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

Laurent Condat

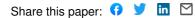**Institutions:** Grenoble Institute of Technology

Related papers:

- A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging

- A splitting algorithm for dual monotone inclusions involving cocoercive operators

- Convex Analysis and Monotone Operator Theory in Hilbert Spaces

- Primal-Dual Splitting Algorithm for Solving Inclusions with Mixtures of Composite, Lipschitzian, and Parallel-Sum Type Monotone Operators

- A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems

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

Laurent Condat

A primal–dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms

Laurent Condat*

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 applied 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 and Douglas–Rachford methods, as well as the recent primal–dual method of Chambolle and Pock designed for problems with linear composite terms.

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

Mathematics Subject Classification (2010) 47H05, 49M29, 49M27, 90C25.

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. Solving ill-posed inverse problems in medical imaging by regularization is an example of high practical interest [1]. 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

---
*Laurent Condat is with GIPSA-lab, CNRS–University of Grenoble, France. Email: Laurent.Condat@gipsa-lab.grenoble-inp.fr

problems. They proceed by *splitting* [2–7], in that the original problem is decomposed into an iterative sequence of much easier subproblems, which involves the functions individually. A 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 [1, 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. [13–24]. The most classical operator splitting methods to minimize the sum of two convex functions are the forward–backward method, proposed in [2] and further developed in [3, 4, 7, 25–28], and the Douglas–Rachford method [3, 6, 7, 22]. 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 [14]. It is fully split in that the gradient, proximity, and linear operators are applied individually; in particular, there is no implicit operation like an inner loop or a applying the inverse of a linear operator. Moreover, the proposed 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 [29–32]. Equivalently, the method finds saddle-points of convex–concave functions with bilinear coupling, here the Lagrangian associated to the primal and dual problems [33, 34].

The paper is organized as follows. In Section 2, we present the convex optimization problem under investigation, along with the corresponding dual formulation and primal–dual variational inclusion. In Section 3, we present the iterative algorithms 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. Section 4 is devoted to the proofs of convergence. Finally, in Section 5, we present parallelized variants of the algorithms, 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 semicontinuous, 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 $J^*(s) := \sup_{s' \in \mathcal{H}} \left[\langle s, s' \rangle - J(s') \right]$ and its proximity operator by $\mathrm{prox}_J(s) := \mathrm{argmin}_{s' \in \mathcal{H}} \left[J(s') + \frac{1}{2} \| s - s' \|^2 \right]$. 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 $\text{sri}(\Omega)$. Finally, we denote by I the identity operator. For background in convex analysis, we refer the readers to textbooks, like [32].

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

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

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

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

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

The corresponding *dual* formulation of the *primal* problem (1) is [32, Chapters 15 and 19]

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

where we note that $(F + G)^*(-L^* y) = \min_{x' \in \mathcal{X}} \left[F^*(-L^* y - x') + G^*(x') \right]$ is an infimal convolution [32, Proposition 15.2]. Without further assumption, the set of solutions to (4) may be empty.

The proposed algorithms are *primal–dual*, in that they solve both the primal and the dual problems (1) and (4), jointly. Another formulation of these two minimization problems is to combine them into the search of a saddle point of the Lagrangian [32, Section 19.2]:

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

The classical Karush–Kuhn–Tucker theory asserts that, if the pair $(\hat{x}, \hat{y})$ is a solution to the monotone variational inclusion

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

then $\hat{x}$ is a solution to (1), $\hat{y}$ is a solution to (4) [32, Theorem 19.1], and $(\hat{x}, \hat{y})$ is a solution to (5) [32, 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:

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

then the set of solutions to (4) is nonempty [32, 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) [32, 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: (i) the composite function $H \circ L$ has been split; (ii) we obtain not only the primal solution $\hat{x}$ but also the dual solution $\hat{y}$, and the proposed algorithms actually use 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:

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

where the additional function $K \in \Gamma_0(\mathcal{Y})$ is such that K^* is differentiable on $\mathcal{Y}$ with a β'-Lipschitz continuous gradient, for some $\beta' \geq 0$. Also, since the proofs in Section 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 resolvents. This more general framework has been considered in the later work [24]. The intention of the author was to keep the study accessible to the practitioner interested in the optimization problem (1).

3   Proposed Algorithms

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

Algorithm 3.1   Choose the proximal parameters $\tau > 0$, $\sigma > 0$, the sequence of positive relaxation parameters $(\rho_n)_{n \in \mathbb{N}}$ and the initial estimate $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$. Then iterate, for every $n \geq 0$,

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

where the error terms $e_{F,n} \in \mathcal{X}$, $e_{G,n} \in \mathcal{X}$, $e_{H,n} \in \mathcal{Y}$ 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 $\text{prox}_{\sigma H^*}(y) = y - \sigma \text{prox}_{H/\sigma}(y/\sigma)$.

Algorithm 3.1 is not symmetric with respect to the primal and dual variables, 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. If we switch the roles of the primal and dual variables, we obtain the following algorithm:

Algorithm 3.2 Choose the proximal parameters $\sigma > 0$, $\tau > 0$, the sequence of positive relaxation parameters $(\rho_n)_{n\in\mathbb{N}}$ and the initial estimate $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$. Then iterate, for every $n \geq 0$,

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

for some error terms $e_{F,n} \in \mathcal{X}$, $e_{G,n} \in \mathcal{X}$, $e_{H,n} \in \mathcal{Y}$.

Now, we claim at once the convergence results for Algorithms 3.1 and 3.2. The proofs are derived in Section 4.

Theorem 3.1 *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 and Algorithm 3.2. 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* $\delta := 2 - \frac{\beta}{2}\left(\frac{1}{\tau} - \sigma\|L\|^2\right)^{-1} \in [1, 2[$,

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

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

Then there exists a pair $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ solution to (6), such that, in Algorithm 3.1 or in Algorithm 3.2, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge weakly to $\hat{x}$ and $\hat{y}$, respectively.

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 and Algorithm 3.2. Suppose that $F = 0$, that the error terms $e_{F,n}$ are all zero, and that the following hold:*

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

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

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

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

Then there exists a pair $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ solution to (6), such that, in Algorithm 3.1 or in Algorithm 3.2, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge weakly to $\hat{x}$ and $\hat{y}$, respectively.

Theorem 3.3 *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are of finite dimension. 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 and Algorithm 3.2. Suppose that $F = 0$, that the error terms $e_{F,n}$ are all zero, and that the following hold:*

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

(ii) $\forall n \in \mathbb{N}$, $\rho_n \in [\varepsilon, 2-\varepsilon]$, *for some $\varepsilon > 0$,*

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

Then there exists a pair $(\hat{x}, \hat{y}) \in \mathcal{X} \times \mathcal{Y}$ solution to (6), such that, in Algorithm 3.1 or in Algorithm 3.2, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge to $\hat{x}$ and $\hat{y}$, respectively.

Remark 3.1 The difference between Theorems 3.2 and 3.3 is that $\sigma\tau\|L\|^2 = 1$ is allowed in the latter. This is a significant improvement: in practice, one can set $\sigma = 1/(\tau\|L\|^2)$ in the algorithms and have only one parameter left, namely τ, to tune, like in the Douglas–Rachford method.

Remark 3.2 If β, the Lipschitz constant defined in (2), is zero, then 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 two equivalent formulations of the same problem and exactly the same calculations in Algorithms 3.1 and 3.2. Thus, the convergence results of Theorems 3.2 and 3.3 also hold if the assumption $F = 0$ is relaxed to $\beta = 0$.

3.1   Relationship to Existing Optimization Methods

The proposed Algorithms 3.1 and 3.2 are able to solve the general problem (1) iteratively, by applying the operators ∇F, $\text{prox}_{\tau G}$, $\text{prox}_{\sigma H^*}$, L and L^*, without any other implicit (inverse) operator or inner loop. In particular, no inverse operator of the form $(I + \alpha L^* L)^{-1}$ is required. To the author's knowledge, the only existing method of the literature having this feature is the recent one of Combettes and Pesquet [21], based on a different splitting of (6). Their algorithm requires two applications of ∇F, L, L^* per iteration, against only one with the proposed algorithm. The convergence speed of the two algorithms has not been compared in practical applications, yet.

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 [35–37]. 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, Algorithms 3.1 and 3.2 revert to classical splitting methods of the literature.

3.1.1 The Case $F = 0$

If the smooth term F is absent of the problem, Algorithms 3.1 and 3.2 exactly revert to the primal–dual algorithms of Chambolle and Pock [14], which have been proposed in other forms in [15, 16]. Chambolle and Pock proved the convergence in the finite-dimensional case, assuming that $\tau\sigma\|L\|^2 < 1$ and $\rho_n \equiv 1$, and without error terms [14]. The convergence has been proved in a different way in [17] with a constant relaxation parameter $\rho_n \equiv \rho \in {]}0,2{[}$ and the same other hypotheses. Thus, Theorems 3.2 and 3.3 extend the domain of convergence of these primal–dual algorithms.

When $F = 0$, the primal–dual method in [19] and the method in [38] can be used to solve (1), as well. They yield algorithms different from Algorithms 3.1 and 3.2.

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 $\frac{\lambda}{2}\|M \cdot -b\|^2$, for some bounded linear operator M, element b and positive real λ, the proximity operator of G requires applying 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 $\frac{\lambda}{2}\|M \cdot -b\|^2$ as the function F in (1) yields an algorithm with simple applications of M and M^*.

If $F = 0$ and $L = I$, let us discuss the relationship of Algorithm 3.1 with the classical Douglas–Rachford splitting method [3, 6, 7, 22]. For simplicity, we suppose 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 3.3 Choose the proximal parameters $\tau > 0$, $\sigma > 0$ and the initial estimate $(x_0, s_0) \in \mathcal{X}^2$. Then iterate, for every $n \geq 0$,

$$
\begin{array}{ll}
1. & x_{n+1} := \mathrm{prox}_{\tau G}\big((1 - \tau\sigma)x_n + \tau\sigma s_n\big), \\
2. & s_{n+1} := s_n - x_{n+1} + \mathrm{prox}_{\frac{1}{\sigma}H}(2x_{n+1} - s_n).
\end{array}
\tag{11}
$$

If, additionally, we set $\sigma = 1/\tau$, Algorithm 3.3 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 in [22] that x_n converges weakly to a primal solution $\hat{x} = \mathrm{prox}_{\tau G}(\hat{s})$ and $y_n = \sigma(x_n - s_n)$ converges weakly to a dual solution $\hat{y}$. The convergence also follows from Theorem 3.3.

Interestingly, in the case $\sigma\tau < 1$, Algorithm 3.3 was studied in [18] 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 [18] and Algorithm 3.1.

3.1.2 The Case $H \circ L = 0$

If the term $H \circ L$ is removed from the formulation, the primal and dual problems are uncoupled. Then, in Algorithm 3.1, we can keep only the computations involving the primal variable, so that we obtain the iteration

$$x_{n+1} = \rho_n \text{prox}_{\tau G}\big(x_n - \tau(\nabla F(x_n) + e_{F,n})\big) + e_{G,n} + (1 - \rho_n)x_n, \tag{12}$$

which is exactly the forward–backward splitting method [7, 26, 28]. 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 $[\varepsilon, 1]$ for some $\varepsilon > 0$, and the error terms are absolutely summable. The convergence also follows from Theorem 3.1 under weaker assumptions. Indeed, 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.

On the other hand, if we set $G = 0$ and $L = I$, Algorithms 3.1 and 3.2 are different from the forward–backward method, 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 3.1, 3.1 and 3.3.

First, we recall some definitions and properties of operator theory. We refer the readers to [32] for more details. In the following, $\mathcal{H}$ is a real Hilbert space. Let $M : \mathcal{H} \rightrightarrows \mathcal{H}$ 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 iff $\forall (u, u') \in \mathcal{H}^2, \forall (v, v') \in Mu \times Mu', \langle u - u', v - v' \rangle \geq 0$ and maximally monotone iff 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 $\text{fix}(T) := \{x \in \mathcal{H} : Tx = x\}$. T is said to be nonexpansive iff it is 1-Lipschitz continuous on $\mathcal{H}$, see (2), and firmly nonexpansive iff $2T - I$ is nonexpansive. Let $\alpha \in]0, 1]$. T is said to be α-averaged iff 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 $M : \mathcal{H} \rightrightarrows \mathcal{H}$ 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 $(I + \partial J)^{-1} = \text{prox}_J$.

Lemma 4.1 (Krasnosel'skii–Mann iteration) [7, Lemma 5.1] *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator, $(\rho_n)_{n \in \mathbb{N}}$ be a sequence in $]0, 1[$ and $(e_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $\text{fix}(T) \neq \emptyset$, $\sum_{n \in \mathbb{N}} \rho_n(1 - \rho_n) = +\infty$, and $\sum_{n \in \mathbb{N}} \rho_n \|e_n\| < +\infty$. Let $s_0 \in \mathcal{H}$ and $(s_n)_{n \in \mathbb{N}}$ be the sequence in $\mathcal{H}$ such that, for every $n \in \mathbb{N}$,*

$$s_{n+1} = s_n + \rho_n\big(T(s_n) + e_n - s_n\big). \tag{13}$$

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

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

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

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

Proof Set $T := (I + M)^{-1}$, $T' := 2T - I$, 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 (14) as

$$s_{n+1} = s_n + \rho'_n\big(T'(s_n) + e'_n - s_n\big). \tag{15}$$

Moreover, $T \in \mathscr{A}(\mathcal{H}, \frac{1}{2})$ [32, Corollary 23.8], so that $T' \in \mathscr{A}(\mathcal{H}, 1)$, and $\mathrm{fix}(T') = \mathrm{fix}(T) = \mathrm{zer}(M)$ [32, Proposition 23.38]. Thus, we obtain the desired result by applying Lemma 4.1 to the iteration (15). $\qquad\square$

Lemma 4.3 (Composition of averaged operators) [39, Theorem 3] *Let $\alpha_1 \in]0,1[$, $\alpha_2 \in]0,1]$, $T_1 \in \mathscr{A}(\mathcal{H}, \alpha_1)$, and $T_2 \in \mathscr{A}(\mathcal{H}, \alpha_2)$. Then $T_1 \circ T_2 \in \mathscr{A}(\mathcal{H}, \alpha')$, where*

$$\alpha' := \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}. \tag{16}$$

Lemma 4.4 (forward–backward iteration) *Let $M_1 : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $\kappa \in]0, +\infty[$, let $M_2 : \mathcal{H} \to \mathcal{H}$ be κ-cocoercive; that is, $\kappa M_2 \in \mathscr{A}(\mathcal{H}, \frac{1}{2})$. Suppose that $\mathrm{zer}(M_1 + M_2) \neq \emptyset$. Let $\gamma \in]0, 2\kappa]$, and set $\delta := 2 - \frac{\gamma}{2\kappa}$. 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 $\sum_{n \in \mathbb{N}} \rho_n \|e_{2,n}\| < +\infty$. Let $s_0 \in \mathcal{H}$ and $(s_n)_{n \in \mathbb{N}}$ be the sequence in $\mathcal{H}$ such that, for every $n \in \mathbb{N}$,*

$$s_{n+1} = \rho_n\Big((I + \gamma M_1)^{-1}\big(s_n - \gamma(M_2(s_n) + e_{2,n})\big) + e_{1,n}\Big) + (1 - \rho_n)s_n. \tag{17}$$

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

Proof Set $T_1 := (I + \gamma M_1)^{-1}$, $T_2 := I - \gamma M_2$, and $T := T_1 \circ T_2$. Then $T_1 \in \mathscr{A}(\mathcal{H}, \frac{1}{2})$ [32, Corollary 23.8] and $T_2 \in \mathscr{A}(\mathcal{H}, \frac{\gamma}{2\kappa})$ [32, 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 (17) as (15). Moreover, T' is nonexpansive, $\mathrm{fix}(T') = \mathrm{fix}(T) = \mathrm{zer}(M_1 + M_2)$ [32, Proposition 25.1(iv)] and $\sum_{n \in \mathbb{N}} \rho'_n(1 - \rho'_n) = +\infty$. From the nonexpansiveness of T_1, we obtain, $\forall n \in \mathbb{N}$,

$$\|e'_n\| \leq \delta \|e_{1,n}\| + \delta \big\| T_1\big(T_2(s_n) + e_{2,n}\big) - T_1\big(T_2(s_n)\big)\big\| \leq \delta \|e_{1,n}\| + \delta \|e_{2,n}\|. \tag{18}$$

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

Lemma 4.5 (Baillon–Haddad theorem) [32, Corollary 18.16]. *Let $J : \mathcal{H} \to \mathbb{R}$ be convex, differentiable on $\mathcal{H}$ 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.*

Lemma 4.6 [40, Lemma 3]   *Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$ be sequences of reals, such that the following hold:*

(i) $0 \le a_n$ *and* $0 \le c_n < 1$ *and* $0 \le b_n$, $\forall n \in \mathbb{N}$,

(ii) $a_{n+1} \le c_n a_n + b_n$, $\forall n \in \mathbb{N}$,

(iii) $\sum_{n\in\mathbb{N}}(1 - c_n) = +\infty$,

(iv) $b_n/(1 - c_n) \to 0$.

Then $a_n \to 0$.

Proof of Theorem 3.1 for Algorithm 3.1   In essence, we show that Algorithm 3.1 has the structure of a forward–backward iteration, when expressed in terms of nonexpansive operators on $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$, equipped with a particular inner product.

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

$$\langle z, z' \rangle_I := \langle x, x' \rangle + \langle y, y' \rangle, \text{ for every } z = (x, y), z' = (x', y') \in \mathcal{Z}. \tag{19}$$

By endowing $\mathcal{Z}$ with this inner product, we obtain the Hilbert space denoted by $\mathcal{Z}_I$. Let us define the bounded linear operator on $\mathcal{Z}$,

$$P : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\tau}I & -L^* \\ -L & \frac{1}{\sigma}I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{20}$$

In $\mathcal{Z}_I$, P is bounded, self-adjoint, and, from the assumptions $\beta > 0$ and (i), strictly positive; that is, $\langle z, Pz \rangle_I > 0$, for every $z \ne 0$. 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

$$\langle z, z' \rangle_P := \langle z, Pz' \rangle_I, \text{ for every } (z, z') \in \mathcal{Z}^2. \tag{21}$$

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 $z \in \mathcal{Z}$, $\langle z_n - \hat{z}, Pz \rangle_I \to 0$ 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:

$$-\underbrace{\begin{pmatrix} \nabla F(x_n) \\ 0 \end{pmatrix}}_{B(z_n)} \in \underbrace{\begin{pmatrix} \partial G(\tilde{x}_{n+1}) + L^* \tilde{y}_{n+1} \\ -L\tilde{x}_{n+1} + \partial H^*(\tilde{y}_{n+1}) \end{pmatrix}}_{A(\tilde{z}_{n+1})} + \underbrace{\begin{pmatrix} \frac{1}{\tau}I & -L^* \\ -L & \frac{1}{\sigma}I \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} \tilde{x}_{n+1} - x_n \\ \tilde{y}_{n+1} - y_n \end{pmatrix}}_{(\tilde{z}_{n+1}-z_n)}, \tag{22}$$

or equivalently,

$$\tilde{z}_{n+1} = (I + P^{-1} \circ A)^{-1} \circ (I - P^{-1} \circ B)(z_n). \tag{23}$$

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

$$z_{n+1} = \rho_n\big((I + P^{-1} \circ A)^{-1}\big(z_n - P^{-1} \circ B(z_n) - e_{2,n}\big) + e_{1,n}\big) + (1 - \rho_n)z_n, \tag{24}$$

10

where $e_{1,n} := (e_{G,n}, e_{H,n})$ and $e_{2,n} := P^{-1}(e_{F,n}, -2Le_{G,n})$, and we recognize the structure of the forward–backward iteration (17). Thus, it is sufficient to check the conditions of application of Lemma 4.4 for the convergence result of Theorem 3.1 to follow. To that end, set $\mathscr{H} := \mathscr{Z}_P$, $M_1 := P^{-1} \circ A$, $M_2 := P^{-1} \circ B$, and $\kappa := (\frac{1}{\tau} - \sigma \|L\|^2)/\beta$. Then,

• The operator $(x, y) \mapsto \partial G(x) \times \partial H^*(y)$ is maximally monotone in $\mathscr{Z}_I$, by Theorem 20.40, Corollary 16.24, Propositions 20.22 and 20.23 of [32]. Moreover, the skew operator $(x, y) \mapsto (L^* y, -Lx)$ is maximally monotone in $\mathscr{Z}_I$ [32, Example 20.30] and has full domain. Hence, A is maximally monotone [32, Corollary 24.4(i)]. Thus, M_1 is monotone in $\mathscr{Z}_P$ and, from the injectivity of P, M_1 is maximally monotone in $\mathscr{Z}_P$.

• Let us show the cocoercivity of M_2. For every $z = (x, y)$, $z' = (x', y') \in \mathscr{Z}$, we have

$$\|M_2(z) - M_2(z')\|_P^2 = \left\langle P^{-1} \circ B(z) - P^{-1} \circ B(z'), B(z) - B(z') \right\rangle_I \tag{25}$$

$$= \left\langle \frac{1}{\sigma} \left(\frac{1}{\sigma\tau} I - L^* L \right)^{-1} \left(\nabla F(x) - \nabla F(x') \right), \nabla F(x) - \nabla F(x') \right\rangle \tag{26}$$

$$\leq \frac{1}{\sigma} \left(\frac{1}{\sigma\tau} - \|L\|^2 \right)^{-1} \|\nabla F(x) - \nabla F(x')\|^2 \tag{27}$$

$$\leq \frac{\beta^2}{\sigma} \left(\frac{1}{\sigma\tau} - \|L\|^2 \right)^{-1} \|x - x'\|^2 = \frac{\beta}{\kappa} \|x - x'\|^2. \tag{28}$$

We define the linear operator $Q : (x, y) \mapsto (x, 0)$ of $\mathscr{Z}$. Since $P - \beta\kappa Q$ is positive in $\mathscr{Z}_I$, we have

$$\beta\kappa \|x - x'\|^2 = \beta\kappa \langle (z - z'), Q(z - z') \rangle_I \leq \langle (z - z'), P(z - z') \rangle_I = \|z - z'\|_P^2. \tag{29}$$

Putting together (28) and (29), we get

$$\kappa \|M_2(z) - M_2(z')\|_P \leq \|z - z'\|_P, \tag{30}$$

so that κM_2 is nonexpansive in $\mathscr{Z}_P$. Let us define on $\mathscr{Z}$ the function $J : (x, y) \mapsto F(x)$. Then, in $\mathscr{Z}_P$, $\nabla J = M_2$. Therefore, from Lemma 4.5, κM_2 is firmly nonexpansive in $\mathscr{Z}_P$.

• We set $\gamma := 1$. Since $\kappa \geq \frac{1}{2}$ from (i), $\gamma \in {]0, 2\kappa]}$. Moreover, $\delta = 2 - \frac{\gamma}{2\kappa}$.

• Since P^{-1} and L are bounded and the norms $\|\cdot\|_I$ and $\|\cdot\|_P$ are equivalent, we get from 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. $\qquad\square$

Proof of Theorem 3.2 for Algorithm 3.1 We follow the same line of proof as for Theorem 3.1. From assumption (i), P is still bounded from below and defines a valid scalar product on $\mathscr{Z}$. $F = 0$, so that $B = 0$ and (24) becomes

$$z_{n+1} = \rho_n\big((I + P^{-1} \circ A)^{-1}(z_n) + e_n\big) + (1 - \rho_n)z_n, \tag{31}$$

where $e_n := (e_{G,n}, e_{H,n}) + (I + P^{-1} \circ A)^{-1}\big(z_n + P^{-1}(0, 2Le_{G,n})\big) - (I + P^{-1} \circ A)^{-1}(z_n)$, $\forall n \in \mathbb{N}$. Since P, P^{-1}, L are bounded and $I + P^{-1} \circ A$ is nonexpansive in $\mathscr{Z}_P$, we get from assumption (iv) that $\sum_{n \in \mathbb{N}} \rho_n \|e_n\|_P < +\infty$. We recognize in (31) the structure of the proximal point algorithm (14) in $\mathscr{Z}_P$, as observed in [17]. Thus, the convergence result of Theorem 3.2 follows from Lemma 4.2, whose conditions of application have

been checked in the proof of Theorem 3.1. $\qquad\square$

Proof of Theorem 3.3 for Algorithm 3.1 We keep the definitions and notations of the proof of Theorem 3.1. We introduce S, the orthogonal projector in $\mathcal{Z}_I$ onto the closed subspace $\mathrm{ran}(P)$. From assumption (i), P is only guaranteed to be positive; that is, $\langle z, Pz \rangle_I \geq 0$, $\forall z \in \mathcal{Z}$. So, it does not define a valid inner product on $\mathcal{Z}$. However, $R := P + I - S$ is strictly positive, so that we can define the Hilbert space $\mathcal{Z}_R$ by endowing $\mathcal{Z}$ with the inner product $\langle z, z' \rangle_R := \langle z, Rz' \rangle_I$, $\forall (z, z') \in \mathcal{Z}^2$. R is bounded from below, so that $\langle \cdot, \cdot \rangle_R$ and $\langle \cdot, \cdot \rangle_I$ are equivalent. Let T be the operator defined by the steps 1. and 2. of Algorithm 3.1, in absence of error terms ; that is, $T : (x, y) \mapsto (\tilde{x}, \tilde{y})$ with $(\tilde{x}, \tilde{y}) = \big(\mathrm{prox}_{\tau G}(x - \tau L^* y), \mathrm{prox}_{\sigma H^*}(y + \sigma L(2\tilde{x} - x))\big)$. Let $n \in \mathbb{N}$. We have $\tilde{z}_{n+1} = T(z_n) + e_n$, where $e_n := \big(e_{G,n}, e_{H,n} + \mathrm{prox}_{\sigma H^*}(y_n + \sigma L(2\tilde{x}_{n+1} - x_n)) - \mathrm{prox}_{\sigma H^*}(y_n + \sigma L(2\tilde{x}_{n+1} - 2e_{G,n} - x_n))\big)$. From the nonexpansiveness of $\mathrm{prox}_{\sigma H^*}$ and by assumption (iii), $\sum_{n \in \mathbb{N}} \|e_n\|_R < +\infty$. Further on, we can rewrite the steps 1. and 2. of Algorithm 3.1 as

$$\tilde{x}_{n+1} = \mathrm{prox}_{\tau G}(\tau u_n) + e_{G,n}, \tag{32}$$

$$\tilde{y}_{n+1} = \mathrm{prox}_{\sigma H^*}\big(2\sigma L\tilde{x}_{n+1} + \sigma v_n\big) + e_{H,n}, \tag{33}$$

where $(u_n, v_n) := P(x_n, y_n)$. Since $P \circ S = P$, this shows that $T \circ S = T$. For every $n \in \mathbb{N}$, set $z'_n := Sz_n$, $e'_n := Se_n$, $T' := S \circ T$, so that we have

$$z'_{n+1} = \rho_n\big(T'(z'_n) + e'_n\big) + (1 - \rho_n)z'_n. \tag{34}$$

Let $z, z' \in \mathcal{Z}$. From (22), we have

$$0 \in A(T(z)) + PT(z) - Pz, \tag{35}$$

$$0 \in A(T(z')) + PT(z') - Pz', \tag{36}$$

so that $(T(z), Pz - PT(z))$ and $(T(z'), Pz' - PT(z'))$ belong to $\mathrm{gra}(A)$. Hence, by monotonicity of A in $\mathcal{Z}_I$, we get

$$0 \leq \langle T(z) - T(z'), Pz - PT(z) - Pz' + PT(z') \rangle_I \tag{37}$$

$$= \langle ST(z) - ST(z'), Pz - PT(z) - Pz' + PT(z') \rangle_I \tag{38}$$

$$\text{(because the right argument belongs to } \mathrm{ran}(P))$$

$$= \langle ST(z) - ST(z'), P(z - z') - PST(z) + PST(z') \rangle_I \tag{39}$$

$$= -\|ST(z) - ST(z')\|_R^2 + \langle ST(z) - ST(z'), z - z' \rangle_R. \tag{40}$$

Thus, by [32, Prop. 4.2], $T' = S \circ T$ is firmly nonexpansive in $\mathcal{Z}_R$. Let $z \in \mathrm{zer}(A) \neq \emptyset$. Then $(T(z), Pz - PT(z))$ and $(z, 0)$ belong to $\mathrm{gra}(A)$, and by monotonicity of A in $\mathcal{Z}_I$, we get

$$0 \leq \langle T(z) - z, Pz - PT(z) \rangle_I. \tag{41}$$

But since P is positive, we also have

$$0 \leq \langle T(z) - z, PT(z) - Pz \rangle_I. \tag{42}$$

Combining (41) and (42) yields $\langle T(z) - z, PT(z) - Pz \rangle_I = 0$ and, by [32, Corollary 18.17], $PT(z) - Pz = 0$. This implies that $ST(z) = ST(Sz) = Sz$, so that Sz is a fixed point of T' and $\text{fix}(T') \neq \emptyset$. Conversely, by (35), $z \in \text{fix}(T') \Rightarrow PT(z) = Pz \Rightarrow T(z) \in \text{zer}(A)$. All together with assumption (ii), the conditions are met to apply Lemma 4.1 to the iteration (34), after a change of variables like in the proof of Lemma 4.2, so that $(z'_n)_{n \in \mathbb{N}}$ converges to $\hat{z}' \in \text{fix}(T')$.

T is continuous, by continuity of the proximity operators from which it is defined. Therefore, since $\tilde{z}_{n+1} = T(z'_n) + e_n$, $\forall n \in \mathbb{N}$, and $(e_n)_{n \in \mathbb{N}}$ converges to zero, $(\tilde{z}_n)_{n \in \mathbb{N}}$ converges to $\hat{z} := T(\hat{z}') \in \text{zer}(A)$. Moreover,

$$\|z_{n+1} - \hat{z}\| \leq \rho_n \|\tilde{z}_{n+1} - \hat{z}\| + |1 - \rho_n| \|z_n - \hat{z}\|, \quad \forall n \in \mathbb{N}. \tag{43}$$

For every $n \in \mathbb{N}$, set $a_n := \|z_n - \hat{z}\|$, $b_n := \rho_n \|\tilde{z}_{n+1} - \hat{z}\|$, $c_n := |1 - \rho_n|$. Then $b_n \to 0$ and, by assumption (ii), $c_n \in [0, 1 - \varepsilon]$ $\forall n \in \mathbb{N}$, $\sum_{n \in \mathbb{N}} (1 - c_n) = +\infty$, and $b_n / (1 - c_n) \to 0$. So, by Lemma 4.6, $a_n \to 0$, so that $(z_n)_{n \in \mathbb{N}}$ converges to $\hat{z} \in \text{zer}(A)$. $\square$

Proof of Theorems 3.1, 3.2, 3.3 for Algorithm 3.2 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 3.2:

$$-\underbrace{\begin{pmatrix} \nabla F(x_n) \\ 0 \end{pmatrix}}_{B(z_n)} \in \underbrace{\begin{pmatrix} \partial G(\tilde{x}_{n+1}) + L^* \tilde{y}_{n+1} \\ -L\tilde{x}_{n+1} + \partial H^* (\tilde{y}_{n+1}) \end{pmatrix}}_{A(\tilde{z}_{n+1})} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & L^* \\ L & \frac{1}{\sigma} I \end{pmatrix}}_{P'} \underbrace{\begin{pmatrix} \tilde{x}_{n+1} - x_n \\ \tilde{y}_{n+1} - y_n \end{pmatrix}}_{(\tilde{z}_{n+1} - z_n)}. \tag{44}$$

Thus, after comparison of (22) and (44), it appears that the whole convergence analysis for Algorithm 3.1 applies to Algorithm 3.2, just replacing P by P' in the derivations. Therefore, both algorithms converge under the same assumptions.

5 Extension to Several Composite Functions

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

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\arg\min} \left[F(x) + G(x) + \sum_{i=1}^{m} H_i(L_i x) \right], \tag{45}$$

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 $L_i : \mathcal{X} \to \mathcal{Y}_i$. At the same time, we consider the dual problem

$$\text{Find } (\hat{y}_1, \ldots, \hat{y}_m) \in \underset{y_1 \in \mathcal{Y}_1, \ldots, y_m \in \mathcal{Y}_m}{\arg \min} \left[(F + G)^* \left(-\sum_{i=1}^{m} L_i^* y_i \right) + \sum_{i=1}^{m} H_i^* (y_i) \right] \tag{46}$$

$$\equiv \text{Find } (\hat{y}_1, \ldots, \hat{y}_m) \in \underset{y_1 \in \mathcal{Y}_1, \ldots, y_m \in \mathcal{Y}_m}{\arg \min} \underset{x' \in \mathcal{X}}{\min} \left[F^* \left(-\sum_{i=1}^{m} L_i^* y_i - x' \right) + G^*(x') + \sum_{i=1}^{m} H_i^* (y_i) \right], \tag{47}$$

13

and the corresponding monotone variational inclusion:

$$\text{Find } (\hat{x}, \hat{y}_1, \ldots, \hat{y}_m) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m \text{ such that } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G(\hat{x}) + \sum_{i=1}^m L_i^* \hat{y}_i + \nabla F(\hat{x}) \\ -L_1 \hat{x} + \partial H_1^* (\hat{y}_1) \\ \vdots \\ -L_m \hat{x} + \partial H_m^* (\hat{y}_m) \end{pmatrix}. \tag{48}$$

Thus, if $(\hat{x}, \hat{y}_1, \ldots, \hat{y}_m)$ is solution to (48), then $\hat{x}$ is solution to (45) and $(\hat{y}_1, \ldots, \hat{y}_m)$ is solution to (46). In the following, we suppose that the set of solutions to (48) is nonempty.

To the author's knowledge, the only method in the literature able to solve (48), in whole generality and by full splitting (that is, without applications of operators other than ∇F, prox_G, $\text{prox}_{H_i^*}$, L_i, L_i^*), is the one in [21]. We note that another recent method has been proposed in [23] to solve (45), in the more restrictive setting where $G = 0$ and $L_i = I$, for every $i = 1, \ldots, m$.

Although the primal and dual problems (45) and (46) 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 $\boldsymbol{e}_{H,n} := (e_{H_1,n}, \ldots, e_{H_m,n})$, $\forall n \in \mathbb{N}$. We have the following properties, $\forall \boldsymbol{y} \in \boldsymbol{\mathcal{Y}}$:

$$\boldsymbol{H}^*(\boldsymbol{y}) = \sum_{i=1}^m H_i^*(y_i), \tag{49}$$

$$\boldsymbol{L}^* \boldsymbol{y} = \sum_{i=1}^m L_i^* y_i, \tag{50}$$

$$\text{prox}_{\sigma \boldsymbol{H}^*}(\boldsymbol{y}) = \left(\text{prox}_{\sigma H_1^*}(y_1), \ldots, \text{prox}_{\sigma H_m^*}(y_m) \right), \tag{51}$$

$$\|\boldsymbol{L}\|^2 = \left\| \sum_{i=1}^m L_i^* L_i \right\|. \tag{52}$$

Thus, we can rewrite (45) and (46) as

$$\text{Find } \hat{x} \in \underset{x \in \mathcal{X}}{\arg\min} \left[F(x) + G(x) + \boldsymbol{H}(\boldsymbol{L}x) \right], \tag{53}$$

$$\text{Find } \hat{\boldsymbol{y}} \in \underset{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}}{\arg\min} \left[(F + G)^*(-\boldsymbol{L}^* \boldsymbol{y}) + \boldsymbol{H}^*(\boldsymbol{y}) \right], \tag{54}$$

which exactly take the form of (1) and (4), respectively. Accordingly, we can rewrite Algorithms 3.1 and 3.2 by doing the appropriate substitutions, and we obtain the two following algorithms, respectively:

Algorithm 5.1 Choose the proximal parameters $\tau > 0$, $\sigma > 0$, the sequence of positive relaxation parameters $(\rho_n)_{n\in\mathbb{N}}$ 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 $n \geq 0$,

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

$$\tag{55}$$

for some error terms $e_{F,n} \in \mathcal{X}$, $e_{G,n} \in \mathcal{X}$, $e_{H_i,n} \in \mathcal{Y}_i$.

Algorithm 5.2 Choose the proximal parameters $\sigma > 0$, $\tau > 0$, the sequence of positive relaxation parameters $(\rho_n)_{n\in\mathbb{N}}$ 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 $n \geq 0$,

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

$$\tag{56}$$

for some error terms $e_{F,n} \in \mathcal{X}$, $e_{G,n} \in \mathcal{X}$, $e_{H_i,n} \in \mathcal{Y}_i$.

By doing the same substitutions in Theorems 3.1, 3.2 and 3.3, we obtain the following spin-off theorems, respectively:

Theorem 5.1 *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) $\quad \dfrac{1}{\tau} - \sigma \left\| \sum_{i=1}^m L_i^* L_i \right\| \geq \dfrac{\beta}{2}$,

(ii) $\quad \forall n \in \mathbb{N}, \ \rho_n \in \,]0, \delta[, \ \text{where we set } \delta := 2 - \dfrac{\beta}{2}\left(\frac{1}{\tau} - \sigma \|\sum_{i=1}^m L_i^* L_i\|\right)^{-1} \in [1, 2[,$

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

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

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 (48) *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 5.2 *Let $\tau > 0$, $\sigma > 0$ and the sequences $(\rho_n)_{n\in\mathbb{N}}$, $(e_{F,n})_{n\in\mathbb{N}}$, $(e_{G,n})_{n\in\mathbb{N}}$, $(e_{H_i,n})_{\{i=1,\ldots,m,n\in\mathbb{N}\}}$, be the parameters of Algorithm 5.1 and Algorithm 5.2. Suppose that $F = 0$, that the error terms $e_{F,n}$ are all zero, and that the following hold:*

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

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

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

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

Then there exists $(\hat{x}, \hat{y}_1,\ldots,\hat{y}_m) \in \mathscr{X} \times \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_m$ solution to (48) 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 5.3 *Suppose that $\mathscr{X}$ and $\mathscr{Y}$ are of finite dimension. 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 $F = 0$, that the error terms $e_{F,n}$ are all zero, and that the following hold:*

(i) $\sigma\tau\Big\| \sum\limits_{i=1}^{m} L_i^* L_i \Big\| \leq 1,$

(ii) $\forall n \in \mathbb{N},\ \rho_n \in [\varepsilon, 2-\varepsilon],$ *for some $\varepsilon > 0$,*

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

Then there exists $(\hat{x}, \hat{y}_1,\ldots,\hat{y}_m) \in \mathscr{X} \times \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_m$ solution to (48) such that, in Algorithm 5.1 or in Algorithm 5.2, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to $\hat{x}$ and, for every $i = 1,\ldots,m$, the sequence $(y_{i,n})_{n\in\mathbb{N}}$ converges to $\hat{y}_i$.

Remark 5.1   If one of the functions in (45) is $\|M\cdot -b\|^2$, for some bounded linear operator M and element b, assigning this term to F or to one of the $H_i \circ L_i$ yields different algorithms. Which one is the most efficient depends on the problem at hand and on the way the algorithms are implemented. But considering a smooth functional as the term F, whenever one can, should be beneficial for the convergence speed. Indeed, 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. For the same reason, considering a functional as the term G, instead of another $H_i \circ I$, should be beneficial as well, especially because we have the feasibility $\tilde{x}_n \in \text{dom}(G)$ for every $n \in \mathbb{N}$, e.g. to force $\tilde{x}_n$ to belong to some closed convex subset of $\mathscr{X}$.

16

6 Conclusion

The class of optimization problems captured by the template problem (45) is quite large, as it covers the presence of a smooth function and several proximable functions, composed or not with linear operators. The proposed algorithms proceed by full splitting, as they only apply the gradient or proximity operators of the functions and the linear operators or their adjoints, with parsimony—only once per iteration. The auxiliary variables are restricted to a minimum number, eventually zero when no relaxation is performed. 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 large-scale problems of practical interest. Also, it is desirable to make the proximal parameters σ and τ variable through the iterations, and to provide rules to find their "best" values for the problem at hand.

Acknowledgements

References

[1] Sidky, E.Y., Jørgensen, J.H., Pan, X.: Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. Physics in Medicine and Biology **57**(10), 3065–3091 (2012)

[2] Mercier, B.: Topics in finite element solution of elliptic problems. No. 63 in lectures on mathematics. Tata Institute of Fundamental Research, Bombay (1979)

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

[4] 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)

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

[6] 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)

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

[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] 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)

[15] 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)

[16] 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)

[17] He, B.S., Yuan, X.M.: Convergence analysis of primal–dual algorithms for a saddle-point problem: from contraction perspective. SIAM Journal on Imaging Sciences **5**, 119–149 (2012)

[18] Pesquet, J.C., Pustelnik, N.: A parallel inertial proximal optimization method. Pacific Journal of Optimization **8**(2), 273–305 (2012)

[19] 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)

[20] 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)

[21] Combettes, P.L., Pesquet, J.C.: Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued and Variational Analysis **20**(2), 307–330 (2012)

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

[23] Raguet, H., Fadili, J., Peyré, G.: Generalized forward–backward splitting (2011). Preprint arXiv:1108.4404

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

[25] 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)

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

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

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

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

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

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

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

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

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

[35] 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)

[36] 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)

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

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

[39] Ogura, N., Yamada, I.: Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping. Numer. Funct. Anal. and Optimiz. **23**(1–2), 113–137 (2002)

[40] Polyak, B.T.: Introduction to Optimization. Optimization Software, Inc., Publications Division, New York, NY, USA (1987)