Infimal post-composition approach for composite convex optimization. Application to image restoration.

L M Briceño-Arias, Nelly Pustelnik

Infimal post-composition approach for composite convex optimization. Application to image restoration.

L. M. Briceño-Arias and N. Pustelnik *†‡

November 16, 2023

Abstract

In this paper we introduce a new approach for solving image restoration problems by using the infimal postcomposition of a convex function by a linear operator. We derive this formulation for general linear composite convex problems in Hilbert spaces and we provide globally weakly convergent algorithms based on the Douglas-Rachford splitting. We apply our algorithms to the image restoration problem, giving an explicit closed expression for the proximity operator of the infimal postcomposition. Comprehensive numerical experiments are performed in order to serve two key objectives: first, to highlight the advantages of the proposed procedure over a wide array of state-of-the-art methods, considering diverse levels of image degradations; and second, to assess the impact of TV-l2 penalization, which introduces strong convexity while maintaining high performance, contrary to conventional beliefs.

Douglas-Rachford algorithm, infimal post-composition, composite convex optimization problems, image restoration, splitting algorithms, strong convexity.

*The work of L. M. Briceño-Arias was supported by Centro de Modelamiento Matemático (CMM), FB210005, BASAL funds for centers of excellence, FONDECYT 1190871, and FONDECYT 1230257 from ANID-Chile. The work of N. Pustelnik is supported by the ANR (Agence Nationale de la Recherche) from France ANR-19-CE48-0009 Multisc'In and Fondation Simone et Cino Del Duca.

†Luis Briceño-Arias is with Department of Mathematics, Universidad Técnica Federico Santa María, Chile (e-mail: luis.briceno@usm.cl).

‡Nelly Pustelnik is with Univ Lyon, Ens de Lyon, Univ Lyon 1, CNRS, Laboratoire de Physique, Lyon, 69342, France (e-mail: nelly.pustelnik@lens-lyon.fr).

1 Introduction

Image reconstruction is a crucial task in image processing. It is based on the formulation of the direct model involving the observation $y = A\overline{x} + \eta$, where A is often considered as a linear operator and η is a stochastic perturbation, i.e., a noise, and consists in recovering an image $\widehat{x}$ close from the original image $\overline{x}$.

The resolution of the image reconstruction problem by standard variational approaches, including total variation (TV) reconstruction, relies on optimization algorithms solving linearly composite convex optimization problems of the form

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \frac{1}{2}\|Ax - y\|^2 + g(Lx), \tag{1}$$

where $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, $g\colon \mathcal{G} \to]-\infty, +\infty]$ is proper, lower semicontinuous, and convex, and $L\colon \mathcal{H} \to \mathcal{G}$ is linear and bounded (see [1, 2, 3] for examples of $g \circ L$). The specific case of isotropic TV is obtained when $g = \|\cdot\|_{2,1}$ and L models the first order finite difference operator.

A wide literature is dedicated to proximal splitting methods for solving (1) (see, e.g., [4, 5, 6, 7, 3, 8]). The choice of the most appropriate algorithm will depend on the properties of the involved functions and linear operators. A first class of approaches are the primal algorithms such as forward-backward (FB) algorithm [9], Douglas-Rachford (DR) algorithm [10, 11, 12], and Peaceman-Rachford (PR) algorithm [13, 14, 12], whose use depends essentially on the Lipschitz differentiability or proximability[1] of the functions. In the context of strongly convex problems when $g \circ L$ is proximable and differentiable with a Lipschitz gradient, a theoretical comparison among these first order methods is provided in [15]. This work leads to preferable choices depending on the minimum and maximum eigenvalues of A as well as the Lipschitz constant of $\nabla(g \circ L)$. In this context, DR and PR are often preferable choices in some specific instances. Another theoretical comparison relaxing the hypothesis of differentiability of $g \circ L$ is developed in [16].

However, for many penalization terms encountered in image processing, $g \circ L$ is not proximable, starting with the isotropic TV. In its general form, the problem in (1) can be solved using FB or DR/PR algorithms, where the proximity operator of $g \circ L$ is computed with inner iterations. This approach has the numerical disadvantage that each iteration may require important computational time to reach an accurate precision in the computation of the proximity operator. Another well-known approach is the alternating direction method of multipliers (ADMM), which needs the inversion of matrices of the form $L^\top L + A^\top A$. This can be efficiently computed when L and A have compatible structures leading to a simple inversion (e.g., via discrete Fourier transform), but can be very costly in general, especially when the dimension is big. On the other hand, primal-dual algorithms can alternate between

[1] A function is *proximable* if there is a closed-form expression of its proximity operator.

the primal and the dual spaces to solve (1) by splitting all their functionals without needing any matrix inversion [17, 6, 18, 19]). The disadvantage in this case is the need of keep track of iterates in a higher dimensional primal-dual space.

Contributions – In this work, we introduce an alternative approach for solving general linear composite convex problems in Hilbert spaces including (1), by using the infimal postcomposition of a convex function by a linear operator. We first provide an equivalent variational formulation of the convex problem involving the infimal postcomposition and we propose a weakly convergent algorithm derived from DR. This algorithm requires the computation of the proximity operator of the infimal postcomposition derived in [20]. In the context of image restoration we provide an explicit formula of this proximity operator. Numerical experiments are conducted to illustrate the efficiency of the proposed algorithm. We show the benefit of the proposed procedure compared to a large panel of state-of-the-art methods, for different level of degradations and different image types. Moreover, we evaluate the impact of a TV-l2 penalization, highlighting the modelling impact of including strong convexity and, contrary to common thought, without degrading performance.

Outline – The paper is organized as follows. In Section 2 we introduce a variational formulation using the infimal postcomposition of a general linear composite optimization problem in Hilbert spaces, we study its theoretical properties, and we provide two weakly convergent algorithms based on DR. These results are applied to the context of image restoration in Section 3, including the explicit formula for the proximity operator of the infimal postcomposition. In Section 4 we illustrate the efficiency of the proposed methods with respect to concurrent state-of-the-art algorithms and we study the advantages of including a strongly convex term on the isotropic TVl2 minimization problem.

Notations – Throughout this paper, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces with inner products $\langle \cdot \mid \cdot \rangle$ and associated norm $\| \cdot \|$, $\rightharpoonup$ denotes the convergence in the weak topology. Given a function $f \colon \mathcal{H} \to]-\infty, +\infty]$, the domain of f is $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\}$ and f is proper if and only if $\operatorname{dom} f \neq \varnothing$. Moreover, $\Gamma_0(\mathcal{H})$ denotes the class of convex, lower semi-continuous, and proper functions from $\mathcal{H}$ to $]-\infty, +\infty]$.

Given a linear operator $L \colon \mathcal{H} \to \mathcal{G}$, L^* denotes its adjoint and its range is $\operatorname{ran} L = \big\{u \in \mathcal{G} \;\big|\; (\exists x \in \mathcal{H})\ Lx = u\big\}$. If $\mathcal{H} = \mathcal{G}$ is finite dimensional, and $L = L^\top$, denote by $\lambda_{\min}(L)$ the smallest eigenvalue of L.

For every $f \in \Gamma_0(\mathcal{H})$ and a linear operator $L \colon \mathcal{H} \to \mathcal{G}$, define

$$\operatorname{prox}_{f,L} \colon \mathcal{G} \to 2^{\mathcal{H}} \colon u \mapsto \arg\min_{x \in \mathcal{H}} f(x) + \frac{1}{2}\|Lx - u\|^2. \tag{2}$$

Note that $\operatorname{prox}_{f,\mathrm{Id}} = \operatorname{prox}_f$ is the classical proximity operator [21]. As for the classical proximity operator, in several instances $\operatorname{prox}_{f,L}$ has closed expression [20]. Given a nonempty set $C \subset \mathcal{H}$, the strong relative interior of C is $\operatorname{sri}(C) = \big\{x \in \mathcal{H} \;\big|\; \overline{\operatorname{span}}(C - x) = \operatorname{cone}(C - x)\big\}$, where $\overline{\operatorname{span}}(A)$ denotes the smallest closed linear subspace containing the set A and $\operatorname{cone}(A)$

denotes its conicall hull, i.e., the smallest cone containing A.

2 Optimization problems with infimal postcomposition

This contribution relies on the concept of infimal postcomposition of a function by a linear operator [4, 20] defined below.

Definition 1. *Let $\mathcal{G}$ be a real Hilbert space. Given $f\colon \mathcal{H} \to]-\infty, +\infty]$ and a linear bounded operator $L\colon \mathcal{H} \to \mathcal{G}$, the infimal postcomposition of f by L is defined by*

$$L \triangleright f\colon \mathcal{G} \to]-\infty, +\infty] : u \mapsto \inf_{\substack{x\in\mathcal{H}\\ Lx=u}} f(x). \tag{3}$$

We say that $L \triangleright f$ is exact at $u \in \mathcal{G}$, if

$$\arg\min_{\substack{x\in\mathcal{H}\\ Lx=u}} f(x) \neq \varnothing.$$

Note that $\operatorname{dom}(L \triangleright f) \subset \operatorname{ran} L$.

A general formulation of (1) can be rewritten considering previous definition.

Problem 1. *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{G})$, and let $L\colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator. We consider the problems*

$$\inf_{x\in\mathcal{H}} f(x) + g(Lx), \tag{P}$$

and

$$\inf_{u\in\mathcal{G}} g(u) + (L \triangleright f)(u). \tag{Q}$$

We denote $\mathcal{S}(P)$ and $\mathcal{S}(Q)$ the sets of solutions to (P) and (Q), respectively.

The following result connects problems (P) and (Q), whose proof can be found in the Appendix 6.1.

Proposition 1. *In the context of Problem 1, the following hold:*

1. $\inf_{x\in\mathcal{H}} f(x) + g(Lx) = \inf_{u\in\mathcal{G}} g(u) + (L \triangleright f)(u)$.
2. *Suppose that $\widehat{x} \in \mathcal{S}(P)$. Then $L\widehat{x} \in \mathcal{S}(Q)$.*
3. *Suppose that $\widehat{u} \in \mathcal{S}(Q)$. Then*

$$\arg\min_{\substack{x\in\mathcal{H}\\ Lx=\widehat{u}}} f(x) \subset \mathcal{S}(P).$$

Remark 1. *Note that Proposition 1-2) implies that $\mathcal{S}(P) \neq \varnothing$, then $\mathcal{S}(Q) \neq \varnothing$. The converse is not true in general. Indeed, if $\widehat{u}$ is a solution to (Q), Proposition 1-3) asserts that $\mathcal{S}(P)$ is nonempty if $L \triangleright f$ is exact at some solution $\widehat{u} \in \mathcal{S}(Q)$. Let us give a counterexample of the existence of a solution to (P) where, for every $\widehat{u} \in \mathcal{S}(Q)$, $\arg\min_{\substack{x \in \mathcal{H} \\ Lx = \widehat{u}}} f(x) = \varnothing$. Set $\mathcal{H} = \mathbb{R}^2$, $\mathcal{G} = \mathbb{R}$, $f \colon (x_1, x_2) \mapsto e^{x_2}$, $g \colon u \mapsto u^2/2$, and $L \colon (x_1, x_2) \mapsto x_1$. Then $(L \triangleright f) \colon u \mapsto \inf_{x_1 = u} e^{x_2} = 0$ and (Q) reduces to $\inf_{u \in \mathbb{R}} u^2/2$, which has value 0 and a unique solution $\widehat{u} = 0$. On the other hand, (P) reduces to $\inf_{(x_1, x_2) \in \mathbb{R}^2} e^{x_2} + x_1^2/2$, whose infimum is also 0 but it does not admit solutions. Note that, in this case, $\arg\min_{\substack{x \in \mathcal{H} \\ Lx = \widehat{u}}} f(x) = \varnothing$.*

Proposition 1 provides an alternative formulation of (P) depending on the infimal postcomposition. We aim at solving this formulation via proximal splitting methods, which need $L \triangleright f \in \Gamma_0(\mathcal{G})$ and an explicit expression of its proximity operator. Next result gives sufficient conditions in this direction.

Proposition 2. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $f \in \Gamma_0(\mathcal{H})$, and let $L \colon \mathcal{H} \to \mathcal{G}$ be a linear bounded operator such that*

$$0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*). \tag{4}$$

Then $L \triangleright f \in \Gamma_0(\mathcal{G})$, it is exact in $\mathcal{G}$, and

$$(\forall \gamma > 0) \quad \operatorname{prox}_{\gamma (L \triangleright f)} = L \operatorname{prox}_{\gamma f, L}, \tag{5}$$

where $\operatorname{dom} \operatorname{prox}_{\gamma f, L} = \mathcal{G}$.

The proof can be found in the Appendix 6.2. An interesting feature of the identity in (5) is that, even if $\operatorname{prox}_{f,L}$ may be a set-valued operator, $L\operatorname{prox}_{f,L}$ is single-valued.

Under the assumptions on Proposition 2, (Q) is an optimization problem involving the sum of two proper lower semicontinuous convex functions. Moreover, if the proximity operator of g is available, this alternative formulation can be solved, for instance, by using the Douglas-Rachford algorithm using the closed form expression of the proximity operator of $L \triangleright f$. We now describe two *infimal postcomposition Douglas-Rachford* (IPCDR) algorithms, whose difference comes from the order in which the proximity operators of $L \triangleright f$ and g are activated.

Theorem 1. *In the context of Problem 1, suppose that $\mathcal{S}(P) \neq \varnothing$ and that*

$$0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} L^*). \tag{6}$$

Moreover, let $(x_n)_{n \in \mathbb{N}}$ be a sequence defined by either Algorithm 1 or Algorithm 2. Then, the sequences $(u_n)_{n \in \mathbb{N}}$ and $(Lx_n)_{n \in \mathbb{N}}$ converge weakly to some $\widehat{u} \in \mathcal{S}(Q)$. In addition, for every $\widehat{x} \in \mathcal{H}$,

$$\widehat{x} \in \arg\min_{\substack{x \in \mathcal{H} \\ Lx = \widehat{u}}} f(x) \quad \Rightarrow \quad \widehat{x} \in \mathcal{S}(P). \tag{7}$$

The proof can be found in the Appendix 6.3.

Algorithm 1: IPCDR1

Let $z_0 \in \mathcal{H}$, let $\gamma > 0$, and consider the routine

For $n = 0, 1, \ldots$
$$\begin{bmatrix} x_n = \operatorname{prox}_{\gamma f, L} z_n \\ u_n = \operatorname{prox}_{\gamma g}\big(2Lx_n - z_n\big) \\ z_{n+1} = z_n + u_n - Lx_n \end{bmatrix}$$

Algorithm 2: IPCDR2

Let $z_0 \in \mathcal{H}$, let $\gamma > 0$, and consider the routine

For $n = 0, 1, \ldots$
$$\begin{bmatrix} u_n = \operatorname{prox}_{\gamma g} z_n \\ x_n = \operatorname{prox}_{\gamma f, L}\big(2u_n - z_n\big) \\ z_{n+1} = z_n + Lx_n - u_n. \end{bmatrix}$$

3 Image restoration

We place our study in a standard image restoration context where we consider the following finite dimensional problem.

Problem 2. *Let A and D be $N \times N$ and $M \times N$ real matrices, let $\epsilon \geq 0$, let $y \in \mathbb{R}^N$, and let $g \in \Gamma_0(\mathbb{R}^M)$ be a positive, positively homogeneous function vanishing only at 0 such that* $\operatorname{ran} D \cap \operatorname{dom} g \neq \varnothing$. *The problem is to*

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \;\; F(x) = \frac{1}{2}\|Ax - y\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2 + g(Dx). \tag{8}$$

The minimisation problem (8) models an image restoration problem where the degradation process is given by $y = A\overline{x} + \eta \in \mathbb{R}^N$, where $\overline{x} \in \mathbb{R}^N$ is the original image to recover, A is an $N \times N$ real matrix, and η models an additive white Gaussian noise. Moreover, D is a linear transform such as the discrete finite difference operator, involving horizontal and vertical differences, which induces piecewise constant images when combined with a particular function g. The standard anisotropic total variation is obtained by choosing $g = \lambda\|\cdot\|_1$ while the isotropic one is obtained with $g = \lambda\|\cdot\|_{2,1}$ the hybrid $\ell_{2,1}$-norm. The regularization parameter $\lambda > 0$ permits a tradeoff between the fidelity and the prior.

The introduction of the euclidean norm insures strongly convexity when $\epsilon > 0$ and allows us to recover more standard non strongly convex procedure when $\epsilon = 0$.

By setting $f = \frac{1}{2}\|A \cdot -y\|^2 + \frac{\epsilon}{2}\|\cdot\|_2^2$, $L = D$, $\mathcal{H} = \mathbb{R}^N$, and $\mathcal{G} = \mathbb{R}^M$, (8) can be written as a particular case of (P) by noting that f is convex, differentiable with a $(\|A\|^2+\epsilon)$-Lipschitz gradient. In addition, f is $(\lambda_{\min}(A^*A) + \epsilon)$-strongly convex, where $\lambda_{\min}$ associates to any symmetric $N \times N$ real matrix its lowest eigenvalue. The following result provides existence of solutions to Problem 2 and its proof can be found in Appendix 6.4.

Proposition 3. *Problem 2 admits at least a solution.*

Note that, Proposition 3 and Proposition 1(2) imply that the equivalent optimization problem (Q) associated to Problem 2 also admits solutions. Our strategy is thus to find a solution $\widehat{u}$ to (Q) and recover a solution $\widehat{x}$ to (P) by showing that $L \rhd f$ is exact at $\widehat{u}$ and using Proposition 1(3). We estimate $\widehat{u}$ by considering IPCDR1 (Algorithm 1) or IPCDR2 (Algorithm 2), which requires to compute $\operatorname{prox}_{\gamma f, D}$ provided in Proposition 4. Its proof can be found in the Appendix 6.5..

Proposition 4. *In the setting of Problem 2, let*

$$f = \frac{1}{2}\|A \cdot -y\|_2^2 + \frac{\epsilon}{2}\|\cdot\|_2^2,$$

let $\gamma > 0$*, and set*

$$\Phi = \gamma(\epsilon \operatorname{Id} + A^\top A) + D^\top D.$$

Then the following hold:

1. $\operatorname{dom} f^* = \operatorname{ran}(A^\top A + \epsilon \operatorname{Id})$.
2. $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} D^\top)$.
3. $\operatorname{prox}_{\gamma f, D} \colon u \mapsto \big\{p \in \mathcal{H} \;\big|\; \Phi p = \gamma A^\top y + D^\top u\big\}$.

In the following proposition, we provide conditions ensuring the invertibility of Φ defined in Proposition 4. This result allows us to obtain a single-valued proximity operator $\operatorname{prox}_{\gamma f, D}$.

Proposition 5. *In the context of Problem 2, let* $\gamma > 0$ *and set* $\Phi = \gamma(A^\top A + \epsilon \operatorname{Id}) + D^\top D$*. Moreover, suppose that one of the following holds.*

1. $\lambda_{\min}(\Phi) > 0$.
2. $\epsilon > 0$.
3. $\ker A \cap \ker D = \{0\}$.
4. $\ker D = \langle\{\mathbf{1}\}\rangle$ *and* $\mathbf{1} \notin \ker A$.

Then, Φ *is invertible and* $\operatorname{prox}_{\gamma f,D}\colon u \mapsto \Phi^{-1}(\gamma A^\top y + D^\top u)$.

The proof of Proposition 5 can be found in Appendix 6.6. Note that condition $\ker A \cap \ker D = \{0\}$ is satisfied, e.g., if $\ker A = \{0\}$. On the other hand, when D is the discrete difference operator, $\ker D = \langle\{\mathbf{1}\}\rangle$, i.e., the kernel of D is the set of constant vectors in $\mathbb{R}^N$. In this case, Proposition 5(4) ensures the inversibility of Φ when $\mathbf{1} \notin \ker A$. In thoses cases, several efficient inversion techniques can be used to compute Φ^{-1}. For instance, if A and D share similar block-circulant structures, the inversion can be efficiently computed by considering the Fourier transform and $\operatorname{prox}_{\gamma f,D}$ can be computed from Proposition 5.

The following results combine Theorem 1 and Proposition 4 in order to provide explicit implementations of Algorithms 1 and 2 for solving Problem 2. The proof of Proposition 6 can be found in Appendix 6.7 and that of Proposition 7 is analogous.

Proposition 6. *In the context of Problem 2, suppose that one of the assumptions 1), 2), 3), or 4) in Proposition 5 holds. Let* $z_0 \in \mathbb{R}^M$*, let* $\gamma > 0$*, and consider the routine*

$$\begin{array}{l} \textit{For } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} x_n = \big(\gamma(A^\top A + \epsilon\,\mathrm{Id}) + D^\top D\big)^{-1}(\gamma A^\top y + D^\top z_n) \\ u_n = \operatorname{prox}_{\gamma g}\big(2Dx_n - z_n\big) \\ z_{n+1} = z_n + u_n - Dx_n \end{array} \right. \end{array} \tag{9}$$

Then, there exists $\widehat{x} \in \mathbb{R}^N$ *such that* $x_n \to \widehat{x}$ *and*

$$\widehat{x} + \frac{\mathbf{1}^\top\big(A^\top y - (A^\top A + \epsilon\,\mathrm{Id})\widehat{x}\big)}{\mathbf{1}^\top(A^\top A + \epsilon\,\mathrm{Id})\mathbf{1}}\mathbf{1} \tag{10}$$

is a solution to Problem 2.

Proposition 7. *In the context of Problem 2, suppose that one of the assumptions 1), 2), 3), or 4) in Proposition 5 holds. Let* $z_0 \in \mathbb{R}^M$*, let* $\gamma > 0$*, and consider the routine*

$$\begin{array}{l} \textit{For } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} u_n = \operatorname{prox}_{\gamma g} z_n \\ x_n = (\gamma(A^\top A + \epsilon\,\mathrm{Id}) + D^\top D)^{-1}\big(\gamma A^\top y + D^\top(2u_n - z_n)\big) \\ z_{n+1} = z_n + Dx_n - u_n. \end{array} \right. \end{array} \tag{11}$$

Then, there exists $\widehat{x} \in \mathbb{R}^N$ *such that* $x_n \to \widehat{x}$ *and*

$$\widehat{x} + \frac{\mathbf{1}^\top\big(A^\top y - (A^\top A + \epsilon\,\mathrm{Id})\widehat{x}\big)}{\mathbf{1}^\top(A^\top A + \epsilon\,\mathrm{Id})\mathbf{1}}\mathbf{1} \tag{12}$$

is a solution to Problem 2.

4 Numerical experiments

4.1 Algorithmic comparison

In this section we highlight the benefit of the infimal postcomposition combined with DR procedures. To this purpose, we compare several algorithms for solving Problem 2 in the case when $\epsilon > 0$.

[**IPCDR1**] Infimal postcomposition combined with standard DR described in Proposition 6.

[**IPCDR2**] Infimal postcomposition combined with standard DR described in Proposition 7, where the proximal steps are swapped compared to [IPCDR1].

[**DR-kerL1**] Formulation by means of an auxiliary variable combined to standard DR procedure considering kernel projection as described in [22].

[**DR-kerL2**] Formulation by means of an auxiliary variable to apply standard DR procedure considering kernel projection as described in [22] and where the proximal steps are inverted compared to [DR-kerL1].

[**ADMM**] ADMM procedure [23, 24, 25].

[**CP**] Chambolle-Pock algorithm [17].

[**CV**] Condat-Vũ algorithm [6, 18].

Each of these methods are specified below.

• **DR-kerL1** Following [22], the minimization problem (8) can be formulated as

$$\underset{\substack{(x,v)\in\mathbb{R}^N\times\mathbb{R}^M \\ Dx=v}}{\text{minimize}} \quad \frac{1}{2}\|Ax-y\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2 + g(v) \tag{13}$$

or equivalently as

$$\underset{(x,v)\in\mathbb{R}^N\times\mathbb{R}^M}{\text{minimize}} \quad \frac{1}{2}\|Ax-y\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2 + g(v) + \iota_{\ker \widetilde{L}}(x,v) \tag{14}$$

with $\widetilde{L} = \begin{bmatrix} D & -\mathrm{Id} \end{bmatrix}$. Since

$$P_{\ker \widetilde{L}} = \mathrm{Id} - \widetilde{L}^\top(\widetilde{L}\widetilde{L}^\top)^{-1}\widetilde{L} = \mathrm{Id} - \widetilde{L}^\top(DD^\top + \mathrm{Id})^{-1}\widetilde{L},$$

the iterations resulting by applying Douglas-Rachford splitting as in [22, Section 5.3.1] are provided below and the sequence $(x_n)_{n\in\mathbb{N}}$ converges to the minimizer of (8).

Let $z_0 = x_{\text{init}} \in \mathbb{R}^N$ and $\tilde{z}_0 = Dz_0$ and $\gamma > 0$
For $n = 0, 1, \ldots$
$x_n = (\gamma A^\top A + (\gamma\epsilon + 1)\,\mathrm{Id})^{-1}(\gamma A^\top y + z_n)$
$v_n = \mathrm{prox}_{\gamma g}\tilde{z}_n$
$u_n = (DD^\top + \mathrm{Id})^{-1}(D(2x_n - z_n) - 2v_n + \tilde{z}_n)$
$z_{n+1} = x_n - D^\top u_n$
$\tilde{z}_{n+1} = v_n + u_n$.

• **DR-kerL2** Relying on the same principle but swapping the two proximal operator computations leads to the resulting iterations described below for which the sequence $(x_n)_{n\in\mathbb{N}}$ also converges to a minimizer of (8).

Let $z_0 = x_{\text{init}} \in \mathbb{R}^N$ and $\tilde{z}_0 = Dz_0$ and $\gamma > 0$
For $n = 0, 1, \ldots$
$t_n = (DD^\top + \mathrm{Id})^{-1}(Dz_n - \tilde{z}_n)$
$u_n = z_n - D^\top t_n$
$\tilde{u}_n = \tilde{z}_n + t_n$
$x_n = (\gamma A^\top A + (\gamma\epsilon + 1)\,\mathrm{Id})^{-1}(\gamma A^\top y + 2u_n - z_n)$
$v_n = \mathrm{prox}_{\gamma g}(2\tilde{u}_n - \tilde{z}_n)$
$z_{n+1} = z_n + x_n - u_n$
$\tilde{z}_{n+1} = \tilde{z}_n + v_n - \tilde{u}_n$.

• **ADMM** Another well-established algorithm is ADMM [8] whose resulting iterations are described below and are such that the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a minimizer of (8).

Let $u_0 = y_0 = Dx_{\text{init}}$ and $\gamma > 0$
For $n = 0, 1, \ldots$
$x_{n+1} = (A^\top A + \epsilon\mathrm{Id} + \gamma D^\top D)^{-1}(A^\top y + D^\top(\gamma u_n - y_n))$
$u_{n+1} = \mathrm{prox}_{\gamma^{-1}g}(Dx_{n+1} + \gamma^{-1}y_n)$
$y_{n+1} = y_n + \gamma(Dx_{n+1} - u_{n+1})$.

• **CP** The iterations of Chambolle-Pock [17] are described below and are such that the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a minimizer of (8).

Let $\widetilde{x}_0 = x_0 = x_{\text{init}} \in \mathbb{R}^N$ and $u_0 = Dx_0$
Set $\tau > 0$ and $\gamma < (\tau\|D\|^2)^{-1}$
For $n = 0, 1, \ldots$
$x_{n+1} = (\tau A^\top A + (\tau\epsilon + 1)\mathrm{Id})^{-1}(x_n - \tau D^\top u_n + \tau A^\top y)$
$u_{n+1} = \mathrm{prox}_{\gamma g^*}\big(u_n + \gamma D(2x_{n+1} - x_n)\big)$.

• **CV** The iterations of Condat-Vũ [6, 18] are described below and are such that the sequence $(x_n)_{n\in\mathbb{N}}$ also converges to a minimizer of (8).

Let $x_0 = x_{\text{init}} \in \mathbb{R}^N$ and $u_0 = Dx_0$
Set $\tau > 0$ and $\gamma < ((\|A\|^2 + \epsilon)/2 + \tau\|D\|^2)^{-1}$
For $n = 0, 1, \ldots$
$\quad x_{n+1} = x_n - \tau\big(A^\top(Ax_n - y) + \epsilon x_n + D^\top u_n\big)$
$\quad u_{n+1} = \text{prox}_{\gamma g^*}\big(u_n + \gamma D(2x_{n+1} - x_n)\big)$.

4.2 Numerical settings

For our numerical experiments, we consider the standard framework of image restoration when the acquisition process A is associated with a blur having periodic boundary effects as encountered in many optics systems leading to a bloc circulant matrix [2]. This matrix is diagonalizable in the Fourier domain, which makes the computations, especially inversion, very efficient. $\eta \sim \mathcal{N}(0, \sigma^2\text{Id})$ models an additive white Gaussian noise of standard deviation σ. In our experiments, $D = [D_h, D_v]$ denotes horizontal and vertical differences operators associated with the filter $h = [1/2, -1/2]$ and $v = h^\top$ with periodic boundary effects also leading to a bloc circulant matrix, and $g = \|\cdot\|_{1,2}$.

We implement each algorithm for a large panel of values γ (for all four algorithms based on DR and ADMM) or values (τ, γ) (for CP and CV).

In Figures 1, 2, and 3, we display, the original image, the degraded one, and its reconstruction, for different choices of blur and noise. The evolution of the objective function for each method (right figures) are displayed for the step-size parameter(s) leading to the fastest procedure in terms of convergence of the objective function. As all the methods converge to the same solution, the provided reconstructed image is the one obtained with the fastest algorithm and for the parameters (λ, ϵ) leading to the best signal-to-noise ratio (SNR) represented with the red box.

We can observe that CV is systematically the slowest algorithm followed by CP and DR-kerL procedures. ADMM and IPCDR1/IPCDR2 have very close behaviour even if IPCDR1 is always faster. In terms of image reconstruction, we can observe that the best results are not necessarily obtained for the smallest ϵ leading to the conclusion that adding strong convexity does not impact badly in the reconstruction while it can help for the optimization procedure.

Moreover, the profil in terms of convergence seems very close for different images but similar degradation. IPCDR1/IPCDR2 appears to have a better convergence profile for larger blur but not larger noise degradation.

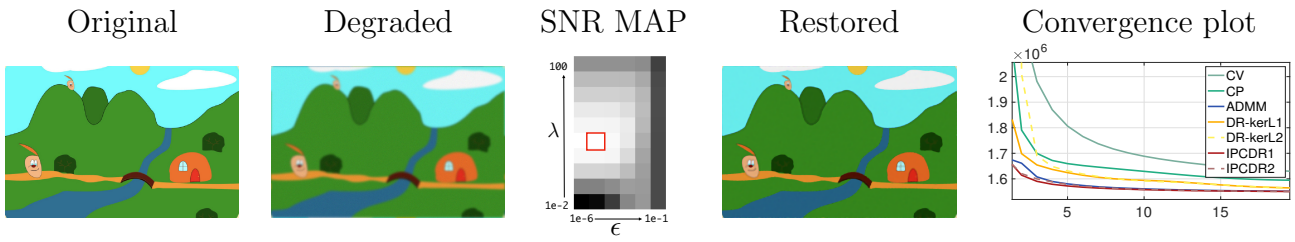

Figure 1: From left to right: original image extracted from [26], degraded image with a uniform blur of size 3×3 and a Gaussian noise with standard deviation $\sigma = 0.012$, SNR map w.r.t. (ϵ, λ) where the red box displays the highest SNR value, restored image, evolution of $F(x_n)$ w.r.t the iterations n.

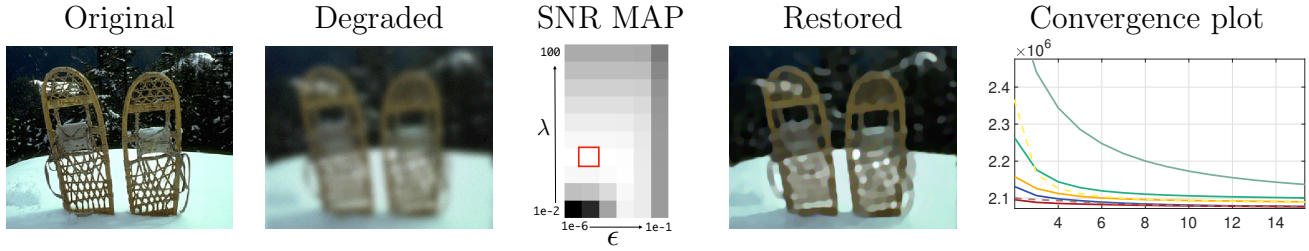

Figure 2: From left to right: original Image '2018' extracted from BSD dataset [27], degraded image with a uniform blur of size 5×5 and a Gaussian noise with standard deviation $\sigma = 0.02$, SNR map w.r.t. (ϵ, λ) where the red box displays the highest SNR value, restored image, evolution of $F(x_n)$ w.r.t the iterations n.

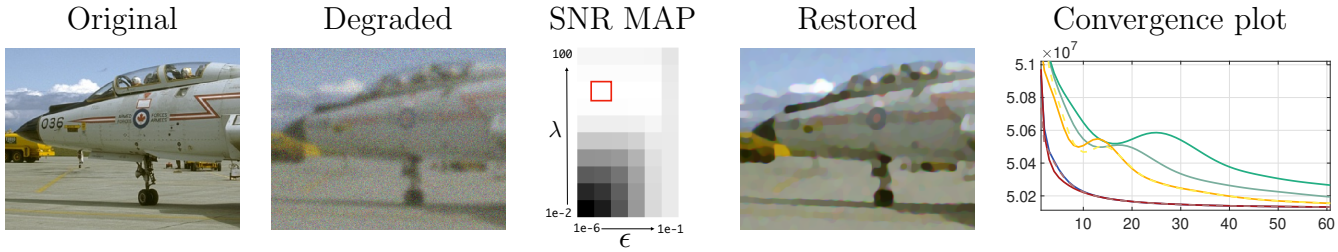

Figure 3: From left to right: original Image '10081' extracted from BSD dataset [27], degraded image with a uniform blur of size 3×3 and a Gaussian noise with standard deviation $\sigma = 0.1$, SNR map w.r.t. (ϵ, λ) where the red box displays the highest SNR value, restored image, evolution of $F(x_n)$ w.r.t the iterations n.

5 Conclusion

In this paper we introduce a new formulation for linear composite convex optimization problems using the infimal postcomposition of a convex function by a linear operator. We study theoretical properties connecting the original convex optimization problem and our formulation. This new formulation allows for the efficient application of first-order proximal splitting methods as the Douglas-Rachford splitting, given a recent implementation of the proximity operator of the infimal postcomposition in [20]. We provide numerical implementations of the proposed method in the context of image restoration problems, where the proximity operator of the infimal postcomposition is computed explicitly. We illustrate the efficiency of our proposed method by comparing it with respect to the state-of-the-art methods in the literature.

6 Appendix

6.1 Proof of Proposition 1

Proof. 1: We have

$$\begin{aligned}\inf_{u\in\mathcal{G}} g(u) + (L \triangleright f)(u) &= \inf_{u\in\mathcal{G}} g(u) + \inf_{\substack{x\in\mathcal{H}\\ Lx=u}} f(x) \\ &= \inf_{x\in\mathcal{H}} \inf_{\substack{u\in\mathcal{G}\\ Lx=u}} f(x) + g(u) \\ &= \inf_{x\in\mathcal{H}} f(x) + g(Lx). \end{aligned} \tag{15}$$

2: Suppose that $\widehat{x} \in \mathcal{S}(P)$. Then it follows from (15) and Definition 1 that

$$\begin{aligned} f(\widehat{x}) + g(L\widehat{x}) &= \min_{x\in\mathcal{H}} f(x) + g(Lx) \\ &= \inf_{u\in\mathcal{G}} g(u) + (L \triangleright f)(u) \\ &\leq g(L\widehat{x}) + (f \triangleright L)(L\widehat{x}) \\ &\leq g(L\widehat{x}) + f(\widehat{x}), \end{aligned} \tag{16}$$

which yields

$$g(L\widehat{x}) + (f \triangleright L)(L\widehat{x}) = \min_{u\in\mathcal{G}} g(u) + (L \triangleright f)(u) \tag{17}$$

and, hence, $\widehat{u} = L\widehat{x} \in \mathcal{S}(Q)$.

3: Suppose that $\widehat{u} \in \mathcal{S}(Q)$ and let $\widehat{x} \in L^{-1}\widehat{u}$ be such that $f(\widehat{x}) = (L \triangleright f)(\widehat{u})$. Hence,

since $L\widehat{x} = \widehat{u}$, we deduce from (15) that

$$\begin{aligned} g(L\widehat{x}) + f(\widehat{x}) &= g(\widehat{u}) + (f \triangleright L)(\widehat{u}) \\ &= \min_{u \in \mathcal{G}} g(u) + (f \triangleright L)(u) \\ &= \inf_{x \in \mathcal{H}} f(x) + g(Lx), \end{aligned} \tag{18}$$

and the result follows. □

6.2 Proof of Proposition 2

In view of [20, Proposition 3], we only need to prove the exactness of $L \triangleright f$ (cf. Definition 1). Let $u \in \mathcal{G}$ and note that

$$(L \triangleright f)(u) = \inf_{x \in \mathcal{H}} f(x) + \iota_{\{u\}}(Lx), \tag{19}$$

that $\iota^*_{\{u\}} = \langle u \mid \cdot \rangle$, and that $\operatorname{dom} \iota^*_{\{u\}} = \mathcal{G}$. Therefore, it follows from (4) and [4, Theorem 15.23] that

$$-\inf_{v \in \mathcal{G}} \langle u \mid v \rangle + f^*(-L^*v) = \min_{x \in \mathcal{H}} f(x) + \iota_{\{u\}}(Lx), \tag{20}$$

and the result follows from (19).

6.3 Proof of Theorem 1

Proof. First note that $\mathcal{S}(P) \neq \varnothing$ and Proposition 1(2) yield $\mathcal{S}(Q) \neq \varnothing$. Moreover, (6) and Proposition 2 imply that $(L \triangleright f) \in \Gamma_0(\mathcal{G})$ and $\operatorname{prox}_{\gamma(L \triangleright f)} = L \operatorname{prox}_{\gamma f, L}$. Hence, since Algorithm 1 can be written equivalently as

$$z_{n+1} = \operatorname{prox}_{\gamma g}(2\operatorname{prox}_{\gamma(L \triangleright f)} z_n - z_n) + z_n - \operatorname{prox}_{\gamma(L \triangleright f)} z_n, \tag{21}$$

[4, Theorem 26.11] implies the existence of $\widehat{u} \in \mathcal{S}(Q)$ such that $Lx_n = \operatorname{prox}_{\gamma(L \triangleright f)} z_n \rightharpoonup \widehat{u}$ and $u_n = \operatorname{prox}_{\gamma g}(2\operatorname{prox}_{\gamma(L \triangleright f)} z_n - z_n) \rightharpoonup \widehat{u}$. The result is analogous for Algorithm 2, because it reduces to

$$z_{n+1} = \operatorname{prox}_{\gamma(L \triangleright f)}(2\operatorname{prox}_{\gamma g} z_n - z_n) + z_n - \operatorname{prox}_{\gamma g} z_n. \tag{22}$$

The last assertion follows from the exactness guaranteed in Proposition 2 and Proposition 1(3). □

6.4 Proof of Proposition 3

Proof. Note that $F = \Psi(\Lambda \cdot - \ell)$, where $\Psi \colon (a, b, c) \mapsto \|a\|_2^2/2 + \|b\|_2^2/2 + g(c)$, $\Lambda \colon x \mapsto (Ax, \sqrt{\epsilon}x, Dx)$, and $\ell = (z, 0, 0)$. Moreover, since $g \geq 0$ is a positively homogeneous function

vanishing only at 0, Ψ is a proper, convex, lower semicontinuous, coercive function. Hence, it follows from [28, Remark 2.2, Remark 2.3, p. 774, and Theorem 2.1] that the solution set of Problem 2 is nonempty. □

6.5 Proof of Proposition 4

Proof. 1: Note that $f = \Psi \circ \Lambda$, where $\Psi\colon (a,b) \mapsto \|a-y\|_2^2/2 + \|b\|_2^2/2$, and $\Lambda\colon x \mapsto (Ax, \sqrt{\epsilon}x)$. Moreover, $\Psi^*\colon (u,v) \mapsto \frac{1}{2}\|v\|^2 + \frac{1}{2}\|u+y\|^2 - \frac{1}{2}\|y\|^2$ and $\Lambda^\top\colon (u,v) \mapsto A^\top u + \sqrt{\epsilon}v$. Since $0 \in \mathbb{R}^N \times \mathbb{R}^N = \operatorname{sri}(\operatorname{dom}\Psi - \operatorname{ran}\Lambda)$, it follows from [4, Corollary 15.28] that, for every $x^* \in \mathbb{R}^N$,

$$\begin{aligned} f^*(x^*) &= (\Psi \circ \Lambda)^*(x^*) \\ &= (\Lambda^\top \triangleright \Psi^*)(x^*) \\ &= \inf_{\substack{(u,v)\in\mathbb{R}^N\times\mathbb{R}^N \\ A^\top u + \sqrt{\epsilon}v = x^*}} \frac{1}{2}\|v\|^2 + \frac{1}{2}\|u+y\|^2 - \frac{1}{2}\|y\|^2. \end{aligned} \tag{23}$$

We explore two cases. If $\epsilon > 0$, then (23) reduces to

$$\inf_{u\in\mathbb{R}^N} \frac{1}{2\epsilon}\|x^* - A^\top u\|^2 + \frac{1}{2}\|u+y\|^2 - \frac{1}{2}\|y\|^2,$$

which has always a unique solution and a real value, because it is a strongly convex optimization problem. Hence, $\operatorname{dom} f^* = \mathbb{R}^N = \operatorname{ran}(A^\top A + \epsilon \operatorname{Id})$. On the other hand, suppose that $\epsilon = 0$. Then, if $x^* \notin \operatorname{ran} A^\top$, $f^*(x^*) = +\infty$, while if $x^* \in \operatorname{ran} A^\top$, there exists $u^* \in \mathbb{R}^N$ such that $x^* = A^\top u^*$ and (23) reduces to

$$\begin{aligned} &\inf_{\substack{u\in\mathbb{R}^N \\ A^\top(u-u^*)=0}} \frac{1}{2}\|u+y\|^2 - \frac{1}{2}\|y\|^2 \\ &\qquad = \inf_{\substack{w\in\mathbb{R}^N \\ A^\top w = 0}} \frac{1}{2}\|u^*+y-w\|^2 - \frac{1}{2}\|y\|^2 \\ &\qquad = \frac{1}{2} d^2_{\ker A^\top}(u^*+y) - \frac{1}{2}\|y\|^2, \end{aligned}$$

which is a real number. Therefore, $\operatorname{dom} f^* = \operatorname{ran} A^\top = \operatorname{ran}(A^\top A + \epsilon \operatorname{Id})$.

2: Note that 1 yields $\operatorname{dom} f^* = \operatorname{ran}(A^\top A + \epsilon \operatorname{Id})$ and, since $\operatorname{ran}(A^\top A + \epsilon \operatorname{Id})$ and $\operatorname{ran}(D^\top)$ are linear subspaces of $\mathbb{R}^N$, the finite dimensional context of Problem 2 and [4, Proposition 6.19(ii)] imply the result.

3: For every $u \in \mathcal{G}$, it follows from (2) and Fermat's theorem [4, Theorem 16.3] that $p = \operatorname{prox}_{\gamma f, D} u$ is a solution to the linear system

$$\Phi p = (\gamma(A^\top A + \epsilon \operatorname{Id}) + D^\top D)p = \gamma A^\top y + D^\top u. \tag{24}$$

The proof is complete. □

6.6 Proof of Proposition 5

Proof. First note that Φ is symmetric and, for every $x \in \mathbb{R}^N$, we have

$$x^\top \Phi x = \gamma \|Ax\|^2 + \gamma\epsilon\|x\|^2 + \|Dx\|^2 \geq 0. \tag{25}$$

Therefore, Φ is a symmetric positive semidefinite $N \times N$ real matrix.

1: This implies that Φ is a symmetric positive definite real matrix and, thus, it is invertible.

2: It follows from (25) that $\lambda_{\min}(\Phi) \geq \gamma\epsilon > 0$ and the result follows from 1.

3: Suppose that $\Phi x = 0$ for some $x \in \mathbb{R}^N$. Then, (25) implies that

$$\gamma \|Ax\|^2 + \gamma\epsilon\|x\|^2 + \|Dx\|^2 = 0, \tag{26}$$

which implies that $x \in \ker A \cap \ker D = \{0\}$ and, hence, $\ker \Phi = \{0\}$. Therefore Φ is invertible.

4: Clear from 3.

The last assertion follows from Proposition 4(3). □

6.7 Proof of Proposition 6

Proof. First observe that the existence of solutions to Problem 2 is guaranteed by Proposition 3. Moreover, Proposition 4(2) asserts that $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{ran} D^\top)$. In addition, it follows from Proposition 5 that (9) is a particular instance of Algorithm 1. Therefore, Theorem 1 asserts that sequences $(u_n)_{n\in\mathbb{N}}$ and $(Dx_n)_{n\in\mathbb{N}}$ converge to some $\widehat{u} \in \mathcal{S}(Q)$ in the specific case of Problem 2. We recall that weak and strong topologies coincide in finite dimensions.

In order to recover a solution to Problem 2 from $\widehat{u} \in \mathcal{S}(Q)$, it follows from (7) in Theorem 1 that we need to solve

$$\min_{\substack{x\in\mathcal{H} \\ Dx=\widehat{u}}} \frac{1}{2}\|Ax - z\|_2^2 + \frac{\epsilon}{2}\|x\|_2^2. \tag{27}$$

Since D is a discrete gradient, we have that the kernel of D is the space of constant images, or, equivalently, $\ker D = \langle\{\mathbf{1}\}\rangle$, where $\mathbf{1} \in \mathbb{R}^N$ is the image constant and equal to 1. Moreover, since $\widehat{u} \in \mathcal{S}(Q) \subset \operatorname{dom}(D \rhd f) \subset \operatorname{ran} D$, we obtain that $\{x \in \mathbb{R}^N \mid Dx = \widehat{u}\} \neq \varnothing$. Let $\widehat{x} \in \mathbb{R}^N$ such that $D\widehat{x} = \widehat{u}$. Then

$$\begin{aligned} (\forall x \in \mathcal{H}) \quad Dx = \widehat{u} \;&\Leftrightarrow\; D(x - \widehat{x}) = 0 \\ &\Leftrightarrow\; x - \widehat{x} \in \ker D = \langle\{\mathbf{1}\}\rangle \\ &\Leftrightarrow\; (\exists \lambda \in \mathbb{R})\ x = \widehat{x} + \lambda\mathbf{1}. \end{aligned} \tag{28}$$

Therefore, (27) reduces to

$$\min_{\lambda\in\mathbb{R}} \frac{1}{2}\|A(\widehat{x}+\lambda\mathbf{1})-z\|_2^2+\frac{\epsilon}{2}\|\widehat{x}+\lambda\mathbf{1}\|_2^2, \tag{29}$$

whose explicit solution is

$$\widehat{\lambda}=\frac{\mathbf{1}^\top\big(A^\top z-(A^\top A+\epsilon\,\mathrm{Id})\widehat{x}\big)}{\mathbf{1}^\top(A^\top A+\epsilon\,\mathrm{Id})\mathbf{1}}. \tag{30}$$

We conclude that $\widehat{x}+\widehat{\lambda}\mathbf{1}$ is a solution to Problem 2. The proof is thus complete. □

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[2] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based image deconvolution and reconstruction," *Wiley Encyclopedia of EEE*, 2016.

[3] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

[4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2017.

[5] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.

[6] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[7] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2014.

[8] A. Beck, *First-order methods in optimization*, SIAM, 2017.

[9] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. and Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.

[10] J. Douglas and H. H. Rachford, "On the numerical solution of the heat conduction problem in two and three space variables," *Trans. Amer. Math. Soc.*, vol. 82, no. 2, pp. 421–439, Jul. 1956.

[11] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Selected Topics Signal Process.*, vol. 1, pp. 564–574, 2007.

[12] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, 1979.

[13] D. W. Peaceman and H. H. Rachford, "The numerical solution of parabolic and elliptic differential equations," *Journal of the Society for Industrial and Applied Mathematics*, vol. 3, pp. 28–41, 1955.

[14] D. Davis and W. Yin, "Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions," *Mathematics of Operations Research*, vol. 42, no. 3, pp. 783–805, 2017.

[15] L. M Briceño-Arias and N. Pustelnik, "Theoretical and numerical comparison of first order algorithms for cocoercive equations and smooth convex optimization," *Signal Processing*, vol. 206, pp. 108900, 2023.

[16] L. M. Briceño-Arias and N. Pustelnik, "Convergence rate comparison of proximal algorithms for non-smooth convex optimization with an application to texture segmentation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1337–1341, 2022.

[17] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[18] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

[19] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal? Dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, 2015.

[20] L. M. Briceño-Arias and F. Roldán, "Resolvent of the parallel composition and the proximity operator of the infimal postcomposition," *Optimization Letters*, vol. 17, no. 2, pp. 399–412, 2023.

[21] J. J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[22] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, "Proximal algorithms for multicomponent image recovery problems," *J. Math. Imag. Vis.*, vol. 41, no. 1, pp. 3–22, Sep. 2011.

[23] R. Glowinski and A. Marrocco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non

linéaires," *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér.*, vol. 9, no. R-2, pp. 41–76, 1975.

[24] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[25] D. Gabay, "Chapter IX applications of the method of multipliers to variational inequalities," in *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Michel Fortin and Roland Glowinski, Eds., vol. 15 of *Studies in Mathematics and Its Applications*, pp. 299 – 331. Elsevier, 1983.

[26] Michau G., *Opérateur Dérivé va à la Montagne*, PhyLab Editions, 2016.

[27] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[28] Alfred Auslender, "Noncoercive optimization problems," *Math. Oper. Res.*, vol. 21, no. 4, pp. 769–782, 1996.