\left|\begin{array}{ll} 1. & \tilde{y}_{n+1} = \mathrm{prox}_{\sigma H^*}\big(y_n + \sigma L x_n\big) + e_{H,n} \\ 2. & \tilde{x}_{n+1} = \mathrm{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau L^*(2\tilde{y}_{n+1} - y_n)\big) + e_{G,n} \\ 3. & (x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1-\rho_n)(x_n, y_n) \end{array}\right. \tag{35}$$

Remark 4.8 on Algorithm 4.7. Algorithm 3.1 is not symmetric with respect to the primal and dual variable, 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. Thus, we can switch the update steps of the two variables and propose Algorithm 4.7 above. For every $n \in \mathbb{N}$, if the error terms are zero, the following inclusion is satisfied by $\tilde{z}_{n+1} = (\tilde{x}_{n+1}, \tilde{y}_{n+1})$ computed by Algorithm 4.7:

$$-\underbrace{\begin{pmatrix} \nabla F(x_n) \\ 0 \end{pmatrix}}_{B(z_n)} \in \underbrace{\begin{pmatrix} \partial G(\tilde{x}_{n+1}) + L^*\tilde{y}_{n+1} \\ -L\tilde{x}_{n+1} + \partial H^*(\tilde{y}_{n+1}) \end{pmatrix}}_{A(\tilde{z}_{n+1})} + \underbrace{\begin{pmatrix} \frac{1}{\tau}I & L^* \\ L & \frac{1}{\sigma}I \end{pmatrix}}_{P'} \underbrace{\begin{pmatrix} \tilde{x}_{n+1} - x_n \\ \tilde{y}_{n+1} - y_n \end{pmatrix}}_{(\tilde{z}_{n+1} - z_n)}. \tag{36}$$

Thus, after comparison of (24) and (36), it appears that the whole convergence analysis for Algorithm 3.1 applies to Algorithm 4.7, just replacing P by P', and both algorithms converge under the same assumptions, given by Theorems 3.2 and 3.3.

5 Extension to Several Composite Functions

In this section, we focus on the following primal problem with $m \geq 2$ composite functions:

$$\text{Find } \hat{x} \in \arg\min_{x\in\mathscr{X}} F(x) + G(x) + \sum_{i=1}^{m} H_i(L_i x), \tag{37}$$

with the same assumptions on F and G as in the problem (1), m functions $H_i \in \Gamma_0(\mathscr{Y}_i)$ defined on real Hilbert spaces $\mathscr{Y}_i$, and m bounded linear functions $L_i : \mathscr{X} \to \mathscr{Y}_i$. At the same time, we consider the dual problem

$$\text{Find } (\hat{y}_1, \dots, \hat{y}_m) \in \arg\min_{y_1\in\mathscr{Y}_1,\dots,y_m\in\mathscr{Y}_m} (F+G)^*(-\textstyle\sum_{i=1}^m L_i^* y_i) + \displaystyle\sum_{i=1}^{m} H_i^*(y_i) \tag{38}$$

$$\equiv \text{ Find } (\hat{y}_1, \dots, \hat{y}_m) \in \arg\min_{y_1\in\mathscr{Y}_1,\dots,y_m\in\mathscr{Y}_m} \min_{x'\in\mathscr{X}} F^*(-\textstyle\sum_{i=1}^m L_i^* y_i - x') + G^*(x') + \displaystyle\sum_{i=1}^{m} H_i^*(y_i), \tag{39}$$

and the corresponding monotone variational inclusion:

$$\text{Find } (\hat{x}, \hat{y}_1, \dots, \hat{y}_m) \in \mathscr{X} \times \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_m \text{ such that } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G(\hat{x}) + \sum_{i=1}^m L_i^* \hat{y}_i + \nabla F(\hat{x}) \\ -L_1\hat{x} + \partial H_1^*(\hat{y}_1) \\ \vdots \\ -L_m\hat{x} + \partial H_m^*(\hat{y}_m) \end{pmatrix}. \tag{40}$$

Thus, if $(\hat{x}, \hat{y}_1, \ldots, \hat{y}_m)$ is solution to (40), then $\hat{x}$ is solution to (37) and $(\hat{y}_1, \ldots, \hat{y}_m)$ is solution to (38). In the following, we suppose that the set of solutions to (40) is nonempty.

To our knowledge, the only existing method able to solve (40), in whole generality and by full splitting (that is, without calls to implicit operators), is the one in [20]. We note that another recent method has been proposed in [21] to solve (37) in the more restrictive setting where $G = 0$ and $L_i = I$ for every $i = 1, \ldots, m$.

Although the primal and dual problems (37) and (38) are more general than the problems (1) and (4), respectively, they can be recast as particular cases of them using product spaces. To that end, we introduce the bold notation $\boldsymbol{y} = (y_1, \ldots, y_m)$ for an element of the Hilbert space $\boldsymbol{\mathcal{Y}} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ equipped with the inner product $\langle \boldsymbol{y}, \boldsymbol{y}' \rangle = \sum_{i=1}^m \langle y_i, y_i' \rangle$. We define the function $\boldsymbol{H} \in \Gamma_0(\boldsymbol{\mathcal{Y}})$ by $\boldsymbol{H}(\boldsymbol{y}) = \sum_{i=1}^m H_i(y_i)$ and the linear function $\boldsymbol{L} : \mathcal{X} \to \boldsymbol{\mathcal{Y}}$ by $\boldsymbol{L}x = (L_1 x, \ldots, L_m x)$. We also define the error term $\boldsymbol{e}_{\boldsymbol{H},n} = (e_{H_1,n}, \ldots, e_{H_m,n})$, $\forall n \in \mathbb{N}$. We have the following properties, $\forall \boldsymbol{y} \in \boldsymbol{\mathcal{Y}}$:

$$\boldsymbol{H}^*(\boldsymbol{y}) = \sum_{i=1}^m H_i^*(y_i), \tag{41}$$

$$\boldsymbol{L}^* \boldsymbol{y} = \sum_{i=1}^m L_i^* y_i, \tag{42}$$

$$\mathrm{prox}_{\sigma \boldsymbol{H}^*}(\boldsymbol{y}) = \big(\mathrm{prox}_{\sigma H_1}(y_1), \ldots, \mathrm{prox}_{\sigma H_m}(y_m)\big), \tag{43}$$

$$\|\boldsymbol{L}\|^2 = \| \sum_{i=1}^m L_i^* L_i \|. \tag{44}$$

Thus, we can rewrite (37) and (38) as

$$\text{Find } \hat{x} \in \operatorname*{arg\,min}_{x \in \mathcal{X}} F(x) + G(x) + \boldsymbol{H}(\boldsymbol{L}x), \tag{45}$$

$$\text{Find } \hat{\boldsymbol{y}} \in \operatorname*{arg\,min}_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}} (F+G)^*(-\boldsymbol{L}^* \boldsymbol{y}) + \boldsymbol{H}^*(\boldsymbol{y}), \tag{46}$$

which exactly take the form of (1) and (4), respectively. Accordingly, we can rewrite Algorithms 3.1 and 4.7 by doing the appropriate substitutions, and we obtain the two following algorithms, respectively:

Algorithm 5.1. Choose the parameters $\tau > 0$, $\sigma > 0$, the sequence $(\rho_n)_{n \in \mathbb{N}}$ in $]0, +\infty[$, and the initial estimate $(x_0, y_{1,0}, \ldots, y_{m,0}) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$, then iterate, for every $n \geq 0$,

$$\left\lfloor \begin{array}{ll} 1. & \tilde{x}_{n+1} = \mathrm{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau \sum_{i=1}^m L_i^* y_{i,n}\big) + e_{G,n} \\ 2. & x_{n+1} = \rho_n \tilde{x}_{n+1} + (1-\rho_n) x_n \\ 3. & \forall i = 1, \ldots, m,\ \tilde{y}_{i,n+1} = \mathrm{prox}_{\sigma H_i^*}\big(y_{i,n} + \sigma L_i(2\tilde{x}_{n+1} - x_n)\big) + e_{H_i,n} \\ 4. & \forall i = 1, \ldots, m,\ y_{i,n+1} = \rho_n \tilde{y}_{i,n+1} + (1-\rho_n) y_{i,n} \end{array} \right. \tag{47}$$

Algorithm 5.2. Choose the parameters $\sigma > 0$, $\tau > 0$, the sequence $(\rho_n)_{n\in\mathbb{N}}$ in $]0,+\infty[$, and the initial estimate $(x_0, y_{1,0}, \ldots, y_{m,0}) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$, then iterate, for every $n \geq 0$,

$$\left\lfloor \begin{array}{ll} 1. & \forall i = 1,\ldots,m,\ \tilde{y}_{i,n+1} = \operatorname{prox}_{\sigma H_i^*}\big(y_{i,n} + \sigma L_i x_n\big) + e_{H_i,n} \\ 2. & \forall i = 1,\ldots,m,\ y_{i,n+1} = \rho_n \tilde{y}_{i,n+1} + (1-\rho_n) y_{i,n} \\ 3. & \tilde{x}_{n+1} = \operatorname{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n}) - \tau \sum_{i=1}^m L_i^*(2\tilde{y}_{n+1} - y_n)\big) + e_{G,n} \\ 4. & x_{n+1} = \rho_n \tilde{x}_{n+1} + (1-\rho_n) x_n \end{array} \right. \tag{48}$$

Hence, by doing the same substitutions in Theorems 3.2 and 3.3, we obtain the following spin-off theorems, respectively:

Theorem 5.3. *Let $\tau > 0$, $\sigma > 0$ and the sequences $(\rho_n)_{n\in\mathbb{N}}$, $(e_{F,n})_{n\in\mathbb{N}}$, $(e_{G,n})_{n\in\mathbb{N}}$, $(e_{H_i,n})_{\{i=1,\ldots,m,n\in\mathbb{N}\}}$, be the parameters of Algorithm 5.1 and Algorithm 5.2. Let β be the Lipschitz constant defined in* (2)*. Suppose that $\beta > 0$ and that the following hold:*

(i) $\dfrac{1}{\tau} - \sigma \|\sum_{i=1}^m L_i^* L_i\| \geq \dfrac{\beta}{2}$,

(ii) $\forall n \in \mathbb{N},\ \rho_n \in]0,\delta[$, *where we set*

$$\delta = \max\left(\frac{4\kappa}{2\kappa+1}, \min\left(\frac{3}{2}, \frac{1}{2} + \kappa\right)\right) \in [1,2[\quad \textit{and} \tag{49}$$

$$\kappa = \frac{1}{\beta}\left(\frac{1}{\tau} - \sigma \|\sum_{i=1}^m L_i^* L_i\|\right) \in [\tfrac{1}{2}, +\infty[, \tag{50}$$

(iii) $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$,

(iv) $\sum_{n\in\mathbb{N}} \rho_n \|e_{F,n}\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \rho_n \|e_{G,n}\| < +\infty$ *and, for* $i = 1,\ldots,m$, $\sum_{n\in\mathbb{N}} \rho_n \|e_{H_i,n}\| < +\infty$.

Then, there exists $(\hat{x}, \hat{y}_1, \ldots, \hat{y}_m) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ solution to (40) *such that, in Algorithm 5.1 or in Algorithm 5.2, the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{x}$ and, for every $i = 1,\ldots,m$, the sequence $(y_{i,n})_{n\in\mathbb{N}}$ converges weakly to $\hat{y}_i$.*

Theorem 5.4. *Let $\tau > 0$, $\sigma > 0$ and the sequences $(\rho_n)_{n\in\mathbb{N}}$, $(e_{F,n})_{n\in\mathbb{N}}$, $(e_{G,n})_{n\in\mathbb{N}}$, $(e_{H_i,n})_{\{i=1,\ldots,m,n\in\mathbb{N}\}}$, be the parameters of Algorithm 5.1 and Algorithm 5.2. Suppose that $F = 0$ and that the error terms $e_{F,n}$ are all zero. Moreover, suppose that the following hold:*

(i) $\sigma\tau \|\sum_{i=1}^m L_i^* L_i\| < 1$,

(ii) $\forall n \in \mathbb{N},\ \rho_n \in]0,2[$,

(iii) $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$,

(iv) $\sum_{n\in\mathbb{N}} \rho_n \|e_{G,n}\| < +\infty$ *and, for* $i = 1,\ldots,m$, $\sum_{n\in\mathbb{N}} \rho_n \|e_{H_i,n}\| < +\infty$.

Then, there exists $(\hat{x}, \hat{y}_1, \ldots, \hat{y}_m) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ *solution to* (40) *such that, in Algorithm 5.1 or in Algorithm 5.2, the sequence* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to* $\hat{x}$ *and, for every* $i = 1, \ldots, m$*, the sequence* $(y_{i,n})_{n\in\mathbb{N}}$ *converges weakly to* $\hat{y}_i$.

Remark 5.5. If one of the functions in (37) is $\|Mx - b\|^2$ for some bounded linear operator M, assigning this term to F or to one of the $H_i \circ L_i$ yields different algorithms. Which one is the most efficient probably depends on the problem at hand and on the way the algorithms are implemented. The author likes to think, without any proof, that considering a smooth functional as the term F whenever one can would yield faster convergence. Also, considering a functional as the term G, instead of another $H_i \circ I$, should be beneficial, especially because we have the feasibility $\tilde{x}_n \in \mathrm{dom}(G)$ for every $n \in \mathbb{N}$, e.g. to force $\tilde{x}_n$ to belong to some closed convex subset of $\mathcal{X}$. Other said, the proposed algorithms are serial in the sense that the gradient descent with respect to F updates the primal estimate before the proximal step with respect to G, which itself "feeds" information into the proximal step with respect to the H_i, in the Gauss-Seidel spirit, within the same iteration. By contrast, the dual variables are updated with respect to the H_i independently, and then combined to form the next primal estimate.

6 Conclusion

The proposed algorithms are able to find minimizers of sums of convex functions. The class of optimization problems captured by the generic template (37) is quite large, since it covers the presence of a smooth function and several proximable functions, composed or not with linear operators. The algorithms proceed by full splitting, as they only make calls to the gradient or proximity operators of the functions and to the linear operators or their adjoints, with parsimony—only once per iteration. As can be seen from the compact form of the algorithms, the auxiliary variables are also restricted to a minimum number, eventually zero when no relaxation is performed. Together with the fact that the variables are updated in a serial way, this promises good convergence speed. Future work will consist in studying the convergence rates and potential accelerations, probably by exhibiting bounds on the primal-dual gap, and in validating experimentally the performances of the algorithms in real large-scale problems of signal and image processing. Also, it is desirable to make the stepsizes σ and τ variable through the iterations, and to provide rules to find their "best" values for the problem at hand.

Acknowledgements

The author wants to thank Jalal Fadili for stimulating discussions.

References

[1] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)

[2] Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. Journal of Mathematical Analysis and Applications **72**, 383–390 (1979)

[3] Combettes, P.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**(5-6), 475–504 (2004)

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

[5] Svaiter, B.F.: On weak convergence of the Douglas-Rachford method. SIAM J. Control Optim. **49**(1), 280–287 (2011)

[6] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)

[7] Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, MIT, Cambridge, MA (1989)

[8] Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Sér. A Math. **255**, 2897–2899 (1962)

[9] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: H.H. Bauschke, R. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer-Verlag, New York (2010)

[10] Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. SIAM Journal on Imaging Sciences **2**(2), 323–343 (2009)

[11] Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. In: Theoretical Foundations and Numerical Methods for Sparse Recovery, vol. 9, pp. 263–340. De Gruyter, Radon Series Comp. Appl. Math. (2010)

[12] Becker, S., Candès, E., Grant, M.: Templates for convex cone problems with applications to sparse signal recovery. Mathematical Programming Computation **3**(3), 165–218 (2011)

[13] Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented Lagrangian method (2010). Preprint, submitted to Mathematical Programming.

[14] Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010)

[15] Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. Journal of Scientific Computing **46**(1), 20–46 (2010)

[16] He, B.S., Yuan, X.M.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective (2010). To appear in SIAM Journal on Imaging Sciences.

[17] Pesquet, J.C., Pustelnik, N.: A parallel inertial proximal optimization method (2010). To appear in Pacific Journal of Optimization.

[18] Briceño-Arias, L.M., Combettes, P.L.: A monotone+skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**(4), 1230–1250 (2011)

[19] Combettes, P.L., Dũng, D., Vũ, B.C.: Proximity for sums of composite functions. Journal of Mathematical Analysis and Applications **380**(2), 680–688 (2011)

[20] Combettes, P.L., Pesquet, J.C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators (2011). Preprint arXiv:1107.0081, to appear in Set-Valued and Variational Analysis.

[21] Raguet, H., Fadili, J., Peyré, G.: Generalized forward-backward splitting (2011). Preprint arXiv:1108.4404v2, submitted to SIAM J. Opt.

[22] Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators (2011). Preprint arXiv:1110.1697, to appear in Adv. Comput. Math.

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

[24] Tseng, P.: Applications of splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**, 119–138 (1991)

[25] Chen, H.G.: Forward-backward splitting techniques: Theory and applications. Ph.D. thesis, University of Washington, Seattle, WA (1994)

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

[27] Rockafellar, R.T.: Conjugate Duality and Optimization. SIAM, Philadelphia, PA (1974)

[28] Robinson, S.M.: Composition duality and maximal monotonicity. Math. Programming **85**, 1–13 (1999)

[29] Pennanen, T.: Dualization of generalized equations of maximal monotone type. SIAM J. Optim. **10**, 809–835 (2000)

[30] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

[31] Rockafellar, R.T.: Minimax theorems and conjugate saddle-functions. Math. Scand. **14**, 151–173 (1964)

[32] McLinden, L.: An extension of Fenchel's duality theorem to saddle functions and dual minimax problems. Pacific J. Math. **50**, 135–158 (1974)

[33] Condat, L.: A generic first-order primal-dual method for convex optimization involving Lipschitzian, proximable and linear composite terms (2011). Working paper hal-00609728.

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

[35] Dupé, F.X., Fadili, M.J., Starck, J.L.: A proximal iteration for deconvolving Poisson noisy images using sparse representations. IEEE Trans. Image Processing **18**(2), 310–321 (2009)

[36] Fadili, J.M., Peyré, G.: Total variation projection with first order schemes. IEEE Trans. Image Processing **20**(3), 657–669 (2011)

[37] Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Programming **64**, 81–101 (1994)