A PROXIMAL DECOMPOSITION METHOD FOR SOLVING CONVEX VARIATIONAL INVERSE PROBLEMS[*]

Patrick L. Combettes[1] and Jean-Christophe Pesquet[2]

[1]UPMC Université Paris 06
Laboratoire Jacques-Louis Lions – UMR 7598
75005 Paris, France
plc@math.jussieu.fr

[2]Université Paris-Est
Institut Gaspard Monge and UMR CNRS 8049
77454 Marne la Vallée Cedex 2, France
jean-christophe.pesquet@univ-paris-est.fr

Abstract

A broad range of inverse problems can be abstracted into the problem of minimizing the sum of several convex functions in a Hilbert space. We propose a proximal decomposition algorithm for solving this problem with an arbitrary number of nonsmooth functions and establish its weak convergence. The algorithm fully decomposes the problem in that it involves each function individually via its own proximity operator. A significant improvement over the methods currently in use in the area of inverse problems is that it is not limited to two nonsmooth functions. Numerical applications to signal and image processing problems are demonstrated.

---

1   Introduction

Throughout this paper, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, norm $\|\cdot\|$, and distance d. Moreover, $(f_i)_{1 \le i \le m}$ are proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$. We consider inverse problems that can be formulated as decomposed optimization problems of the form

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x). \tag{1.1}$$

In this flexible variational formulation, each potential function f_i may represent a prior constraint on the ideal solution $\overline{x}$ or on the data acquisition model. The purpose of this paper is to propose a decomposition method that, under rather general conditions, will provide solutions to (1.1).

To place our investigation in perspective, let us review some important special cases of (1.1) for which globally convergent numerical methods are available. These examples encompass a variety of inverse problems in areas such as signal denoising [25, 44], signal deconvolution [17], Bayesian image recovery [16], intensity-modulated radiation therapy [10, 13], image restoration [5, 6, 15], linear inverse problems with sparsity constraints [24, 29, 32, 48], signal reconstruction from Fourier phase information [37], and tomographic reconstruction [2, 10, 46].

(a) If the functions $(f_i)_{1 \le i \le m}$ are the indicator functions (see (2.1)) of closed convex sets $(C_i)_{1 \le i \le m}$ in $\mathcal{H}$, (1.1) reduces to the convex feasibility problem [10, 13, 18, 46, 50]

$$\text{find } x \in \bigcap_{i=1}^{m} C_i, \tag{1.2}$$

which can be solved by projection techniques, e.g., [4, 12, 19, 36].

(b) The constraint sets in (a) are based on information or measurements that can be inaccurate. As a result, the feasibility set $\bigcap_{i=1}^{m} C_i$ may turn out to be empty. An approximate solution can be obtained by setting, for every $i \in \{1, \ldots, m\}$, $f_i = \omega_i d_{C_i}^2$, where d_{C_i} is the distance function to C_i (see (2.2)) and where $\omega_i \in]0, 1]$. Thus, (1.1) becomes

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^{m} \omega_i d_{C_i}^2(x). \tag{1.3}$$

This approach is proposed in [17], where it is solved by a parallel projection method. Finite-dimensional variants based on Bregman distances are investigated in [11].

(c) If the functions $(f_i)_{1 \le i \le m-1}$ are the indicator functions of closed convex sets $(C_i)_{1 \le i \le m-1}$ in $\mathcal{H}$ and $f_m \colon x \mapsto \|x - r\|^2$ for some $r \in \mathcal{H}$, then (1.1) reduces to the best approximation problem [2, 21]

$$\underset{x \in \bigcap_{i=1}^{m-1} C_i}{\text{minimize}} \quad \|x - r\|^2. \tag{1.4}$$

Several algorithms are available to solve this problem [7, 21, 33, 34, 49]. There are also methods that are applicable in the presence of a more general strictly convex potential f_m; see [20] and the references therein.

(d) In [26], the special instance of (1.1) in which $m = 2$ and f_2 is Lipschitz-differentiable on $\mathcal{H}$ is shown to cover a variety of seemingly unrelated inverse problem formulations such as Fourier regularization problems, constrained least-squares problems, split feasibility problems, multiresolution sparse regularization problems, geometry/texture image decomposition problems, hard-constrained inconsistent feasibility problems, as well as certain maximum *a posteriori* problems (see also [6, 8, 9, 16, 24, 29, 32] for further developments within this framework). The forward-backward splitting algorithm proposed in [26] is governed by the updating rule

$$x_{n+1} = x_n + \lambda_n \Big(\operatorname{prox}_{\gamma_n f_1} \big(x_n - \gamma_n(\nabla f_2(x_n) + b_n) \big) + a_n - x_n \Big), \tag{1.5}$$

where $\lambda_n \in \,]0,1]$ and $\gamma_n \in \,]0, +\infty[$, where

$$\operatorname{prox}_{\gamma_n f_1} : x \mapsto \operatorname*{argmin}_{y \in \mathcal{H}} \; \gamma_n f_1(y) + \frac{1}{2}\|x - y\|^2 \tag{1.6}$$

is the proximity operator of $\gamma_n f_1$, and where the vectors a_n and b_n model tolerances in the implementation of $\operatorname{prox}_{\gamma_n f_1}$ and ∇f_2, respectively. Naturally, this 2-function framework can be extended to (1.1) under the severe restriction that the functions $(f_i)_{2 \le i \le m}$ be Lipschitz-differentiable. Indeed, in this case, $\widetilde{f}_2 = \sum_{i=2}^{m} f_i$ also enjoys this property and it can be used in lieu of f_2 in (1.5).

(e) The problem considered in [25] corresponds to $m = 2$ in (1.1). In other words, the smoothness assumption on f_2 in (d) is relaxed. The algorithm adopted in [25] is based on the Douglas-Rachford splitting method [22, 38] and operates via the updating rule

$$\left\lfloor \begin{aligned} y_{n+\frac{1}{2}} &= \operatorname{prox}_{\gamma f_2} y_n + a_n \\ y_{n+1} &= y_n + \lambda_n \Big(\operatorname{prox}_{\gamma f_1} \big(2y_{n+\frac{1}{2}} - y_n \big) + b_n - y_{n+\frac{1}{2}} \Big), \end{aligned} \right. \tag{1.7}$$

where $\lambda_n \in \,]0,2[$ and $\gamma \in \,]0, +\infty[$, and where the vectors a_n and b_n model tolerances in the implementation of the proximity operators. Under suitable assumptions, the sequence $(y_n)_{n \in \mathbb{N}}$ converges weakly to a point $y \in \mathcal{H}$ and $\operatorname{prox}_{\gamma f_2} y \in \operatorname{Argmin} f_1 + f_2$. In this approach, the smoothness assumption made on f_2 in (d) is replaced by the practical assumption that $\operatorname{prox}_{\gamma f_2}$ be implementable (to within some error).

Some important scenarios are not covered by the above settings, namely the formulations of type (1.1) that feature three or more potentials, at least two of which are nonsmooth. In this paper, we investigate a reformulation of (1.7) in a product space that allows us to capture instances of (1.1) in which none of the functions need be differentiable. The resulting algorithm proceeds by decomposition in that each function is involved individually via its own proximity operator. Since proximity operators can be implemented for a wide variety of potentials, the proposed framework is applicable to a broad array of problems.

In section 2, we set our notation and provide some background on convex analysis and proximity operators. We also obtain closed-form formulas for new examples of proximity operators that will be used subsequently. In section 3, we introduce our algorithm and prove its weak convergence.

Applications to signal and image processing problems are detailed in section 4, where numerical results are also provided. These results show that complex nonsmooth variational inverse problems, that were beyond the reach of the methods reviewed above, can be decomposed and solved efficiently within the proposed framework. Section 5 concludes the paper with some remarks.

2 Notation and background

2.1 Convex analysis

We provide here some basic elements; for proofs and complements see [51] and, for the finite dimensional setting, [43].

Let C be a nonempty convex subset of $\mathcal{H}$. The indicator function of C is

$$\iota_C \colon x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C, \end{cases} \tag{2.1}$$

its distance function is

$$d_C \colon \mathcal{H} \to [0, +\infty[\, : x \mapsto \inf_{y \in C} \|x - y\|, \tag{2.2}$$

its support function is

$$\sigma_C \colon \mathcal{H} \to \,]-\infty, +\infty] : u \mapsto \sup_{x \in C} \langle x \mid u \rangle, \tag{2.3}$$

and its conical hull is

$$\operatorname{cone} C = \bigcup_{\lambda > 0} \{ \lambda x \mid x \in C \}. \tag{2.4}$$

Moreover, $\operatorname{span} C$ denotes the span of C and $\overline{\operatorname{span}}\, C$ the closure of $\operatorname{span} C$. The strong relative interior of C is

$$\operatorname{sri} C = \left\{ x \in C \mid \operatorname{cone}(C - x) = \overline{\operatorname{span}}\, (C - x) \right\} \tag{2.5}$$

and its relative interior is

$$\operatorname{ri} C = \left\{ x \in C \mid \operatorname{cone}(C - x) = \operatorname{span}(C - x) \right\}. \tag{2.6}$$

We have

$$\operatorname{int} C \subset \operatorname{sri} C \subset \operatorname{ri} C \subset C. \tag{2.7}$$

Lemma 2.1 [43, Section 6] *Suppose that $\mathcal{H}$ is finite-dimensional, and let C and D be convex subsets of $\mathcal{H}$. Then the following hold.*

(i) *Suppose that $C \neq \varnothing$. Then $\operatorname{sri} C = \operatorname{ri} C \neq \varnothing$.*

(ii) $\operatorname{ri}(C - D) = \operatorname{ri} C - \operatorname{ri} D$.

(iii) *Suppose that D is an affine subspace and that $(\operatorname{ri} C) \cap D \neq \varnothing$. Then $\operatorname{ri}(C \cap D) = (\operatorname{ri} C) \cap D$.*

4

Now let C be a nonempty closed and convex subset of $\mathcal{H}$. The projection of a point x in $\mathcal{H}$ onto C is the unique point $P_C x$ in C such that $\|x - P_C x\| = d_C(x)$. We have

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = P_C x \;\Leftrightarrow\; \big[\, p \in C \quad \text{and} \quad (\forall y \in C) \quad \langle y - p \mid x - p \rangle \leq 0 \,\big]. \tag{2.8}$$

Moreover, d_C is Fréchet differentiable on $\mathcal{H} \smallsetminus C$ and

$$(\forall x \in \mathcal{H} \smallsetminus C) \quad \nabla d_C(x) = \frac{x - P_C x}{d_C(x)}. \tag{2.9}$$

The domain of a function $f \colon \mathcal{H} \to {]-\infty, +\infty]}$ is $\operatorname{dom} f = \{ x \in \mathcal{H} \mid f(x) < +\infty \}$ and its set of global minimizers is denoted by $\operatorname{Argmin} f$; if f possesses a unique global minimizer, it is denoted by $\operatorname{argmin}_{y \in \mathcal{H}} f(y)$. The class of lower semicontinuous convex functions from $\mathcal{H}$ to ${]-\infty, +\infty]}$ which are proper (i.e., with nonempty domain) is denoted by $\Gamma_0(\mathcal{H})$. Now let $f \in \Gamma_0(\mathcal{H})$. The conjugate of f is the function $f^* \in \Gamma_0(\mathcal{H})$ defined by $f^* \colon \mathcal{H} \to {]-\infty, +\infty]} : u \mapsto \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - f(x)$, and the subdifferential of f is the set-valued operator

$$\partial f \colon \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \big\{ u \in \mathcal{H} \mid (\forall y \in \operatorname{dom} f) \;\; \langle y - x \mid u \rangle + f(x) \leq f(y) \big\}. \tag{2.10}$$

We have

$$(\forall x \in \mathcal{H}) \quad x \in \operatorname{Argmin} f \;\Leftrightarrow\; 0 \in \partial f(x) \tag{2.11}$$

and

$$(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad \begin{cases} f(x) + f^*(u) \geq \langle x \mid u \rangle \\ f(x) + f^*(u) = \langle x \mid u \rangle \;\Leftrightarrow\; u \in \partial f(x). \end{cases} \tag{2.12}$$

Moreover, if f is Gâteaux-differentiable at $x \in \mathcal{H}$, then $\partial f(x) = \{\nabla f(x)\}$.

Lemma 2.2 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \colon \mathbb{R} \to \mathbb{R}$ be an even convex function, and set $f = \phi \circ d_C$. Then $f \in \Gamma_0(\mathcal{H})$ and $f^* = \sigma_C + \phi^* \circ \| \cdot \|$.*

Proof. Since $\phi \colon \mathbb{R} \to \mathbb{R}$ is convex and even, it is continuous and increasing on $[0, +\infty[$. On the other hand, since C is convex, d_C is convex. Hence, $\phi \circ d_C$ is a finite continuous convex function, which shows that $f \in \Gamma_0(\mathcal{H})$. Moreover, $\phi \circ d_C = \phi(\inf_{y \in C} \| \cdot - y \|) = \inf_{y \in C} \phi \circ \| \cdot - y \|$. Therefore,

$$\begin{aligned} (\forall u \in \mathcal{H}) \quad f^*(u) &= \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - \inf_{y \in C} \phi(\|x - y\|) \\ &= \sup_{y \in C} \langle y \mid u \rangle + \sup_{x \in \mathcal{H}} \langle x - y \mid u \rangle - (\phi \circ \| \cdot \|)(x - y) \\ &= \sup_{y \in C} \langle y \mid u \rangle + (\phi \circ \| \cdot \|)^*(u) \\ &= \sigma_C(u) + (\phi \circ \| \cdot \|)^*(u). \end{aligned} \tag{2.13}$$

Since $(\phi \circ \| \cdot \|)^* = \phi^* \circ \| \cdot \|$ [31, Proposition I.4.2], the proof is complete. $\square$

2.2 Proximity operators

For detailed accounts of the theory of proximity operators, see [26, Section 2] and [39].

The proximity operator of a function $f \in \Gamma_0(\mathcal{H})$ is the operator $\operatorname{prox}_f \colon \mathcal{H} \to \mathcal{H}$ which maps every $x \in \mathcal{H}$ to the unique minimizer of the function $f + \|x - \cdot\|^2/2$, i.e.,

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_f x = \operatorname*{argmin}_{y \in \mathcal{H}} \ f(y) + \frac{1}{2}\|x - y\|^2. \tag{2.14}$$

We have

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \operatorname{prox}_f x \ \Leftrightarrow \ x - p \in \partial f(p). \tag{2.15}$$

In other words, $\operatorname{prox}_f = (\operatorname{Id} + \partial f)^{-1}$.

Lemma 2.3 *Let $f \in \Gamma_0(\mathcal{H})$. Then the following hold.*

(i) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \ \| \operatorname{prox}_f x - \operatorname{prox}_f y\|^2 \le \langle x - y \mid \operatorname{prox}_f x - \operatorname{prox}_f y \rangle.$

(ii) $(\forall x \in \mathcal{H})(\forall \gamma \in \,]0, +\infty[) \ x = \operatorname{prox}_{\gamma f} x + \gamma \operatorname{prox}_{f^*/\gamma}(x/\gamma).$

Lemma 2.4 [25, Proposition 11] *Let $\mathcal{G}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{G})$, and let $L \colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator such that $L \circ L^* = \kappa \operatorname{Id}$, for some $\kappa \in \,]0, +\infty[$. Then $f \circ L \in \Gamma_0(\mathcal{H})$ and*

$$\operatorname{prox}_{f \circ L} = \operatorname{Id} + \frac{1}{\kappa} L^* \circ (\operatorname{prox}_{\kappa f} - \operatorname{Id}) \circ L. \tag{2.16}$$

2.3 Examples of proximity operators

Closed-form formulas for various proximity operators are provided in [16, 24, 25, 26, 39]. The following examples will be of immediate use subsequently.

Proposition 2.5 [16, Proposition 2.10 and Remark 3.2(ii)] *Set*

$$f \colon \mathcal{H} \to \,]-\infty, +\infty] \colon x \mapsto \sum_{k \in \mathbb{K}} \phi_k(\langle x \mid e_k \rangle), \tag{2.17}$$

where:

(i) $\varnothing \ne \mathbb{K} \subset \mathbb{N}$;

(ii) $(e_k)_{k \in \mathbb{K}}$ *is an orthonormal basis of $\mathcal{H}$;*

(iii) $(\phi_k)_{k \in \mathbb{K}}$ *are functions in $\Gamma_0(\mathbb{R})$;*

(iv) *Either $\mathbb{K}$ is finite, or there exists a subset $\mathbb{L}$ of $\mathbb{K}$ such that:*

    (a) $\mathbb{K} \smallsetminus \mathbb{L}$ *is finite;*

(b) $(\forall k \in \mathbb{L}) \; \phi_k \geq \phi_k(0) = 0$.

Then $f \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f\, x = \sum_{k \in \mathbb{K}} \big(\mathrm{prox}_{\phi_k} \langle x \mid e_k \rangle \big) e_k. \tag{2.18}$$

We shall also require the following results, which appear to be new.

Proposition 2.6 *Let $(\mathcal{G}, \|\cdot\|)$ be a real Hilbert space, let $L\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $z \in \mathcal{G}$, let $\gamma \in\,]0, +\infty[$, and set $f = \gamma\|L \cdot -z\|^2/2$. Then $f \in \Gamma_0(\mathcal{H})$ and*

$$(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f\, x = (\mathrm{Id} + \gamma L^* L)^{-1}(x + \gamma L^* z). \tag{2.19}$$

Proof. It is clear that f is a finite continuous convex function. Now, take x and p in $\mathcal{H}$. Then (2.15) yields $p = \mathrm{prox}_f\, x \Leftrightarrow x - p = \nabla\big(\gamma\|L \cdot -z\|^2/2\big)(p) \Leftrightarrow x - p = \gamma L^*(Lp - z) \Leftrightarrow p = (\mathrm{Id} + \gamma L^* L)^{-1}(x + \gamma L^* z)$. $\square$

Proposition 2.7 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even convex function which is differentiable on $\mathbb{R} \smallsetminus \{0\}$, and set $f = \phi \circ d_C$. Then*

$$(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f\, x = \begin{cases} x + \dfrac{\mathrm{prox}_{\phi^*}\, d_C(x)}{d_C(x)}(P_C x - x), & \text{if } d_C(x) > \max \partial\phi(0); \\ P_C x, & \text{if } d_C(x) \leq \max \partial\phi(0). \end{cases} \tag{2.20}$$

Proof. As seen in Lemma 2.2, $f \in \Gamma_0(\mathcal{H})$. Now let $x \in \mathcal{H}$ and set $p = \mathrm{prox}_f\, x$. Since ϕ is a finite even convex function, $\partial\phi(0) = [-\beta, \beta]$ for some $\beta \in [0, +\infty[$ [43, Theorem 23.4]. We consider two alternatives.

(a) $p \in C$: Let $y \in C$. Then $f(y) = \phi(d_C(y)) = \phi(0)$ and, in particular, $f(p) = \phi(0)$. Hence, it follows from (2.15) and (2.10) that

$$\langle y - p \mid x - p \rangle + \phi(0) = \langle y - p \mid x - p \rangle + f(p) \leq f(y) = \phi(0). \tag{2.21}$$

Consequently, $\langle y - p \mid x - p \rangle \leq 0$ and, in view of (2.8), we get $p = P_C x$. Thus,

$$p \in C \quad \Leftrightarrow \quad p = P_C x. \tag{2.22}$$

Now, let $u \in \partial f(p)$. Since $p \in C$, $d_C(p) = 0$ and, by (2.3), $\sigma_C(u) \geq \langle p \mid u \rangle$. Hence, (2.12) and Lemma 2.2 yield

$$-0\,\|u\| = 0 \leq \sigma_C(u) - \langle p \mid u \rangle = \sigma_C(u) - f(p) - f^*(u) = -\phi(0) - \phi^*(\|u\|). \tag{2.23}$$

We therefore deduce from (2.12) that $\|u\| \in \partial\phi(0)$. Thus, $u \in \partial f(p) \Rightarrow \|u\| \leq \beta$. Since (2.15) asserts that $x - p \in \partial f(p)$, we obtain $\|x - p\| \leq \beta$ and hence, since $p \in C$, $d_C(x) \leq \|x - p\| \leq \beta$. As a result,

$$p \in C \quad \Rightarrow \quad d_C(x) \leq \beta. \tag{2.24}$$

7

(b) $p \notin C$: Since C is closed, $d_C(p) > 0$ and ϕ is therefore differentiable at $d_C(p)$. It follows from (2.15), the Fréchet chain rule, and (2.9) that

$$x - p = f'(p) = \frac{\phi'(d_C(p))}{d_C(p)}(p - P_C p). \tag{2.25}$$

Since $\phi' \geq 0$ on $]0, +\infty[$, upon taking the norm, we obtain

$$\|p - x\| = \phi'(d_C(p)) \tag{2.26}$$

and therefore

$$p - x = \frac{\|p - x\|}{d_C(p)}(P_C p - p). \tag{2.27}$$

In turn, appealing to Lemma 2.3(i) (with $f = \iota_C$) and (2.8), we obtain

$$\|P_C p - P_C x\|^2 \leq \langle p - x \mid P_C p - P_C x \rangle = \frac{\|p - x\|}{d_C(p)} \langle P_C p - p \mid P_C p - P_C x \rangle \leq 0, \tag{2.28}$$

from which we deduce that

$$P_C p = P_C x. \tag{2.29}$$

Hence, (2.27) becomes

$$p - x = \frac{\|p - x\|}{\|p - P_C x\|}(P_C x - p), \tag{2.30}$$

which can be rewritten as

$$p - x = \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|}(P_C x - x). \tag{2.31}$$

Taking the norm yields

$$\|p - x\| = \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|} d_C(x), \tag{2.32}$$

and it follows from (2.29) that

$$d_C(x) = \|p - x\| + \|p - P_C x\| = \|p - x\| + d_C(p). \tag{2.33}$$

Therefore, in the light of (2.26), we obtain

$$d_C(x) - d_C(p) = \|p - x\| = \phi'(d_C(p)) \tag{2.34}$$

and we derive from (2.15) that

$$d_C(p) = \operatorname{prox}_\phi d_C(x). \tag{2.35}$$

Thus, Lemma 2.3(ii) yields

$$d_C(x) - d_C(p) = d_C(x) - \operatorname{prox}_\phi d_C(x) = \operatorname{prox}_{\phi^*} d_C(x) \tag{2.36}$$

and, in turn, (2.34) results in

$$\|p - x\| = d_C(x) - d_C(p) = \operatorname{prox}_{\phi^*} d_C(x). \tag{2.37}$$

8

To sum up, coming back to (2.31) and invoking (2.33) and (2.37), we obtain

$$p \notin C \quad \Rightarrow \quad p = x + \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|}(P_C x - x)$$

$$= x + \frac{\mathrm{prox}_{\phi^*} d_C(x)}{d_C(x)}(P_C x - x). \tag{2.38}$$

Furthermore, we derive from (2.35) and (2.15) that

$$p \notin C \Rightarrow d_C(p) > 0 \Rightarrow \mathrm{prox}_\phi d_C(x) \neq 0 \Rightarrow d_C(x) \notin \partial\phi(0) \Rightarrow d_C(x) > \beta. \tag{2.39}$$

Upon combining (2.24) and (2.39), we obtain

$$p \in C \quad \Leftrightarrow \quad d_C(x) \leq \beta. \tag{2.40}$$

Altogether, (2.20) follows from (2.22), (2.38), and (2.40). □

The above proposition shows that a nice feature of the proximity operator of $\phi \circ d_C$ is that it can be decomposed in terms of prox_{ϕ^*} and P_C. Here is an application of this result.

Proposition 2.8 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\alpha \in \,]0, +\infty[$, let $p \in [1, +\infty[$, and set $f = \alpha d_C^p$. Then the following hold.*

(i) *Suppose that $p = 1$. Then*

$$(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f x = \begin{cases} x + \dfrac{\alpha}{d_C(x)}(P_C x - x), & \text{if } d_C(x) > \alpha; \\ P_C x, & \text{if } d_C(x) \leq \alpha. \end{cases} \tag{2.41}$$

(ii) *Suppose that $p > 1$. Then*

$$(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f x = \begin{cases} x + \dfrac{\nu(x)}{d_C(x)}(P_C x - x), & \text{if } x \notin C; \\ x, & \text{if } x \in C, \end{cases} \tag{2.42}$$

where $\nu(x)$ is the unique real number in $[0, +\infty[$ that satisfies $\nu(x) + (\nu(x)/(\alpha p))^{1/(p-1)} = d_C(x)$.

Proof. (i): Set $\phi = \alpha|\cdot|$. Then $\max \partial\phi(0) = \max[-\alpha, \alpha] = \alpha$ and $\phi^* = \iota_{[-\alpha, \alpha]}$. Therefore, $\mathrm{prox}_{\phi^*} = P_{[-\alpha, \alpha]}$ and hence $(\forall \mu \in \,]\alpha, +\infty[)$ $\mathrm{prox}_{\phi^*} \mu = \alpha$. In view of (2.20), we obtain (2.41).

(ii): Let $x \in \mathcal{H}$ and note that, since C is closed, $d_C(x) > 0 \Leftrightarrow x \notin C$. Now set $\phi = \alpha|\cdot|^p$. Then $\max \partial\phi(0) = \max\{0\} = 0$ and $\phi^* \colon \mu \mapsto (p-1)(\alpha p)^{1/(1-p)}|\mu|^{p/(p-1)}/p$. Hence, it follows from (2.15) and [24, Corollary 2.5] that $\mathrm{prox}_{\phi^*} d_C(x)$ is the unique solution $\nu(x) \in [0, +\infty[$ to the equation $d_C(x) - \nu(x) = \phi^{*\prime}(\nu(x)) = (\nu(x)/(\alpha p))^{1/(p-1)}$. Appealing to (2.20), we obtain (2.42). □

Let us note that explicit expressions can be obtained for several values of p in Proposition 2.8(ii). Here is an example that will be used subsequently.

Example 2.9 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\alpha \in]0, +\infty[$, and set $f = \alpha d_C^{3/2}$. Then

$$
(\forall x \in \mathcal{H}) \quad \mathrm{prox}_f\, x = \begin{cases} x + \dfrac{9\alpha^2\left(\sqrt{1 + 16 d_C(x)/(9\alpha^2)} - 1\right)}{8 d_C(x)}(P_C x - x), & \text{if } x \notin C; \\ x, & \text{if } x \in C. \end{cases}
\tag{2.43}
$$

Proof. Set $p = 3/2$ in Proposition 2.8(ii). □

3 Algorithm and convergence

The main algorithm is presented in section 3.1. In section 3.2, we revisit the Douglas-Rachford algorithm in the context of minimization problems (Proposition 3.2), with special emphasis on its convergence in a specific case (Proposition 3.3). These results are transcribed in a product space in section 3.3 to prove the weak convergence of Algorithm 3.1.

3.1 Algorithm

We propose the following proximal method to solve (1.1). In this splitting algorithm, each function f_i is used separately by means of its own proximity operator.

Algorithm 3.1 For every $i \in \{1, \ldots, m\}$, let $(a_{i,n})_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$. A sequence $(x_n)_{n \in \mathbb{N}}$ is generated by the following routine.

$$
\begin{aligned}
&\text{Initialization} \\
&\left\lfloor \begin{aligned}
&\gamma \in]0, +\infty[\\
&(\omega_i)_{1 \le i \le m} \in]0, 1]^m \text{ satisfy } \textstyle\sum_{i=1}^m \omega_i = 1 \\
&(y_{i,0})_{1 \le i \le m} \in \mathcal{H}^m \\
&x_0 = \sum_{i=1}^m \omega_i y_{i,0}
\end{aligned} \right. \\
&\text{For } n = 0, 1, \ldots \\
&\left\lfloor \begin{aligned}
&\text{For } i = 1, \ldots, m \\
&\quad \left\lfloor\, p_{i,n} = \mathrm{prox}_{\gamma f_i/\omega_i}\, y_{i,n} + a_{i,n} \right. \\
&p_n = \sum_{i=1}^m \omega_i p_{i,n} \\
&\lambda_n \in]0, 2[\\
&\text{For } i = 1, \ldots, m \\
&\quad \left\lfloor\, y_{i,n+1} = y_{i,n} + \lambda_n\big(2p_n - x_n - p_{i,n}\big) \right. \\
&x_{n+1} = x_n + \lambda_n(p_n - x_n).
\end{aligned} \right.
\end{aligned}
\tag{3.1}
$$

At iteration n, the proximal vectors $(p_{i,n})_{1 \le i \le m}$, as well as the auxiliary vectors $(y_{i,n})_{1 \le i \le m}$, can be computed simultaneously, hence the parallel structure of Algorithm 3.1. Another feature of the algorithm is that some error $a_{i,n}$ is tolerated in the computation of the ith proximity operator.

3.2 The Douglas-Rachford algorithm for minimization problems

To ease our presentation, we introduce in this section a second real Hilbert space $(\mathcal{H}, ||| \cdot |||)$. As usual, $\rightharpoonup$ denotes weak convergence.

The (nonlinear) Douglas-Rachford splitting method was initially developed for the problem of finding a zero of the sum of two maximal monotone operators in [38] (see [22] for recent refinements). In the case when the maximal monotone operators are subdifferentials, it provides an algorithm for minimizing the sum of two convex functions. In this section, we develop this point of view, starting with the following result.

Proposition 3.2 *Let $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ be functions in $\Gamma_0(\mathcal{H})$, let $(\boldsymbol{a}_n)_{n \in \mathbb{N}}$ and $(\boldsymbol{b}_n)_{n \in \mathbb{N}}$ be sequences in $\mathcal{H}$, and let $(\boldsymbol{y}_n)_{n \in \mathbb{N}}$ be a sequence generated by the following routine.*

$$
\begin{array}{l}
\text{Initialization} \\
\left\lfloor
\begin{array}{l}
\gamma \in \,]0, +\infty[\\
\boldsymbol{y}_0 \in \mathcal{H}
\end{array}
\right. \\
\text{For } n = 0, 1, \dots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{y}_{n+\frac{1}{2}} = \text{prox}_{\gamma \boldsymbol{f}_2} \boldsymbol{y}_n + \boldsymbol{a}_n \\
\lambda_n \in \,]0, 2[\\
\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n \Big(\text{prox}_{\gamma \boldsymbol{f}_1} \big(2\boldsymbol{y}_{n+\frac{1}{2}} - \boldsymbol{y}_n \big) + \boldsymbol{b}_n - \boldsymbol{y}_{n+\frac{1}{2}} \Big).
\end{array}
\right.
\end{array}
\tag{3.2}
$$

Set

$$
\boldsymbol{G} = \text{Argmin} \, \boldsymbol{f}_1 + \boldsymbol{f}_2 \quad \text{and} \quad \boldsymbol{T} = 2 \, \text{prox}_{\gamma \boldsymbol{f}_1} \circ (2 \, \text{prox}_{\gamma \boldsymbol{f}_2} - \mathbf{Id}) - 2 \, \text{prox}_{\gamma \boldsymbol{f}_2} + \mathbf{Id}, \tag{3.3}
$$

and suppose that the following hold.

(i) $\displaystyle \lim_{||| \boldsymbol{x} ||| \to +\infty} \boldsymbol{f}_1(\boldsymbol{x}) + \boldsymbol{f}_2(\boldsymbol{x}) = +\infty.$

(ii) $\boldsymbol{0} \in \text{sri}(\text{dom} \, \boldsymbol{f}_1 - \text{dom} \, \boldsymbol{f}_2).$

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

(iv) $\sum_{n \in \mathbb{N}} \lambda_n (||| \boldsymbol{a}_n ||| + ||| \boldsymbol{b}_n |||) < +\infty.$

Then $\boldsymbol{G} \neq \varnothing$, $(\boldsymbol{y}_n)_{n \in \mathbb{N}}$ converges weakly to a fixed point $\boldsymbol{y}$ of $\boldsymbol{T}$, and $\text{prox}_{\gamma \boldsymbol{f}_2} \boldsymbol{y} \in \boldsymbol{G}$.

Proof. It follows from (ii) that $\operatorname{dom}(\boldsymbol{f}_1 + \boldsymbol{f}_2) = \operatorname{dom}\boldsymbol{f}_1 \cap \operatorname{dom}\boldsymbol{f}_2 \neq \varnothing$. Hence, since $\boldsymbol{f}_1 + \boldsymbol{f}_2$ is lower semicontinuous and convex as the sum of two such functions, we have $\boldsymbol{f}_1 + \boldsymbol{f}_2 \in \Gamma_0(\mathcal{H})$. In turn, we derive from (i) and [51, Theorem 2.5.1(ii)] that

$$\boldsymbol{G} \neq \varnothing. \tag{3.4}$$

Next, let us set $\boldsymbol{A}_1 = \partial\boldsymbol{f}_1$, $\boldsymbol{A}_2 = \partial\boldsymbol{f}_2$, and $\boldsymbol{Z} = \{\boldsymbol{x} \in \mathcal{H} \mid \boldsymbol{0} \in \boldsymbol{A}_1\boldsymbol{x} + \boldsymbol{A}_2\boldsymbol{x}\}$. Then $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are maximal monotone operators [51, Theorem 3.1.11]. In addition, in view of (2.15), the resolvents of $\gamma\boldsymbol{A}_1$ and $\gamma\boldsymbol{A}_2$ are respectively

$$J_{\gamma\boldsymbol{A}_1} = (\operatorname{Id} + \gamma\boldsymbol{A}_1)^{-1} = \operatorname{prox}_{\gamma\boldsymbol{f}_1} \quad \text{and} \quad J_{\gamma\boldsymbol{A}_2} = (\operatorname{Id} + \gamma\boldsymbol{A}_2)^{-1} = \operatorname{prox}_{\gamma\boldsymbol{f}_2}. \tag{3.5}$$

Thus, the iteration in (3.2) can be rewritten as

$$\left| \begin{array}{l} \boldsymbol{y}_{n+\frac{1}{2}} = J_{\gamma\boldsymbol{A}_2}\boldsymbol{y}_n + \boldsymbol{a}_n \\[4pt] \lambda_n \in\,]0, 2[\\[4pt] \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\Big(J_{\gamma\boldsymbol{A}_1}\big(2\boldsymbol{y}_{n+\frac{1}{2}} - \boldsymbol{y}_n\big) + \boldsymbol{b}_n - \boldsymbol{y}_{n+\frac{1}{2}}\Big). \end{array} \right. \tag{3.6}$$

Moreover, it follows from (2.11), (ii), and [51, Theorem 2.8.3] that

$$\boldsymbol{G} = \{\boldsymbol{x} \in \mathcal{H} \mid \boldsymbol{0} \in \partial(\boldsymbol{f}_1 + \boldsymbol{f}_2)(\boldsymbol{x})\} = \{\boldsymbol{x} \in \mathcal{H} \mid \boldsymbol{0} \in \partial\boldsymbol{f}_1(\boldsymbol{x}) + \partial\boldsymbol{f}_2(\boldsymbol{x})\} = \boldsymbol{Z}. \tag{3.7}$$

Thus, (3.4) yields $\boldsymbol{Z} \neq \varnothing$ and it follows from (iii), (iv), and the results of [22, Section 5] that $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ converges weakly to a fixed point $\boldsymbol{y}$ of the operator $2J_{\gamma\boldsymbol{A}_1} \circ (2J_{\gamma\boldsymbol{A}_2} - \operatorname{\mathbf{Id}}) - 2J_{\gamma\boldsymbol{A}_2} + \operatorname{\mathbf{Id}}$, and that $J_{\gamma\boldsymbol{A}_2}\boldsymbol{y} \in \boldsymbol{Z}$. In view of (3.3), (3.5), and (3.7), the proof is complete. $\square$

It is important to stress that algorithm (3.2) provides a minimizer indirectly: the sequence $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ is first constructed, and then a minimizer of $\boldsymbol{f}_1 + \boldsymbol{f}_2$ is obtained as the image of the weak limit $\boldsymbol{y}$ of $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ under $\operatorname{prox}_{\gamma\boldsymbol{f}_2}$. In general, nothing is known about the weak convergence of the sequences $(\operatorname{prox}_{\gamma\boldsymbol{f}_1}\boldsymbol{y}_n)_{n\in\mathbb{N}}$ and $(\operatorname{prox}_{\gamma\boldsymbol{f}_2}\boldsymbol{y}_n)_{n\in\mathbb{N}}$. The following result describes a remarkable situation in which $(\operatorname{prox}_{\gamma\boldsymbol{f}_1}\boldsymbol{y}_n)_{n\in\mathbb{N}}$ does converges weakly and its weak limit turns out to be a minimizer of $\boldsymbol{f}_1 + \boldsymbol{f}_2$.

Proposition 3.3 *Let $\boldsymbol{D}$ be a closed vector subspace of $\mathcal{H}$, let $\boldsymbol{f} \in \Gamma_0(\mathcal{H})$, let $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, and let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ be a sequence generated by the following routine.*

$$\begin{array}{l} \text{Initialization} \\ \left| \begin{array}{l} \gamma \in\,]0, +\infty[\\ \boldsymbol{y}_0 \in \mathcal{H} \\ \boldsymbol{x}_0 = P_{\boldsymbol{D}}\boldsymbol{y}_0 \end{array} \right. \\ \text{For } n = 0, 1, \dots \\ \left| \begin{array}{l} \boldsymbol{y}_{n+\frac{1}{2}} = \operatorname{prox}_{\gamma\boldsymbol{f}}\boldsymbol{y}_n + \boldsymbol{a}_n \\[4pt] \boldsymbol{p}_n = P_{\boldsymbol{D}}\boldsymbol{y}_{n+\frac{1}{2}} \\[4pt] \lambda_n \in\,]0, 2[\\[4pt] \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\big(2\boldsymbol{p}_n - \boldsymbol{x}_n - \boldsymbol{y}_{n+\frac{1}{2}}\big) \\[4pt] \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n(\boldsymbol{p}_n - \boldsymbol{x}_n). \end{array} \right. \end{array} \tag{3.8}$$

Let $\boldsymbol{G}$ be the set of minimizers of $\boldsymbol{f}$ over $\boldsymbol{D}$ and suppose that the following hold.

(i) $\displaystyle\lim_{\boldsymbol{x}\in\boldsymbol{D},\,|||\boldsymbol{x}|||\to+\infty} \boldsymbol{f}(\boldsymbol{x}) = +\infty$.

(ii) $\boldsymbol{0}\in\mathrm{sri}(\boldsymbol{D}-\mathrm{dom}\,\boldsymbol{f})$.

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

(iv) $\sum_{n\in\mathbb{N}}\lambda_n|||\boldsymbol{a}_n||| < +\infty$.

Then $\boldsymbol{G}\neq\varnothing$ and $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\boldsymbol{G}$.

Proof. Set $\boldsymbol{f}_1 = \iota_{\boldsymbol{D}}$, $\boldsymbol{f}_2 = \boldsymbol{f}$, and $(\forall n\in\mathbb{N})\;\boldsymbol{b}_n = \boldsymbol{0}$. Then (2.1) and (2.14) yield $\mathrm{prox}_{\gamma\boldsymbol{f}_1} = P_{\boldsymbol{D}}$ and, since $\boldsymbol{D}$ is a closed vector subspace, $P_{\boldsymbol{D}}$ is a linear operator. Hence, proceeding by induction, we can rewrite the update equation for $\boldsymbol{x}_n$ in (3.8) as

$$
\begin{aligned}
\boldsymbol{x}_{n+1} &= \boldsymbol{x}_n + \lambda_n(\boldsymbol{p}_n - \boldsymbol{x}_n) \\
&= P_{\boldsymbol{D}}\,\boldsymbol{y}_n + \lambda_n\big(2P_{\boldsymbol{D}}\,\boldsymbol{p}_n - P_{\boldsymbol{D}}\,\boldsymbol{x}_n - P_{\boldsymbol{D}}\,\boldsymbol{y}_{n+\frac{1}{2}}\big) \\
&= P_{\boldsymbol{D}}\Big(\boldsymbol{y}_n + \lambda_n\big(2\boldsymbol{p}_n - \boldsymbol{x}_n - \boldsymbol{y}_{n+\frac{1}{2}}\big)\Big) \\
&= P_{\boldsymbol{D}}\,\boldsymbol{y}_{n+1}.
\end{aligned}
\tag{3.9}
$$

As a result, (3.8) is equivalent to

$$
\begin{aligned}
&\text{Initialization} \\
&\left\lfloor
\begin{aligned}
&\gamma\in\,]0,+\infty[\\
&\boldsymbol{y}_0\in\mathcal{H}
\end{aligned}
\right. \\
&\text{For } n = 0, 1, \ldots \\
&\left\lfloor
\begin{aligned}
&\boldsymbol{x}_n = P_{\boldsymbol{D}}\,\boldsymbol{y}_n \\
&\boldsymbol{y}_{n+\frac{1}{2}} = \mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}_n + \boldsymbol{a}_n \\
&\boldsymbol{p}_n = P_{\boldsymbol{D}}\,\boldsymbol{y}_{n+\frac{1}{2}} \\
&\lambda_n\in\,]0,2[\\
&\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\big(2\boldsymbol{p}_n - \boldsymbol{x}_n - \boldsymbol{y}_{n+\frac{1}{2}}\big).
\end{aligned}
\right.
\end{aligned}
\tag{3.10}
$$

Thus, since

$$
(\forall\boldsymbol{x}\in\mathcal{H})(\forall\boldsymbol{y}\in\mathcal{H})\quad P_{\boldsymbol{D}}(2\boldsymbol{y}-\boldsymbol{x}) = 2P_{\boldsymbol{D}}\boldsymbol{y} - P_{\boldsymbol{D}}\boldsymbol{x},
\tag{3.11}
$$

(3.10) appears as a special case of (3.2) in which we have introduced the auxiliary variables $\boldsymbol{x}_n$ and $\boldsymbol{p}_n$. In addition, the operator $\boldsymbol{T}$ of (3.3) becomes

$$
\boldsymbol{T} = 4(P_{\boldsymbol{D}}\circ\mathrm{prox}_{\gamma\boldsymbol{f}}) - 2P_{\boldsymbol{D}} - 2\,\mathrm{prox}_{\gamma\boldsymbol{f}} + \mathbf{Id}\,.
\tag{3.12}
$$

Since (i)–(iv) are specializations of their respective counterparts in Proposition 3.2, it follows from Proposition 3.2 that $\boldsymbol{G}\neq\varnothing$ and that there exists a fixed point $\boldsymbol{y}$ of $\boldsymbol{T}$ such that $\boldsymbol{y}_n\rightharpoonup\boldsymbol{y}$ and $\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}\in\boldsymbol{G}$. Note that, since $\boldsymbol{G}\subset\boldsymbol{D}$, $\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}\in\boldsymbol{D}$ and, in turn, $P_{\boldsymbol{D}}(\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}) = \mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}$. Thus, in view of (3.12), we obtain

$$
\begin{aligned}
\boldsymbol{T}\boldsymbol{y} = \boldsymbol{y} \;&\Leftrightarrow\; 4P_{\boldsymbol{D}}(\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}) - 2P_{\boldsymbol{D}}\boldsymbol{y} - 2\,\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y} + \boldsymbol{y} = \boldsymbol{y} & (3.13) \\
&\Leftrightarrow\; 2P_{\boldsymbol{D}}(\mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y}) - P_{\boldsymbol{D}}\boldsymbol{y} = \mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y} & \\
&\Leftrightarrow\; \mathrm{prox}_{\gamma\boldsymbol{f}}\,\boldsymbol{y} = P_{\boldsymbol{D}}\boldsymbol{y}. & (3.14)
\end{aligned}
$$

13

Hence, since $\operatorname{prox}_{\gamma f} \boldsymbol{y} \in \boldsymbol{G}$, we also have $P_{\boldsymbol{D}} \boldsymbol{y} \in \boldsymbol{G}$. On the other hand, since $P_{\boldsymbol{D}}$ is linear and continuous, it is weakly continuous and therefore $\boldsymbol{y}_n \rightharpoonup \boldsymbol{y} \Rightarrow P_{\boldsymbol{D}} \boldsymbol{y}_n \rightharpoonup P_{\boldsymbol{D}} \boldsymbol{y} \in \boldsymbol{G}$. We conclude that $\boldsymbol{x}_n \rightharpoonup P_{\boldsymbol{D}} \boldsymbol{y} \in \boldsymbol{G}$. □

3.3 Convergence of Algorithm 3.1

The convergence of the main algorithm can now be demonstrated.

Theorem 3.4 *Let G be the set of solutions to (1.1) and let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 3.1 under the following assumptions.*

(i) $\lim\limits_{\|x\|\to+\infty} f_1(x) + \cdots + f_m(x) = +\infty.$

(ii) $(0,\ldots,0) \in \operatorname{sri}\left\{(x - x_1, \ldots, x - x_m) \mid x \in \mathcal{H}, \, x_1 \in \operatorname{dom} f_1, \ldots, \, x_m \in \operatorname{dom} f_m\right\}.$

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

(iv) $(\forall i \in \{1,\ldots,m\}) \; \sum_{n\in\mathbb{N}} \lambda_n \|a_{i,n}\| < +\infty.$

Then $G \neq \varnothing$ and $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in G.

Proof. Let $\mathcal{H}$ be the real Hilbert space obtained by endowing the m-fold Cartesian product $\mathcal{H}^m$ with the scalar product

$$\langle\langle \cdot \mid \cdot \rangle\rangle \colon (\boldsymbol{x}, \boldsymbol{y}) \mapsto \sum_{i=1}^{m} \omega_i \langle x_i \mid y_i \rangle, \tag{3.15}$$

where $(\omega_i)_{1 \le i \le m}$ is defined in (3.1), and where $\boldsymbol{x} = (x_i)_{1\le i\le m}$ and $\boldsymbol{y} = (y_i)_{1\le i\le m}$ denote generic elements in $\mathcal{H}$. The associated norm is denoted by $|||\cdot|||$, i.e.,

$$|||\cdot||| \colon \boldsymbol{x} \mapsto \sqrt{\sum_{i=1}^{m} \omega_i \|x_i\|^2}. \tag{3.16}$$

Furthermore, set

$$\boldsymbol{D} = \left\{(x,\ldots,x) \in \mathcal{H} \mid x \in \mathcal{H}\right\} \tag{3.17}$$

and

$$\boldsymbol{f} \colon \mathcal{H} \to \left]-\infty, +\infty\right] : \boldsymbol{x} \mapsto \sum_{i=1}^{m} f_i(x_i). \tag{3.18}$$

It follows from (3.16) that $\boldsymbol{D}$ is a closed vector subspace of $\mathcal{H}$ with projector

$$P_{\boldsymbol{D}} \colon \boldsymbol{x} \mapsto \left(\sum_{i=1}^{m} \omega_i x_i, \ldots, \sum_{i=1}^{m} \omega_i x_i\right), \tag{3.19}$$

and that the operator

$$\boldsymbol{j} \colon \mathcal{H} \to \boldsymbol{D} \colon x \mapsto (x,\ldots,x) \tag{3.20}$$

14

is an isomorphism. In addition, $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$ and we derive from (2.14), (3.16), and (3.18) that

$$\operatorname{prox}_{\boldsymbol{f}} \colon \boldsymbol{x} \mapsto \left(\operatorname{prox}_{f_1/\omega_1} x_1, \ldots, \operatorname{prox}_{f_m/\omega_m} x_m \right). \tag{3.21}$$

From the sequences $(x_n)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$, $((y_{i,n})_{n\in\mathbb{N}})_{1\leq i\leq m}$, $((p_{i,n})_{n\in\mathbb{N}})_{1\leq i\leq m}$, and $((a_{i,n})_{n\in\mathbb{N}})_{1\leq i\leq m}$ of Algorithm 3.1 we define, for every $n \in \mathbb{N}$,

$$\boldsymbol{x}_n = \boldsymbol{j}(x_n), \ \ \boldsymbol{p}_n = \boldsymbol{j}(p_n), \ \ \boldsymbol{y}_n = (y_{i,n})_{1\leq i\leq m}, \ \ \boldsymbol{y}_{n+1/2} = (p_{i,n})_{1\leq i\leq m}, \ \text{and} \ \boldsymbol{a}_n = (a_{i,n})_{1\leq i\leq m}. \tag{3.22}$$

It follows from (3.19), (3.20), and (3.21) that the sequences defined in (3.22) are precisely those involved in (3.8), and that the set of minimizers $\boldsymbol{G}$ in Proposition 3.3 is precisely

$$\boldsymbol{G} = \boldsymbol{j}(G). \tag{3.23}$$

On the other hand, it follows from (3.16), (3.17), and (3.18) that the properties (i)–(iv) above yield their respective counterparts in Proposition 3.3. Thus, we deduce from Proposition 3.3 and (3.23) that $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point $\boldsymbol{j}(x)$ for some $x \in G$. Thus, $(x_n)_{n\in\mathbb{N}} = (\boldsymbol{j}^{-1}(\boldsymbol{x}_n))_{n\in\mathbb{N}}$ converges weakly to $x \in G$. $\square$

Remark 3.5

(i) We have conveniently obtained Algorithm 3.1 as a direct transcription of a special case (see Proposition 3.3) of the Douglas-Rachford algorithm transposed in a product space. A similar decomposition method could be obtained by using the theory of partial inverses for monotone operators [45].

(ii) When $m = 2$, Algorithm 3.1 does not revert to the standard Douglas-Rachford iteration (1.7). Actually, even in this case, it seems better to use the former to the extent that, as seen in Theorem 3.4, it produces directly a sequence that converges weakly to a minimizer of $f_1 + f_2$.

To conclude this section, we describe some situations in which condition (ii) in Theorem 3.4 is satisfied.

Proposition 3.6 *Set* $C = \left\{ (x - x_1, \ldots, x - x_m) \mid x \in \mathcal{H}, \, x_1 \in \operatorname{dom} f_1, \ldots, x_m \in \operatorname{dom} f_m \right\}$ *and suppose that any of the following holds.*

(i) $\boldsymbol{C}$ *is a closed vector subspace.*

(ii) $\bigcap_{i=1}^{m} \operatorname{dom} f_i \neq \varnothing$ *and* $(\operatorname{dom} f_i)_{1\leq i\leq m}$ *are affine subspaces of finite dimensions.*

(iii) $\bigcap_{i=1}^{m} \operatorname{dom} f_i \neq \varnothing$ *and* $(\operatorname{dom} f_i)_{1\leq i\leq m}$ *are closed affine subspaces of finite codimensions.*

(iv) $\boldsymbol{0} \in \operatorname{int} \boldsymbol{C}$.

(v) $\operatorname{dom} f_1 \cap \bigcap_{i=2}^{m} \operatorname{int} \operatorname{dom} f_i \neq \varnothing$.

(vi) $\mathcal{H}$ *is finite-dimensional and* $\bigcap_{i=1}^{m} \operatorname{ri} \operatorname{dom} f_i \neq \varnothing$.

Then $\mathbf{0} \in \operatorname{sri} \boldsymbol{C}$.

Proof. We use the notation of the proof of Theorem 3.4, hence $\boldsymbol{C} = \boldsymbol{D} - \operatorname{dom} \boldsymbol{f}$.

(i): We have $\boldsymbol{C} = \overline{\operatorname{span}} \, \boldsymbol{C}$. Since $\boldsymbol{C} \subset \operatorname{cone} \boldsymbol{C} \subset \operatorname{span} \boldsymbol{C} \subset \overline{\operatorname{span}} \, \boldsymbol{C}$, we therefore obtain $\operatorname{cone} \boldsymbol{C} = \overline{\operatorname{span}} \, \boldsymbol{C}$. Appealing to (2.5), we conclude that $\mathbf{0} \in \operatorname{sri} \boldsymbol{C}$.

(ii)$\Rightarrow$(i): The assumption implies that $\operatorname{dom} \boldsymbol{f} = \operatorname{dom} f_1 \times \cdots \times \operatorname{dom} f_m$ is a finite-dimensional affine subspace of $\boldsymbol{\mathcal{H}}$ and that $\boldsymbol{D} \cap \operatorname{dom} \boldsymbol{f} \neq \varnothing$. Since $\boldsymbol{D}$ is closed vector subspace, it follows from [30, Lemma 9.36] that $\boldsymbol{D} - \operatorname{dom} \boldsymbol{f}$ is a closed vector subspace.

(iii)$\Rightarrow$(i): Here $\operatorname{dom} \boldsymbol{f} = \operatorname{dom} f_1 \times \cdots \times \operatorname{dom} f_m$ is a closed affine subspace of $\boldsymbol{\mathcal{H}}$ of finite codimension and that $\boldsymbol{D} \cap \operatorname{dom} \boldsymbol{f} \neq \varnothing$. Appealing to [30, Theorem 9.35 and Corollary 9.37], we conclude that $\boldsymbol{D} - \operatorname{dom} \boldsymbol{f}$ is a closed vector subspace.

(iv): See (2.7).

(v)$\Rightarrow$(iv): See the proof of [3, Theorem 6.3].

(vi): Using Lemma 2.1(i)&(ii), we obtain $\mathbf{0} \in \operatorname{sri} \boldsymbol{C} \Leftrightarrow \mathbf{0} \in \operatorname{sri}(\boldsymbol{D} - \operatorname{dom} \boldsymbol{f}) \Leftrightarrow \mathbf{0} \in \operatorname{ri}(\boldsymbol{D} - \operatorname{dom} \boldsymbol{f})$ $\Leftrightarrow \mathbf{0} \in \operatorname{ri} \boldsymbol{D} - \operatorname{ri} \operatorname{dom} \boldsymbol{f} = \boldsymbol{D} - \operatorname{ri} \operatorname{dom} \boldsymbol{f} \Leftrightarrow \boldsymbol{D} \cap \operatorname{ri} \operatorname{dom} \boldsymbol{f} \neq \varnothing \Leftrightarrow \bigcap_{i=1}^{m} \operatorname{ri} \operatorname{dom} f_i \neq \varnothing$. $\square$

4 Applications to signal and image processing

To illustrate the versatility of the proposed framework, we present three applications in signal and image processing. In each experiment, Algorithm 3.1 is implemented with $\omega_i \equiv 1/m$, $\lambda_n \equiv 1.5$, and, since the proximity operators required by the algorithm will be computable in closed form, we can dispense with errors and set $a_{i,n} \equiv 0$ in (3.1). As a result, conditions (iii) and (iv) in Theorem 3.4 are straightforwardly satisfied. In each experiment, the number of iterations of the algorithm is chosen large enough so that no significant improvement is gained by letting the algorithm run further.

4.1 Experiment 1

This first experiment is an image restoration problem in the standard Euclidean space $\mathcal{H} = \mathbb{R}^{N^2}$, where $N = 512$. The original vignetted $N \times N$ image $\overline{x}$ is shown in figure 1 (the vignetting is modeled by a black area in the image corners). The degraded image z shown in figure 2 is obtained via the degradation model

$$z = L\overline{x} + w, \tag{4.1}$$

where L is the two-dimensional convolution operator induced by a 15×15 uniform kernel, and where w is a realization of a zero-mean white Gaussian noise. The blurred image-to-noise ratio is $20 \log_{10}(\|L\overline{x}\|/\|w\|) = 31.75$ dB and the relative quadratic error with respect to the original image is $20 \log_{10}(\|z - \overline{x}\|/\|\overline{x}\|) = -19.98$ dB.

Figure 1: Experiment 1. Original image.

Figure 2: Experiment 1. Degraded image.

Figure 3: Experiment 1. Image restored with 300 iterations of Algorithm 3.1 ($\gamma = 1/4$).

The pixel values are known to fall in the interval $[0, 255]$. In addition, the vignetting area $\mathbb{S}$ of the original image is known. This information leads to the constraint set

$$C_1 = [0, 255]^{N^2} \cap \left\{ x \in \mathcal{H} \mid x \, 1_{\mathbb{S}} = \underline{0} \right\}, \tag{4.2}$$

where $x 1_{\mathbb{S}}$ denotes the coordinatewise multiplication of x with the characteristic vector $1_{\mathbb{S}}$ of $\mathbb{S}$ (its kth coordinate is 1 or 0 according as $k \in \mathbb{S}$ or $k \notin \mathbb{S}$), and where $\underline{0}$ the zero image. The mean value $\mu \in \left]0, 255\right[$ of $\overline{x}$ is also known, which corresponds to the constraint set

$$C_2 = \left\{ x \in \mathcal{H} \mid \langle x \mid \underline{1} \rangle = N^2 \mu \right\}, \tag{4.3}$$

where $\underline{1} = [1, \ldots, 1]^{\top} \in \mathbb{R}^{N^2}$. In addition, the phase of the discrete Fourier transform of the original image is measured over some frequency range $\mathbb{D} \subset \{0, \ldots, N^2 - 1\}$ [17, 37, 42]. If we denote by $\widehat{x} = \left(|\chi_k| \exp(i \angle \chi_k) \right)_{0 \le k \le N^2 - 1}$ the discrete Fourier transform of an image $x \in \mathcal{H}$ and by $(\phi_k)_{k \in \mathbb{D}}$ the known phase values, we obtain the constraint set

$$C_3 = \left\{ x \in \mathcal{H} \mid (\forall k \in \mathbb{D}) \ \angle \chi_k = \phi_k \right\}. \tag{4.4}$$

A constrained least-squares formulation of the problem is

$$\underset{x \in C_1 \cap C_2 \cap C_3}{\text{minimize}} \ \|Lx - z\|^2 \tag{4.5}$$

or, equivalently,

$$\underset{x \in C_1 \cap C_2}{\text{minimize}} \ \iota_{C_3}(x) + \|Lx - z\|^2. \tag{4.6}$$

18

However, in most instances, the phase cannot be measured exactly. This is simulated by introducing a 5 % perturbation on each of the phase components $(\phi_k)_{k \in \mathbb{D}}$. To take these uncertainties into account in (4.6), we replace the "hard" potential ι_{C_3} by a smoothed version, namely $\alpha d_{C_3}^p$, for some $\alpha \in]0, +\infty[$ and $p \in [1, +\infty[$. This leads to the variational problem

$$\underset{x \in C_1 \cap C_2}{\text{minimize}} \ \alpha d_{C_3}^p(x) + \|Lx - z\|^2, \tag{4.7}$$

which is a special case of (1.1), with $m = 4$, $f_1 = \iota_{C_1}$, $f_2 = \iota_{C_2}$, $f_3 = \alpha d_{C_3}^p$, and $f_4 = \|L \cdot -z\|^2$. Let us note that, since C_1 is bounded, condition (i) in Theorem 3.4 is satisfied. In addition, it follows from Proposition 3.6(vi) that condition (ii) in Theorem 3.4 also holds. Indeed, set $E =]0, 255[^{N^2} \cap A \cap C_2$, where $A = \{x \in \mathcal{H} \mid x1_{\mathbb{S}} = \underline{0}\}$. Then it follows from (4.3) that

$$\frac{N^2 \mu}{N^2 - \text{card} \, \mathbb{S}} \left(\underline{1} - 1_{\mathbb{S}} \right) \in E. \tag{4.8}$$

Hence, since A and C_2 are affine subspaces, (4.2) and Lemma 2.1(iii) yield

$$\bigcap_{i=1}^4 \text{ri} \, \text{dom} \, f_i = \text{ri} \, C_1 \cap \text{ri} \, C_2 = (\text{ri} \, C_1) \cap C_2 = (\text{ri} \, [0, 255]^{N^2}) \cap A \cap C_2 = E \neq \varnothing. \tag{4.9}$$

Problem (4.7) is solved for the following scenario: $\mathbb{D}$ corresponds to a low frequency band including about 80 % of the frequency components, $p = 3/2$, and $\alpha = 10$. The proximity operators required by Algorithm 3.1 are obtained as follows. First, prox_{f_1} and prox_{f_2} are respectively the projectors onto C_1 and C_2, which can be obtained explicitly [18]. Next, prox_{f_3} is given in Example 2.9. It involves P_{C_3}, which can be found in [18]. Finally, prox_{f_4} is supplied by Proposition 2.6. Note that, since L is a two-dimensional convolutional blur, it can be approximated by a block circulant matrix and hence (2.19) can be efficiently implemented in the frequency domain via the fast Fourier transform [1]. The restored image, shown in figure 3, is much sharper than the degraded image z and it achieves a relative quadratic error of -23.25 dB with respect to the original image $\overline{x}$.

4.2 Experiment 2

In image recovery, variational formulations involving total variation [15, 44, 47] or sparsity promoting potentials [5, 8, 14, 28] are popular. The objective of the present experiment is to show that it is possible to employ more sophisticated, hybrid potentials.

In order to simplify our presentation, we place ourselves in the Hilbert space $\mathcal{G}$ of periodic discrete images $y = (\eta_{k,l})_{(k,l) \in \mathbb{Z}^2}$ with horizontal and vertical periods equal to N ($N = 512$), endowed with the standard Euclidean norm

$$y \mapsto \sqrt{\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |\eta_{k,l}|^2}. \tag{4.10}$$

As usual, images of size $N \times N$ are viewed as elements of this space through periodization [1]. The original 8-bit satellite image $\overline{y} \in \mathcal{G}$ displayed in figure 4 is degraded through the linear model

$$z = L\overline{y} + w, \tag{4.11}$$

where L is the two-dimensional periodic convolution operator with a 7×7 uniform kernel, and w is a realization of a periodic zero-mean white Gaussian noise. The resulting degraded image $z \in \mathcal{G}$ is shown in figure 5. The blurred image-to-noise ratio is $20 \log_{10}(\|L\overline{y}\|/\|w\|) = 20.71$ dB and the relative quadratic error with respect to the original image is $20 \log_{10}(\|z - \overline{y}\|/\|\overline{y}\|) = -12.02$ dB.

In the spirit of a number of recent investigations (see [16] and the references therein), we use a tight frame representation of the images under consideration. This representation is defined through a synthesis operator F^*, which is a linear operator from $\mathcal{H} = \mathbb{R}^K$ to $\mathcal{G}$ (with $K \geq N^2$) such that

$$F^* \circ F = \kappa \, \mathrm{Id} \tag{4.12}$$

for some $\kappa \in]0, +\infty[$. Thus, the original image can be written as $\overline{y} = F^*\overline{x}$, where $\overline{x} \in \mathcal{H}$ is a vector of frame coefficients to be estimated. The rationale behind this approach is that, by appropriately choosing the frame, a sparse representation $\overline{x}$ of $\overline{y}$ can be achieved.

The restoration problem is posed in the frame coefficient space $\mathcal{H}$. We use the constraint set imposing the range of the pixel values of the original image $\overline{y}$, namely

$$C = \big\{x \in \mathcal{H} \mid F^*x \in D\big\}, \quad \text{where } D = \big\{y \in \mathcal{G} \mid (\forall(k,l) \in \{0,\ldots,N-1\}^2) \, \eta_{k,l} \in [0,255]\big\}, \tag{4.13}$$

as well as three potentials. The first potential is the standard least-squares data fidelity term $x \mapsto \|LF^*x - z\|^2$. The second potential is the ℓ^1 norm, which aims at promoting a sparse frame representation [16, 28, 48]. Finally, the third potential is the discrete total variation tv, which aims at preserving piecewise smooth areas and sharp edges [15, 44, 47]. Using the notation $(\eta_{k,l})^\top_{(k,l)\in\mathbb{Z}^2} = (\eta_{l,k})_{(k,l)\in\mathbb{Z}^2}$, the discrete total variation of $y \in \mathcal{G}$ is defined as

$$\mathrm{tv}(y) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \varrho_{k,l}\big(\nabla_1 y, (\nabla_1(y^\top))^\top\big), \tag{4.14}$$

where $\nabla_1 \colon \mathcal{G} \to \mathbb{R}^{N\times N}$ is a discrete vertical gradient operator and where, for every $\{k,l,q,r\} \subset \{0,\ldots,N-1\}$, we set

$$\varrho_{k,l} = \varrho_{k,l,k,l}, \tag{4.15}$$

with

$$\varrho_{k,l,q,r} \colon \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N} \to \mathbb{R} \colon \Big([\nu_{a,b}]_{0\leq a,b\leq N-1}, [\tilde{\nu}_{a,b}]_{0\leq a,b\leq N-1}\Big) \mapsto \sqrt{|\nu_{k,l}|^2 + |\tilde{\nu}_{q,r}|^2}. \tag{4.16}$$

A common choice for the gradient operator is $\nabla_1 \colon y \mapsto [\eta_{k+1,l} - \eta_{k,l}]_{0\leq k,l\leq N-1}$. As is customary in image processing [35, Section 9.4], we adopt here a horizontally smoothed version of this operator, namely,

$$\nabla_1 \colon \mathcal{G} \to \mathbb{R}^{N\times N} \colon y \mapsto \frac{1}{2}\big[\eta_{k+1,l+1} - \eta_{k,l+1} + \eta_{k+1,l} - \eta_{k,l}\big]_{0\leq k,l\leq N-1}. \tag{4.17}$$

Figure 4: Experiment 2. Original image.

We thus arrive at a variational formulation of the form (1.1), namely

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \iota_C(x) + \|LF^*x - z\|^2 + \alpha\|x\|_{\ell^1} + \beta\mathrm{tv}(F^*x), \qquad (4.18)$$

where α and β are in $]0, +\infty[$. Since C is bounded, condition (i) in Theorem 3.4 is satisfied. In addition, it is clear from Proposition 3.6(vi) that condition (ii) in Theorem 3.4 also holds. Indeed, all the potentials in (4.18) have full domain, except ι_C. However, Lemma 2.1(i) implies that $\mathrm{ri}\,\mathrm{dom}\,\iota_C = \mathrm{ri}\,C \neq \varnothing$ since $\underline{0} \in C$.

Although (4.18) assumes the form of (1.1), it is not directly exploitable by Algorithm 3.1 because the proximity operator of $\mathrm{tv}\circ F^*$ cannot be computed explicitly. To circumvent this numerical hurdle, the total variation potential (4.14) is split in four terms and (4.18) is rewritten as

$$\underset{x\in C}{\text{minimize}} \;\; \|LF^*x - z\|^2 + \alpha\|x\|_{\ell^1} + \beta\sum_{i=0}^{3} \mathrm{tv}_i(F^*x), \qquad (4.19)$$

where

$$(\forall(q,r)\in\{0,1\}^2) \quad \mathrm{tv}_{q+2r}\colon \mathcal{G}\to\mathbb{R}\colon y\mapsto \sum_{k=0}^{N/2-1}\sum_{l=0}^{N/2-1} \varrho_{2k+q,2l+r}\big(\nabla_1 y, (\nabla_1(y^\top))^\top\big). \qquad (4.20)$$

For every q and r in $\{0,1\}$, let $\downarrow_{q,r}$ be the decimation operator given by

$$\downarrow_{q,r}\colon \mathbb{R}^{2N\times 2N}\to\mathbb{R}^{N\times N}\colon v = [\nu_{k,l}]_{0\leq k,l\leq 2N-1} \mapsto [\nu_{2k+q,2l+r}]_{0\leq k,l\leq N-1}, \qquad (4.21)$$
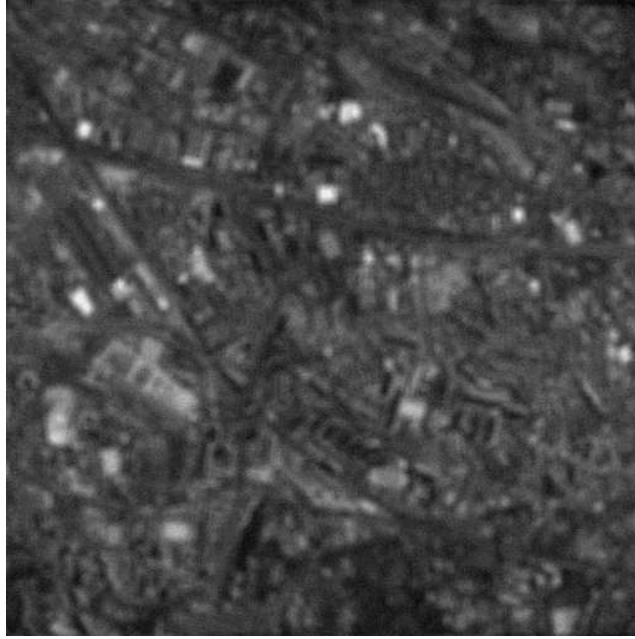
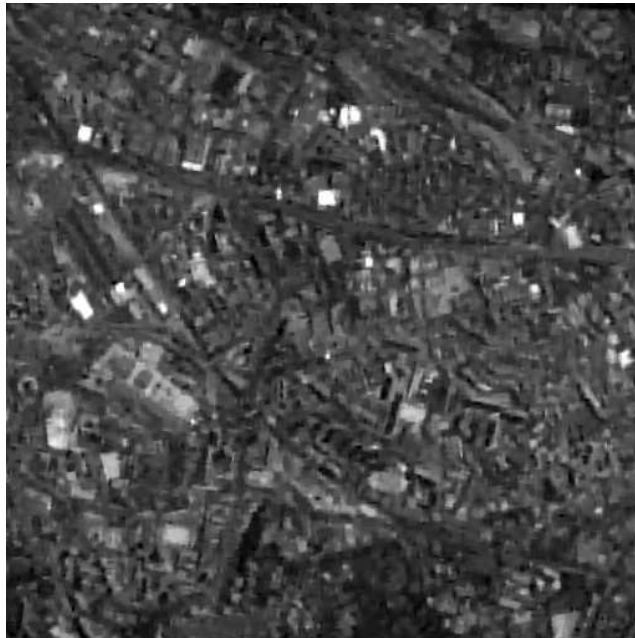Figure 5: Experiment 2. Degraded image.

Figure 6: Experiment 2. Image restored by (4.27), using 350 iterations of Algorithm 3.1 with $\gamma = 150$.
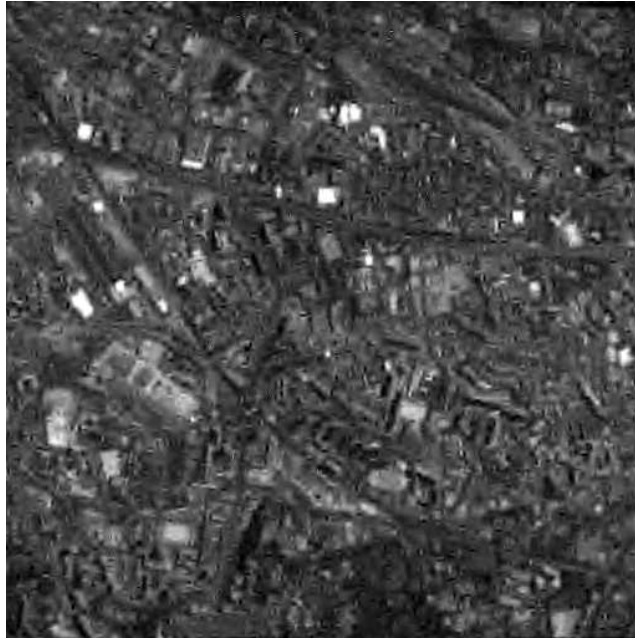
Figure 7: Experiment 2. Image restored without the total variation potential in (4.27), using 350 iterations of Algorithm 3.1 with $\gamma = 150$.
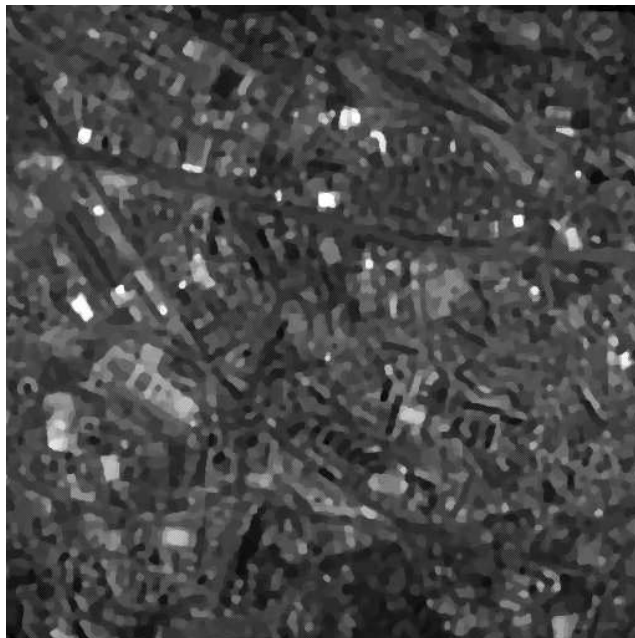

Figure 8: Experiment 2. Image restored without the ℓ^1 potential in (4.27), using 350 iterations of Algorithm 3.1 with $\gamma = 150$.

and set

$$U_{q+2r} \colon \mathcal{G} \to \mathbb{R}^{N \times N} \colon y \mapsto \downarrow_{q,r} \begin{bmatrix} \nabla_0 y & \nabla_1 y \\ (\nabla_1(y^\top))^\top & \nabla_2 y \end{bmatrix}, \tag{4.22}$$

where ∇_1 is defined in (4.17),

$$\nabla_0 \colon \mathcal{G} \to \mathbb{R}^{N \times N} \colon y \mapsto \frac{1}{2} \big[\eta_{k+1,l+1} + \eta_{k,l+1} + \eta_{k+1,l} + \eta_{k,l} \big]_{0 \le k,l \le N-1}, \tag{4.23}$$

and

$$\nabla_2 \colon \mathcal{G} \to \mathbb{R}^{N \times N} \colon y \mapsto \frac{1}{2} \big[\eta_{k+1,l+1} - \eta_{k,l+1} - \eta_{k+1,l} + \eta_{k,l} \big]_{0 \le k,l \le N-1}. \tag{4.24}$$

Moreover, set

$$h \colon \mathbb{R}^{N \times N} \to \mathbb{R} \colon v \mapsto \sum_{k=0}^{N/2-1} \sum_{l=0}^{N/2-1} \varrho_{k,l+N/2,k+N/2,l}(v,v). \tag{4.25}$$

Then it follows from (4.20) and (4.22) that

$$(\forall i \in \{0,1,2,3\}) \quad \mathrm{tv}_i = h \circ U_i. \tag{4.26}$$

Hence, (4.19) becomes

$$\underset{x \in C}{\text{minimize}} \ \|LF^*x - z\|^2 + \alpha \|x\|_{\ell^1} + \beta \sum_{i=0}^{3} h(U_i F^* x). \tag{4.27}$$

Problem (4.27) is a specialization of (1.1), in which $m = 7$, $f_1 = \iota_C$, $f_2 = \|LF^* \cdot - z\|^2$, $f_3 = \alpha \| \cdot \|_{\ell^1}$, and $f_{i+4} = \beta\, h \circ U_i \circ F^*$ for $i \in \{0,1,2,3\}$. To implement Algorithm 3.1, we need the expressions of the proximity operators of these functions. The proximity operator of f_1 can be calculated by first observing that the projection onto the set D of (4.13) is explicit, and by then applying Lemma 2.4, which states that (4.12) and (4.13) imply that

$$\mathrm{prox}_{f_1} = \mathrm{prox}_{\iota_D \circ F^*} = \mathrm{Id} + \frac{1}{\kappa} F \circ (\mathrm{prox}_{\iota_D} - \mathrm{Id}) \circ F^* = \mathrm{Id} + \frac{1}{\kappa} F \circ (P_D - \mathrm{Id}) \circ F^*. \tag{4.28}$$

On the other hand, the proximity operator of f_2 can be derived from Proposition 2.6 using a frequency domain implementation (as in section 4.1), and by again invoking Lemma 2.4. Next, the proximity operator of f_3 can be found in [26, Example 2.20]. Finally, the operators $(\mathrm{prox}_{f_i})_{4 \le i \le 7}$ are provided by the following fact.

Proposition 4.1 *Set* $\Pi \colon \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N} \colon v = [\nu_{k,l}]_{0 \le k,l \le N-1} \mapsto [\pi_{k,l}]_{0 \le k,l \le N-1}$, *where*

$$\big(\forall (k,l) \in \{0,\ldots,N/2-1\}^2\big) \quad \begin{cases} \pi_{k,l} = \nu_{k,l} \\ \pi_{k+N/2,l+N/2} = \nu_{k+N/2,l+N/2} \\ \pi_{k,l+N/2} = \sigma_{k,l}(v)\, \nu_{k,l+N/2} \\ \pi_{k+N/2,l} = \sigma_{k,l}(v)\, \nu_{k+N/2,l} \end{cases} \quad \text{with}$$

$$\sigma_{k,l} \colon v \mapsto \begin{cases} 1 - \dfrac{\kappa\beta}{\sqrt{|\nu_{k,l+N/2}|^2 + |\nu_{k+N/2,l}|^2}}, & \text{if } \sqrt{|\nu_{k,l+N/2}|^2 + |\nu_{k+N/2,l}|^2} \ge \kappa\beta; \\ 0, & \text{otherwise.} \end{cases} \tag{4.29}$$

Then, for every $i \in \{0, 1, 2, 3\}$,

$$\text{prox}_{f_{i+4}} = \text{Id} + \frac{1}{\kappa} F \circ (U_i^* \circ \Pi \circ U_i - \text{Id}) \circ F^*. \tag{4.30}$$

Proof. Set $\varphi \colon \mathbb{R}^2 \to \mathbb{R} \colon (\xi_1, \xi_2) \mapsto \kappa\beta\sqrt{|\xi_1|^2 + |\xi_2|^2}$. By applying Proposition 2.8(i) in $\mathbb{R}^2$ with the set $\{(0, 0)\}$, we obtain

$$(\forall (\xi_1, \xi_2) \in \mathbb{R}^2) \quad \text{prox}_\varphi(\xi_1, \xi_2) = \begin{cases} \left(1 - \dfrac{\kappa\beta}{\sqrt{|\xi_1|^2 + |\xi_2|^2}}\right)(\xi_1, \xi_2), & \text{if } \sqrt{|\xi_1|^2 + |\xi_2|^2} \geq \kappa\beta; \\ 0, & \text{otherwise.} \end{cases} \tag{4.31}$$

Now set $p = [\pi_{k,l}]_{0 \leq k, l \leq N-1} = \text{prox}_{\kappa\beta h} v$. In view of (2.14), (4.25), and (4.16), p minimizes over $\tilde{p} \in \mathbb{R}^{N \times N}$ the cost

$$\begin{aligned}
\kappa\beta h(\tilde{p}) + \frac{1}{2}\|v - \tilde{p}\|^2 &= \kappa\beta \sum_{k=0}^{N/2-1} \sum_{l=0}^{N/2-1} \varrho_{k,l+N/2,k+N/2,l}(\tilde{p}, \tilde{p}) + \frac{1}{2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} |\nu_{k,l} - \tilde{\pi}_{k,l}|^2 \\
&= \sum_{k=0}^{N/2-1} \sum_{l=0}^{N/2-1} \left(\kappa\beta\sqrt{|\tilde{\pi}_{k,l+N/2}|^2 + |\tilde{\pi}_{k+N/2,l}|^2} \right. \\
&\qquad\qquad\qquad \left. + \frac{1}{2}\left(|\nu_{k,l+N/2} - \tilde{\pi}_{k,l+N/2}|^2 + |\nu_{k+N/2,l} - \tilde{\pi}_{k+N/2,l}|^2\right) \right) \\
&\quad + \frac{1}{2}\sum_{k=0}^{N/2-1}\sum_{l=0}^{N/2-1} \left(|\nu_{k,l} - \tilde{\pi}_{k,l}|^2 + |\nu_{k+N/2,l+N/2} - \tilde{\pi}_{k+N/2,l+N/2}|^2\right). \tag{4.32}
\end{aligned}$$

Therefore,

$$(\forall (k, l) \in \{0, \ldots, N/2 - 1\}^2) \quad \begin{cases} (\pi_{k,l+N/2}, \pi_{k+N/2,l}) = \text{prox}_\varphi(\nu_{k,l+N/2}, \nu_{k+N/2,l}), \\ \pi_{k,l} = \nu_{k,l}, \\ \pi_{k+N/2,l+N/2} = \nu_{k+N/2,l+N/2}. \end{cases} \tag{4.33}$$

Appealing to (4.29) and (4.31), we obtain $\Pi = \text{prox}_{\kappa\beta h}$. Now, let $i \in \{0, 1, 2, 3\}$. It follows from (4.22) that U_i is a separable two-dimensional Haar-like orthogonal operator [35, Section 5.9]. Hence, appealing to (4.12), we obtain $(U_i \circ F^*) \circ (U_i \circ F^*)^* = \kappa\,\text{Id}$. In turn, Lemma 2.4 yields

$$\begin{aligned}
\text{prox}_{f_{i+4}} &= \text{prox}_{\beta h \circ (U_i \circ F^*)} \\
&= \text{Id} + \frac{1}{\kappa}(U_i \circ F^*)^* \circ (\text{prox}_{\kappa\beta h} - \text{Id}) \circ (U_i \circ F^*) \\
&= \text{Id} + \frac{1}{\kappa} F \circ (U_i^* \circ \Pi \circ U_i - \text{Id}) \circ F^*, \tag{4.34}
\end{aligned}$$

which completes the proof. □

In (4.27), we employ a tight frame ($\kappa = 4$) resulting from the concatenation of four shifted separable dyadic orthonormal wavelet decompositions [41] carried out over 4 resolution levels. The

shift parameters are $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$. In addition, symlet filters [27] of length 8 are used. The parameters α and β have been adjusted so as to minimize the error with respect to the original image $\overline{y}$. The restored image we obtain is displayed in figure 6. It achieves a relative mean-square error with respect to $\overline{y}$ of -14.82 dB. For comparison, the result obtained without the total variation potential in (4.27) is shown in figure 7 (error of -14.06 dB), and the result obtained without the ℓ^1 potential in (4.27) is shown in figure 8 (error of -13.70 dB). It can be observed that the image of figure 7 suffers from small visual artifacts, whereas the details in figure 8 are not sharp. This shows the advantage of combining an ℓ^1 potential and a total variation potential.

4.3   Experiment 3

We revisit via the variational formulation (1.1) a pulse shape design problem investigated in [23] in a more restrictive setting (see also [40] for the original two-constraint formulation). This problem illustrates further ramifications of the proposed algorithm.

The problem is to design a pulse shape for digital communications. The signal space is the standard Euclidean space $\mathcal{H} = \mathbb{R}^N$, where $N = 1024$ is the number of samples of the discrete pulse (the underlying sampling rate is 2560 Hz). Five constraints arise from engineering specifications. We denote by $x = (\xi_k)_{0 \leq k \leq N-1}$ a signal in $\mathcal{H}$ and by $\widehat{x} = (\chi_k)_{0 \leq k \leq N-1}$ its discrete Fourier transform.

- The Fourier transform of the pulse should vanish at the zero frequency and at integer multiples of 50 Hz. This constraint is associated with the set

$$C_1 = \left\{ x \in \mathcal{H} \mid \widehat{x} 1_{\mathbb{D}_1} = \underline{0} \right\}, \qquad (4.35)$$

  where $\mathbb{D}_1$ is the set of discrete frequencies at which $\widehat{x}$ should vanish.

- The modulus of the Fourier transform of the pulse should no exceed a prescribed bound $\rho > 0$ beyond 300 Hz. This constraint is associated with the set

$$C_2 = \left\{ x \in \mathcal{H} \mid (\forall k \in \mathbb{D}_2) \ |\chi_k| \leq \rho \right\}, \qquad (4.36)$$

  where $\mathbb{D}_2$ represents frequencies beyond 300 Hz.

- The energy of the pulse should not exceed a prescribed bound $\mu^2 > 0$ in order not to interfere with other systems. The associated set is

$$C_3 = \left\{ x \in \mathcal{H} \mid \|x\| \leq \mu \right\}. \qquad (4.37)$$

- The pulse should be symmetric about its mid-point, where its value should be equal to 1. This corresponds to the set

$$C_4 = \left\{ x \in \mathcal{H} \mid \xi_{N/2} = 1 \text{ and } (\forall k \in \{0, \ldots, N/2\}) \ \xi_k = \xi_{N-1-k} \right\}. \qquad (4.38)$$

- The duration of the pulse should be 50 ms and it should have periodic zero crossings every 3.125 ms. This leads to the set

$$C_5 = \left\{ x \in \mathcal{H} \mid x \, 1_{\mathbb{S}} = \underline{0} \right\}, \tag{4.39}$$

where $\mathbb{S}$ is the set of time indices in the zero areas.

In this problem, C_1, C_2, and C_3 are hard constraints that must be satisfied, whereas the other constraints are soft ones that are incorporated via powers of distance potentials. This leads to the variational formulation

$$\underset{x \in C_1 \cap C_2 \cap C_3}{\text{minimize}} \quad d_{C_4}^{p_4}(x) + d_{C_5}^{p_5}(x), \tag{4.40}$$

where p_4 and p_5 are in $[1, +\infty[$. The design problem is thus cast in the general form of (1.1), with $m = 5$, $f_i = \iota_{C_i}$ for $i \in \{1, 2, 3\}$, and $f_i = d_{C_i}^{p_i}$ for $i \in \{4, 5\}$. Since C_3 is bounded, condition (i) in Theorem 3.4 holds. In addition, it follows from Proposition 3.6(vi) that condition (ii) in Theorem 3.4 is satisfied. Indeed,

$$0 \in C_1 \cap \left\{ x \in \mathcal{H} \mid (\forall k \in \mathbb{D}_2) \, |\chi_k| < \rho \right\} \cap \left\{ x \in \mathcal{H} \mid \|x\| < \mu \right\} = \bigcap_{i=1}^{5} \operatorname{ri} \operatorname{dom} f_i. \tag{4.41}$$

Let us emphasize that our approach is applicable to any value of $(p_4, p_5) \in [1, +\infty[^2$. The proximity operators of f_4 and f_5 are supplied by Proposition 2.8, whereas the other proximity operators are the projectors onto $(C_i)_{1 \leq i \leq 3}$, which are straightforward [23]. A solution to (4.40) when $p_4 = p_5 = 2$, $\rho = 10^{-3/2}$, and $\mu = 2$ is shown in figure 9 and its Fourier transform is shown in figure 10. As is apparent in figure 9, the constraints corresponding to C_4 and C_5 are not satisfied. Forcing $C_4 \cap C_5$ as a hard constraint would therefore result in an infeasible problem. Finally, figure 10 shows that C_2 induces a 30 dB attenuation in the stop-band (beyond 300 Hz), in agreement with the value chosen for ρ.

5 Concluding remarks

We have proposed a proximal method for solving inverse problems that can be decomposed into the minimization of a sum of lower semicontinuous convex potentials. The algorithms currently in use in inverse problems are restricted to at most two nonsmooth potentials, which excludes many important scenarios and offers limited flexibility in terms of numerical implementation. By contrast, the algorithm proposed in the paper can handle an arbitrary number of nonsmooth potentials. It involves each potential by means of its own proximity operator, and activates these operators in parallel at each iteration. The versatility of the method is demonstrated through applications in signal and image recovery that illustrate various decomposition schemes, including one in which total variation is mixed up with other nonsmooth potentials.

References

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.
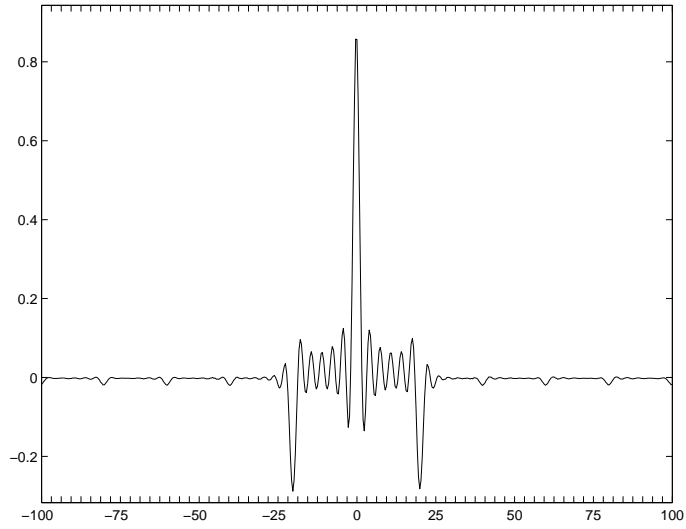
Figure 9: Experiment 3. Pulse (amplitude versus time in ms) synthesized using 100 iterations of Algorithm 3.1 with $\gamma = 1/5$.
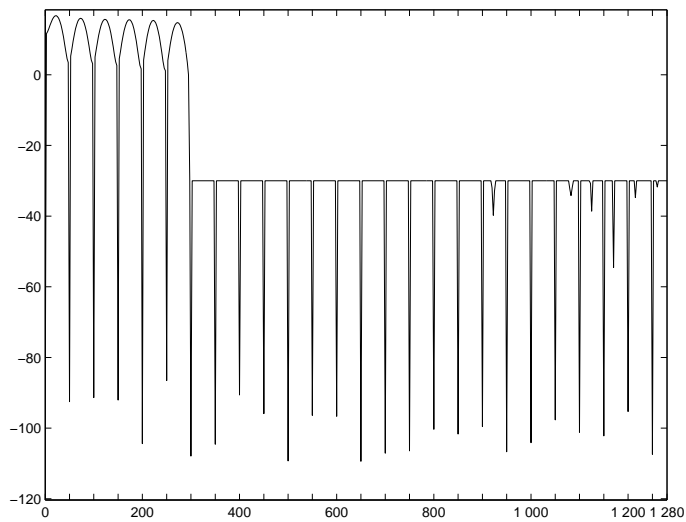

Figure 10: Experiment 3. Fourier transform (amplitude in dB versus frequency in Hz) of the pulse of figure 9.

[2] E. Artzy, T. Elfving, and G. T. Herman, Quadratic optimization for image reconstruction II, *Comput. Graph. Image Processing,* vol. 11, pp. 242–261, 1979.

[3] H. H. Bauschke and J. M. Borwein, On the convergence of von Neumann's alternating projection algorithm for two sets, *Set-Valued Anal.,* vol. 1, pp. 185–212, 1993.

[4] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Rev.,* vol. 38, pp. 367–426, 1996.

[5] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, A ℓ^1 unified variational framework for image restoration, in *Proc. Eighth Europ. Conf. Comput. Vision,* Prague, 2004, T. Pajdla and J. Matas, eds., Lecture Notes in Comput. Sci. 3024, Springer-Verlag, New York, 2004, pp. 1–13.

[6] J. M. Bioucas-Dias and M. A. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Trans. Image Process.,* vol. 16, pp. 2992–3004, 2007.

[7] J. P. Boyle and R. L. Dykstra, A method for finding projections onto the intersection of convex sets in Hilbert spaces, *Lecture Notes Statist.,* vol. 37, pp. 28–47, 1986.

[8] K. Bredies and D. A. Lorenz, Iterated hard shrinkage for minimization problems with sparsity constraints, *SIAM J. Sci. Comput.,* vol. 30, pp. 657–683, 2008.

[9] K. Bredies and D. A. Lorenz, Linear convergence of iterative soft-thresholding, *J. Fourier Anal. Appl.,* to appear.

[10] C. L. Byrne, *Signal Processing – A Mathematical Approach,* A. K. Peters, Wellesley, MA, 2005.

[11] C. L. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Ann. Oper. Res.*, vol. 105, pp. 77–98, 2001.

[12] Y. Censor, Iterative methods for the convex feasibility problem, *Annals Discrete Math.,* vol. 20, pp. 83–91, 1984.

[13] Y. Censor, T. Bortfeld, B. Martin, and A. Tromov, A unified approach for inversion problems in intensity-modulated radiation therapy, *Phys. Med. Biol.,* vol. 51, pp. 2353–2365, 2006.

[14] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process.,* vol. 7, pp. 319–335, 1998.

[15] A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and related problems, *Numer. Math.,* vol. 76, pp. 167–188, 1997.

[16] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, A variational formulation for frame-based inverse problems, *Inverse Problems,* vol. 23, pp. 1495–1518, 2007.

[17] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.,* vol. 42, pp. 2955–2966, 1994.

[18] P. L. Combettes, The convex feasibility problem in image recovery, in *Advances in Imaging and Electron Physics* (P. Hawkes, Ed.), vol. 95, pp. 155–270, Academic Press, New York, 1996.

[19] P. L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Trans. Image Process.,* vol. 6, pp. 493–506, 1997.

[20] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.,* vol. 38, pp. 538–565, 2000.

[21] P. L. Combettes, A block-iterative surrogate constraint splitting method for quadratic signal recovery, *IEEE Trans. Signal Process.,* vol. 51, pp. 1771–1782, 2003.

[22] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization,* vol. 53, pp. 475–504, 2004.

[23] P. L. Combettes and P. Bondon, Hard-constrained inconsistent signal feasibility problems, *IEEE Trans. Signal Process.,* vol. 47, pp. 2460–2468, 1999.

[24] P. L. Combettes and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM J. Optim.,* vol. 18, pp. 1351–1376, 2007.

[25] 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.

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

[27] I. Daubechies, *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992.

[28] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.,* vol. 57, pp. 1413–1457, 2004.

[29] I. Daubechies, G. Teschke, and L. Vese, Iteratively solving linear inverse problems under general convex constraints, *Inverse Problems and Imaging,* vol. 1, pp. 29–46, 2007.

[30] F. Deutsch, *Best Approximation in Inner Product Spaces.* Springer-Verlag, New York, 2001.

[31] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels,* Dunod, Paris, 1974; *Convex Analysis and Variational Problems,* SIAM, Philadelphia, PA, 1999.

[32] M. Fornasier, Domain decomposition methods for linear inverse problems with sparsity constraints, *Inverse Problems,* vol. 23, pp. 2505–2526, 2007.

[33] N. Gaffke and R. Mathar, A cyclic projection algorithm via duality, *Metrika,* vol. 36, pp. 29–54, 1989.

[34] S. A. Hirstoaga, Iterative selection methods for common fixed point problems, *J. Math. Anal. Appl.,* vol. 324, pp. 1020–1035, 2006.

[35] A. K. Jain, *Fundamentals of Digital Image Processing.* Prentice-Hall, London, 1989.

[36] K. C. Kiwiel and B. Łopuch, Surrogate projection methods for finding fixed points of firmly inonexpansive mappings, *SIAM J. Optim.,* vol. 7, pp. 1084–1102, 1997.

[37] A. Levi and H. Stark, Signal reconstruction from phase by projection onto convex sets, *J. Opt. Soc. Amer.,* vol. 73, pp. 810–822, 1983.

[38] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.,* vol. 16, pp. 964–979, 1979.

[39] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France,* vol. 93, pp. 273-299, 1965.

[40] R. A. Nobakht and M. R. Civanlar, Optimal pulse shape design for digital communication systems by projections onto convex sets, *IEEE Trans. Communications,* vol. 43, pp. 2874–2877, 1995.

[41] J.-C. Pesquet, H. Krim, and H. Carfantan, Time-invariant orthonormal wavelet representations, *IEEE Trans. Signal Process.*, vol. 44, pp. 1964–1970, 1996.

[42] M. Porat and G. Shachor, Signal representation in the combined phase-spatial space: Reconstruction and criteria for uniqueness, *IEEE Trans. Signal Process.,* vol. 47, pp. 1701–1707, 1999.

[43] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[44] L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D,* vol. 60, pp. 259–268, 1992.

[45] J. E. Spingarn, Applications of the method of partial inverses to convex programming: Decomposition, *Math. Programming.*, vol. 32, pp. 199–223, 1985.

[46] H. Stark (Ed.), *Image Recovery: Theory and Application.* Academic Press, San Diego, CA, 1987.

[47] D. Strong and T. Chan, Edge-preserving and scale-dependent properties of total variation regularization, *Inverse Problems,* vol. 19, pp. S165–S187, 2003.

[48] J. A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory,* vol. 52, pp. 1030–1051, 2006.

[49] I. Yamada, N. Ogura, Y. Yamashita, and K. Sakaniwa, Quadratic optimization of fixed points of nonexpansive mappings in Hilbert space, *Numer. Funct. Anal. Optim.,* vol. 19, pp. 165–190, 1998.

[50] D. C. Youla and H. Webb, Image restoration by the method of convex projections: Part 1 – theory, *IEEE Trans. Medical Imaging,* vol. 1, pp. 81–94, 1982.

[51] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific, River Edge, NJ, 2002.