Dualization of Signal Recovery Problems*

Patrick L. Combettes,[1] Đinh Dũng,[2] and Bằng Công Vũ[3]

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

[2]Vietnam National University
Information Technology Institute
Hanoi, Vietnam
dinhdung@vnu.edu.vn

[3]Vietnam National University
Department of Mathematics – Mechanics – Informatics
Hanoi, Vietnam
bangvc@vnu.edu.vn

Abstract

In convex optimization, duality theory can sometimes lead to simpler solution methods than those resulting from direct primal analysis. In this paper, this principle is applied to a class of composite variational problems arising in particular in signal recovery. These problems are not easily amenable to solution by current methods but they feature Fenchel-Moreau-Rockafellar dual problems that can be solved by forward-backward splitting. The proposed algorithm produces simultaneously a sequence converging weakly to a dual solution, and a sequence converging strongly to the primal solution. Our framework is shown to capture and extend several existing duality-based signal recovery methods and to be applicable to a variety of new problems beyond their scope.

Keywords Convex optimization, Denoising, Dictionary, Dykstra-like algorithm, Duality, Forward-backward splitting, Image reconstruction, Image restoration, Inverse problem, Signal recovery, Primal-dual algorithm, Proximity operator, Total variation

Mathematics Subject Classifications (2010) 90C25, 49N15, 94A12, 94A08

*Contact author: P. L. Combettes, plc@math.jussieu.fr, phone: +33 1 4427 6319, fax: +33 1 4427 7200. The work of P. L. Combettes was supported the Agence Nationale de la Recherche under grant ANR-08-BLAN-0294-02. The work of Đinh Dũng and Bằng Công Vũ was supported by the Vietnam National Foundation for Science and Technology Development.

1 Introduction

Over the years, several structured frameworks have been proposed to unify the analysis and the numerical solution methods of classes of signal (including image) recovery problems. An early contribution was made by Youla in 1978 [76]. He showed that several signal recovery problems, including those of [46, 62], shared a simple common geometrical structure and could be reduced to the following formulation in a Hilbert space $\mathcal{H}$ with scalar product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$: find the signal in a closed vector subspace C which admits a known projection r onto a closed vector subspace V, and which is at minimum distance from some reference signal z. This amounts to solving the variational problem

$$\underset{\substack{x \in C \\ P_V x = r}}{\text{minimize}} \ \frac{1}{2}\|x - z\|^2, \tag{1.1}$$

where P_V denotes the projector onto V. Abstract Hilbert space signal recovery problems have also been investigated by other authors. For instance, in 1965, Levi [52] considered the problem of finding the minimum energy band-limited signal fitting N linear measurements. In the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$, the underlying variational problem is to

$$\underset{\substack{x \in C \\ \langle x \mid s_1 \rangle = \rho_1 \\ \vdots \\ \langle x \mid s_N \rangle = \rho_N}}{\text{minimize}} \ \frac{1}{2}\|x\|^2, \tag{1.2}$$

where C is the subspace of band-limited signals, $(s_i)_{1 \leq i \leq N} \in \mathcal{H}^N$ are the measurement signals, and $(\rho_i)_{1 \leq i \leq N} \in \mathbb{R}^N$ are the measurements. In [64], Potter and Arun observed that, for a general closed convex set C, the formulation (1.2) models a variety of problems, ranging from spectral estimation [10, 70] and tomography [54], to other inverse problems [11]. In addition, they employed an elegant duality framework to solve it, which led to the following result.

Proposition 1.1 [64, Theorems 1 and 3] *Set $r = (\rho_i)_{1 \leq i \leq N}$ and $L \colon \mathcal{H} \to \mathbb{R}^N \colon x \mapsto (\langle x \mid s_i \rangle)_{1 \leq i \leq N}$, and let $\gamma \in \left]0, 2\right[$. Suppose that $\sum_{i=1}^N \|s_i\|^2 \leq 1$ and that r lies in the relative interior of $L(C)$. Set*

$$w_0 \in \mathbb{R}^N \quad \textit{and} \quad (\forall n \in \mathbb{N}) \quad w_{n+1} = w_n + \gamma\big(r - LP_C L^* w_n\big), \tag{1.3}$$

where $L^* \colon \mathbb{R}^N \to \mathcal{H} \colon (\nu_i)_{1 \leq i \leq N} \mapsto \sum_{i=1}^N \nu_i s_i$ is the adjoint of L. Then $(w_n)_{n \in \mathbb{N}}$ converges to a point w such that $LP_C L^* w = r$ and $P_C L^* w$ is the solution to (1.2).

Duality theory plays a central role in convex optimization [42, 58, 67, 79] and it has been used, in various forms and with different objectives, in several places in signal recovery, e.g., [10, 13, 22, 24, 35, 39, 43, 47, 49, 51, 75]; let us add that, since the completion of the present paper [30], other aspects of duality in imaging have been investigated in [14]. For our purposes, the most suitable type of duality is the so-called Fenchel-Moreau-Rockafellar duality, which associates to a composite minimization problem a "dual" minimization problem involving the conjugates of the functions and the adjoint of the linear operator acting in the primal problem. In general, the dual problem sheds a new light on the properties of the primal problem and enriches its analysis. Moreover, in certain specific situations, it is actually possible to solve the dual problem and to recover a solution to the primal problem from any dual solution. Such a scenario underlies Proposition 1.1: the primal problem (1.2) is difficult to solve but, if C is simple enough, the dual problem can be solved efficiently and, furthermore, a primal solution can be recovered explicitly. This principle is also explicitly or implicitly present in other signal recovery problems. For instance, the variational denoising problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; g(Lx)+\frac{1}{2}\|x-z\|^2, \tag{1.4}$$

where z is a noisy observation of an ideal signal, L is a bounded linear operator from $\mathcal{H}$ to some Hilbert space $\mathcal{G}$, and $g\colon \mathcal{G}\to\,]-\infty,+\infty]$ is a proper lower semicontinuous convex function, can often be approached efficiently using duality arguments [35]. A popular development in this direction is the total variation denoising algorithm proposed in [22] and refined in [23].

The objective of the present paper is to devise a duality framework that captures problems such as (1.1), (1.2), and (1.4) and leads to improved algorithms and convergence results, in an effort to standardize the use of duality techniques in signal recovery and extend their range of potential applications. More specifically, we focus on a class of convex variational problems which satisfy the following.

(a) They cover the above minimization problems.

(b) They are not easy to solve directly, but they admit a Fenchel-Moreau-Rockafellar dual which can be solved reliably in the sense that an implementable algorithm is available with proven weak or strong convergence to a solution of the sequences of iterates it generates. Here "implementable" is taken in the classical sense of [63]: the algorithm does not involve subprograms (e.g., "oracles" or "black-boxes") which are not guaranteed to converge in a finite number of steps.

(c) They allow for the construction of a primal solution from any dual solution.

A problem formulation which complies with these requirements is the following, where we denote by $\operatorname{sri} C$ the strong relative interior of a convex set C (see (2.5) and Remark 2.1).

Problem 1.2 (primal problem) Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $z\in\mathcal{H}$, let $r\in\mathcal{G}$, let $f\colon\mathcal{H}\to\,]-\infty,+\infty]$ and $g\colon\mathcal{G}\to\,]-\infty,+\infty]$ be lower semicontinuous convex functions, and let $L\colon\mathcal{H}\to\mathcal{G}$ be a nonzero linear bounded operator such that the qualification condition

$$r\in\operatorname{sri}\big(L(\operatorname{dom} f)-\operatorname{dom} g\big) \tag{1.5}$$

holds. The problem is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+g(Lx-r)+\frac{1}{2}\|x-z\|^2. \tag{1.6}$$

In connection with (a), it is clear that (1.6) covers (1.4) for $f=0$. Moreover, if we let f and g be the indicator functions (see (2.1)) of closed convex sets $C\subset\mathcal{H}$ and $D\subset\mathcal{G}$, respectively, then (1.6) reduces to the best approximation problem

$$\underset{\substack{x\in C\\ Lx-r\in D}}{\text{minimize}}\;\; \frac{1}{2}\|x-z\|^2, \tag{1.7}$$

which captures both (1.1) and (1.2) in the case when C is a closed vector subspace and $D = \{0\}$. Indeed, (1.1) corresponds to $\mathcal{G} = \mathcal{H}$ and $L = P_V$, while (1.2) corresponds to $\mathcal{G} = \mathbb{R}^N$, $L\colon \mathcal{H} \to \mathbb{R}^N\colon x \mapsto (\langle x \mid s_i\rangle)_{1\le i\le N}$, $r = (\rho_i)_{1\le i\le N}$, and $z = 0$. As will be seen in Section 4, Problem 1.2 models a broad range of additional signal recovery problems.

In connection with (b), it is natural to ask whether the minimization problem (1.6) can be solved reliably by existing algorithms. Let us set

$$h\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto f(x) + g(Lx - r). \tag{1.8}$$

Then it follows from (1.5) that h is a proper lower semicontinuous convex function. Hence its proximity operator prox_h, which maps each $y \in \mathcal{H}$ to the unique minimizer of the function $x \mapsto h(x) + \|y - x\|^2/2$, is well defined (see Section 2.3). Accordingly, Problem 1.2 possesses a unique solution, which can be concisely written as

$$x = \operatorname{prox}_h z. \tag{1.9}$$

Since no-closed form expression exists for the proximity operator of composite functions such as h, one can contemplate the use of splitting strategies to construct $\operatorname{prox}_h z$ since (1.6) is of the form

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f_1(x) + f_2(x), \tag{1.10}$$

where

$$f_1\colon x \mapsto f(x) + \frac{1}{2}\|x - z\|^2 \quad \text{and} \quad f_2\colon x \mapsto g(Lx - r) \tag{1.11}$$

are lower semicontinuous convex functions from $\mathcal{H}$ to $\left]-\infty, +\infty\right]$. To tackle (1.10), a first splitting framework is that described in [35], which requires the additional assumption that f_2 be Lipschitz-differentiable on $\mathcal{H}$ (see also [12, 15, 19, 18, 25, 31, 38, 45] for recent work within this setting). In this case, (1.10) can be solved by the proximal forward-backward algorithm, which is governed by the updating rule

$$\left\lfloor\begin{array}{ll} x_{n+\frac{1}{2}} &= \nabla f_2(x_n) + a_{2,n} \\ x_{n+1} &= x_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n f_1}\big(x_n - \gamma_n x_{n+\frac{1}{2}}\big) + a_{1,n} - x_n\Big), \end{array}\right. \tag{1.12}$$

where $\lambda_n > 0$ and $\gamma_n > 0$, and where $a_{1,n}$ and $a_{2,n}$ model respectively tolerances in the approximate implementation of the proximity operator of f_1 and the gradient of f_2. Precise convergence results for the iterates $(x_n)_{n\in\mathbb{N}}$ can be found in Theorem 3.6. Let us add that there exist variants of this splitting method, which do not guarantee convergence of the iterates but do provide an optimal (in the sense of [59]) $O(1/n^2)$ rate of convergence of the objective values [7]. A limitation of this first framework is that it imposes that g be Lipschitz-differentiable and therefore excludes key problems such as (1.7). An alternative framework, which does not demand any smoothness assumption in (1.10), is investigated in [32]. It employs the Douglas-Rachford splitting algorithm, which revolves around the updating rule

$$\left\lfloor\begin{array}{ll} x_{n+\frac{1}{2}} &= \operatorname{prox}_{\gamma f_2} x_n + a_{2,n} \\ x_{n+1} &= x_n + \lambda_n\Big(\operatorname{prox}_{\gamma f_1}\big(2x_{n+\frac{1}{2}} - x_n\big) + a_{1,n} - x_{n+\frac{1}{2}}\Big), \end{array}\right. \tag{1.13}$$

where $\lambda_n > 0$ and $\gamma > 0$, and where $a_{1,n}$ and $a_{2,n}$ model tolerances in the approximate implementation of the proximity operators of f_1 and f_2, respectively (see [32, Theorem 20] for precise convergence results and [26] for further applications). However, this approach requires that the proximity operator of the composite function f_2 in (1.11) be computable to within some quantifiable error. Unfortunately, this is not possible in general, as explicit expressions of $\operatorname{prox}_{g\circ L}$ in terms of prox_g require stringent assumptions, for instance $L\circ L^* = \kappa\,\mathrm{Id}$ for some $\kappa > 0$ (see Example 2.8), which does not hold in the case of (1.2) and many other important problems. A third framework that appears to be relevant is that of [5], which is tailored for problems of the form

$$\underset{x\in\mathcal{H}}{\operatorname{minimize}}\;\; h_1(x) + h_2(x) + \frac{1}{2}\|x-z\|^2, \tag{1.14}$$

where h_1 and h_2 are lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$ such that $\operatorname{dom} h_1 \cap \operatorname{dom} h_2 \neq \varnothing$. This formulation coincides with our setting for $h_1 = f$ and $h_2 \colon x \mapsto g(Lx-r)$. The Dykstra-like algorithm devised in [5] to solve (1.14) is governed by the iteration

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
y_0 = z\\
q_0 = 0\\
p_0 = 0
\end{array}\right.\\
\text{For } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{lcl}
x_n & = & \operatorname{prox}_{h_2}(y_n+q_n)\\
q_{n+1} & = & y_n + q_n - x_n\\
y_{n+1} & = & \operatorname{prox}_{h_1}(x_n+p_n)\\
p_{n+1} & = & x_n + p_n - y_{n+1}
\end{array}\right.
\end{array} \tag{1.15}$$

and therefore requires that the proximity operators of h_1 and h_2 be computable explicitly. As just discussed, this is seldom possible in the case of the composite function h_2. To sum up, existing splitting techniques do not offer satisfactory options to solve Problem 1.2 and alternative routes must be explored. The cornerstone of our paper is that, by contrast, Problem 1.2 can be solved reliably via Fenchel-Moreau-Rockafellar duality so long as the operators prox_f and prox_g can be evaluated to within some quantifiable error, which will be shown to be possible in a wide variety of problems.

The paper is organized as follows. In Section 2 we provide the convex analytical background required in subsequent sections and, in particular, we review proximity operators. In Section 3, we show that Problem 1.2 satisfies properties (b) and (c). We then derive the Fenchel-Moreau-Rockafellar dual of Problem 1.2 and then show that it is amenable to solution by forward-backward splitting. The resulting primal-dual algorithm involves the functions f and g, as well as the operator L, separately and therefore achieves full splitting of the constituents of the primal problem. We show that the primal sequence produced by the algorithm converges strongly to the solution to Problem 1.2, and that the dual sequence converges weakly to a solution to the dual problem. Finally, in Section 4, we highlight applications of the proposed duality framework to best approximation problems, denoising problems using dictionaries, and recovery problems involving support functions. In particular, we extend and provide formal convergence results for the total variation denoising algorithm proposed in [23]. Although signal recovery applications are emphasized in the present paper, the proposed duality framework is applicable to any variational problem conforming to the format described in Problem 1.2.

2 Convex-analytical tools

2.1 General notation

Throughout the paper, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, and $\mathcal{B}(\mathcal{H},\mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$. The identity operator is denoted by Id, the adjoint of an operator $T \in \mathcal{B}(\mathcal{H},\mathcal{G})$ by T^*, the scalar products of both $\mathcal{H}$ and $\mathcal{G}$ by $\langle \cdot \mid \cdot \rangle$ and the associated norms by $\|\cdot\|$. Moreover, $\rightharpoonup$ and $\to$ denote respectively weak and strong convergence. Finally, we denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex functions $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ which are proper in the sense that $\operatorname{dom}\varphi = \big\{x \in \mathcal{H} \;\big|\; \varphi(x) < +\infty\big\} \neq \varnothing$.

2.2 Convex sets and functions

We provide some background on convex analysis; for a detailed account, see [79] and, for finite-dimensional spaces, [66].

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

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

the distance function of C is

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

the support function of C is

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

and the conical hull of C is

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

If C is also closed, 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 denote by $\operatorname{int} C$ the interior of C, by $\operatorname{span} C$ the span of C, and by $\overline{\operatorname{span}}\, C$ the closure of $\operatorname{span} C$. The core of C is $\operatorname{core} C = \big\{x \in C \;\big|\; \operatorname{cone}(C-x) = \mathcal{H}\big\}$, the strong relative interior of C is

$$\operatorname{sri} C = \big\{x \in C \;\big|\; \operatorname{cone}(C-x) = \overline{\operatorname{span}}\,(C-x)\big\}, \tag{2.5}$$

and the relative interior of C is $\operatorname{ri} C = \big\{x \in C \;\big|\; \operatorname{cone}(C-x) = \operatorname{span}(C-x)\big\}$. We have

$$\operatorname{int} C \subset \operatorname{core} C \subset \operatorname{sri} C \subset \operatorname{ri} C \subset C. \tag{2.6}$$

The strong relative interior is therefore an extension of the notion of an interior. This extension is particularly important in convex analysis as many useful sets have empty interior infinite-dimensional spaces.

Remark 2.1 The qualification condition (1.5) in Problem 1.2 is rather mild. In view of (2.6), it is satisfied in particular when r belongs to the core and, a fortiori, to the interior of $L(\operatorname{dom} f) - \operatorname{dom} g$; the latter is for instance satisfied when $L(\operatorname{dom} f) \cap (r + \operatorname{int} \operatorname{dom} g) \neq \varnothing$. If f and g are proper, then (1.5) is also satisfied when $L(\operatorname{dom} f) - \operatorname{dom} g = \mathcal{H}$ and, a fortiori, when f is finite-valued and L is surjective, or when g is finite-valued. If $\mathcal{G}$ is finite-dimensional, then (1.5) reduces to [66, Section 6]

$$r \in \operatorname{ri}\big(L(\operatorname{dom} f) - \operatorname{dom} g\big) = (\operatorname{ri} L(\operatorname{dom} f)) - \operatorname{ri} \operatorname{dom} g, \tag{2.7}$$

i.e., $(\operatorname{ri} L(\operatorname{dom} f)) \cap (r + \operatorname{ri} \operatorname{dom} g) \neq \varnothing$.

Let $\varphi \in \Gamma_0(\mathcal{H})$. The conjugate of φ is the function $\varphi^* \in \Gamma_0(\mathcal{H})$ defined by

$$(\forall u \in \mathcal{H}) \quad \varphi^*(u) = \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - \varphi(x). \tag{2.8}$$

The Fenchel-Moreau theorem states that $\varphi^{**} = \varphi$. The subdifferential of φ is the set-valued operator

$$\partial\varphi \colon \mathcal{H} \to 2^{\mathcal{H}} \colon x \mapsto \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H}) \;\; \langle y - x \mid u \rangle + \varphi(x) \leq \varphi(y)\big\}. \tag{2.9}$$

We have

$$(\forall (x,u) \in \mathcal{H} \times \mathcal{H}) \quad u \in \partial\varphi(x) \quad \Leftrightarrow \quad x \in \partial\varphi^*(u). \tag{2.10}$$

Moreover, if φ is Gâteaux differentiable at x, then

$$\partial\varphi(x) = \{\nabla\varphi(x)\}. \tag{2.11}$$

Fermat's rule states that

$$(\forall x \in \mathcal{H}) \quad x \in \operatorname{Argmin} \varphi = \big\{x \in \operatorname{dom} \varphi \;\big|\; (\forall y \in \mathcal{H}) \;\; \varphi(x) \leq \varphi(y)\big\} \quad \Leftrightarrow \quad 0 \in \partial\varphi(x). \tag{2.12}$$

If $\operatorname{Argmin} \varphi$ is a singleton, we denote by $\operatorname{argmin}_{y \in \mathcal{H}} \varphi(y)$ the unique minimizer of φ.

Lemma 2.2 [79, Theorem 2.8.3] *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\psi \in \Gamma_0(\mathcal{G})$, and let $M \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $0 \in \operatorname{sri}(M(\operatorname{dom} \varphi) - \operatorname{dom} \psi)$. Then $\partial(\varphi + \psi \circ M) = \partial\varphi + M^* \circ (\partial\psi) \circ M$.*

2.3 Moreau envelopes and proximity operators

Essential to this paper is the notion of a proximity operator, which is due to Moreau [56] (see [35, 57] for detailed accounts and Section 2.4 for closed-form examples). The Moreau envelope of φ is the continuous convex function

$$\widetilde{\varphi} \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto \min_{y \in \mathcal{H}} \varphi(y) + \frac{1}{2}\|x - y\|^2. \tag{2.13}$$

For every $x \in \mathcal{H}$, the function $y \mapsto \varphi(y) + \|x - y\|^2/2$ admits a unique minimizer, which is denoted by $\operatorname{prox}_\varphi x$. The proximity operator of φ is defined by

$$\operatorname{prox}_\varphi \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \operatorname*{argmin}_{y \in \mathcal{H}} \varphi(y) + \frac{1}{2}\|x - y\|^2 \tag{2.14}$$

and characterized by

$$(\forall (x,p) \in \mathcal{H} \times \mathcal{H}) \quad p = \operatorname{prox}_\varphi x \quad \Leftrightarrow \quad x - p \in \partial\varphi(p). \tag{2.15}$$

Lemma 2.3 [57] *Let $\varphi \in \Gamma_0(\mathcal{H})$. Then the following hold.*

(i) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\operatorname{prox}_\varphi x - \operatorname{prox}_\varphi y\|^2 \leq \big\langle x - y \mid \operatorname{prox}_\varphi x - \operatorname{prox}_\varphi y \big\rangle$.

(ii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\operatorname{prox}_\varphi x - \operatorname{prox}_\varphi y\| \leq \|x - y\|$.

(iii) $\widetilde{\varphi} + \widetilde{\varphi^*} = \|\cdot\|^2/2$.

(iv) $\widetilde{\varphi^*}$ *is Fréchet differentiable and* $\nabla\widetilde{\varphi^*} = \operatorname{prox}_\varphi = \operatorname{Id} - \operatorname{prox}_{\varphi^*}$.

The identity $\operatorname{prox}_\varphi = \operatorname{Id} - \operatorname{prox}_{\varphi^*}$ can be stated in a slightly extended context.

Lemma 2.4 [35, Lemma 2.10] *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and let $\gamma \in \left]0, +\infty\right[$. Then $x = \operatorname{prox}_{\gamma\varphi} x + \gamma \operatorname{prox}_{\gamma^{-1}\varphi^*}(\gamma^{-1}x)$.*

The following fact will also be required.

Lemma 2.5 *Let $\psi \in \Gamma_0(\mathcal{H})$, let $w \in \mathcal{H}$, and set $\varphi\colon x \mapsto \psi(x) + \|x - w\|^2/2$. Then $\varphi^*\colon u \mapsto \widetilde{\psi^*}(u + w) - \|w\|^2/2$.*

Proof. Let $u \in \mathcal{H}$. It follows from (2.8) and Lemma 2.3(iii) that

$$\begin{aligned}
\varphi^*(u) &= -\inf_{x\in\mathcal{H}} \psi(x) + \frac{1}{2}\|x - w\|^2 - \langle x \mid u\rangle \\
&= \frac{1}{2}\|u\|^2 + \langle w \mid u\rangle - \inf_{x\in\mathcal{H}} \psi(x) + \frac{1}{2}\|x - (w + u)\|^2 \\
&= \frac{1}{2}\|u + w\|^2 - \frac{1}{2}\|w\|^2 - \widetilde{\psi}(u + w) \\
&= \widetilde{\psi^*}(u + w) - \frac{1}{2}\|w\|^2,
\end{aligned} \tag{2.16}$$

which yields the desired identity. □

2.4 Examples of proximity operators

To solve Problem 1.2, our algorithm will use (approximate) evaluations of the proximity operators of the functions f and g^* (or, equivalently, of g by Lemma 2.3(iv)). In this section, we supply examples of proximity operators which admit closed-form expressions.

Example 2.6 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then the following hold.

(i) Set $\varphi = \iota_C$. Then $\operatorname{prox}_\varphi = P_C$ [57, Example 3.d].

(ii) Set $\varphi = \sigma_C$. Then $\operatorname{prox}_\varphi = \operatorname{Id} - P_C$ [35, Example 2.17].

(iii) Set $\varphi = d_C^2/(2\alpha)$. Then $(\forall x \in \mathcal{H})\ \operatorname{prox}_\varphi x = x + (1 + \alpha)^{-1}(P_C x - x)$ [35, Example 2.14].

(iv) Set $\varphi = (\|\cdot\|^2 - d_C^2)/(2\alpha)$. Then $(\forall x \in \mathcal{H})\ \operatorname{prox}_\varphi x = x - \alpha^{-1}P_C(\alpha(\alpha+1)^{-1}x)$ [35, Lemma 2.7].

Example 2.7 [35, Lemma 2.7] Let $\psi \in \Gamma_0(\mathcal{H})$ and set $\varphi = \|\cdot\|^2/2 - \widetilde{\psi}$. Then $\varphi \in \Gamma_0(\mathcal{H})$ and $(\forall x \in \mathcal{H})$ $\operatorname{prox}_\varphi x = x - \operatorname{prox}_{\psi/2}(x/2)$.

Example 2.8 [32, Proposition 11] Let $\mathcal{G}$ be a real Hilbert space, let $\psi \in \Gamma_0(\mathcal{G})$, let $M \in \mathcal{B}\,(\mathcal{H}, \mathcal{G})$, and set $\varphi = \psi \circ M$. Suppose that $M \circ M^* = \kappa \operatorname{Id}$, for some $\kappa \in \left]0, +\infty\right[$. Then $\varphi \in \Gamma_0(\mathcal{H})$ and

$$\operatorname{prox}_\varphi = \operatorname{Id} + \frac{1}{\kappa} M^* \circ (\operatorname{prox}_{\kappa\psi} - \operatorname{Id}) \circ M. \tag{2.17}$$

Example 2.9 [25, Proposition 2.10 and Remark 3.2(ii)] Set

$$\varphi \colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \sum_{k \in \mathbb{K}} \phi_k(\langle x \mid o_k \rangle), \tag{2.18}$$

where:

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

(ii) $(o_k)_{k \in \mathbb{K}}$ 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 $\varphi \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_\varphi x = \sum_{k \in \mathbb{K}} \big(\operatorname{prox}_{\phi_k} \langle x \mid o_k \rangle\big) o_k. \tag{2.19}$$

Example 2.10 [16, Proposition 2.1] Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \in \Gamma_0(\mathbb{R})$ be even, and set $\varphi = \phi \circ d_C$. Then $\varphi \in \Gamma_0(\mathcal{H})$. Moreover, $\operatorname{prox}_\varphi = P_C$ if $\phi = \iota_{\{0\}} + \eta$ for some $\eta \in \mathbb{R}$ and, otherwise,

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

Remark 2.11 Taking $C = \{0\}$ and $\phi \neq \iota_{\{0\}} + \eta$ $(\eta \in \mathbb{R})$ in Example 2.10 yields the proximity operator of $\phi \circ \|\cdot\|$, namely (using Lemma 2.3(iv))

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_\varphi x = \begin{cases} \dfrac{\operatorname{prox}_\phi \|x\|}{\|x\|} x, & \text{if } \|x\| > \max \partial\phi(0); \\ 0, & \text{if } \|x\| \leq \max \partial\phi(0). \end{cases} \tag{2.21}$$

On the other hand, if ϕ is differentiable at 0 in Example 2.10, then $\partial\phi(0) = \{0\}$ and (2.20) yields

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_\varphi x = \begin{cases} x + \dfrac{\operatorname{prox}_{\phi^*} d_C(x)}{d_C(x)}(P_C x - x), & \text{if } x \notin C; \\ x, & \text{if } x \in C. \end{cases} \tag{2.22}$$

Example 2.12 [16, Proposition 2.2] Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \in \Gamma_0(\mathbb{R})$ be even and nonconstant, and set $\varphi = \sigma_C + \phi \circ \|\cdot\|$. Then $\varphi \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_\varphi x = \begin{cases} \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x - P_C x), & \text{if } \ d_C(x) > \max \operatorname{Argmin} \phi; \\ x - P_C x, & \text{if } \ x \notin C \text{ and } \ d_C(x) \leq \max \operatorname{Argmin} \phi; \\ 0, & \text{if } \ x \in C. \end{cases} \tag{2.23}$$

Example 2.13 Let $A \in \mathcal{B}(\mathcal{H})$ be positive and self-adjoint, let $b \in \mathcal{H}$, let $\alpha \in \mathbb{R}$, and set $\varphi\colon x \mapsto \langle Ax \mid x\rangle/2 + \langle x \mid b\rangle + \alpha$. Then $\varphi \in \Gamma_0(\mathcal{H})$ and $(\forall x \in \mathcal{H})$ $\operatorname{prox}_\varphi x = (\operatorname{Id} + A)^{-1}(x - b)$.

Proof. It is clear that φ is a finite-valued continuous convex function. Now fix $x \in \mathcal{H}$ and set $\psi\colon y \mapsto \|x-y\|^2/2 + \langle Ay \mid y\rangle/2 + \langle y \mid b\rangle + \alpha$. Then $\nabla\psi\colon y \mapsto y - x + Ay + b$. Hence, $(\forall y \in \mathcal{H})$ $\nabla\psi(y) = 0 \Leftrightarrow y = (\operatorname{Id} + A)^{-1}(x - b)$. ☐

Example 2.14 For every $i \in \{1, \ldots, m\}$, let $(\mathcal{G}_i, \|\cdot\|)$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let $T_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$, and let $\alpha_i \in \left]0, +\infty\right[$. Set $(\forall x \in \mathcal{H})$ $\varphi(x) = (1/2)\sum_{i=1}^m \alpha_i \|T_i x - r_i\|^2$. Then $\varphi \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_\varphi x = \left(\operatorname{Id} + \sum_{i=1}^m \alpha_i T_i^* T_i\right)^{-1}\left(x + \sum_{i=1}^m \alpha_i T_i^* r_i\right). \tag{2.24}$$

Proof. We have $\varphi\colon x \mapsto \sum_{i=1}^m \alpha_i \langle T_i x - r_i \mid T_i x - r_i\rangle/2 = \langle Ax \mid x\rangle/2 + \langle x \mid b\rangle + \alpha$, where $A = \sum_{i=1}^m \alpha_i T_i^* T_i$, $b = -\sum_{i=1}^m \alpha_i T_i^* r_i$, and $\alpha = \sum_{i=1}^m \alpha_i \|r_i\|^2/2$. Hence, (2.24) follows from Example 2.13. ☐

As seen in Example 2.9, Example 2.10, Remark 2.11, and Example 2.12, some important proximity operators can be decomposed in terms of those of functions in $\Gamma_0(\mathbb{R})$. Here are explicit expressions for the proximity operators of such functions.

Example 2.15 [25, Examples 4.2 and 4.4] Let $p \in \left[1, +\infty\right[$, let $\alpha \in \left]0, +\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}\colon \eta \mapsto \alpha|\eta|^p$, let $\xi \in \mathbb{R}$, and set $\pi = \operatorname{prox}_\phi \xi$. Then the following hold.

(i) $\pi = \operatorname{sign}(\xi)\max\{|\xi| - \alpha, 0\}$, if $p = 1$;

(ii) $\pi = \xi + \dfrac{4\alpha}{3 \cdot 2^{1/3}}\left(|\rho - \xi|^{1/3} - |\rho + \xi|^{1/3}\right)$, where $\rho = \sqrt{\xi^2 + 256\alpha^3/729}$, if $p = 4/3$;

(iii) $\pi = \xi + 9\alpha^2 \operatorname{sign}(\xi)\big(1 - \sqrt{1 + 16|\xi|/(9\alpha^2)}\,\big)/8$, if $p = 3/2$;

(iv) $\pi = \xi/(1 + 2\alpha)$, if $p = 2$;

(v) $\pi = \operatorname{sign}(\xi)\big(\sqrt{1 + 12\alpha|\xi|} - 1\big)/(6\alpha)$, if $p = 3$;

(vi) $\pi = \left|\dfrac{\rho + \xi}{8\alpha}\right|^{1/3} - \left|\dfrac{\rho - \xi}{8\alpha}\right|^{1/3}$, where $\rho = \sqrt{\xi^2 + 1/(27\alpha)}$, if $p = 4$.

Example 2.16 [35, Example 2.18] Let $\alpha \in]0, +\infty[$ and set

$$\phi\colon \xi \mapsto \begin{cases} -\alpha \ln(\xi), & \text{if } \xi > 0; \\ +\infty, & \text{if } \xi \le 0. \end{cases} \tag{2.25}$$

Then $(\forall \xi \in \mathbb{R})$ $\operatorname{prox}_\phi \xi = (\xi + \sqrt{\xi^2 + 4\alpha})/2$.

Example 2.17 [31, Example 3.5] Let $\omega \in]0, +\infty[$ and set

$$\phi\colon \mathbb{R} \to]-\infty, +\infty] : \xi \mapsto \begin{cases} \ln(\omega) - \ln(\omega - |\xi|), & \text{if } |\xi| < \omega; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.26}$$

Then

$$(\forall \xi \in \mathbb{R}) \quad \operatorname{prox}_\phi \xi = \begin{cases} \operatorname{sign}(\xi)\, \dfrac{|\xi| + \omega - \sqrt{\big||\xi| - \omega\big|^2 + 4}}{2}, & \text{if } |\xi| > 1/\omega; \\ 0 & \text{otherwise.} \end{cases} \tag{2.27}$$

Example 2.18 [25, Example 4.5] Let $\omega \in]0, +\infty[$, $\tau \in]0, +\infty[$, and set

$$\phi\colon \mathbb{R} \to]-\infty, +\infty] : \xi \mapsto \begin{cases} \tau\xi^2, & \text{if } |\xi| \le \omega/\sqrt{2\tau}; \\ \omega\sqrt{2\tau}|\xi| - \omega^2/2, & \text{otherwise.} \end{cases} \tag{2.28}$$

Then

$$(\forall \xi \in \mathbb{R}) \quad \operatorname{prox}_\phi \xi = \begin{cases} \dfrac{\xi}{2\tau + 1}, & \text{if } |\xi| \le \omega(2\tau + 1)/\sqrt{2\tau}; \\ \xi - \omega\sqrt{2\tau}\operatorname{sign}(\xi), & \text{if } |\xi| > \omega(2\tau + 1)/\sqrt{2\tau}. \end{cases} \tag{2.29}$$

Further examples can be constructed via the following rules.

Lemma 2.19 [31, Proposition 3.6] *Let $\phi = \psi + \sigma_\Omega$, where $\psi \in \Gamma_0(\mathbb{R})$ and $\Omega \subset \mathbb{R}$ is a nonempty closed interval. Suppose that ψ is differentiable at 0 with $\psi'(0) = 0$. Then $\operatorname{prox}_\phi = \operatorname{prox}_\psi \circ \operatorname{soft}_\Omega$, where*

$$\operatorname{soft}_\Omega : \mathbb{R} \to \mathbb{R} \colon \xi \mapsto \begin{cases} \xi - \underline{\omega}, & \textit{if } \xi < \underline{\omega}; \\ 0, & \textit{if } \xi \in \Omega; \\ \xi - \overline{\omega}, & \textit{if } \xi > \overline{\omega}, \end{cases} \qquad \textit{with} \quad \begin{cases} \underline{\omega} = \inf \Omega, \\ \overline{\omega} = \sup \Omega. \end{cases} \tag{2.30}$$

Lemma 2.20 [32, Proposition 12(ii)] *Let $\phi = \iota_C + \psi$, where $\psi \in \Gamma_0(\mathbb{R})$ and where C is a closed interval in $\mathbb{R}$ such that $C \cap \operatorname{dom} \psi \neq \varnothing$. Then $\operatorname{prox}_{\iota_C + \psi} = P_C \circ \operatorname{prox}_\psi$.*

3 Dualization and algorithm

3.1 Fenchel-Moreau-Rockafellar duality

Our analysis will revolve around the following version of the Fenchel-Moreau-Rockafellar duality formula (see [44], [58], and [65] for historical work). It will also exploit various aspects of the Baillon-Haddad theorem [6].

Lemma 3.1 [79, Corollary 2.8.5] *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\psi \in \Gamma_0(\mathcal{G})$, and let $M \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $0 \in \operatorname{sri}(M(\operatorname{dom}\varphi) - \operatorname{dom}\psi)$. Then*

$$\inf_{x \in \mathcal{H}} \varphi(x) + \psi(Mx) = -\min_{v \in \mathcal{G}} \varphi^*(-M^*v) + \psi^*(v). \tag{3.1}$$

The problem of minimizing $\varphi + \psi \circ M$ on $\mathcal{H}$ in (3.1) is referred to as the primal problem, and that of minimizing $\varphi^* \circ (-M^*) + \psi^*$ on $\mathcal{G}$ as the dual problem. Lemma 3.1 gives conditions under which a dual solution exists and the value of the dual problem coincides with the opposite of the value of the primal problem. We can now introduce the dual of Problem 1.2.

Problem 3.2 (dual problem) Under the same assumptions as in Problem 1.2,

$$\underset{v \in \mathcal{G}}{\text{minimize}} \; \widetilde{f^*}(z - L^*v) + g^*(v) + \langle v \mid r \rangle. \tag{3.2}$$

Proposition 3.3 *Problem 3.2 is the dual of Problem 1.2 and it admits at least one solution. Moreover, every solution v to Problem 3.2 is characterized by the inclusion*

$$L\big(\operatorname{prox}_f(z - L^*v)\big) - r \in \partial g^*(v). \tag{3.3}$$

Proof. Let us set $w = z$, $\varphi = f + \|\cdot - w\|^2/2$, $M = L$, and $\psi = g(\cdot - r)$. Then $(\forall x \in \mathcal{H})$ $\varphi(x) + \psi(Mx) = f(x) + g(Lx - r) + \|x - z\|^2/2$. Hence, it results from (3.1) and Lemma 2.5 that the dual of Problem 1.2 is to minimize the function

$$\begin{aligned} \varphi^* \circ (-M^*) + \psi^* \colon v \mapsto & \widetilde{f^*}(-M^*v + w) - \frac{1}{2}\|w\|^2 + \psi^*(v) \\ &= \widetilde{f^*}(z - L^*v) - \frac{1}{2}\|z\|^2 + g^*(v) + \langle v \mid r \rangle \end{aligned} \tag{3.4}$$

or, equivalently, the function $v \mapsto \widetilde{f^*}(z - L^*v) + g^*(v) + \langle v \mid r \rangle$. In view of (1.5), the first two claims therefore follow from Lemma 3.1. To establish the last claim, note that (2.13) asserts that $\operatorname{dom} \widetilde{f^*} \circ (z - L^*\cdot) = \mathcal{G}$. Hence, using (2.12), Lemma 2.2, (2.11), and Lemma 2.3(iv), we get

$$\begin{aligned} v \text{ solves (3.2)} \quad &\Leftrightarrow \quad 0 \in \partial\Big(\widetilde{f^*} \circ (z - L^*\cdot) + g^* + \langle \cdot \mid r \rangle\Big)(v) \\ &\Leftrightarrow \quad 0 \in -L\big(\nabla\widetilde{f^*}(z - L^*v)\big) + \partial g^*(v) + r \\ &\Leftrightarrow \quad 0 \in -L\big(\operatorname{prox}_f(z - L^*v)\big) + \partial g^*(v) + r, \end{aligned} \tag{3.5}$$

which yields (3.3). ☐

A key property underlying our setting is that the primal solution can actually be recovered from any dual solution (this is property (c) in the Introduction).

Proposition 3.4 *Let v be a solution to Problem 3.2 and set*

$$x = \operatorname{prox}_f(z - L^*v). \tag{3.6}$$

Then x is the solution to Problem 1.2.

Proof. We derive from (3.6) and (2.15) that $z - L^*v - x \in \partial f(x)$. Therefore

$$-L^*v \in \partial f(x) + x - z. \tag{3.7}$$

On the other hand, it follows from (3.3), (3.6), and (2.10) that

$$\begin{aligned} v \text{ solves } (3.2) \quad &\Leftrightarrow \quad Lx - r \in \partial g^*(v) \\ &\Leftrightarrow \quad v \in \partial g(Lx - r) \\ &\Rightarrow \quad L^*v \in L^*\big(\partial g(Lx - r)\big). \end{aligned} \tag{3.8}$$

Upon adding (3.7) and (3.8), invoking Lemma 2.2, and then (2.12) we obtain

$$\begin{aligned} v \text{ solves } (3.2) \quad \Rightarrow \quad & 0 = L^*v - L^*v \\ & \in \partial f(x) + L^*\big(\partial g(Lx - r)\big) + x - z \\ & = \partial f(x) + L^*\big(\partial g(Lx - r)\big) + \nabla\Big(\frac{1}{2}\|\cdot - z\|^2\Big)(x) \\ & = \partial\Big(f + g(L \cdot - r) + \frac{1}{2}\|\cdot - z\|^2\Big)(x) \\ \Leftrightarrow \quad & x \text{ solves } (1.6), \end{aligned} \tag{3.9}$$

which completes the proof. $\square$

3.2 Algorithm

As seen in (1.9), the unique solution to Problem 1.2 is $\operatorname{prox}_h z$, where h is defined in (1.8). Since $\operatorname{prox}_h z$ cannot be computed directly, it will be constructed iteratively by the following algorithm, which produces a primal sequence $(x_n)_{n\in\mathbb{N}}$ as well as a dual sequence $(v_n)_{n\in\mathbb{N}}$.

Algorithm 3.5 Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}} \|a_n\| < +\infty$ and let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$. Sequences $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are generated by the following routine.

$$\begin{array}{l} \text{Initialization} \\ \left\lfloor \begin{array}{l} \varepsilon \in \left]0, \min\{1, \|L\|^{-2}\}\right[\\ v_0 \in \mathcal{G} \end{array} \right. \\ \text{For } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{prox}_f(z - L^*v_n) + b_n \\ \gamma_n \in \left[\varepsilon, 2\|L\|^{-2} - \varepsilon\right] \\ \lambda_n \in [\varepsilon, 1] \\ v_{n+1} = v_n + \lambda_n\big(\operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n - r)) + a_n - v_n\big). \end{array} \right. \end{array} \tag{3.10}$$

It is noteworthy that each iteration of Algorithm 3.5 achieves full splitting with respect to the operators L, prox_f, and $\operatorname{prox}_{g^*}$, which are used at separate steps. In addition, (3.10) incorporates tolerances a_n and b_n in the computation of the proximity operators at iteration n.

3.3 Convergence

Our main convergence result will be a consequence of Proposition 3.4 and the following results on the convergence of the forward-backward splitting method.

Theorem 3.6 [35, Theorem 3.4] *Let f_1 and f_2 be functions in $\Gamma_0(\mathcal{G})$ such that the set G of minimizers of f_1+f_2 is nonempty and such that f_2 is differentiable on $\mathcal{G}$ with a $1/\beta$-Lipschitz continuous gradient for some $\beta \in \,]0,+\infty[$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2\beta[$ such that $\inf_{n\in\mathbb{N}}\gamma_n > 0$ and $\sup_{n\in\mathbb{N}}\gamma_n < 2\beta$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\inf_{n\in\mathbb{N}}\lambda_n > 0$, and let $(a_{1,n})_{n\in\mathbb{N}}$ and $(a_{2,n})_{n\in\mathbb{N}}$ be sequences in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}}\|a_{1,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}}\|a_{2,n}\| < +\infty$. Fix $v_0 \in \mathcal{G}$ and, for every $n \in \mathbb{N}$, set*

$$v_{n+1} = v_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n f_1}\big(v_n - \gamma_n(\nabla f_2(v_n) + a_{2,n})\big) + a_{1,n} - v_n\Big). \tag{3.11}$$

Then $(v_n)_{n\in\mathbb{N}}$ converges weakly to a point $v \in G$ and $\sum_{n\in\mathbb{N}}\|\nabla f_2(v_n) - \nabla f_2(v)\|^2 < +\infty$.

The following theorem describes the asymptotic behavior of Algorithm 3.5.

Theorem 3.7 *Let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 3.5, and let x be the solution to Problem 1.2. Then the following hold.*

(i) *$(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to Problem 3.2 and $x = \operatorname{prox}_f(z - L^*v)$.*

(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to x.*

Proof. Let us define two functions f_1 and f_2 on $\mathcal{G}$ by $f_1\colon v \mapsto g^*(v)+\langle v \mid r\rangle$ and $f_2\colon v \mapsto \widetilde{f^*}(z-L^*v)$. Then (3.2) amounts to minimizing $f_1 + f_2$ on $\mathcal{G}$. Let us first check that all the assumptions specified in Theorem 3.6 are satisfied. First, f_1 and f_2 are in $\Gamma_0(\mathcal{G})$ and, by Proposition 3.3, $\operatorname{Argmin} f_1 + f_2 \neq \varnothing$. Moreover, it follows from Lemma 2.3(iv) that f_2 is differentiable on $\mathcal{G}$ with gradient

$$\nabla f_2\colon v \mapsto -L\big(\operatorname{prox}_f(z - L^*v)\big). \tag{3.12}$$

Hence, we derive from Lemma 2.3(ii) that

$$\begin{aligned}(\forall v \in \mathcal{G})(\forall w \in \mathcal{G}) \quad \|\nabla f_2(v) - \nabla f_2(w)\| &\leq \|L\|\,\|\operatorname{prox}_f(z - L^*v) - \operatorname{prox}_f(z - L^*w)\|\\ &\leq \|L\|\,\|L^*v - L^*w\|\\ &\leq \|L\|^2\,\|v - w\|.\end{aligned} \tag{3.13}$$

The reciprocal of the Lipschitz constant of ∇f_2 is therefore $\beta = \|L\|^{-2}$. Now set

$$(\forall n \in \mathbb{N}) \quad a_{1,n} = a_n \quad \text{and} \quad a_{2,n} = -Lb_n. \tag{3.14}$$

Then $\sum_{n\in\mathbb{N}}\|a_{1,n}\| = \sum_{n\in\mathbb{N}}\|a_n\| < +\infty$ and $\sum_{n\in\mathbb{N}}\|a_{2,n}\| \leq \|L\|\sum_{n\in\mathbb{N}}\|b_n\| < +\infty$. Moreover, for every $n \in \mathbb{N}$, (3.10) yields

$$x_n = \operatorname{prox}_f(z - L^*v_n) + b_n \tag{3.15}$$

and, together with [35, Lemma 2.6(i)],

$$\begin{aligned} v_{n+1} &= v_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n g^*}\big(v_n + \gamma_n(Lx_n - r)\big) + a_n - v_n\Big) \\ &= v_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n g^* + \langle \cdot \mid \gamma_n r\rangle}\big(v_n + \gamma_n L x_n\big) + a_n - v_n\Big) \\ &= v_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n (g^* + \langle \cdot \mid r\rangle)}\big(v_n + \gamma_n L(\operatorname{prox}_f(z - L^* v_n) + b_n)\big) + a_n - v_n\Big) \\ &= v_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n f_1}\big(v_n - \gamma_n(\nabla f_2(v_n) + a_{2,n})\big) + a_{1,n} - v_n\Big). \end{aligned} \tag{3.16}$$

This provides precisely the update rule (3.11), which allows us to apply Theorem 3.6.

(i): In view of the above, we derive from Theorem 3.6 that $(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to (3.2). The second assertion follows from Proposition 3.4.

(ii): Let us set

$$(\forall n \in \mathbb{N}) \quad y_n = x_n - b_n = \operatorname{prox}_f(z - L^* v_n). \tag{3.17}$$

As seen in (i), $v_n \rightharpoonup v$, where v is a solution to (3.2), and $x = \operatorname{prox}_f(z - L^* v)$. Now set $\rho = \sup_{n\in\mathbb{N}} \|v_n - v\|$. Then $\rho < +\infty$ and, using Lemma 2.3(i) and (3.12), we obtain

$$\begin{aligned} \|y_n - x\|^2 &= \|\operatorname{prox}_f(z - L^* v_n) - \operatorname{prox}_f(z - L^* v)\|^2 \\ &\leq \big\langle L^* v - L^* v_n \mid \operatorname{prox}_f(z - L^* v_n) - \operatorname{prox}_f(z - L^* v)\big\rangle \\ &= \big\langle v_n - v \mid -L\big(\operatorname{prox}_f(z - L^* v_n)\big) + L\big(\operatorname{prox}_f(z - L^* v)\big)\big\rangle \\ &= \langle v_n - v \mid \nabla f_2(v_n) - \nabla f_2(v)\rangle \\ &\leq \rho\|\nabla f_2(v_n) - \nabla f_2(v)\|. \end{aligned} \tag{3.18}$$

However, as seen in Theorem 3.6, $\|\nabla f_2(v_n) - \nabla f_2(v)\| \to 0$. Hence, we derive from (3.18) that $y_n \to x$. In turn, since $b_n \to 0$, (3.17) yields $x_n \to x$. ☐

Remark 3.8 (Dykstra-like algorithm) Suppose that, in Problem 1.2, $\mathcal{G} = \mathcal{H}$, $L = \operatorname{Id}$, and $r = 0$. Then it follows from Theorem 3.7(ii) that the sequence $(x_n)_{n\in\mathbb{N}}$ produced by Algorithm 3.5 converges strongly to $x = \operatorname{prox}_{f+g} z$. Now let us consider the special case when Algorithm 3.5 is implemented with $v_0 = 0$, $\gamma_n \equiv 1$, $\lambda_n \equiv 1$, and no errors, i.e., $a_n \equiv 0$ and $b_n \equiv 0$. Then it follows from Lemma 2.3(iv) that (3.10) simplifies to

$$\begin{array}{l} \text{Initialization} \\ \left\lfloor \; v_0 = 0 \right. \\ \text{For } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{prox}_f(z - v_n) \\ v_{n+1} = x_n + v_n - \operatorname{prox}_g(x_n + v_n). \end{array} \right. \end{array} \tag{3.19}$$

Using [5, Eq. (2.10)] it can then easily be shown by induction that the resulting sequence $(x_n)_{n\in\mathbb{N}}$ coincides with that produced by the Dykstra-like algorithm (1.15) (with $h_1 = g$ and $h_2 = f$) and that the sequence $(v_n)_{n\in\mathbb{N}}$ coincides with the sequence $(p_n)_{n\in\mathbb{N}}$ of (1.15). The fact that $x_n \to \operatorname{prox}_{f+g} z$ was established in [5, Theorem 3.3(i)] using different tools. Thus, Algorithm 3.5 can be regarded as a generalization of the Dykstra-like algorithm (1.15).

Remark 3.9 Theorem 3.7 remains valid if we introduce explicitly errors in the implementation of the operators L and L^* in Algorithm 3.5. More precisely, we can replace the steps defining x_n and v_n in (3.10) by

$$\left\lfloor \begin{array}{l} x_n = \operatorname{prox}_f(z - L^*v_n - d_{2,n}) + d_{1,n} \\ v_{n+1} = v_n + \lambda_n\big(\operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n + c_{2,n} - r)) + c_{1,n} - v_n\big), \end{array} \right. \tag{3.20}$$

where $(d_{1,n})_{n\in\mathbb{N}}$ and $(d_{2,n})_{n\in\mathbb{N}}$ are sequences in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|d_{1,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}}\|d_{2,n}\| < +\infty$, and where $(c_{1,n})_{n\in\mathbb{N}}$ and $(c_{2,n})_{n\in\mathbb{N}}$ are sequences in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}}\|c_{1,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}}\|c_{2,n}\| < +\infty$. Indeed set, for every $n \in \mathbb{N}$,

$$\begin{cases} a_n = c_{1,n} + \operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n + c_{2,n} - r)) - \operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n - r)) \\ b_n = d_{1,n} + \operatorname{prox}_f(z - L^*v_n - d_{2,n}) - \operatorname{prox}_f(z - L^*v_n). \end{cases} \tag{3.21}$$

Then (3.20) reverts to

$$\left\lfloor \begin{array}{l} x_n = \operatorname{prox}_f(z - L^*v_n) + b_n \\ v_{n+1} = v_n + \lambda_n\big(\operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n - r)) + a_n - v_n\big), \end{array} \right. \tag{3.22}$$

as in (3.10). Moreover, by Lemma 2.3(ii),

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad \|a_n\| &\le \|c_{1,n}\| + \|\operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n + c_{2,n} - r)) - \operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n - r))\| \\ &\le \|c_{1,n}\| + \gamma_n\|c_{2,n}\| \\ &\le \|c_{1,n}\| + 2\|L\|^{-2}\|c_{2,n}\|. \end{aligned} \tag{3.23}$$

Thus, $\sum_{n\in\mathbb{N}}\|a_n\| < +\infty$. Likewise, we have $\sum_{n\in\mathbb{N}}\|b_n\| < +\infty$.

4 Application to specific signal recovery problems

In this section, we present a few applications of the duality framework presented in Section 3, which correspond to specific choices of $\mathcal{H}$, $\mathcal{G}$, L, f, g, r, and z in Problem 1.2.

4.1 Best feasible approximation

A standard feasibility problem in signal recovery is to find a signal in the intersection of two closed convex sets modeling constraints on the ideal solution [29, 69, 72, 78]. A more structured variant of this problem, is the so-called split feasibility problem [17, 20, 21], which requires to find a signal in a closed convex set $C \subset \mathcal{H}$ and such that some affine transformation of it lies in a closed convex set $D \subset \mathcal{G}$. Such problems typically admit infinitely many solutions and one often seeks to find the solution that lies closest to a nominal signal $z \in \mathcal{H}$ [27, 64]. This leads to the formulation (1.7), which consists in finding the best approximation to a reference signal $z \in \mathcal{H}$ from the feasibility set $C \cap L^{-1}(r + D)$.

Problem 4.1 Let $z \in \mathcal{H}$, let $r \in \mathcal{G}$, let $C \subset \mathcal{H}$ and $D \subset \mathcal{G}$ be closed convex sets, and let L be a nonzero operator in $\mathcal{B}(\mathcal{H}, \mathcal{G})$ such that

$$r \in \operatorname{sri}\big(L(C) - D\big). \tag{4.1}$$

The problem is to

$$\underset{\substack{x \in C \\ Lx - r \in D}}{\text{minimize}} \ \frac{1}{2}\|x - z\|^2, \tag{4.2}$$

and its dual is to

$$\underset{v \in \mathcal{G}}{\text{minimize}} \ \frac{1}{2}\|z - L^*v\|^2 - \frac{1}{2}d_C^2(z - L^*v) + \sigma_D(v) + \langle v \mid r\rangle. \tag{4.3}$$

Proposition 4.2 *Let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|b_n\| < +\infty$, let $(c_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}}\|c_n\| < +\infty$, and let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences generated by the following routine.*

$$\begin{array}{l}
\text{Initialization} \\
\left\lfloor \begin{array}{l}
\varepsilon \in \left]0, \min\{1, \|L\|^{-2}\}\right[\\
v_0 \in \mathcal{G}
\end{array}\right. \\
\text{For } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
x_n = P_C(z - L^*v_n) + b_n \\
\gamma_n \in \left[\varepsilon, 2\|L\|^{-2} - \varepsilon\right] \\
\lambda_n \in [\varepsilon, 1] \\
v_{n+1} = v_n + \lambda_n\gamma_n\big(Lx_n - r - P_D(\gamma_n^{-1}v_n + Lx_n - r) + c_n\big).
\end{array}\right.
\end{array} \tag{4.4}$$

Then the following hold, where x designates the primal solution to Problem 4.1.

(i) *$(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to (4.3) and $x = P_C(z - L^*v)$.*

(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to x.*

Proof. Set $f = \iota_C$ and $g = \iota_D$. Then (1.6) reduces to (4.2) and (1.5) reduces to (4.1). In addition, we derive from Lemma 2.3(iii) that $\widetilde{f^*} = \|\cdot\|^2/2 - \widetilde{\iota_C} = (\|\cdot\|^2 - d_C^2)/2$. Hence, in view of (3.2), (4.3) in indeed the dual of (4.2). Furthermore, items (i) and (ii) in Example 2.6 yield $\operatorname{prox}_f = P_C$ and

$$(\forall n \in \mathbb{N}) \quad \operatorname{prox}_{\gamma_n g^*} = \operatorname{prox}_{\gamma_n \sigma_D} = \operatorname{prox}_{\sigma_{\gamma_n D}} = \operatorname{Id} - P_{\gamma_n D} = \operatorname{Id} - \gamma_n P_D(\cdot/\gamma_n). \tag{4.5}$$

Finally, set $(\forall n \in \mathbb{N})$ $a_n = \gamma_n c_n$. Then $\sum_{n\in\mathbb{N}}\|a_n\| \leq 2\|L\|^{-2}\sum_{n\in\mathbb{N}}\|c_n\| < +\infty$ and, altogether, (3.10) reduces to (4.4). Hence, the results follow from Theorem 3.7. ☐

Our investigation was motivated in the Introduction by the duality framework of [64]. In the next example we recover and sharpen Proposition 1.1.

Example 4.3 Consider the special case of Problem 4.1 in which $z = 0$, $\mathcal{G} = \mathbb{R}^N$, $D = \{0\}$, $r = (\rho_i)_{1\leq i\leq N}$, and $L\colon x \mapsto (\langle x \mid s_i\rangle)_{1\leq i\leq N}$, where $(s_i)_{1\leq i\leq N} \in \mathcal{H}^N$ satisfies $\sum_{i=1}^N \|s_i\|^2 \leq 1$. Then, by (2.7), (4.1) reduces to $r \in \operatorname{ri} L(C)$ and (4.2) to (1.2). Since $\|L\| \leq 1$, specializing (4.4) to the case when $c_n \equiv 0$ and $\lambda_n \equiv 1$, and introducing the sequence $(w_n)_{n\in\mathbb{N}} = (-v_n)_{n\in\mathbb{N}}$ for convenience yields the following routine.

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\varepsilon \in \left]0,1\right[\\
w_0 \in \mathbb{R}^N
\end{array}
\right.\\
\text{For } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_n = P_C(L^* w_n) + b_n\\
\gamma_n \in \big[\varepsilon, 2\|L\|^{-2} - \varepsilon\big]\\
w_{n+1} = w_n + \gamma_n\big(r - Lx_n\big).
\end{array}
\right.
\end{array}
\tag{4.6}
$$

Thus, if $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$, we deduce from Proposition 4.2(i) and Proposition 3.3 the weak convergence of $(w_n)_{n\in\mathbb{N}}$ to a point w such that $v = -w$ satisfies (3.3), i.e., $L(P_C(-L^*v)) - r \in \partial\iota^*_{\{0\}}(v) = \{0\}$ or, equivalently, $L(P_C(L^*w)) = r$, and such that $P_C(-L^*v) = P_C(L^*w)$ is the solution to (1.2). In addition, we derive from Proposition 4.2(ii), the strong convergence of $(x_n)_{n\in\mathbb{N}}$ to the solution to (1.2). These results sharpen the conclusion of Proposition 1.1 (note that (1.3) corresponds to setting $b_n \equiv 0$ and $\gamma_n \equiv \gamma \in \left]0,2\right[$ in (4.6)).

Example 4.4 We consider the standard linear inverse problem of recovering an ideal signal $\overline{x} \in \mathcal{H}$ from an observation

$$r = L\overline{x} + s \tag{4.7}$$

in $\mathcal{G}$, where $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ and where $s \in \mathcal{G}$ models noise. Given an estimate x of $\overline{x}$, the residual $r - Lx$ should ideally behave like the noise process. Thus, any known probabilistic attribute of the noise process can give rise to a constraint. This observation was used in [34, 72] to construct various constraints of the type $Lx - r \in D$, where D is closed and convex. In this context, (4.2) amounts to finding the signal which is closest to some nominal signal z and which satisfies a noise-based constraint and some convex constraint on $\overline{x}$ represented by C. Such problems were considered for instance in [27], where they were solved by methods that require the projection onto the set $\big\{x \in \mathcal{H} \;\big|\; Lx - r \in D\big\}$, which is typically hard to compute, even in the simple case when D is a closed Euclidean ball [72]. By contrast, the iterative method (4.4) requires only the projection onto D to enforce such constraints.

4.2 Soft best feasible approximation

It follows from (4.1) that the underlying feasibility set $C \cap L^{-1}(r + D)$ in Problem 4.1 is nonempty. In many situations, feasibility may not guaranteed due to, for instance, imprecise prior information or unmodeled dynamics in the data formation process [28, 77]. In such instances, one can relax the hard constraints $x \in C$ and $Lx - r \in D$ in (4.2) by merely forcing that x be close to C and $Lx - r$ be close to D. Let us formulate this problem within the framework of Problem 1.2.

Problem 4.5 Let $z \in \mathcal{H}$, let $r \in \mathcal{G}$, let $C \subset \mathcal{H}$ and $D \subset \mathcal{G}$ be nonempty closed convex sets, let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be a nonzero operator, and let ϕ and ψ be even functions in $\Gamma_0(\mathbb{R}) \smallsetminus \{\iota_{\{0\}}\}$ such that

$$r \in \operatorname{sri}\Big(L\big(\big\{x \in \mathcal{H} \;\big|\; d_C(x) \in \operatorname{dom}\phi\big\}\big) - \big\{y \in \mathcal{G} \;\big|\; d_D(y) \in \operatorname{dom}\psi\big\}\Big). \tag{4.8}$$

The problem is to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \phi\big(d_C(x)\big) + \psi\big(d_D(Lx - r)\big) + \frac{1}{2}\|x - z\|^2, \tag{4.9}$$

and its dual is to

$$\underset{v\in\mathcal{G}}{\text{minimize}}\;\; \frac{1}{2}\|z-L^*v\|^2-(\phi\circ d_C)^{\sim}(z-L^*v)+\sigma_D(v)+\psi^*(\|v\|)+\langle v\mid r\rangle. \tag{4.10}$$

Since ϕ and ψ are even functions in $\Gamma_0(\mathbb{R})\smallsetminus\{\iota_{\{0\}}\}$, we can use Example 2.10 to get an explicitly expression of the proximity operators involved and solve the minimization problems (4.9) and (4.10) as follows.

Proposition 4.6 *Let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|b_n\|<+\infty$, let $(c_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}}\|c_n\|<+\infty$, and let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences generated by the following routine.*

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor\begin{array}{l}
\varepsilon\in\left]0,\min\{1,\|L\|^{-2}\}\right[\\
v_0\in\mathcal{G}
\end{array}\right.\\
\text{For } n=0,1,\ldots\\
\left\lfloor\begin{array}{l}
y_n=z-L^*v_n\\
\text{if}\;\; d_C(y_n)>\max\partial\phi(0)\\
\left\lfloor\; x_n=y_n+\dfrac{\operatorname{prox}_{\phi^*}d_C(y_n)}{d_C(y_n)}(P_Cy_n-y_n)+b_n\right.\\
\text{if}\;\; d_C(y_n)\leq\max\partial\phi(0)\\
\left\lfloor\; x_n=P_Cy_n+b_n\right.\\
\gamma_n\in\left[\varepsilon,2\|L\|^{-2}-\varepsilon\right]\\
w_n=\gamma_n^{-1}v_n+Lx_n-r\\
\text{if}\;\; d_D(w_n)>\gamma_n^{-1}\max\partial\psi(0)\\
\left\lfloor\; p_n=\dfrac{\operatorname{prox}_{(\gamma_n^{-1}\psi)^*}d_D(w_n)}{d_D(w_n)}(w_n-P_Dw_n)+c_n\right.\\
\text{if}\;\; d_D(w_n)\leq\gamma_n^{-1}\max\partial\psi(0)\\
\left\lfloor\; p_n=w_n-P_Dw_n+c_n\right.\\
\lambda_n\in[\varepsilon,1]\\
v_{n+1}=v_n+\lambda_n\big(\gamma_np_n-v_n\big).
\end{array}\right.
\end{array} \tag{4.11}$$

Then the following hold, where x designates the primal solution to Problem 4.5.

(i) *$(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to (4.10) and, if we set $y=z-L^*v$,*

$$x=\begin{cases} y+\dfrac{\operatorname{prox}_{\phi^*}d_C(y)}{d_C(y)}(P_Cy-y), & \text{if } d_C(y)>\max\partial\phi(0);\\ P_Cy, & \text{if } d_C(y)\leq\max\partial\phi(0).\end{cases} \tag{4.12}$$

(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to x.*

Proof. Set $f=\phi\circ d_C$ and $g=\psi\circ d_D$. Since d_C and d_D are continuous convex functions, $f\in\Gamma_0(\mathcal{H})$ and $g\in\Gamma_0(\mathcal{G})$. Moreover, (4.8) implies that (1.5) holds. Thus, Problem 4.5 is a special case of

Problem 1.2. On the other hand, it follows from Lemma 2.3(iii) that $\widetilde{f^*} = \|\cdot\|^2/2 - (\phi\circ d_C)^\sim$ and from [16, Lemma 2.2] that $g^* = \sigma_D + \psi^*\circ\|\cdot\|$. This shows that (4.10) is the dual of (4.9). Let us now examine iteration n of the algorithm. In view of Example 2.10, the vector x_n in (4.11) is precisely the vector $x_n = \operatorname{prox}_f(z - L^*v_n) + b_n$ of (3.10). Moreover, using successively the definition of w_n in (4.11), Lemma 2.4, Example 2.10, and the definition of p_n in (4.11), we obtain

$$\begin{aligned}
&\gamma_n^{-1}\operatorname{prox}_{\gamma_n g^*}(v_n + \gamma_n(Lx_n - r))\\
&\qquad\qquad = \gamma_n^{-1}\operatorname{prox}_{\gamma_n g^*}(\gamma_n w_n)\\
&\qquad\qquad = w_n - \operatorname{prox}_{\gamma_n^{-1}g} w_n\\
&\qquad\qquad = w_n - \operatorname{prox}_{(\gamma_n^{-1}\psi)\circ d_D} w_n\\
&\qquad\qquad = \begin{cases} \dfrac{\operatorname{prox}_{(\gamma_n^{-1}\psi)^*} d_D(w_n)}{d_D(w_n)}(w_n - P_D w_n) & \text{if } d_D(w_n) > \gamma_n^{-1}\max\partial\psi(0)\\ w_n - P_D w_n & \text{if } d_D(w_n) \leq \gamma_n^{-1}\max\partial\psi(0)\end{cases}\\
&\qquad\qquad = p_n - c_n.
\end{aligned}\tag{4.13}$$

Altogether, (4.11) is a special instance of (3.10) in which $(\forall n\in\mathbb{N})$ $a_n = \gamma_n c_n$. Therefore, since $\sum_{n\in\mathbb{N}}\|a_n\| \leq 2\|L\|^{-2}\sum_{n\in\mathbb{N}}\|c_n\| < +\infty$, the assertions follow from Theorem 3.7, where we have used (2.20) to get (4.12). ☐

Example 4.7 We can obtain a soft-constrained version of the Potter-Arun problem (1.2) revisited in Example 4.3 by specializing Problem 4.5 as follows: $z = 0$, $\mathcal{G} = \mathbb{R}^N$, $D = \{0\}$, $r = (\rho_i)_{1\leq i\leq N}$, and $L\colon x\mapsto (\langle x \mid s_i\rangle)_{1\leq i\leq N}$, where $(s_i)_{1\leq i\leq N}\in\mathcal{H}^N$ satisfies $\sum_{i=1}^N\|s_i\|^2\leq 1$. We thus arrive at the relaxed version of (1.2)

$$\underset{x\in\mathcal{H}}{\operatorname{minimize}}\;\; \phi(d_C(x)) + \psi\Big(\sqrt{\textstyle\sum_{i=1}^N|\langle x\mid s_i\rangle - \rho_i|^2}\Big) + \frac{1}{2}\|x\|^2. \tag{4.14}$$

Since $D = \{0\}$, we can replace each occurrence of $d_D(w_n)$ by $\|w_n\|$ and each occurrence of $w_n - P_D w_n$ by w_n in (4.11). Proposition 4.6(ii) asserts that any sequence $(x_n)_{n\in\mathbb{N}}$ produced by the resulting algorithm converges strongly to the solution to (4.14). For the sake of illustration, let us consider the case when $\phi = \alpha|\cdot|^{4/3}$ and $\psi = \beta|\cdot|$, for some α and β in $]0,+\infty[$. Then $\operatorname{dom}\psi = \mathbb{R}$ and (4.8) is trivially satisfied. In addition, (4.14) becomes

$$\underset{x\in\mathcal{H}}{\operatorname{minimize}}\;\; \alpha d_C^{4/3}(x) + \beta\sqrt{\textstyle\sum_{i=1}^N|\langle x\mid s_i\rangle - \rho_i|^2} + \frac{1}{2}\|x\|^2. \tag{4.15}$$

Since $\phi^*\colon\mu\mapsto 27|\mu|^4/(256\alpha^3)$, $\operatorname{prox}_{\phi^*}$ in (4.11) can be derived from Example 2.15(vi). On the other hand, since $\psi^* = \iota_{[-\beta,\beta]}$, Example 2.6(i) yields $\operatorname{prox}_{\psi^*} = P_{[-\beta,\beta]}$. Thus, upon setting, for simplicity, $b_n\equiv 0$, $c_n\equiv 0$, $\lambda_n\equiv 1$, and $\gamma_n\equiv 1$ (note that $\|L\|\leq 1$) in (4.11) and observing that $\partial\phi(0) = \{0\}$

and $\partial\psi(0) = [-\beta, \beta]$, we obtain the following algorithm, where $L^*\colon (\nu_i)_{1\leq i\leq N} \mapsto \sum_{i=1}^N \nu_i s_i$.

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\tau = 3/(2\alpha 4^{1/3}),\ \sigma = 256\alpha^3/729\\
v_0 \in \mathbb{R}^N
\end{array}
\right.\\
\text{For } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
y_n = z - L^* v_n\\
\text{if } \ y_n \notin C\\
\left\lfloor
x_n = y_n + \dfrac{\left|\sqrt{d_C^2(y_n) + \sigma} + d_C(y_n)\right|^{1/3} - \left|\sqrt{d_C^2(y_n) + \sigma} - d_C(y_n)\right|^{1/3}}{\tau d_C(y_n)}(P_C y_n - y_n)
\right.\\
\text{if } \ y_n \in C\\
\left\lfloor x_n = y_n \right.\\
w_n = v_n + L x_n - r\\
\text{if } \ \|w_n\| > \beta\\
\left\lfloor v_{n+1} = \dfrac{\beta}{\|w_n\|} w_n \right.\\
\text{if } \ \|w_n\| \leq \beta\\
\left\lfloor v_{n+1} = w_n. \right.
\end{array}
\right.
\end{array}
$$

As shown above, the sequence $(x_n)_{n\in\mathbb{N}}$ converges strongly to the solution to (4.15).

Remark 4.8 Alternative relaxations of (1.2) can be derived from Problem 1.2. For instance, given an even function $\phi \in \Gamma_0(\mathbb{R}) \smallsetminus \{\iota_{\{0\}}\}$ and $\alpha \in \left]0, +\infty\right[$, an alternative to (4.14) is

$$\underset{x\in\mathcal{H}}{\text{minimize}} \ \ \phi(d_C(x)) + \alpha \max_{1\leq i\leq N} |\langle x \mid s_i\rangle - \rho_i| + \frac{1}{2}\|x\|^2. \tag{4.16}$$

This formulation results from (1.6) with $z = 0$, $f = \phi \circ d_C$, $\mathcal{G} = \mathbb{R}^N$, $r = (\rho_i)_{1\leq i\leq N}$, $L\colon x \mapsto (\langle x \mid s_i\rangle)_{1\leq i\leq N}$, and $g = \alpha\|\cdot\|_\infty$ (note that (1.5) holds since $\operatorname{dom} g = \mathcal{G}$). Since $g^* = \iota_D$, where $D = \big\{(\nu_i)_{1\leq i\leq N} \in \mathbb{R}^N \bigm| \sum_{i=1}^N |\nu_i| \leq \alpha\big\}$, the dual problem (3.2) therefore assumes the form

$$\underset{(\nu_i)_{1\leq i\leq N}\in D}{\text{minimize}} \ \ \frac{1}{2}\bigg\|\sum_{i=1}^N \nu_i s_i\bigg\|^2 - (\phi\circ d_C)^{\sim}\bigg(-\sum_{i=1}^N \nu_i s_i\bigg) + \sum_{i=1}^N \rho_i\nu_i. \tag{4.17}$$

The proximity operators of $f = \phi \circ d_C$ and $\gamma_n g^* = \iota_D$ required by Algorithm 3.5 are supplied by Example 2.10 and Example 2.6(i), respectively. Strong convergence of the resulting sequence $(x_n)_{n\in\mathbb{N}}$ to the solution to (4.16) is guaranteed by Theorem 3.7(ii).

4.3 Denoising over dictionaries

In denoising problems, the goal is to recover the original form of an ideal signal $\overline{x} \in \mathcal{H}$ from a corrupted observation

$$z = \overline{x} + s, \tag{4.18}$$

where $s \in \mathcal{H}$ is the realization of a noise process which may for instance model imperfections in the data recording instruments, uncontrolled dynamics, or physical interferences. A common approach to solve this problem is to minimize the least-squares data fitting functional $x \mapsto \|x-z\|^2/2$ subject to some constraints on x that represent a priori knowledge on the ideal solution $\overline{x}$ and some affine transformation $L\overline{x}-r$ thereof, where $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ and $r \in \mathcal{G}$. By measuring the degree of violation of these constraints via potentials $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$, we arrive at (1.6). In this context, L can be a gradient [22, 41, 50, 68], a low-pass filter [2, 73], a wavelet or a frame decomposition operator [32, 40, 74]. Alternatively, the vector $r \in \mathcal{G}$ may arise from the availability of a second observation in the form of a noise-corrupted linear measurement of $\overline{x}$, as in (4.7) [25].

In this section, the focus is placed on models in which information on the scalar products $(\langle \overline{x} \mid e_k\rangle)_{k\in\mathbb{K}}$ of the original signal $\overline{x}$ against a finite or infinite a sequence of reference unit norm vectors $(e_k)_{k\in\mathbb{K}}$ of $\mathcal{H}$, called a dictionary, is available. In practice, such information can take various forms, e.g., sparsity, distribution type, statistical properties [25, 31, 37, 45, 53, 71], and they can often be modeled in a variational framework by introducing a sequence of convex potentials $(\phi_k)_{k\in\mathbb{K}}$. If we model the rest of the information available about $\overline{x}$ via a potential f, we obtain the following formulation.

Problem 4.9 Let $z \in \mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$, let $(e_k)_{k\in\mathbb{K}}$ be a sequence of unit norm vectors in $\mathcal{H}$ such that

$$(\exists\, \delta \in \left]0,+\infty\right[)(\forall x \in \mathcal{H}) \quad \sum_{k\in\mathbb{K}} |\langle x \mid e_k\rangle|^2 \leq \delta\|x\|^2, \tag{4.19}$$

and let $(\phi_k)_{k\in\mathbb{K}}$ be functions in $\Gamma_0(\mathbb{R})$ such that

$$(\forall k \in \mathbb{K}) \quad \phi_k \geq \phi_k(0) = 0 \tag{4.20}$$

and

$$0 \in \operatorname{sri}\left\{ \big(\langle x \mid e_k\rangle - \xi_k\big)_{k\in\mathbb{K}} \;\middle|\; (\xi_k)_{k\in\mathbb{K}} \in \ell^2(\mathbb{K}),\ \sum_{k\in\mathbb{K}} \phi_k(\xi_k) < +\infty,\ \text{ and } x \in \operatorname{dom} f \right\}. \tag{4.21}$$

The problem is to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle) + \frac{1}{2}\|x-z\|^2, \tag{4.22}$$

and its dual is to

$$\underset{(\nu_k)_{k\in\mathbb{K}}\in\ell^2(\mathbb{K})}{\text{minimize}} \;\; \widetilde{f^*}\Big(z - \sum_{k\in\mathbb{K}} \nu_{n,k} e_k\Big) + \sum_{k\in\mathbb{K}} \phi_k^*(\nu_k). \tag{4.23}$$

Problems (4.22) and (4.23) can be solved by the following algorithm, where $\alpha_{n,k}$ stands for a numerical tolerance in the implementation of the operator $\operatorname{prox}_{\gamma_n\phi_k^*}$. Let us note that closed-form expressions for the proximity operators of a wide range of functions in $\Gamma_0(\mathbb{R})$ are available [25, 31, 35], in particular in connection with Bayesian formulations involving log-concave densities, and with problems involving sparse representations (see also Examples 2.15–2.18 and Lemmas 2.19–2.20).

Proposition 4.10 *Let $((\alpha_{n,k})_{n\in\mathbb{N}})_{k\in\mathbb{K}}$ be sequences in $\mathbb{R}$ such that $\sum_{n\in\mathbb{N}} \sqrt{\sum_{k\in\mathbb{K}} |\alpha_{n,k}|^2} < +\infty$, let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$, and let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ =*

$((\nu_{n,k})_{k\in\mathbb{K}})_{n\in\mathbb{N}}$ *be sequences generated by the following routine.*

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\varepsilon \in \left]0, \min\{1, \delta^{-1}\}\right[\\
(\nu_{0,k})_{k\in\mathbb{K}} \in \ell^2(\mathbb{K})
\end{array}
\right.\\
\text{For } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
x_n = \operatorname{prox}_f \big(z - \sum_{k\in\mathbb{K}} \nu_{n,k} e_k\big) + b_n\\
\gamma_n \in \left[\varepsilon, 2\delta^{-1} - \varepsilon\right]\\
\lambda_n \in [\varepsilon, 1]\\
\text{For every } k \in \mathbb{K}\\
\left\lfloor
\begin{array}{l}
\nu_{n+1,k} = \nu_{n,k} + \lambda_n \big(\operatorname{prox}_{\gamma_n \phi_k^*}(\nu_{n,k} + \gamma_n \langle x_n \mid e_k\rangle) + \alpha_{n,k} - \nu_{n,k}\big).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{4.24}
$$

Then the following hold, where x designates the primal solution to Problem 4.9.

(i) $(v_n)_{n\in\mathbb{N}}$ *converges weakly to a solution* $(\nu_k)_{k\in\mathbb{K}}$ *to* (4.23) *and* $x = \operatorname{prox}_f(z - \sum_{k\in\mathbb{K}} \nu_k e_k)$.

(ii) $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* x.

Proof. Set $\mathcal{G} = \ell^2(\mathbb{K})$ and $r = 0$. Define

$$
L\colon \mathcal{H} \to \mathcal{G}\colon x \mapsto (\langle x \mid e_k\rangle)_{k\in\mathbb{K}} \quad \text{and} \quad g\colon \mathcal{G} \to \left]-\infty, +\infty\right] : (\xi_k)_{k\in\mathbb{K}} \mapsto \sum_{k\in\mathbb{K}} \phi_k(\xi_k). \tag{4.25}
$$

Then $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ and its adjoint is the operator $L^* \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ defined by

$$
L^*\colon (\xi_k)_{k\in\mathbb{K}} \mapsto \sum_{k\in\mathbb{K}} \xi_k e_k. \tag{4.26}
$$

On the other hand, it follows from our assumptions that $g \in \Gamma_0(\mathcal{G})$ (Example 2.9) and that

$$
g^*\colon \mathcal{G} \to \left]-\infty, +\infty\right] : (\nu_k)_{k\in\mathbb{K}} \mapsto \sum_{k\in\mathbb{K}} \phi_k^*(\nu_k). \tag{4.27}
$$

In addition, (4.21) implies that (1.5) holds. This shows that (4.22) is a special case of (1.6) and that (4.23) is a special case of (3.2). We also observe that (4.19) and (4.25) yield

$$
\|L\|^2 = \sup_{\|x\|=1} \|Lx\|^2 = \sup_{\|x\|=1} \sum_{k\in\mathbb{K}} |\langle x \mid e_k\rangle|^2 \leq \delta. \tag{4.28}
$$

Hence, $\left[\varepsilon, 2\delta^{-1} - \varepsilon\right] \subset \left[\varepsilon, 2\|L\|^{-2} - \varepsilon\right]$. Next, we derive from (2.8) and (4.20) that, for every $k \in \mathbb{K}$, $\phi_k^*(0) = \sup_{\xi\in\mathbb{R}} -\phi_k(\xi) = -\inf_{\xi\in\mathbb{R}} \phi_k(\xi) = \phi_k(0) = 0$ and that $(\forall \nu \in \mathbb{R})$ $\phi_k^*(\nu) = \sup_{\xi\in\mathbb{R}} \xi\nu - \phi_k(\xi) \geq -\phi_k(0) = 0$. In turn, we derive from (4.27) and Example 2.9 (applied to the canonical orthonormal basis of $\ell^2(\mathbb{K})$) that

$$
(\forall \gamma \in \left]0, +\infty\right[)(\forall v = (\nu_k)_{k\in\mathbb{K}} \in \mathcal{G}) \quad \operatorname{prox}_{\gamma g^*} v = \big(\operatorname{prox}_{\gamma\phi_k^*} \nu_k\big)_{k\in\mathbb{K}}. \tag{4.29}
$$

Altogether, (4.24) is a special case of Algorithm 3.5 with $(\forall n \in \mathbb{N})$ $a_n = (\alpha_{n,k})_{k\in\mathbb{K}}$. Hence, the assertions follow from Theorem 3.7. ◻

Remark 4.11 Using (4.25), we can write the potential on the dictionary coefficients in Problem 4.9 as

$$g \circ L \colon x \mapsto \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle). \tag{4.30}$$

(i) If $(e_k)_{k\in\mathbb{K}}$ were an orthonormal basis in Problem 4.9, we would have $L^{-1} = L^*$ and $\operatorname{prox}_{g\circ L}$ would be decomposable as $L^* \circ \operatorname{prox}_g \circ L$ [35, Lemma 2.8]. As seen in the Introduction, we could then approach (4.22) directly via forward-backward, Douglas-Rachford, or Dykstra-like splitting, depending on the properties of f. Our duality framework allows us to solve (4.22) for the much broader class of dictionaries satisfying (4.19) and, in particular, for frames [36].

(ii) Suppose that each ϕ_k in Problem 4.9 is of the form $\phi_k = \psi_k + \sigma_{\Omega_k}$, where $\psi_k \in \Gamma_0(\mathbb{R})$ satisfies $\psi_k \geq \psi_k(0) = 0$ and is differentiable at 0 with $\psi_k'(0) = 0$, and where Ω_k is a nonempty closed interval. In this case, (4.30) aims at promoting the sparsity of the solution in the dictionary $(e_k)_{k\in\mathbb{K}}$ [31] (a standard case is when, for every $k \in \mathbb{K}$, $\psi_k = 0$ and $\Omega_k = [-\omega_k, \omega_k]$, which gives rise to the standard weighted ℓ^1 potential $x \mapsto \sum_{k\in\mathbb{K}} \omega_k|\langle x \mid e_k\rangle|$). Moreover, the proximity operator $\operatorname{prox}_{\gamma_n\phi_k^*}$ in (4.24) can be evaluated via Lemma 2.4 and Lemma 2.19.

4.4 Denoising with support functions

Suppose that g in Problem 1.2 is positively homogeneous, i.e.,

$$(\forall \lambda \in \left]0, +\infty\right[)(\forall y \in \mathcal{G}) \quad g(\lambda y) = \lambda g(y). \tag{4.31}$$

Instances of such functions arising in denoising problems can be found in [1, 7, 8, 23, 31, 35, 38, 61, 68, 75] and in the examples below. It follows from (4.31) and [4, Theorem 2.4.2] that g is the support function of a nonempty closed convex set $D \subset \mathcal{G}$, namely

$$g = \sigma_D = \sup_{v\in D} \langle \cdot \mid v\rangle, \quad \text{where} \quad D = \partial g(0) = \big\{v \in \mathcal{G} \bigm| (\forall y \in \mathcal{G}) \;\; \langle y \mid v\rangle \leq g(y)\big\}. \tag{4.32}$$

If we denote by $\operatorname{bar} D = \big\{y \in \mathcal{G} \bigm| \sup_{v\in D} \langle y \mid v\rangle < +\infty\big\}$ the barrier cone of D, we thus obtain the following instance of Problem 1.2.

Problem 4.12 Let $z \in \mathcal{H}$, $r \in \mathcal{G}$, let $f \in \Gamma_0(\mathcal{H})$, let D be a nonempty closed convex subset of $\mathcal{G}$, and let L be a nonzero operator in $\mathcal{B}(\mathcal{H}, \mathcal{G})$ such that

$$r \in \operatorname{sri}\big(L(\operatorname{dom} f) - \operatorname{bar} D\big). \tag{4.33}$$

The problem is to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sigma_D(Lx - r) + \frac{1}{2}\|x - z\|^2, \tag{4.34}$$

and its dual is to

$$\underset{v\in D}{\text{minimize}} \;\; \widetilde{f^*}(z - L^*v) + \langle v \mid r\rangle. \tag{4.35}$$

Proposition 4.13 *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}$ such that $\sum_{n\in\mathbb{N}}\|a_n\| < +\infty$, let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|b_n\| < +\infty$, and let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences generated by the following routine.*

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\varepsilon \in \left]0, \min\{1, \|L\|^{-2}\}\right[\\
v_0 \in \mathcal{G}
\end{array}
\right.\\
\text{For } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
x_n = \operatorname{prox}_f(z - L^* v_n) + b_n\\
\gamma_n \in \left[\varepsilon, 2\|L\|^{-2} - \varepsilon\right]\\
\lambda_n \in [\varepsilon, 1]\\
v_{n+1} = v_n + \lambda_n\big(P_D(v_n + \gamma_n(Lx_n - r)) + a_n - v_n\big).
\end{array}
\right.
\end{array}
\tag{4.36}
$$

Then the following hold, where x designates the primal solution to Problem 4.12.

(i) *$(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to (4.35) and $x = \operatorname{prox}_f(z - L^* v)$.*

(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to x.*

Proof. The assertions follow from Theorem 3.7 with $g = \sigma_D$. Indeed, $g^* = \iota_D$ and, therefore, $(\forall \gamma \in \left]0, +\infty\right[)\ \operatorname{prox}_{\gamma g^*} = P_D$. □

Remark 4.14 Condition (4.33) is trivially satisfied when D is bounded, in which case bar $D = \mathcal{G}$.

In the remainder of this section, we focus on examples that feature a bounded set D onto which projections are easily computed.

Example 4.15 In Problem 4.12, let D be the closed unit ball of $\mathcal{G}$. Then $P_D : y \mapsto y/\max\{\|y\|, 1\}$ and $\sigma_D = \|\cdot\|$. Hence, (4.34) becomes

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ f(x) + \|Lx - r\| + \frac{1}{2}\|x - z\|^2, \tag{4.37}$$

and the dual problem (4.35) becomes

$$\underset{v\in\mathcal{G},\, \|v\|\leq 1}{\text{minimize}}\ \ \widetilde{f^*}(z - L^* v) + \langle v \mid r\rangle. \tag{4.38}$$

In signal recovery, variational formulations involving positively homogeneous functionals to control the behavior of the gradient of the solutions play a prominent role, e.g., [3, 13, 48, 61, 68]. In the context of image recovery, such a formulation can be obtained by revisiting Problem 4.12 with $\mathcal{H} = H_0^1(\Omega)$, where Ω is a bounded open domain in $\mathbb{R}^2$, $\mathcal{G} = L^2(\Omega) \oplus L^2(\Omega)$, $L = \nabla$, $D = \big\{y \in \mathcal{G} \;\big|\; |y|_2 \leq \mu \text{ a.e.}\big\}$ where $\mu \in \left]0, +\infty\right[$, and $r = 0$. With this scenario, (4.34) is equivalent to

$$\underset{x\in H_0^1(\Omega)}{\text{minimize}}\ \ f(x) + \mu\,\mathrm{tv}(x) + \frac{1}{2}\|x - z\|^2, \tag{4.39}$$

where $\mathrm{tv}(x) = \int_\Omega |\nabla x(\omega)|_2 d\omega$. In mechanics, such minimization problems have been studied extensively for certain potentials f [42]. For instance, $f = 0$ yields Mossolov's problem and its dual analysis is carried out in [42, Section IV.3.1]. In image processing, Mossolov's problem corresponds to the total variation denoising problem. Interestingly, in 1980, Mercier [55] proposed a dual projection algorithm to solve Mossolov's problem. This approach was independently rediscovered by Chambolle in a discrete setting [22, 23]. Next, we apply our framework to a discrete version of (4.39) for $N \times N$ images. This will extend the method of [23], which is restricted to $f = 0$, and provide a formal proof for its convergence (see also [75] for an alternative scheme based on Nesterov's algorithm [60]).

By way of preamble, let us introduce some notation. We denote by $y = \big(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l}\big)_{1\le k,l\le N}$ a generic element in $\mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N}$ and by

$$\nabla\colon \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N}\colon \big(\xi_{k,l}\big)_{1\le k,l\le N} \mapsto \big(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l}\big)_{1\le k,l\le N} \tag{4.40}$$

the discrete gradient operator, where

$$(\forall (k,l) \in \{1,\ldots,N\}^2) \quad \begin{cases} \eta^{(1)}_{k,l} = \xi_{k+1,l} - \xi_{k,l}, & \text{if } \ k < N;\\ \eta^{(1)}_{N,l} = 0;\\ \eta^{(2)}_{k,l} = \xi_{k,l+1} - \xi_{k,l}, & \text{if } \ l < N;\\ \eta^{(2)}_{k,N} = 0. \end{cases} \tag{4.41}$$

Now let $p \in [1,+\infty]$. Then p^* is the conjugate index of p, i.e., $p^* = +\infty$ if $p = 1$, $p^* = 1$ if $p = +\infty$, and $p^* = p/(p-1)$ otherwise. We define the p-th order discrete total variation function as

$$\operatorname{tv}_p\colon \mathbb{R}^{N\times N} \to \mathbb{R}\colon x \mapsto ||\nabla x||_{p,1}\,, \tag{4.42}$$

where

$$(\forall y \in \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N}) \quad \|y\|_{p,1} = \sum_{1\le k,l\le N} \big|(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l})\big|_p, \tag{4.43}$$

with

$$\big(\forall (\eta^{(1)}, \eta^{(2)}) \in \mathbb{R}^2\big) \quad \big|(\eta^{(1)}, \eta^{(2)})\big|_p = \begin{cases} \sqrt[p]{|\eta^{(1)}|^p + |\eta^{(2)}|^p}, & \text{if } \ p < +\infty;\\ \max\big\{|\eta^{(1)}|, |\eta^{(2)}|\big\}, & \text{if } \ p = +\infty. \end{cases} \tag{4.44}$$

In addition, the discrete divergence operator is defined as [22]

$$\operatorname{div}\colon \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}\colon \big(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l}\big)_{1\le k,l\le N} \mapsto \big(\xi^{(1)}_{k,l} + \xi^{(2)}_{k,l}\big)_{1\le k,l\le N}, \tag{4.45}$$

where

$$\xi^{(1)}_{k,l} = \begin{cases} \eta^{(1)}_{1,l} & \text{if } \ k = 1;\\ \eta^{(1)}_{k,l} - \eta^{(1)}_{k-1,l} & \text{if } \ 1 < k < N;\\ -\eta^{(1)}_{N-1,l} & \text{if } \ k = N; \end{cases} \quad \text{and} \quad \xi^{(2)}_{k,l} = \begin{cases} \eta^{(2)}_{k,1} & \text{if } \ l = 1;\\ \eta^{(2)}_{k,l} - \eta^{(2)}_{k,l-1} & \text{if } \ 1 < l < N;\\ -\eta^{(2)}_{k,N-1} & \text{if } \ l = N. \end{cases} \tag{4.46}$$

Problem 4.16 Let $z \in \mathbb{R}^{N\times N}$, let $f \in \Gamma_0(\mathbb{R}^{N\times N})$, let $\mu \in \left]0,+\infty\right[$, let $p \in [1,+\infty]$, and set

$$D_p = \Big\{ \big(\nu^{(1)}_{k,l}, \nu^{(2)}_{k,l}\big)_{1\le k,l\le N} \in \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N} \;\Big|\; \max_{1\le k,l\le N} \big|(\nu^{(1)}_{k,l}, \nu^{(2)}_{k,l})\big|_{p^*} \le 1 \Big\}. \tag{4.47}$$

The problem is to

$$\underset{x\in\mathbb{R}^{N\times N}}{\text{minimize}} \;\; f(x) + \mu\, \operatorname{tv}_p(x) + \frac{1}{2}\|x - z\|^2, \tag{4.48}$$

and its dual is to

$$\underset{v\in D_p}{\text{minimize}}\ \widetilde{f^*}(z+\mu\,\mathrm{div}\,v). \tag{4.49}$$

Proposition 4.17 *Let $(\alpha^{(1)}_{n,k,l})_{n\in\mathbb{N}}$ and $(\alpha^{(2)}_{n,k,l})_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}^{N\times N}$ such that*

$$\sum_{n\in\mathbb{N}}\sqrt{\sum_{1\le k,l\le N}\left|\alpha^{(1)}_{n,k,l}\right|^2+\left|\alpha^{(2)}_{n,k,l}\right|^2}<+\infty, \tag{4.50}$$

let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{N\times N}$ such that $\sum_{n\in\mathbb{N}}\|b_n\|<+\infty$, and let $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences generated by the following routine, where $(\pi_p^{(1)}\mathsf{y},\pi_p^{(2)}\mathsf{y})$ denotes the projection of a point $\mathsf{y}\in\mathbb{R}^2$ onto the closed unit ℓ^{p^*} ball in the Euclidean plane.

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor\begin{array}{l}
\varepsilon\in\left]0,\min\{1,\mu^{-1}/8\}\right[\\
v_0=(\nu^{(1)}_{0,k,l},\nu^{(2)}_{0,k,l})_{1\le k,l\le N}\in\mathbb{R}^{N\times N}\oplus\mathbb{R}^{N\times N}
\end{array}\right.\\
\text{For } n=0,1,\ldots\\
\left\lfloor\begin{array}{l}
x_n=\mathrm{prox}_f(z+\mu\,\mathrm{div}\,v_n)+b_n\\
\tau_n\in\left[\varepsilon,\mu^{-1}/4-\varepsilon\right]\\
(\zeta^{(1)}_{n,k,l},\zeta^{(2)}_{n,k,l})_{1\le k,l\le N}=v_n+\tau_n\nabla x_n\\
\lambda_n\in[\varepsilon,1]\\
\text{For every } (k,l)\in\{1,\ldots,N\}^2\\
\left\lfloor\begin{array}{l}
\nu^{(1)}_{n+1,k,l}=\nu^{(1)}_{n,k,l}+\lambda_n\Big(\pi_p^{(1)}(\zeta^{(1)}_{n,k,l},\zeta^{(2)}_{n,k,l})+\alpha^{(1)}_{n,k,l}-\nu^{(1)}_{n,k,l}\Big)\\
\nu^{(2)}_{n+1,k,l}=\nu^{(2)}_{n,k,l}+\lambda_n\Big(\pi_p^{(2)}(\zeta^{(1)}_{n,k,l},\zeta^{(2)}_{n,k,l})+\alpha^{(2)}_{n,k,l}-\nu^{(2)}_{n,k,l}\Big)
\end{array}\right.\\
v_{n+1}=(\nu^{(1)}_{n+1,k,l},\nu^{(2)}_{n+1,k,l})_{1\le k,l\le N}
\end{array}\right.
\end{array} \tag{4.51}$$

Then $(v_n)_{n\in\mathbb{N}}$ converges to a solution v to (4.49), $x=\mathrm{prox}_f(z+\mu\,\mathrm{div}\,v)$ is the primal solution to Problem 4.16, and $x_n\to x$.

Proof. It follows from (4.43) and (4.47) that $\|\cdot\|_{p,1}=\sigma_{D_p}$. Hence, Problem 4.16 is a special case of Problem 4.12 with $\mathcal{H}=\mathbb{R}^{N\times N}$, $\mathcal{G}=\mathbb{R}^{N\times N}\oplus\mathbb{R}^{N\times N}$, $L=\mu\nabla$ (see (4.40)), $D=D_p$, and $r=0$. Moreover, $L^*=-\mu\,\mathrm{div}$ (see (4.45)), $\|L\|=\mu\|\nabla\|\le 2\sqrt{2}\mu$ [22], and the projection of y onto the set D_p of (4.47) can be decomposed coordinatewise as

$$P_{D_p}y=\Big(\pi_p^{(1)}(\eta^{(1)}_{k,l},\eta^{(2)}_{k,l}),\pi_p^{(2)}(\eta^{(1)}_{k,l},\eta^{(2)}_{k,l})\Big)_{1\le k,l\le N}. \tag{4.52}$$

Altogether, upon setting, for every $n\in\mathbb{N}$, $\tau_n=\mu\gamma_n$ and $a_n=(\alpha^{(1)}_{n,k,l},\alpha^{(2)}_{n,k,l})_{1\le k,l\le N}$, (4.51) appears as a special case of (4.36). The results therefore follow from (4.50) and Proposition 4.13. ☐

Remark 4.18 The inner loop in (4.51) performs the projection step. For certain values of p, this projection can be computed explicitly and we can therefore dispense with errors. Thus, if $p=1$, then $p^* = +\infty$ and the projection loop becomes

$$\begin{array}{l} \text{For every } (k,l) \in \{1,\ldots,N\}^2 \\ \left\lfloor \begin{array}{l} \nu^{(1)}_{n+1,k,l} = \nu^{(1)}_{n,k,l} + \lambda_n \left(\dfrac{\zeta^{(1)}_{n,k,l}}{\max\left\{1, \left|\zeta^{(1)}_{n,k,l}\right|\right\}} - \nu^{(1)}_{n,k,l} \right) \\ \nu^{(2)}_{n+1,k,l} = \nu^{(2)}_{n,k,l} + \lambda_n \left(\dfrac{\zeta^{(2)}_{n,k,l}}{\max\left\{1, \left|\zeta^{(2)}_{n,k,l}\right|\right\}} - \nu^{(2)}_{n,k,l} \right). \end{array} \right. \end{array} \tag{4.53}$$

Likewise, if $p = 2$, then $p^* = 2$ and the projection loop becomes

$$\begin{array}{l} \text{For every } (k,l) \in \{1,\ldots,N\}^2 \\ \left\lfloor \begin{array}{l} \nu^{(1)}_{n+1,k,l} = \nu^{(1)}_{n,k,l} + \lambda_n \left(\dfrac{\zeta^{(1)}_{n,k,l}}{\max\left\{1, \left|\left(\zeta^{(1)}_{n,k,l}, \zeta^{(2)}_{n,k,l}\right)\right|_2\right\}} - \nu^{(1)}_{n,k,l} \right) \\ \nu^{(2)}_{n+1,k,l} = \nu^{(2)}_{n,k,l} + \lambda_n \left(\dfrac{\zeta^{(2)}_{n,k,l}}{\max\left\{1, \left|\left(\zeta^{(1)}_{n,k,l}, \zeta^{(2)}_{n,k,l}\right)\right|_2\right\}} - \nu^{(2)}_{n,k,l} \right). \end{array} \right. \end{array} \tag{4.54}$$

In the special case when $f = 0$, $\lambda_n \equiv 1$, and $\tau_n \equiv \tau \in \left]0, \mu^{-1}/4\right[$ the two resulting algorithms reduce to the popular methods proposed in [23]. Finally, if $p = +\infty$, then $p^* = 1$ and the efficient scheme described in [9] to project onto the ℓ^1 ball can be used.

References

[1] M. Amar and G. Bellettini, A notion of total variation depending on a metric with discontinuous coefficients, *Ann. Inst. H. Poincaré Anal. Non Linéaire,* vol. 11, pp. 91–133, 1994.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration,* Prentice-Hall, Englewood Cliffs, NJ, 1977.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing,* 2nd ed., Springer-Verlag, New York, 2006.

[4] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis,* Birkhäuser, Boston, MA, 1990.

[5] H. H. Bauschke and P. L. Combettes, A Dykstra-like algorithm for two monotone operators, *Pacific J. Optim.,* vol. 4, pp. 383–391, 2008.

[6] H. H. Bauschke and P. L. Combettes, The Baillon-Haddad theorem revisited, *J. Convex Anal.,* vol. 17, 2010.

[7] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.,* vol. 2, pp. 183–202, 2009.

[8] 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.

[9] E. van den Berg and M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.,* vol. 31, 890–912, 2008.

[10] A. Ben-Tal, J. M. Borwein, and M. Teboulle, A dual approach to multidimensional L_p spectral estimation problems, *SIAM J. Control Optim.,* vol. 26, pp. 985–996, 1988.

[11] M. Bertero, C. De Mol, and E. R. Pike, Linear inverse problems with discrete data I – General formulation and singular system analysis, *Inverse Problems,* vol. 1, pp. 301–330, 1985.

[12] 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.

[13] J. M. Borwein, A. S. Lewis, and D. Noll, Maximum entropy reconstruction using derivative information. I: Fisher information and convex duality, *Math. Oper. Res.,* vol. 21, pp. 442–468, 1996.

[14] J. M. Borwein and D. R. Luke, Duality and convex programming, in: *Handbook of Imaging,* O. Scherzer Ed. Springer-Verlag, New York, to appear.

[15] K. Bredies and D. A. Lorenz, Linear convergence of iterative soft-thresholding, *J. Fourier Anal. Appl.,* vol. 14, pp. 813–837, 2008.

[16] L. M. Briceño-Arias and P. L. Combettes, Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery, *Numer. Math. Theory Methods Appl.,* vol. 2, pp. 485–508, 2009.

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

[18] J.-F. Cai, R. H. Chan, L. Shen, and Z. Shen, Convergence analysis of tight framelet approach for missing data recovery, *Adv. Comput. Math.,* vol. 31, 2009.

[19] J.-F. Cai, R. H. Chan, and Z. Shen, A framelet-based image inpainting algorithm, *Appl. Comput. Harm. Anal.,* vol. 24, pp. 131–149, 2008.

[20] Y. Censor and T. Elfving, A multiprojection algorithm using Bregman projections in a product space, *Numer. Algorithms,* vol. 8, pp. 221–239, 1994.

[21] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications,* Oxford University Press, New York, 1997.

[22] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vision,* vol. 20, pp. 89–97, 2004.

[23] A. Chambolle, Total variation minimization and a class of binary MRF model, *Lecture Notes in Comput. Sci.* 3757, pp 136–152, 2005.

[24] T. F. Chan, G. H. Golub, and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, *SIAM J. Sci. Comput.,* vol. 20, 1964–1977, 1999.

[25] 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.

[26] C. Chaux, J.-C. Pesquet, and N. Pustelnik, Nested iterative algorithms for convex contrained image recovery problems, *SIAM J. Imaging Sci.,* vol. 2, pp. 730–762, 2009.

[27] P. L. Combettes, Signal recovery by best feasible approximation, *IEEE Trans. Image Process.,* vol. 2, pp. 269–271, 1993.

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

[29] 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.

[30] P. L. Combettes, Đinh Dũng, and B. C. Vũ, Dualization of signal recovery problems, preprint, July 2, 2009. http://arxiv.org/abs/0907.0436.

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

[32] 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.

[33] P. L. Combettes and J.-C. Pesquet, A proximal decomposition method for solving convex variational inverse problems, *Inverse Problems,* vol. 24, Art. 065014, 27 pp., 2008.

[34] P. L. Combettes and H. J. Trussell, The use of noise properties in set theoretic estimation, *IEEE Trans. Signal Process.,* vol. 39, pp. 1630–1641, 1991.

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

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

[37] 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.

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

[39] P. Destuynder, M. Jaoua, and H. Sellami, A dual algorithm for denoising and preserving edges in image processing, *J. Inverse Ill-Posed Probl.,* vol. 15, pp. 149–165, 2007.

[40] D. Donoho and I. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika,* vol. 81, pp. 425–455, 1994.

[41] S. Durand and M. Nikolova, Denoising of frame coefficients using ℓ^1 data-fidelity term and edge-preserving regularization, *Multiscale Model. Simul.,* vol. 6, pp. 547–576, 2007.

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

[43] J. Fadili and G. Peyré, Total variation projection with first order schemes, preprint (2009). http://hal.archives-ouvertes.fr/hal-00380491

[44] W. Fenchel, *Convex Cones, Sets and Functions,* Lecture notes (mimeograph), Princeton University, 1953.

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

[46] R. W. Gerchberg, Super-resolution through error energy reduction, *Optica Acta,* vol. 21, pp. 709–720, 1974.

[47] M. Hintermüller and G. Stadler, An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, *SIAM J. Sci. Comput.,* vol. 28, pp. 1–23, 2006.

[48] Y. Huang, M. K. Ng, and Y.-W. Wen, A fast total variation minimization method for image restoration, *Multiscale Model. Simul.,* vol. 7, pp. 774–795, 2008.

[49] A. N. Iusem and M. Teboulle, A regularized dual-based iterative method for a class of image reconstruction problems, *Inverse Problems,* vol. 9, pp. 679–696, 1993.

[50] T. Kärkkäinen, K. Majava, and M. M. Mäkelä, Comparison of formulations and solution methods for image restoration problems, *Inverse Problems,* vol. 17, pp. 1977–1995, 2001.

[51] R. M. Leahy and C. E. Goutis, An optimal technique for constraint-based image restoration and reconstruction, *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 34, pp. 1629–1642, 1986.

[52] L. Levi, Fitting a bandlimited signal to given points, *IEEE Trans. Inform. Theory,* vol. 11, pp. 372–376, 1965.

[53] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, New York, 1999.

[54] B. P. Medoff, Image reconstruction from limited data: Theory and applications in computerized tomography, in: H. Stark (Ed.), *Image Recovery: Theory and Application,* pp. 321–368. Academic Press, San Diego, CA, 1987.

[55] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01). Orsay, France, Université de Paris-XI, 1980.

[56] J.-J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, 1962.

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

[58] J.-J. Moreau, *Fonctionnelles Convexes,* Séminaire sur les Équations aux Dérivées Partielles II, Collège de France, Paris, 1966–1967.

[59] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization,* Wiley, New York, 1983.

[60] Yu. Nesterov, Smooth minimization of non-smooth functions, *Math. Program.,* vol. 103, pp. 127–152, 2005.

[61] D. Noll, Reconstruction with noisy data: An approach via eigenvalue optimization, *SIAM J. Optim.,* vol. 8, pp. 82–104, 1998.

[62] A. Papoulis, A new algorithm in spectral analysis and band-limited extrapolation, *IEEE Trans. Circuits Syst.,* vol. 22, pp. 735–742, 1975.

[63] E. Polak, *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York, 1971.

[64] L. C. Potter and K. S. Arun, A dual approach to linear inverse problems with convex constraints, *SIAM J. Control Optim.,* vol. 31, pp. 1080–1092, 1993.

[65] R. T. Rockafellar, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.,* vol. 21, pp. 167–187, 1967.

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

[67] R. T. Rockafellar, *Conjugate Duality and Optimization,* SIAM, Philadelphia, PA, 1974.

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

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

[70] A. O. Steinhardt, R. K. Goodrich, and R. A. Roberts, Spectral estimation via minimum energy correlation extension, *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 33, pp. 1509–1515, 1985.

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

[72] H. J. Trussell and M. R. Civanlar, The feasible solution in signal restoration, *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 32, pp. 201–212, 1984.

[73] S. Twomey, The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements, *J. Franklin Inst.,* vol. 279, pp. 95–109, 1965.

[74] J. B. Weaver, Y. Xu, D. M. Healy Jr., and L. D. Cromwell, Filtering noise from images with wavelet transforms, *Magnetic Resonance in Medicine,* vol. 21, pp. 288–295, 1991.

[75] P. Weiss, G. Aubert, and L. Blanc-Féraud, Efficient schemes for total variation minimization under constraints in image processing, *SIAM J. Sci. Comput.*, vol. 31, pp. 2047–2080, 2009.

[76] D. C. Youla, Generalized image restoration by the method of alternating orthogonal projections, *IEEE Trans. Circuits Syst.,* vol. 25, pp. 694–702, 1978.

[77] D. C. Youla and V. Velasco, Extensions of a result on the synthesis of signals in the presence of inconsistent constraints, *IEEE Trans. Circuits Syst.,* vol. 33, pp. 465–468, 1986.

[78] 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.

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