

Range of the displacement operator of PDHG with applications to quadratic and conic programming

Tao Jiang

Walaa M. Moursi

Stephen A. Vavasis

September 27, 2023

Abstract

Primal-dual hybrid gradient (PDHG) is a first-order method for saddle-point problems and convex programming introduced by Chambolle and Pock. Recently, Applegate et al. analyzed the behavior of PDHG when applied to an infeasible or unbounded instance of linear programming, and in particular, showed that PDHG is able to diagnose these conditions. Their analysis hinges on the notion of the infimal displacement vector in the closure of the range of the displacement mapping of the splitting operator that encodes the PDHG algorithm. In this paper, we develop a novel formula for this range using monotone operator theory. The analysis is then specialized to conic programming and further to quadratic programming (QP) and second-order cone programming (SOCP). A consequence of our analysis is that PDHG is able to diagnose infeasible or unbounded instances of QP and of the ellipsoid-separation problem, a subclass of SOCP.

2010 Mathematics Subject Classification: 49M27, 65K05, 65K10, 90C25; Secondary 47H14, 49M29, 49N15.

Keywords: Chambolle–Pock algorithm, convex optimization problem, inconsistent constrained optimization, primal-dual hybrid gradient method, projection operator, proximal mapping, second-order cone programming, quadratic programming.

1 Introduction

“First-order” methods for convex programming use matrix-vector multiplication as their principal operation. For huge-scale convex programming problems, first-order methods appear to be the only tractable approach. In a recent survey, Lu [21] found that, among first-order methods, primal-dual hybrid gradient (PDHG) appears to be the best in practice for linear programming (LP). PDHG was introduced by Chambolle and Pock [18]. It has been shown by O’Connor and Vandenberghe [24] that PDHG may be regarded as a particular form of the Douglas-Rachford iteration. In the following, we assume that

$$X \text{ and } Y \text{ are real Hilbert spaces} \tag{1}$$

with corresponding inner products $\langle \cdot, \cdot \rangle_X$ (respectively $\langle \cdot, \cdot \rangle_Y$) and induced norms $\|\cdot\|_X$ (respectively $\|\cdot\|_Y$)¹. We also assume that

$$f : X \rightarrow]-\infty, +\infty], \quad g : Y \rightarrow]-\infty, +\infty], \quad (2)$$

are convex lower semicontinuous and proper, and that

$$\mathcal{A} : X \rightarrow Y \text{ is linear and continuous.} \quad (3)$$

PDHG is an algorithm for general saddle-point problems of the form

$$\inf_{\mathbf{x} \in X} \sup_{\mathbf{y} \in Y} f(\mathbf{x}) - g^*(\mathbf{y}) + \langle \mathbf{y}, \mathcal{A}\mathbf{x} \rangle. \quad (4)$$

Here, g^* denotes the *Fenchel–Legendre conjugate* of g . It should be noted that (4) is equivalent to $\inf_{\mathbf{x} \in X} f(\mathbf{x}) + g(\mathcal{A}\mathbf{x})$ since the inner sup of (4) is exactly the formula for conjugation of g^* . We return to this point in [Section 3.1](#). Recall the *proximal mapping* of f at $\mathbf{x} \in X$ is defined by

$$\text{Prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{z} \in X} \left\{ f(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|^2 \right\},$$

and that the operator form for the proximal mapping is

$$\text{Prox}_f \equiv (\text{Id} + \partial f)^{-1}. \quad (5)$$

Let $(\mathbf{x}_0, \mathbf{y}_0) \in X \times Y$. The PDHG iteration updates $(\mathbf{x}_0, \mathbf{y}_0)$ as follows ($\forall k \in \mathbb{N}$):

$$\mathbf{x}_{k+1} := \text{Prox}_{\sigma f}(\mathbf{x}_k - \sigma \mathcal{A}^* \mathbf{y}_k), \quad (6a)$$

$$\mathbf{y}_{k+1} := \text{Prox}_{\tau g^*}(\mathbf{y}_k + \tau \mathcal{A}(2\text{Prox}_{\sigma f}(\mathbf{x}_k - \sigma \mathcal{A}^* \mathbf{y}_k) - \mathbf{x}_k)). \quad (6b)$$

Here, $\sigma, \tau > 0$ are step-size parameters that must be chosen correctly—refer to [Lemma 3.2](#). Thus, the main work on each iteration consists of multiplication by \mathcal{A} and \mathcal{A}^* and two prox operations.

Observe that (6) can be written as $(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) := T(\mathbf{x}_k, \mathbf{y}_k)$ where:

$$T(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \text{Prox}_{\sigma f}(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y}) \\ \text{Prox}_{\tau g^*}(\mathbf{y} + \tau \mathcal{A}(2\text{Prox}_{\sigma f}(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y}) - \mathbf{x})) \end{pmatrix}. \quad (7)$$

Let $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$. Then, assuming a constraint qualification, $(\mathbf{x}^*, \mathbf{y}^*) \in \text{Fix } T := \{(\mathbf{x}, \mathbf{y}) \in X \times Y \mid (\mathbf{x}, \mathbf{y}) = T(\mathbf{x}, \mathbf{y})\}$ or equivalently $(\text{Id} - T)(\mathbf{x}^*, \mathbf{y}^*) = \mathbf{0}$, if and only if $(\mathbf{x}^*, \mathbf{y}^*)$ is a solution to (4) (see, e.g., [Proposition 3.1](#) below). On the other hand, if $\mathbf{0} \notin \text{ran}(\text{Id} - T)$ then $\|(\mathbf{x}_k, \mathbf{y}_k)\| \rightarrow \infty$ (see, e.g., [[2](#), Corollary 2.2]). This motivates the exploration of the set $\overline{\text{ran}}(\text{Id} - T)$ and the corresponding well-defined *infimal displacement vector* (see ([47](#)) below). In passing we point out that the study of the range of the displacement mapping associated with splitting algorithms; namely Douglas–Rachford and forward-backward algorithms, was a key ingredient in exploring the static structure and the dynamic behaviour of these methods in the inconsistent case. In this regard, we refer the reader to [[10, 9, 11, 3, 20, 22, 23, 28](#)].

In the case of PDHG, recently Applegate et al. [[1](#)] showed that in the case of inconsistent linear programming (infeasible or unbounded), the infimal displacement vector characterizes the limiting difference between successive PDHG iterates and also certifies the infeasibility or unboundedness.

Our main results can be summarized as follows:

¹When it is clear from the context, we will drop the subscripts X and Y associated with the inner products and the norms.

- (i) We provide a novel formula for $\overline{\text{ran}}(\text{Id} - T)$ in terms of the domains of the functions f, g, f^*, g^* and \mathcal{A} (see [Theorem 3.5](#) below). Along the way, we obtain a formula for the range of the sum of a skew symmetric operator (of the form [\(9\)](#) below) and a maximally monotone operator with a specific structure (see [Theorem 2.3](#) below).
- (ii) When specializing² $g = \iota_K$, where K is a nonempty closed convex cone of Y , we obtain powerful properties for the infimal displacement vector in $\overline{\text{ran}}(\text{Id} - T)$ (see [Proposition 4.2](#) below).
- (iii) We present a comprehensive analysis of the behavior of PDHG when applied to QP. More specifically, we prove that the infimal displacement vector v in $\overline{\text{ran}}(\text{Id} - T)$ provides certificates of inconsistency (see [Theorem 5.7](#) below). In [Theorem 5.10](#) below we prove that the sequence $((x_k, y_k) + kv)_{k \in \mathbb{N}}$ converges as $k \rightarrow \infty$.
- (iv) We analyze the ellipsoid separation problem, another instance of conic programming. For this problem, we also derive a convergence result (see [Theorem 6.17](#) below), and we establish again that PDHG in the inconsistent case returns a certificate (see [Theorem 6.19](#) below).

Organization. The rest of this paper is organized as follows. In [Section 2](#) we present a formula for the range of the sum of two maximally monotone operators that have particular structures. In [Section 3](#) we develop formulas for $\text{Id} - T$, the displacement operator of PDHG, and its range. In [Section 4](#) and the following sections, we consider the several specializations of PDHG to convex optimization problems. [Section 5](#) presents our analysis of $\text{ran}(\text{Id} - T)$ and the infimal displacement vector in the case of QP. Furthermore, when the problem is inconsistent, v is nonzero and certifies inconsistency. We provide computational experiments that illustrate our conclusions. Another special case of the general conic programming problem is presented in [Section 6](#). An application of this setting to the ellipsoid separation problem is detailed. We also prove that PDHG can diagnose infeasible instances of the ellipsoid separation problem.

2 On the range of the sum of monotone operators

Let $B: X \rightrightarrows X$. Recall that B is *monotone* if $(\forall(x, u) \in \text{gr } B) (\forall(y, v) \in \text{gr } B) \langle x - y, u - v \rangle \geq 0$ and B is *maximally monotone* if it is monotone and its graph does not admit any proper extension (in terms of set inclusion). In this section, we derive a formula for the range of the sum of two maximally monotone operators of the form [\(11\)](#) below. This will play a critical role in our analysis later. In the following, we assume that

$$B_1: X \rightrightarrows X \text{ and } B_2: Y \rightrightarrows Y \text{ are maximally monotone.} \quad (8)$$

For the remainder of the paper, we set

$$S := \begin{pmatrix} 0 & \mathcal{A}^* \\ -\mathcal{A} & 0 \end{pmatrix}. \quad (9)$$

We now define the maximally monotone operator (see, e.g., [[6](#), Proposition 20.23])

$$\mathbf{B}: X \times Y \rightrightarrows X \times Y: (x, y) \mapsto B_1 x \times B_2^{-1} y. \quad (10)$$

²Let $S \subseteq X$. Here and elsewhere we use ι_S to denote the *indicator function* of S defined as: $\iota_S(x) = 0$ if $x \in S$; and $\iota_S(x) = +\infty$ if $x \in X \setminus S$. It is well known that if S is closed, convex, and nonempty, then ι_S is a proper l.s.c. convex function.

Observe that by, e.g., [16, Proposition 2.7(i)–(iii)] we have \mathbf{B} is maximally monotone, $S: X \times Y \rightarrow X \times Y$ is maximally monotone with $S^* = -S$ and

$$\mathbf{B} + S \text{ is maximally monotone.} \quad (11)$$

Let $C: X \rightrightarrows X$ be monotone. Recall that C is 3^* monotone if $(\forall (s, r) \in \text{dom } C \times \text{ran } C)$

$$\inf_{(u,v) \in \text{gr } C} \langle u - s, v - r \rangle > -\infty. \quad (12)$$

Fact 2.1. ∂f is 3^* monotone.

Proof. See, e.g., [6, Example 25.13]. ■

Lemma 2.2. Suppose that $\mathcal{A} \neq 0$. Then S is not 3^* monotone.

Proof. Combine [6, Proposition 25.12] and [4, Example 4.5]. ■

The following result provides a formula for the closure of the range of $\mathbf{B} + S$.

Theorem 2.3. Suppose that B_1 and B_2 are 3^* monotone. Then

$$\overline{\text{ran}}(\mathbf{B} + S) = \overline{(\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2))} \times \overline{(\text{dom } B_2 - \mathcal{A}(\text{dom } B_1))}. \quad (13)$$

Proof. For simplicity, we set $R := (\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2)) \times (\text{dom } B_2 - \mathcal{A}(\text{dom } B_1))$. Let $(u, v) \in \text{ran}(\mathbf{B} + S)$. Then $(\exists (x, y) \in \text{dom } \mathbf{B} = \text{dom } B_1 \times \text{dom } B_2^{-1} = \text{dom } B_1 \times \text{ran } B_2)$ such that $(u, v) \in (B_1 x + \mathcal{A}^* y) \times (B_2^{-1} y - \mathcal{A} x) \subseteq (\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2)) \times (\text{dom } B_2 - \mathcal{A}(\text{dom } B_1))$. This proves that

$$\text{ran}(\mathbf{B} + S) \subseteq R \text{ and hence } \overline{\text{ran}}(\mathbf{B} + S) \subseteq \overline{R}. \quad (14)$$

We now turn to the opposite inclusion. Let $r \in R$. Then there exists $x \in \text{ran } B_1$, $y \in \text{dom } B_2$, $w \in \text{ran } B_2$ and $z \in \text{dom } B_1$ such that $r = (x + \mathcal{A}^* w, y - \mathcal{A} z)$. Recalling (11) and Minty's theorem (see, e.g., [6, Theorem 21.1]), we learn that $(\forall n \geq 1) \frac{1}{n^2} \text{Id} + \mathbf{B} + S$ is surjective. Consequently, $(\forall n \geq 1)$ there exists $x_n \in \text{dom } \mathbf{B}$ such that $r \in (\frac{1}{n^2} \text{Id} + \mathbf{B} + S)(x_n) = \frac{1}{n^2} x_n + \mathbf{B} x_n + S x_n$. Equivalently,

$$\text{the sequence } \left(x_n, r - \frac{1}{n^2} x_n \right)_{n \geq 1} \text{ lies in } \text{gr}(\mathbf{B} + S). \quad (15)$$

This in turn implies that

$$\text{the sequence } \left(r - \frac{1}{n^2} x_n \right)_{n \geq 1} \text{ lies in } \text{ran}(\mathbf{B} + S). \quad (16)$$

We will show that $(x_n/n)_{n \geq 1}$ is bounded. We set $s = (z, w)$. We claim that there exists $K \in \mathbb{R}$ such that

$$\inf_{(u,v) \in \text{gr}(\mathbf{B}+S)} \langle u - s, v - r \rangle \geq K. \quad (17)$$

Indeed, let $(u, v) \in \text{gr}(\mathbf{B} + S)$. Then $(\exists (a, a^*) \in \text{gr } B_1) (\exists (b^*, b) \in \text{gr } B_2^{-1})$ such that $u = (a, b^*)$ and $v = (a^* + \mathcal{A}^* b^*, b - \mathcal{A} a)$. Now,

$$\begin{aligned} & \langle u - s, v - r \rangle \\ &= \langle (a, b^*) - (z, w), (a^* + \mathcal{A}^* b^*, b - \mathcal{A} a) - (x + \mathcal{A}^* w, y - \mathcal{A} z) \rangle \end{aligned} \quad (18a)$$

$$= \langle (a - z, b^* - w), (a^* + \mathcal{A}^* b^* - (x + \mathcal{A}^* w), b - \mathcal{A}a - (y - \mathcal{A}z)) \rangle \quad (18b)$$

$$= \langle a - z, a^* + \mathcal{A}^* b^* - (x + \mathcal{A}^* w) \rangle + \langle b^* - w, b - \mathcal{A}a - (y - \mathcal{A}z) \rangle \quad (18c)$$

$$= \langle a - z, a^* - x \rangle + \langle a - z, \mathcal{A}^*(b^* - w) \rangle + \langle b^* - w, b - y \rangle - \langle b^* - w, \mathcal{A}(a - z) \rangle \quad (18d)$$

$$= \langle a - z, a^* - x \rangle + \langle b^* - w, b - y \rangle + \langle b^* - w, \mathcal{A}(a - z) \rangle - \langle b^* - w, \mathcal{A}(a - z) \rangle \quad (18e)$$

$$= \langle a - z, a^* - x \rangle + \langle b^* - w, b - y \rangle \geq K, \quad (18f)$$

for some $K \in \mathbb{R}$. The inequality in (18f) follows from applying (12) with C replaced by B_1 by noting that $(a, a^*) \in \text{gr } B_1$ and again applying (12) with X replaced by Y , and C replaced by B_2 by noting that $(b, b^*) \in \text{gr } B_2$. This proves (17). Now (15) and (17) imply that $(\forall n \geq 1)$

$$\frac{\|\mathbf{x}_n\|^2}{n^2} = \frac{1}{n^2}(\|\mathbf{x}_n\|^2 - \|\mathbf{s}\|^2) + \frac{\|\mathbf{s}\|^2}{n^2} \quad (19a)$$

$$\leq \frac{1}{n^2}(\|\mathbf{x}_n\|^2 - \|\mathbf{s}\|^2) + \|\mathbf{s}\|^2 \quad (19b)$$

$$= \frac{1}{n^2}(2\|\mathbf{x}_n\|^2 - (\|\mathbf{x}_n\|^2 + \|\mathbf{s}\|^2)) + \|\mathbf{s}\|^2 \quad (19c)$$

$$\leq \frac{1}{n^2}(2\|\mathbf{x}_n\|^2 - 2\langle \mathbf{s}, \mathbf{x}_n \rangle) + \|\mathbf{s}\|^2 \quad (19d)$$

$$= -2\langle \mathbf{x}_n - \mathbf{s}, -\frac{1}{n^2}\mathbf{x}_n \rangle + \|\mathbf{s}\|^2 \quad (19e)$$

$$= -2\langle \mathbf{x}_n - \mathbf{s}, \mathbf{r} - \frac{1}{n^2}\mathbf{x}_n - \mathbf{r} \rangle + \|\mathbf{s}\|^2 \quad (19f)$$

$$\leq -2K + \|\mathbf{s}\|^2. \quad (19g)$$

That is, $(\mathbf{x}_n/n)_{n \geq 1}$ is bounded as claimed. Taking the limit in (16) as $n \rightarrow \infty$ we learn that $\mathbf{r} \in \overline{\text{ran}}(\mathbf{B} + S)$ and hence $R \subseteq \overline{\text{ran}}(\mathbf{B} + S)$. The proof is complete. \blacksquare

Remark 2.4. Suppose that B_1 and B_2 are 3^* monotone. Then \mathbf{B} is 3^* monotone. Some comments are in order.

- (i) Suppose that $\text{dom } \mathbf{B} = X \times Y$. This is equivalent to $\text{dom } B_1 = X$ and $\text{ran } B_2 = \text{dom } B_2^{-1} = Y$. The formula in (13) boils down to

$$\overline{\text{ran}}(\mathbf{B} + S) = \overline{(\text{ran } B_1 + \text{ran } \mathcal{A}^*)} \times \overline{(\text{dom } B_2 - \text{ran } \mathcal{A})} = \overline{\text{ran } \mathbf{B} + \text{ran } S}. \quad (20)$$

The above formula is alternatively obtained using the celebrated Brezis–Haraux theorem, see, e.g., [6, Theorem 25.24(ii)].

- (ii) The assumption that $\text{dom } \mathbf{B} = X \times Y$ is critical to prove that the formula in (13) reduces to the formula in (20) as we illustrate in Example 2.5 below.
- (iii) The assumption that both B_1 and B_2 are 3^* monotone is critical to obtain the conclusion of Theorem 2.3 as we illustrate in Example 2.6 below.

Example 2.5. Suppose that $X = Y = \mathbb{R}$, that $\mathcal{A} = -\text{Id}$, that³ $B_1 = N_{[0, +\infty[}$ and that $B_2 = N_{]-\infty, 0]}$. Then

$$\overline{\text{ran}}(\mathbf{B} + S) =]-\infty, 0] \times \mathbb{R} \subsetneq \mathbb{R}^2 = \overline{\text{ran } \mathbf{B} + \text{ran } S}. \quad (21)$$

Proof. Observe that B_1 and B_2 are subdifferential operators, hence 3^* monotone by Fact 2.1. Moreover, $\mathbf{B} = N_{\mathbb{R}_+^2}$ and $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. On the one hand, clearly $\text{ran } \mathbf{B} = \mathbb{R}_-^2$ and $\text{ran } S = \mathbb{R}^2$ hence $\text{ran } \mathbf{B} + \text{ran } S = \mathbb{R}^2 = \overline{\text{ran } \mathbf{B} + \text{ran } S}$. On the other hand, $\text{dom } B_1 = \text{ran } B_2 = [0, +\infty[$ and $\text{dom } B_2 = \text{ran } B_1 =]-\infty, 0]$, and (13) yields $\overline{\text{ran}}(\mathbf{B} + S) =]-\infty, 0] \times \mathbb{R} \subsetneq \mathbb{R}^2 = \overline{\text{ran } \mathbf{B} + \text{ran } S}$, as claimed. \blacksquare

Example 2.6. Suppose that $X = \mathbb{R}$, that $Y = \mathbb{R}^2$, that $B_1 \equiv 0$, that $B_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and that $\mathcal{A} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then the following hold:

³Let C be a nonempty closed convex subset of X . Here and elsewhere we use N_C to denote the normal cone operator associated with C .

- (i) $\overline{\text{ran}}(\mathbf{B} + S) = \text{span}\{(0, 1, 0)^\top, (-1, 0, 1)^\top\} \subsetneq \mathbb{R}^3$.
- (ii) $\overline{(\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2)) \times (\text{dom } B_2 - \mathcal{A}(\text{dom } B_1))} = \mathbb{R}^3$.
- (iii) $\overline{\text{ran}}(\mathbf{B} + S) \subsetneq \overline{(\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2)) \times (\text{dom } B_2 - \mathcal{A}(\text{dom } B_1))}$.

Proof. It follows from [Lemma 2.2](#) that B_2 is *not* 3^* monotone. Moreover, it is easy to verify that $B_2^{-1} = -B_2$, that $\text{dom } B_1 = \mathbb{R}$, that $\text{ran } B_1 = \{0\}$, and that $\text{dom } B_2 = \text{ran } B_2 = \mathbb{R}^2$. (i): We have

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \text{ and } S = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ hence } \mathbf{B} + S = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \quad (22)$$

This proves (i). (ii): Indeed, we have $\text{ran } B_1 + \mathcal{A}^*(\text{ran } B_2) = \{0\} + \mathcal{A}^*(\mathbb{R}^2) = \mathbb{R}$ and $\text{dom } B_2 - \mathcal{A}(\text{dom } B_1) = \mathbb{R}^2$. (iii): Combine (i) and (ii). ■

3 The PDHG splitting operator, the range of its displacement map, and the infimal displacement vector

In this section, we apply the theory from [Section 2](#) to develop formulas $\text{Id} - T$, the displacement operator of PDHG, and its range. These formulas appear in [Theorem 3.5](#). Then in [Section 3.3](#) we define the infimal displacement vector, which lies in the closure of this range, and state and prove some of its important properties. Let $\sigma > 0$ and let $\tau > 0$. For the remainder of the paper, we set

$$M := \begin{pmatrix} \frac{1}{\sigma} \text{Id}_X & -\mathcal{A}^* \\ -\mathcal{A} & \frac{1}{\tau} \text{Id}_Y \end{pmatrix}, \quad (23)$$

and we set

$$F: X \times Y \rightarrow]-\infty, +\infty]: (\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}) + g^*(\mathbf{y}). \quad (24)$$

Then (see, e.g., [\[6, Proposition 16.9\]](#))

$$\partial F(\mathbf{x}, \mathbf{y}) = \partial f(\mathbf{x}) \times \partial g^*(\mathbf{y}). \quad (25)$$

3.1 PDHG and Fenchel–Rockafellar duality

Consider the primal problem

$$\underset{x \in X}{\text{minimize}} \quad f(x) + g(\mathcal{A}x), \quad (26)$$

which is equivalent to [\(4\)](#), and its Fenchel–Rockafellar dual

$$\underset{y \in Y}{\text{minimize}} \quad f^*(-\mathcal{A}^*y) + g^*(y). \quad (27)$$

Under appropriate constraint qualifications (see, e.g., [\[12\]](#) and [\[13\]](#) or [\[16, Proposition 4.1\(iii\)\]](#)) [\(26\)](#) boils down to solving the primal inclusion:

$$\text{find } x \in X \text{ such that } 0 \in \partial f(x) + \mathcal{A}^* \partial g(\mathcal{A}x) \quad (28)$$

while (27) boils down to solving the dual inclusion:

$$\text{find } \mathbf{y} \in Y \text{ such that } (\exists \mathbf{x} \in X) -\mathcal{A}^*\mathbf{y} \in \partial f(\mathbf{x}) \text{ and } \mathbf{y} \in \partial g(\mathcal{A}\mathbf{x}). \quad (29)$$

Following [14], we say that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in X \times Y$ is a primal-dual solution to both (28) and (29) if

$$-\mathcal{A}^*\bar{\mathbf{y}} \in \partial f(\bar{\mathbf{x}}) \text{ and } \bar{\mathbf{y}} \in \partial g(\mathcal{A}\bar{\mathbf{x}}). \quad (30)$$

One checks that the existence of a solution to (30) implies the existence of a solution to both (28) and (29). Conversely, a solution to either (28) or (29) implies the existence of a solution to (30).

The following result is part of the literature. We include proof for the sake of completeness. Recall that S was defined in (9), M in (23), and F in (24). Recall $\text{Prox}_f = (\text{Id} + \partial f)^{-1}$, and $\partial f^* = (\partial f)^{-1}$.

Proposition 3.1. *Let $(\mathbf{x}, \mathbf{y}) \in X \times Y$. We set*

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = \begin{pmatrix} \text{Prox}_{\sigma f}(\mathbf{x} - \sigma \mathcal{A}^*\mathbf{y}) \\ \text{Prox}_{\tau g^*}(\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x})) \end{pmatrix}. \quad (31)$$

Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in X \times Y$. We set $T = (M + \partial F + S)^{-1}M$. Then the following hold.

(i) *We have*

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = T \begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix}. \quad (32)$$

(ii) *Recalling (30) we have $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a primal-dual solution of (28) if and only if $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \text{Fix } T$.*

Proof. (i): Indeed,

$$(31) \Leftrightarrow \begin{aligned} \mathbf{x}^+ &= \text{Prox}_{\sigma f}(\mathbf{x} - \sigma \mathcal{A}^*\mathbf{y}), \\ \mathbf{y}^+ &= \text{Prox}_{\tau g^*}(\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x})) \end{aligned} \quad (33a)$$

$$\Leftrightarrow \mathbf{x} - \sigma \mathcal{A}^*\mathbf{y} \in \mathbf{x}^+ + \sigma \partial f(\mathbf{x}^+), \quad (33b)$$

$$\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) \in \mathbf{y}^+ + \tau \partial g^*(\mathbf{y}^+) \quad (33b)$$

$$\Leftrightarrow \mathbf{x} - \mathbf{x}^+ \in \sigma \mathcal{A}^*\mathbf{y} + \sigma \partial f(\mathbf{x}^+), \quad (33c)$$

$$\mathbf{y} - \mathbf{y}^+ \in -\tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) + \tau \partial g^*(\mathbf{y}^+) \quad (33c)$$

$$\Leftrightarrow \frac{1}{\sigma}(\mathbf{x} - \mathbf{x}^+) \in \mathcal{A}^*\mathbf{y} + \partial f(\mathbf{x}^+), \quad (33d)$$

$$\frac{1}{\tau}(\mathbf{y} - \mathbf{y}^+) \in -\mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) + \partial g^*(\mathbf{y}^+) \quad (33d)$$

$$\Leftrightarrow \frac{1}{\sigma}(\mathbf{x} - \mathbf{x}^+) - \mathcal{A}^*(\mathbf{y} - \mathbf{y}^+) \in \mathcal{A}^*\mathbf{y} + \partial f(\mathbf{x}^+) - \mathcal{A}^*(\mathbf{y} - \mathbf{y}^+), \quad (33e)$$

$$\frac{1}{\tau}(\mathbf{y} - \mathbf{y}^+) - \mathcal{A}(\mathbf{x} - \mathbf{x}^+) \in -\mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) + \partial g^*(\mathbf{y}^+) - \mathcal{A}(\mathbf{x} - \mathbf{x}^+) \quad (33e)$$

$$\Leftrightarrow \frac{1}{\sigma}(\mathbf{x} - \mathbf{x}^+) - \mathcal{A}^*(\mathbf{y} - \mathbf{y}^+) \in \mathcal{A}^*\mathbf{y} + \partial f(\mathbf{x}^+), \quad (33f)$$

$$\frac{1}{\tau}(\mathbf{y} - \mathbf{y}^+) - \mathcal{A}(\mathbf{x} - \mathbf{x}^+) \in \partial g^*(\mathbf{y}^+) - \mathcal{A}\mathbf{x}^+ \quad (33f)$$

$$\Leftrightarrow \begin{pmatrix} \frac{1}{\sigma} \text{Id}_X & -\mathcal{A}^* \\ -\mathcal{A} & \frac{1}{\tau} \text{Id}_Y \end{pmatrix} \begin{pmatrix} \mathbf{x} - \mathbf{x}^+ \\ \mathbf{y} - \mathbf{y}^+ \end{pmatrix} \in \begin{pmatrix} \partial f(\mathbf{x}^+) \\ \partial g^*(\mathbf{y}^+) \end{pmatrix} + \begin{pmatrix} 0 & \mathcal{A}^* \\ -\mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} \quad (33g)$$

$$\Leftrightarrow M \begin{pmatrix} \mathbf{x} - \mathbf{x}^+ \\ \mathbf{y} - \mathbf{y}^+ \end{pmatrix} \in \begin{pmatrix} \partial f(\mathbf{x}^+) \\ \partial g^*(\mathbf{y}^+) \end{pmatrix} + S \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} \quad (33h)$$

$$\Leftrightarrow M \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \in M \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} + \begin{pmatrix} \partial f(\mathbf{x}^+) \\ \partial g^*(\mathbf{y}^+) \end{pmatrix} + S \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = (M + \partial F + S) \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} \quad (33i)$$

$$\Leftrightarrow \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} \in (M + \partial F + S)^{-1} M \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \quad (33j)$$

$$\Leftrightarrow \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = (M + \partial F + S)^{-1} M \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}. \quad (33k)$$

Note that the last step (33k) follows because (33a) implies that, given (\mathbf{x}, \mathbf{y}) , there is exactly one solution z to the inclusion $(z, M(\mathbf{x}, \mathbf{y})) \in \text{gr}(M + \partial F + S)$. The proof of (i) is complete.

(ii): Indeed,

$$\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix} = T \begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix} \quad (34a)$$

$$\Leftrightarrow \bar{\mathbf{x}} = \text{Prox}_{\sigma f}(\bar{\mathbf{x}} - \sigma \mathcal{A}^* \bar{\mathbf{y}}) \text{ and } \bar{\mathbf{y}} = \text{Prox}_{\tau g^*}(\bar{\mathbf{y}} + \tau \mathcal{A} \bar{\mathbf{x}}) \quad (34b)$$

$$\Leftrightarrow \bar{\mathbf{x}} - \sigma \mathcal{A}^* \bar{\mathbf{y}} \in \bar{\mathbf{x}} + \sigma \partial f(\bar{\mathbf{x}}) \text{ and } \bar{\mathbf{y}} + \tau \mathcal{A} \bar{\mathbf{x}} \in \bar{\mathbf{y}} + \tau \partial g^*(\bar{\mathbf{y}}) \quad (34c)$$

$$\Leftrightarrow -\sigma \mathcal{A}^* \bar{\mathbf{y}} \in \sigma \partial f(\bar{\mathbf{x}}) \text{ and } \tau \mathcal{A} \bar{\mathbf{x}} \in \tau \partial g^*(\bar{\mathbf{y}}) \quad (34d)$$

$$\Leftrightarrow -\mathcal{A}^* \bar{\mathbf{y}} \in \partial f(\bar{\mathbf{x}}) \text{ and } \bar{\mathbf{y}} \in \partial g(\mathcal{A} \bar{\mathbf{x}}) \quad (34e)$$

$$\Leftrightarrow (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \text{ is a primal-dual solution of (28).} \quad (34f)$$

This completes the proof. ■

3.2 The range of $\text{Id} - T$

In this subsection we apply the results of Section 2 to obtain a formula for $\overline{\text{ran}}(\text{Id} - T)$, starting from the formula for T given by Proposition 3.1(i). A key operator in our analysis is $M^{-1}(\partial F + S)$. The significance of this operator becomes apparent below in Theorem 3.5, and hence we set the stage with some preliminary results about M and $\partial F + S$. We start by observing that (11) applied with \mathbf{B} replaced by ∂F implies that

$$\partial F + S \text{ is maximally monotone on } X \times Y. \quad (35)$$

For the remainder of the paper, we impose the assumption that σ, τ are chosen to satisfy

$$\sigma \tau \|\mathcal{A}\|^2 < 1. \quad (36)$$

The following lemma is straightforward to verify. We include the proof for the sake of completeness.

Lemma 3.2. *M is self-adjoint. Moreover, we have:*

- (i) M and M^{-1} are strongly monotone⁴.
- (ii) M and M^{-1} are surjective.
- (iii) M and M^{-1} are injective.
- (iv) M and M^{-1} are bijective.
- (v) M and M^{-1} are maximally monotone.

Proof. (i): Indeed, let $(\mathbf{u}, \mathbf{v}) \in X \times Y$. Then using Cauchy–Schwarz we have

$$\langle (\mathbf{u}, \mathbf{v}), M(\mathbf{u}, \mathbf{v}) \rangle = \langle (\mathbf{u}, \mathbf{v}), \left(\frac{1}{\sigma} \mathbf{u} - \mathcal{A}^* \mathbf{v}, -\mathcal{A} \mathbf{u} + \frac{1}{\tau} \mathbf{v} \right) \rangle \quad (37a)$$

⁴Let $B: X \rightrightarrows X$ be monotone and let $\beta > 0$. Then B is β -strongly monotone if $B - \beta \text{Id}$ is monotone.

$$= \frac{1}{\sigma} \|\mathbf{u}\|^2 + \frac{1}{\tau} \|\mathbf{v}\|^2 - 2 \langle \mathcal{A}\mathbf{u}, \mathbf{v} \rangle \geq \frac{1}{\sigma} \|\mathbf{u}\|^2 + \frac{1}{\tau} \|\mathbf{v}\|^2 - 2 \|\mathcal{A}\| \|\mathbf{u}\| \|\mathbf{v}\| \quad (37b)$$

$$= \frac{1}{\sigma} \|\mathbf{u}\|^2 + \frac{1}{\tau} \|\mathbf{v}\|^2 - 2 \left(\frac{\sqrt[4]{\sigma\tau} \sqrt{\|\mathcal{A}\|}}{\sqrt{\sigma}} \|\mathbf{u}\| \right) \left(\frac{\sqrt[4]{\sigma\tau} \sqrt{\|\mathcal{A}\|}}{\sqrt{\tau}} \|\mathbf{v}\| \right) \quad (37c)$$

$$\geq \frac{1}{\sigma} \|\mathbf{u}\|^2 + \frac{1}{\tau} \|\mathbf{v}\|^2 - \frac{\sqrt{\sigma\tau} \|\mathcal{A}\|}{\sigma} \|\mathbf{u}\|^2 - \frac{\sqrt{\sigma\tau} \|\mathcal{A}\|}{\tau} \|\mathbf{v}\|^2 \quad (37d)$$

$$\geq \min \left\{ \frac{1}{\sigma}, \frac{1}{\tau} \right\} (1 - \sqrt{\sigma\tau} \|\mathcal{A}\|) \|(\mathbf{u}, \mathbf{v})\|^2. \quad (37e)$$

That is, M is $(\min \{ \frac{1}{\sigma}, \frac{1}{\tau} \} (1 - \sqrt{\sigma\tau} \|\mathcal{A}\|))$ -strongly monotone. Consequently, M is 3^* monotone by, e.g., [6, Example 25.15(iv)]. Now combine this with [6, Proposition 25.16 and Example 22.7] to learn that that M^{-1} is strongly monotone. (ii): Combine (i) and [6, Proposition 22.11(ii)]. (iii): This is a direct consequence of (i). (iv): Combine (ii) and (iii). (v): Using (i) we have M and M^{-1} are monotone. Now combine this with [6, Example 20.34]. ■

Recalling that M is positive definite, in the following we let Z be the Hilbert space obtained by endowing $X \times Y$ with the inner product and induced norm

$$\langle \cdot, \cdot \rangle_M : Z \times Z \rightarrow \mathbb{R} : (\mathbf{u}, \mathbf{v}) \mapsto \langle \mathbf{u}, M\mathbf{v} \rangle \quad \text{and} \quad \|\mathbf{u}\|_M = \sqrt{\langle \mathbf{u}, M\mathbf{u} \rangle} \quad (38)$$

respectively.

Lemma 3.3. $M^{-1}(\partial F + S) : Z \rightrightarrows Z$ is maximally monotone.

Proof. Combine (35) with [6, Proposition 20.24] in view of Lemma 3.2(i). ■

Let $D \subseteq X$ and let $L : X \rightarrow X$ be linear and continuous. It is easy to verify that

$$\overline{L(D)} = \overline{L(\overline{D})}. \quad (39)$$

The following result is of central importance in our work.

Theorem 3.4. We have

$$\overline{\text{ran}}(\partial F + S) = \overline{(\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*))} \times \overline{(\text{dom } g - \mathcal{A}(\text{dom } f))}. \quad (40)$$

Proof. Recall that by [6, Corollary 16.30] we have $(\partial f^*, \partial g^*) = ((\partial f)^{-1}, (\partial g)^{-1})$. Moreover, by e.g., [6, Corollary 16.39] we have

$$\overline{\text{dom } \partial f} = \overline{\text{dom } f}. \quad (41)$$

It follows from (39) applied with L replaced by \mathcal{A} and D replaced by $\text{dom } \partial f$ and again by D replaced by $\text{dom } f$ in view of (41) that $\overline{\mathcal{A}(\text{dom } \partial f)} = \overline{\mathcal{A}(\overline{\text{dom } \partial f})} = \overline{\mathcal{A}(\overline{\text{dom } f})} = \overline{\mathcal{A}(\text{dom } f)}$. Similarly, we conclude that $\overline{\mathcal{A}^*(\text{dom } \partial g^*)} = \overline{\mathcal{A}^*(\text{dom } g^*)}$. It therefore follows from Theorem 2.3, applied with (B_1, B_2) replaced by $(\partial f, \partial g)$, in view of Fact 2.1 that

$$\overline{\text{ran}}(\partial F + S) = \overline{(\text{ran } \partial f + \mathcal{A}^*(\text{ran } (\partial g^*)^{-1}))} \times \overline{(\text{dom } \partial g - \mathcal{A}(\text{dom } \partial f))} \quad (42a)$$

$$= \overline{(\text{dom } \partial f^* + \mathcal{A}^*(\text{dom } \partial g^*))} \times \overline{(\text{dom } \partial g - \mathcal{A}(\text{dom } \partial f))} \quad (42b)$$

$$= \overline{(\overline{\text{dom } \partial f^* + \mathcal{A}^*(\text{dom } \partial g^*)})} \times \overline{(\overline{\text{dom } \partial g - \mathcal{A}(\text{dom } \partial f)})} \quad (42c)$$

$$= \overline{(\overline{\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*)})} \times \overline{(\overline{\text{dom } g - \mathcal{A}(\text{dom } f)})} \quad (42d)$$

$$= \overline{(\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*))} \times \overline{(\text{dom } g - \mathcal{A}(\text{dom } f))}. \quad (42e)$$

The proof is complete. ■

We are now ready to prove the main result in this section.

Theorem 3.5. Let $T = (M + \partial F + S)^{-1}M$, that is, the PDHG operator introduced in [Proposition 3.1](#). Then the following hold:

- (i) $T = (\text{Id} + M^{-1}(\partial F + S))^{-1}$.
- (ii) $T: Z \rightarrow Z$ is firmly nonexpansive⁵.
- (iii) $\text{Id} - T = (\text{Id} + (\partial F + S)^{-1}M)^{-1}$.
- (iv) $\text{ran}(\text{Id} - T) = M^{-1}(\text{ran}(\partial F + S))$.
- (v) $\overline{\text{ran}}(\text{Id} - T) = M^{-1}(\overline{\text{ran}}(\partial F + S))$.
- (vi) $\overline{\text{ran}}(\text{Id} - T) = M^{-1}((\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*)) \times \overline{(\text{dom } g - \mathcal{A}(\text{dom } f))})$.

Proof. (i): Indeed, we have $T = (M + \partial F + S)^{-1}M = (M^{-1}(M + \partial F + S))^{-1} = (\text{Id} + M^{-1}(\partial F + S))^{-1}$. (ii): Combine (i), [Lemma 3.3](#), [Lemma 3.2\(i\)](#) and, e.g., [6, Proposition 23.10]. (iii): It follows from [6, Proposition 23.34(iii)] in view of [Lemma 3.2\(i\)](#) that

$$\text{Id} - T = M^{-1}(\text{Id} + M(\partial F + S)^{-1})^{-1}M = M^{-1}(M^{-1} + (\partial F + S)^{-1})^{-1} \quad (43a)$$

$$= (\text{Id} + (\partial F + S)^{-1}M)^{-1}. \quad (43b)$$

(iv): Using (iii) we have

$$\text{ran}(\text{Id} - T) = \text{ran}(\text{Id} + (\partial F + S)^{-1}M)^{-1} = \text{dom}(\text{Id} + (\partial F + S)^{-1}M) \quad (44a)$$

$$= \text{dom}((\partial F + S)^{-1}M) = \text{ran}((\partial F + S)^{-1}M)^{-1} \quad (44b)$$

$$= \text{ran}(M^{-1}(\partial F + S)) = M^{-1}(\text{ran}(\partial F + S)), \quad (44c)$$

where in the last identity we used [Lemma 3.2\(iv\)](#). (v): It follows from (iv) that $\overline{\text{ran}}(\text{Id} - T) = \overline{M^{-1}(\text{ran}(\partial F + S))} = \overline{M^{-1}(\overline{\text{ran}}(\partial F + S))} = M^{-1}(\overline{\text{ran}}(\partial F + S))$. Here the second identity follows from applying (39) with X replaced by $X \times Y$, L replaced by M^{-1} and D replaced by $\text{ran}(\partial F + S)$, while the third identity follows from [Lemma 3.2\(iv\)](#). (vi): Combine (v) and [Theorem 3.4](#). ■

3.3 The infimal displacement vector associated with T

We point out that [Theorem 3.5](#) is a powerful and instrumental tool in analyzing the behaviour of PDHG in the inconsistent case in view [Proposition 3.1\(ii\)](#). Indeed, if $(\mathbf{0}, \mathbf{0}) \notin \text{ran}(\text{Id} - T)$ then T does not have a fixed point. More concretely, in this section, we study the minimal norm element in $\overline{\text{ran}}(\text{Id} - T)$. As we shall see below, this allows us to diagnose if the inconsistency comes from the primal problem or from the dual problem or from both. In this section, we assume that

$$T: Z \rightarrow Z \text{ is defined as in (32)}. \quad (45)$$

Because T is firmly nonexpansive (see [Theorem 3.5\(ii\)](#)) we learn by, e.g., [6, Example 20.29] that

$$\text{Id} - T: Z \rightarrow Z \text{ is maximally monotone}. \quad (46)$$

It, therefore, follows from (46) and, e.g., [6, Corollary 21.14] that $\overline{\text{ran}}(\text{Id} - T)$ is convex. Recalling (38) we now define the *infimal displacement vector* \mathbf{v}

$$\mathbf{v} = \operatorname{argmin}_{\mathbf{w} \in \overline{\text{ran}}(\text{Id} - T)} \|\mathbf{w}\|_M. \quad (47)$$

The vector \mathbf{v} has a beautiful interpretation. Indeed, in some sense, $\|\mathbf{v}\|$ can be viewed as a measurement of how far our problem is from being consistent. We now have the following useful facts.

⁵Let $T: X \rightarrow X$. Then T is firmly nonexpansive if $(\forall (x, y) \in X \times X) \|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2$.

Fact 3.6. Suppose $\mathbf{v} \in \text{ran}(\text{Id} - T)$, equivalently; $\text{Fix}(\mathbf{v} + T) \neq \emptyset$. Let $\mathbf{z}_0 \in X$. ($\forall k \in \mathbb{N}$) update via:

$$\mathbf{z}_{k+1} = T\mathbf{z}_k. \quad (48)$$

Then $(\mathbf{z}_k + k\mathbf{v})_{k \in \mathbb{N}}$ is Fejér monotone⁶ with respect to $\text{Fix}(\mathbf{v} + T)$, hence $(\mathbf{z}_k + k\mathbf{v})_{k \in \mathbb{N}}$ is bounded.

Proof. See [9, Proposition 2.5]. ■

Fact 3.7. Suppose that $\mathbf{z}_0 \in \text{Fix}(\mathbf{v} + T)$. Then $\mathbf{z}_0 - \mathbb{R}_+\mathbf{v} \subseteq \text{Fix}(\mathbf{v} + T)$.

Proof. See [9, Proposition 2.5(i)]. ■

Fact 3.8. Let $\mathbf{z}_0 \in Z$. Update via: $\mathbf{z}_{k+1} = T\mathbf{z}_k$. Then the following hold:

- (i) (Pazy) $\frac{\mathbf{z}_k}{k} \rightarrow -\mathbf{v}$.
- (ii) (Baillon–Bruck–Reich) $\mathbf{z}_k - \mathbf{z}_{k+1} \rightarrow \mathbf{v}$.

Proof. (i): See [25]. (ii): See [2] or [17]. ■

Remark 3.9. Some comments are in order.

- (i) *Fact 3.8(i)&(ii)* reveal that when $\mathbf{v} \neq \mathbf{0}$ PDHG is able to diagnose inconsistent problems. Furthermore, as we shall see in *Theorem 5.7* and again in *Theorem 6.19*, \mathbf{v} may also carry a certificate of inconsistency.
- (ii) On the other hand, if $\mathbf{0} \in \overline{\text{ran}}(\text{Id} - T) \setminus \text{ran}(\text{Id} - T)$ then *Fact 3.8(i)&(ii)* do not tell whether the problem is inconsistent. Alternatively, one could use [2, Corollary 2.2] and monitor the sequence $(\|\mathbf{z}_k\|)_{k \in \mathbb{N}}$.
- (iii) In passing, we point out that in the applications we study in upcoming sections, we prove that $\mathbf{v} \in \text{ran}(\text{Id} - T)$. More precisely, in the QP case (see *Section 5* below) we show that $\text{ran}(\text{Id} - T)$ is closed and the inclusion follows, while in the case of the ellipsoid separation problem in *Section 6* below we proved the inclusion directly.

We conclude this section with the following proposition, which further characterizes and establishes properties of \mathbf{v} used later.

Proposition 3.10. Recalling (47), let $\overline{\mathbf{w}} \in \overline{\text{ran}}(\text{Id} - T)$. Then

$$\overline{\mathbf{w}} = \mathbf{v} \Leftrightarrow (\forall \mathbf{y} \in \overline{\text{ran}}(\partial F + S)) \langle \overline{\mathbf{w}}, M\overline{\mathbf{w}} - \mathbf{y} \rangle \leq 0. \quad (49)$$

In particular

- (i) $(\forall \mathbf{r} \in \overline{(\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*))})$ we have $\langle \mathbf{v}_R, \frac{1}{\sigma}\mathbf{v}_R - \mathcal{A}^*\mathbf{v}_D - \mathbf{r} \rangle \leq 0$.
- (ii) $(\forall \mathbf{d} \in \overline{(\text{dom } g - \mathcal{A}(\text{dom } f)})}$ we have $\langle \mathbf{v}_D, \frac{1}{\tau}\mathbf{v}_D - \mathcal{A}\mathbf{v}_R - \mathbf{d} \rangle \leq 0$.

Proof. Indeed, let $\mathbf{y} \in \overline{\text{ran}}(\partial F + S)$. Recalling *Theorem 3.5(v)*, it follows from the Projection Theorem see, e.g., [6, Theorem 3.16], that

$$\overline{\mathbf{w}} = \mathbf{v} \Leftrightarrow \left\langle \overline{\mathbf{w}}, \overline{\mathbf{w}} - M^{-1}\mathbf{y} \right\rangle_M \leq 0 \Leftrightarrow \left\langle \overline{\mathbf{w}}, M\overline{\mathbf{w}} - MM^{-1}\mathbf{y} \right\rangle \leq 0 \Leftrightarrow \langle \overline{\mathbf{w}}, M\overline{\mathbf{w}} - \mathbf{y} \rangle \leq 0. \quad (50)$$

⁶Let $(\mathbf{x}_k)_{k \in \mathbb{N}}$ be a sequence in X and let C be a nonempty subset of X . Then $(\mathbf{x}_k)_{k \in \mathbb{N}}$ is Fejér monotone with respect to C if $(\forall \mathbf{c} \in C)(\forall k \in \mathbb{N}) \quad \|\mathbf{x}_{k+1} - \mathbf{c}\| \leq \|\mathbf{x}_k - \mathbf{c}\|$.

This verifies (49). Now, let $\mathbf{r} \in \overline{(\text{dom } f^* + \mathcal{A}^*(\text{dom } g^*))}$ and let $\mathbf{d} \in \overline{(\text{dom } g - \mathcal{A}(\text{dom } f))}$. Then $\mathbf{y} := (\mathbf{r}, \mathbf{d}) \in \overline{\text{ran}}(\partial F + S)$ by Theorem 3.4. This and (49) imply that

$$\langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathcal{A}^* \mathbf{v}_D - \mathbf{r} \rangle + \langle \mathbf{v}_D, \frac{1}{\tau} \mathbf{v}_D - \mathcal{A} \mathbf{v}_R - \mathbf{d} \rangle \leq 0. \quad (51)$$

Finally, observe that (47) and Theorem 3.5(v) imply

$$M\mathbf{v} = (\frac{1}{\sigma} \mathbf{v}_R - \mathcal{A}^* \mathbf{v}_D, \frac{1}{\tau} \mathbf{v}_D - \mathcal{A} \mathbf{v}_R) \in \overline{\text{ran}}(\partial F + S). \quad (52)$$

(i): Apply (51) with \mathbf{d} replaced by $\frac{1}{\tau} \mathbf{v}_D - \mathcal{A} \mathbf{v}_R$ in view of (52). (ii): Apply (51) with \mathbf{r} replaced by $\frac{1}{\sigma} \mathbf{v}_R - \mathcal{A}^* \mathbf{v}_D$ in view of (52). ■

4 Specialization to conic problems

From now on we assume that

$$K \text{ is a closed convex cone in } Y \text{ and } \mathbf{b} \in Y. \quad (53)$$

Consider the problem

$$\begin{aligned} & \underset{\mathbf{x} \in X}{\text{minimize}} && f(\mathbf{x}) \\ & \text{subject to} && \mathcal{A}\mathbf{x} - \mathbf{b} \in K. \end{aligned} \quad (54)$$

Problem (54) can be recast as

$$\underset{\mathbf{x} \in X}{\text{minimize}} \quad f(\mathbf{x}) + \iota_K(\mathcal{A}\mathbf{x} - \mathbf{b}). \quad (55)$$

We recover the PDHG framework by setting⁷

$$g = \iota_K(\cdot - \mathbf{b}), \text{ hence } g^* = \iota_{K^\ominus}(\cdot) + \langle \mathbf{b}, \cdot \rangle. \quad (56)$$

It follows from [6, Example 23.4 and Proposition 23.17(ii)] that

$$\text{Prox}_{\tau g^*} = P_{K^\ominus}(\cdot - \tau \mathbf{b}). \quad (57)$$

By combining (31) and (57), the PDHG update to solve (55) is

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = T \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} := \begin{pmatrix} \text{Prox}_{\sigma f}(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y}) \\ P_{K^\ominus}(\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) - \tau \mathbf{b}) \end{pmatrix}. \quad (58)$$

Remark 4.1. The following are special cases of (54).

(i) Let $H: X \rightarrow X$ be linear, monotone, and self-adjoint, i.e., $H = H^*$ and let $\mathbf{c} \in X$. The quadratic optimization problem

$$\begin{aligned} & \underset{\mathbf{x} \in X}{\text{minimize}} && \frac{1}{2} \langle \mathbf{x}, H\mathbf{x} \rangle + \langle \mathbf{c}, \mathbf{x} \rangle \\ & \text{subject to} && \mathcal{A}\mathbf{x} - \mathbf{b} \in K, \end{aligned} \quad (59)$$

is a special case of (55) by setting

$$f(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, H\mathbf{x} \rangle + \langle \mathbf{c}, \mathbf{x} \rangle. \quad (60)$$

⁷Let $K \subseteq X$. Here and elsewhere we use K^\ominus to denote the polar cone of K defined by $K^\ominus := \{\mathbf{u} \in X \mid \sup \langle K, \mathbf{u} \rangle \leq 0\}$.

It follows from [6, Example 24.2] that

$$\text{Prox}_{\sigma f} = J_{\sigma H}(\cdot - \sigma \mathbf{c}) \text{ where } J_{\sigma H} := (\text{Id} + \sigma H)^{-1}. \quad (61)$$

The above formula uses standard notation for the resolvent of an operator defined by $J_A := (\text{Id} + A)^{-1}$.

Let $(\mathbf{x}, \mathbf{y}) \in X \times Y$. By combining (58) and (61), the PDHG update to solve (59) is

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = T \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} := \begin{pmatrix} J_{\sigma H}(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y} - \sigma \mathbf{c}) \\ P_{K^\ominus}(\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) - \tau \mathbf{b}) \end{pmatrix}. \quad (62)$$

We point out that PDHG for standard QP is recovered by taking $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ and $K = \mathbb{R}_-^m$. This special case is the topic of Section 5.

(ii) Let C be a nonempty closed convex subset of X and let $\mathbf{c} \in X$. The problem

$$\begin{aligned} & \underset{\mathbf{x} \in C}{\text{minimize}} && \langle \mathbf{c}, \mathbf{x} \rangle \\ & \text{subject to} && \mathcal{A}\mathbf{x} - \mathbf{b} \in K \end{aligned} \quad (63)$$

is a special case of (55) by setting

$$f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle + \iota_C(\mathbf{x}). \quad (64)$$

It follows from [6, Example 24.2] that

$$\text{Prox}_{\sigma f} = P_C(\cdot - \sigma \mathbf{c}). \quad (65)$$

Let $(\mathbf{x}, \mathbf{y}) \in X \times Y$. By combining (58) and (65), the PDHG update to solve (63) is

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} = T \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} := \begin{pmatrix} P_C(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y} - \sigma \mathbf{c}) \\ P_{K^\ominus}(\mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) - \tau \mathbf{b}) \end{pmatrix}. \quad (66)$$

A further special case of (63) is obtained when C is a cone and $K = \{\mathbf{0}\}$, yielding the problem commonly known as standard conic primal form. We return to standard conic primal form in Section 6.

4.1 The infimal displacement vector in the conic case

In the remainder of this section, we develop additional properties of the infimal displacement vector for (54).

Proposition 4.2. Let T be defined as in (58), let \mathbf{v} be given by (47), and let

$$\mathbf{v} =: (\mathbf{v}_R, \mathbf{v}_D). \quad (67)$$

Then the following hold:

- (i) $-\mathbf{v}_D = P_{K^\ominus}(-\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R)$.
- (ii) $-\tau \mathcal{A}\mathbf{v}_R = P_K(-\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R)$.
- (iii) $\langle \mathcal{A}\mathbf{v}_R, \mathbf{v}_D \rangle = 0$.
- (iv) $\|\mathbf{v}\|_M^2 = \|\mathbf{v}_R\|^2 / \sigma + \|\mathbf{v}_D\|^2 / \tau$.
- (v) $K = \{\mathbf{0}\} \Rightarrow \mathcal{A}\mathbf{v}_R = \mathbf{0}$, i.e., $\mathbf{v}_R \in \ker \mathcal{A}$.

Suppose further that f is defined as in (60). Then (58) reduces to (62) and we additionally have:

- (vi) $\mathcal{A}^* \mathbf{v}_D + H\mathbf{v}_R = \mathbf{0}$.

- (vii) $\langle \mathbf{v}_R, H\mathbf{v}_R \rangle = 0$.
(viii) $J_{\sigma H}(\mathbf{v}_R) = \mathbf{v}_R + \sigma J_{\sigma H}(\mathcal{A}^* \mathbf{v}_D)$.

Proof. Let $(\mathbf{x}_0, \mathbf{y}_0) \in X \times Y$ and let $((\mathbf{x}_k, \mathbf{y}_k))_{k \in \mathbb{N}}$ be the sequence obtained via the update $(\forall k \in \mathbb{N})$ $(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = T(\mathbf{x}_k, \mathbf{y}_k)$. Applying [Fact 3.8\(i\)&\(ii\)](#) with $z_0 = (\mathbf{x}_0, \mathbf{y}_0)$ in view of [Theorem 3.5\(ii\)](#) implies

$$\mathbf{x}_k - \mathbf{x}_{k+1} \rightarrow \mathbf{v}_R \quad \text{and} \quad \mathbf{y}_k - \mathbf{y}_{k+1} \rightarrow \mathbf{v}_D, \quad (68)$$

and

$$\frac{\mathbf{x}_k}{k} \rightarrow -\mathbf{v}_R \quad \text{and} \quad \frac{\mathbf{y}_k}{k} \rightarrow -\mathbf{v}_D. \quad (69)$$

(i): Because K^\ominus is a cone we have P_{K^\ominus} is positively homogeneous by [\[19, Theorem 5.6\(7\)\]](#). Moreover, P_{K^\ominus} is (firmly) nonexpansive by, e.g., [\[6, Proposition 4.16\]](#), hence continuous. It follows from [\(58\)](#) applied with (\mathbf{x}, \mathbf{y}) replaced by $(\mathbf{x}_k, \mathbf{y}_k)$ that

$$\mathbf{y}_{k+1} = P_{K^\ominus}(\mathbf{y}_k + \tau \mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau \mathbf{b}). \quad (70)$$

Dividing both sides of [\(70\)](#) by $k+1$ and taking the limit as $k \rightarrow \infty$ in view of [\(68\)](#), [\(69\)](#) and [\(3\)](#) yield

$$-\mathbf{v}_D \leftarrow \frac{\mathbf{y}_{k+1}}{k+1} = \frac{1}{k+1} P_{K^\ominus}(\mathbf{y}_k + \tau \mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau \mathbf{b}) \quad (71a)$$

$$= \frac{1}{k+1} P_{K^\ominus}(\mathbf{y}_k - \mathbf{y}_{k+1} + \mathbf{y}_{k+1} + \tau \mathcal{A}\mathbf{x}_{k+1} + \tau \mathcal{A}(\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau \mathbf{b}) \quad (71b)$$

$$= P_{K^\ominus}\left(\frac{\mathbf{y}_k - \mathbf{y}_{k+1}}{k+1} + \frac{\mathbf{y}_{k+1}}{k+1} + \tau \mathcal{A} \frac{\mathbf{x}_{k+1}}{k+1} + \tau \mathcal{A}\left(\frac{\mathbf{x}_k - \mathbf{x}_{k+1}}{k+1}\right) - \tau \frac{\mathbf{b}}{k+1}\right) \quad (71c)$$

$$\rightarrow P_{K^\ominus}(-\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R). \quad (71d)$$

This proves the claim. (ii): Indeed, using (i) and the Moreau decomposition, see, e.g., [\[6, Theorem 6.30\(i\)\]](#) we have

$$P_K(-\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R) = -\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R - P_{K^\ominus}(-\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R) \quad (72a)$$

$$= -\mathbf{v}_D - \tau \mathcal{A}\mathbf{v}_R - (-\mathbf{v}_D) = -\tau \mathcal{A}\mathbf{v}_R. \quad (72b)$$

(iii): Combine (i), (ii) and, e.g., [\[6, Theorem 6.30\(ii\)\]](#). (iv): Indeed, recalling (iii) we have

$$\|\mathbf{v}\|_M^2 = \langle \mathbf{v}, M\mathbf{v} \rangle = \langle (\mathbf{v}_R, \mathbf{v}_D), (\mathbf{v}_R/\sigma - \mathcal{A}^* \mathbf{v}_D, \mathbf{v}_D/\tau - \mathcal{A}\mathbf{v}_R) \rangle \quad (73a)$$

$$= \langle \mathbf{v}_R, \mathbf{v}_R/\sigma - \mathcal{A}^* \mathbf{v}_D \rangle + \langle \mathbf{v}_D, \mathbf{v}_D/\tau - \mathcal{A}\mathbf{v}_R \rangle \quad (73b)$$

$$= \|\mathbf{v}_R\|^2/\sigma - \langle \mathbf{v}_R, \mathcal{A}^* \mathbf{v}_D \rangle + \|\mathbf{v}_D\|^2/\tau - \langle \mathbf{v}_D, \mathcal{A}\mathbf{v}_R \rangle \quad (73c)$$

$$= \|\mathbf{v}_R\|^2/\sigma + \|\mathbf{v}_D\|^2/\tau. \quad (73d)$$

(v): This is a direct consequence of (ii). (vi): It follows from [\(62\)](#) applied with (\mathbf{x}, \mathbf{y}) replaced by $(\mathbf{x}_k, \mathbf{y}_k)$ that

$$\mathbf{x}_k - \mathbf{x}_{k+1} - \sigma \mathcal{A}^*(\mathbf{y}_k - \mathbf{y}_{k+1}) = \sigma \mathcal{A}^* \mathbf{y}_{k+1} + \sigma H \mathbf{x}_{k+1} + \sigma \mathbf{c}. \quad (74)$$

Dividing the above equation by $k+1$ and taking the limit as $k \rightarrow \infty$ in view of [\(68\)](#), [\(69\)](#) and [\(3\)](#) yield

$$\mathbf{0} \leftarrow \frac{1}{k+1}(\mathbf{x}_k - \mathbf{x}_{k+1}) - \sigma \mathcal{A}^*\left(\frac{1}{k+1}(\mathbf{y}_k - \mathbf{y}_{k+1})\right) \quad (75a)$$

$$= \sigma \mathcal{A}^*\left(\frac{\mathbf{y}_{k+1}}{k+1}\right) + \sigma H\left(\frac{\mathbf{x}_{k+1}}{k+1}\right) + \sigma\left(\frac{\mathbf{c}}{k+1}\right) \rightarrow -\sigma \mathcal{A}^* \mathbf{v}_D - \sigma H \mathbf{v}_R. \quad (75b)$$

(vii): It follows from (iii) and (vi) that $0 = \langle \mathbf{v}_R, \mathcal{A}^* \mathbf{v}_D + H \mathbf{v}_R \rangle = \langle \mathbf{v}_R, H \mathbf{v}_R \rangle + \langle \mathbf{v}_R, \mathcal{A}^* \mathbf{v}_D \rangle = \langle \mathbf{v}_R, H \mathbf{v}_R \rangle + \langle \mathcal{A} \mathbf{v}_R, \mathbf{v}_D \rangle = \langle \mathbf{v}_R, H \mathbf{v}_R \rangle$. (viii): It follows from (vi) that $(\text{Id} + \sigma H) \mathbf{v}_R = \mathbf{v}_R - \sigma \mathcal{A}^* \mathbf{v}_D$. Hence, $\mathbf{v}_R = J_{\sigma H}(\mathbf{v}_R - \sigma \mathcal{A}^* \mathbf{v}_D) = J_{\sigma H} \mathbf{v}_R - \sigma J_{\sigma H}(\mathcal{A}^* \mathbf{v}_D)$ as claimed. \blacksquare

5 Application to QP

In this section we set

$$X = \mathbb{R}^n, Y = \mathbb{R}^m, K = \mathbb{R}_-^m, \quad (76)$$

and

$$T \text{ is defined as in (62)}. \quad (77)$$

This means that (59) specializes to

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} \langle \mathbf{x}, H\mathbf{x} \rangle + \langle \mathbf{c}, \mathbf{x} \rangle \\ & \text{subject to} && \mathcal{A}\mathbf{x} \leq \mathbf{b}, \end{aligned} \quad (78)$$

whose Lagrangian dual is well known to be

$$\begin{aligned} & \underset{\mathbf{q} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m}{\text{maximize}} && -\frac{1}{2} \langle \mathbf{q}, H\mathbf{q} \rangle - \langle \mathbf{b}, \mathbf{y} \rangle \\ & \text{subject to} && H\mathbf{q} = -(\mathbf{c} + \mathcal{A}^*\mathbf{y}), \\ & && \mathbf{y} \geq \mathbf{0}. \end{aligned} \quad (79)$$

In view of (61), PDHG for QP is efficient in practice only if a fast method is available to solve the equations implicit in (62), e.g., if $\text{Id} + \sigma H$ has a sparse Cholesky factorization. Unlike, e.g., interior-point methods, the coefficient matrix of the linear system to be solved does not vary with iteration counter k .

5.1 PDHG for QP: Static properties

We start with the following key lemma.

Lemma 5.1. *The following hold:*

- (i) $\partial F + S$ is a polyhedral multifunction⁸.
- (ii) $\text{ran}(\partial F + S)$ is a union of finitely many polyhedral⁹ sets.
- (iii) $\text{ran}(\partial F + S)$ is closed and convex.
- (iv) $\text{ran}(\partial F + S) = (\text{ran } H + \mathcal{A}^*(\mathbb{R}_+^m) + \mathbf{c}) \times (\mathbb{R}_-^m + \text{ran } \mathcal{A} + \mathbf{b})$.
- (v) $\text{ran}(\partial F + S)$ is polyhedral.
- (vi) $\text{ran}(\text{Id} - T) = M^{-1}((\text{ran } H + \mathcal{A}^*(\mathbb{R}_+^m) + \mathbf{c}) \times (\mathbb{R}_-^m + \text{ran } \mathcal{A} + \mathbf{b}))$.
- (vii) $\text{ran}(\text{Id} - T)$ is polyhedral. Hence, $\text{ran}(\text{Id} - T)$ is convex and closed.

Proof. (i): It follows from [26, Example on page 207] that $\partial F + S$ is a polyhedral multifunction. That is $\text{gr}(\partial F + S) = P_1 \cup \dots \cup P_l$ and $(\forall i \in \{1, \dots, l\}) P_i$ is polyhedral. (ii): Observe that $\text{ran}(\partial F + S)$ is the canonical projection of $\text{gr}(\partial F + S)$ onto $\mathbb{R}^n \times \mathbb{R}^m$. It follows from [27, Theorem 19.3] that $\text{ran}(\partial F + S)$ is a finite union of polyhedral sets. (iii): The claim of closedness is a consequence of (ii). The convexity of $\text{ran}(\partial F + S)$ follows from combining (35), [6, Corollary 21.14] and the closedness of $\text{ran}(\partial F + S)$. (iv): It follows from (60) and (56) that $\text{dom } f = \mathbb{R}^n$, $\text{dom } f^* = \text{ran } H + \mathbf{c}$, $\text{dom } g = \mathbb{R}_-^m + \mathbf{b}$ and $\text{dom } g^* = \mathbb{R}_+^m$. Now combine this with Theorem 3.4 in view of (iii). (v):

⁸Following [26], we say that a (possibly) set-valued mapping $A: X \rightrightarrows X$ is a polyhedral multifunction if $\text{gr } A$ is a union of finitely many polyhedral subsets of $X \times X$.

⁹Let $C \subseteq X$. We say that C is polyhedral if C is the intersection of finitely many halfspaces.

Combine (iv) and [27, Theorem 19.3 and Corollary 19.3.2]. (vi): Combine Theorem 3.5(iv) and (iv). (vii): It follows from (vi) and [27, Theorem 19.3], in view of (iv), that $\text{ran}(\text{Id} - T)$ is polyhedral, hence it is closed and convex. ■

Remark 5.2 (when K is a general polyhedral cone). One easily checks that the proof of Lemma 5.1 generalizes seamlessly if we replace \mathbb{R}^m by a general polyhedral cone K . In this case we have $\text{ran}(\partial F + S) = (\text{ran} H + \mathcal{A}^*(K^\ominus) + \mathbf{c}) \times (K + \text{ran} \mathcal{A} + \mathbf{b})$ and $\text{ran}(\text{Id} - T) = M^{-1}((\text{ran} H + \mathcal{A}^*(K^\ominus) + \mathbf{c}) \times (K + \text{ran} \mathcal{A} + \mathbf{b}))$.

Finally, we set

$$\mathbf{v} = (\mathbf{v}_R, \mathbf{v}_D) = \underset{\mathbf{w} \in \overline{\text{ran}(\text{Id} - T)}}{\text{argmin}} \|\mathbf{w}\|_M \in \text{ran}(\text{Id} - T), \quad (80)$$

where the inclusion follows from Lemma 5.1(vii).

Lemma 5.3. *The following hold:*

- (i) $-\mathbf{v}_D = P_{\mathbb{R}_+^m}(-\mathbf{v}_D - \tau \mathcal{A} \mathbf{v}_R)$.
- (ii) $-\tau \mathcal{A} \mathbf{v}_R = P_{\mathbb{R}_+^m}(-\mathbf{v}_D - \tau \mathcal{A} \mathbf{v}_R)$.
- (iii) $\langle \mathcal{A} \mathbf{v}_R, \mathbf{v}_D \rangle = 0$.
- (iv) $\mathcal{A}^* \mathbf{v}_D + H \mathbf{v}_R = \mathbf{0}$.
- (v) $H \mathbf{v}_R = \mathbf{0}$, i.e., $\mathbf{v}_R \in \ker H$.
- (vi) $\mathcal{A}^* \mathbf{v}_D = \mathbf{0}$.
- (vii) $J_{\sigma H}(\mathbf{v}_R) = (\text{Id} + \sigma H)^{-1}(\mathbf{v}_R) = \mathbf{v}_R$.
- (viii) Let $i \in \{1, \dots, m\}$. If $(\mathcal{A} \mathbf{v}_R)_i > 0$ then $(\mathbf{v}_D)_i = 0$.
- (ix) Let $i \in \{1, \dots, m\}$. If $(\mathbf{v}_D)_i < 0$ then $(\mathcal{A} \mathbf{v}_R)_i = 0$.

Proof. We apply Proposition 4.2 with (X, Y, K) replaced by $(\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}_+^m)$. (i)–(iv): This follows from Proposition 4.2(i)–(vi). (v): This follows from Proposition 4.2(vii) and the positive semidefiniteness of H in view of (iv). (vi): Combine (iv) and (v). (vii): This is a direct consequence of Proposition 4.2(viii) and (vi). (viii)&(ix): Combine (i), (ii) and (iii). ■

We conclude this section with Example 5.5 below. We first prove the following auxiliary example.

Example 5.4. Suppose that $m = 4$, $n = 2$, $\mathbf{c} = (1, -2)^\top$, $\mathbf{b} = (-2, 1, 0, 0)^\top$ and $\mathcal{A} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$. Let $\mathbf{u} = -0.15 \cdot (1, 1)^\top$, let $\mathbf{w} = -0.15 \cdot (1, 1, 0, 0)^\top$, and we set $\sigma = \tau = 0.3$. Let $\mathbf{r} \in \mathcal{A}^\top(\mathbb{R}_+^4) + \mathbf{c}$ and let $\mathbf{d} \in \mathbb{R}_+^4 + \text{ran} \mathcal{A} + \mathbf{b}$. Then the following hold:

- (i) $\langle \mathbf{w}, \mathcal{A} \mathbf{u} \rangle = 0$.
- (ii) $\|\mathcal{A}\| = \sqrt{5}$. Hence, $\sigma \tau \|\mathcal{A}\|^2 = 0.45 < 1$.
- (iii) $\frac{1}{\sigma} \mathbf{u} \in \mathcal{A}^\top(\mathbb{R}_+^4) + \mathbf{c}$.
- (iv) $\frac{1}{\tau} \mathbf{w} - \mathcal{A} \mathbf{u} \in \mathbb{R}_+^4 + \text{ran} \mathcal{A} + \mathbf{b}$.
- (v) $\frac{1}{\tau} \mathbf{w} \in \mathbb{R}_+^4 + \text{ran} \mathcal{A} + \mathbf{b}$.
- (vi) $\langle \mathbf{u}, \frac{1}{\sigma} \mathbf{u} - \mathbf{r} \rangle \leq 0$.
- (vii) $\langle \mathbf{w}, \frac{1}{\tau} \mathbf{w} - \mathbf{d} \rangle \leq 0$.
- (viii) $\langle \mathbf{w}, \frac{1}{\tau} \mathbf{w} - \mathcal{A} \mathbf{u} - \mathbf{d} \rangle \leq 0$.

Proof. (i): This is clear. (ii): Indeed, a direct calculation yields that $\mathcal{A}^\top \mathcal{A} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$. Hence $\|\mathcal{A}\|^2 = \|\mathcal{A}^\top \mathcal{A}\| = \lambda_{\max}(\mathcal{A}^\top \mathcal{A}) = 5$. (iii): Indeed, let $\mathbf{x} = (1.5, 0, 0, 0)^\top \in \mathbb{R}_+^4$ and observe that $\frac{1}{\sigma} \mathbf{u} = -0.5 \cdot (1, 1)^\top = \mathcal{A}^\top \mathbf{x} + \mathbf{c} \in \mathcal{A}^\top(\mathbb{R}_+^4) + \mathbf{c}$. (iv): Indeed, let $\mathbf{y} = (-2.5, -1)^\top$

and let $z = (0, 0, -2.65, -1.15)^\top \in \mathbb{R}_-^4$ and observe that $\frac{1}{\tau}w - \mathcal{A}u = -(0.5, 0.5, 0.15, 0.15)^\top = z + \mathcal{A}y + \mathbf{b} \in \mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b}$. (v): Indeed, let $y = (-2.5, -1)^\top$ and let $z = (0, 0, -2.5, -1)^\top \in \mathbb{R}_-^4$ and observe that $\frac{1}{\tau}w = (-0.5, -0.5, 0, 0)^\top = z + \mathcal{A}y + \mathbf{b} \in \mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b}$. (vi): Indeed, let $p = (\alpha, \beta, \gamma, \delta)^\top \in \mathbb{R}_+^4$ be such that $r = \mathcal{A}^\top p + \mathbf{c}$. Then $r = (-\alpha + \beta - \gamma + 1, \alpha - \beta - \delta - 2)^\top$. Thus,

$$\langle u, \frac{1}{\sigma}u - r \rangle = -0.15 \langle (1, 1), (-0.5, -0.5) - (-\alpha + \beta - \gamma + 1, \alpha - \beta - \delta - 2) \rangle \quad (81a)$$

$$= -0.15(-0.5 + \alpha - \beta + \gamma - 1 - 0.5 - \alpha + \beta + \delta + 2) \quad (81b)$$

$$= -0.15(\gamma + \delta) \leq 0. \quad (81c)$$

(vii): Indeed, let $q = (\alpha, \beta, \gamma, \delta)^\top \in \mathbb{R}_-^4$ and let $y = (\rho, \eta) \in \mathbb{R}^2$ be such that $d = q + \mathcal{A}y + \mathbf{b}$. Then $d = (\alpha - \rho + \eta - 2, \beta + \rho - \eta + 1, \gamma - \rho, \delta - \eta)^\top$. Therefore,

$$\langle w, \frac{1}{\tau}w - d \rangle \quad (82a)$$

$$= -0.15 \langle (1, 1, 0, 0), (-0.5, -0.5, 0, 0) - (\alpha - \rho + \eta - 2, \beta + \rho - \eta + 1, \gamma - \rho, \delta - \eta) \rangle \quad (82b)$$

$$= -0.15(-0.5 - \alpha + \rho - \eta + 2 - 0.5 - \beta - \rho + \eta - 1) = 0.15(\alpha + \beta) \leq 0. \quad (82c)$$

(viii): Combine (vii) and (i). ■

Example 5.5. Recalling (78), let H be a 2×2 matrix and let $\mathcal{A}, \mathbf{b}, \mathbf{c}$ be given as in Example 5.4. In view of Example 5.4(ii), we set $\sigma = \tau = 0.3$. Recalling (80), the following hold.

(i) Suppose that $H = 0$. Then $v_R = -0.15 \cdot (1, 1)^\top$ and $v_D = -0.15 \cdot (1, 1, 0, 0)^\top$.

(ii) Suppose that $H = \text{Id}$. Then $v_R = (0, 0)^\top$ and $v_D = -0.15(1, 1, 0, 0)^\top$.

Proof. We set $u = -0.15 \cdot (1, 1)^\top$ and set $w = -0.15 \cdot (1, 1, 0, 0)^\top$. (i): Observe that $\text{ran } H = \{0\}$, hence Lemma 5.1(iv) yields $\text{ran } (\partial F + S) = (\mathcal{A}^\top(\mathbb{R}_+^4) + \mathbf{c}) \times (\mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b})$. On the one hand, it follows from Example 5.4(iii)&(iv) and Lemma 5.1(iv) that $M(u, w) = (\frac{1}{\sigma}u, \frac{1}{\tau}w - \mathcal{A}u) \in \text{ran } (\partial F + S) = (\mathcal{A}^\top(\mathbb{R}_+^4) + \mathbf{c}) \times (\mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b})$. Equivalently, $(u, w) \in M^{-1}(\text{ran } (\partial F + S)) = \text{ran } (\text{Id} - T)$ by Lemma 5.1(vi). On the other hand, Example 5.4(vi)&(viii) implies that $(\forall (r, d) \in \text{ran } (\partial F + S))$ we have $\langle (u, w), M(u, w) - (r, d) \rangle \leq 0$. Altogether in view of (49) this yields $(v_R, v_D) = (u, w)$.

(ii): Observe that $\text{ran } H = \mathbb{R}^2$, hence Lemma 5.1(iv) yields $\text{ran } (\partial F + S) = \mathbb{R}^2 \times (\mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b})$. On the one hand, it follows from Example 5.4(iii)&(iv) and Lemma 5.1(iv) that $M(0, w) = (0, \frac{1}{\tau}w) \in \text{ran } (\partial F + S) = \mathbb{R}^2 \times (\mathbb{R}_-^4 + \text{ran } \mathcal{A} + \mathbf{b})$. Equivalently, $(0, w) \in M^{-1}(\text{ran } (\partial F + S)) = \text{ran } (\text{Id} - T)$ by Lemma 5.1(vi). On the other hand, Example 5.4(vii) implies that $(\forall (r, d) \in \text{ran } (\partial F + S))$ we have $\langle (0, w), M(0, w) - (r, d) \rangle = \langle w, \frac{1}{\tau}w - d \rangle \leq 0$. Altogether in view of (49) this yields $(v_R, v_D) = (0, w)$. ■

5.2 Detecting inconsistency of QP

We start with the following useful lemma.

Lemma 5.6. Let $r \in \text{ran } H + \mathcal{A}^*(\mathbb{R}_+^m) + \mathbf{c}$ and let $d \in \mathbb{R}_-^m + \text{ran } \mathcal{A} + \mathbf{b}$. Then the following hold:

(i) $\langle v_R, \frac{1}{\sigma}v_R - r \rangle \leq 0$.

(ii) $\langle v_D, \frac{1}{\tau}v_D - d \rangle \leq 0$.

In particular, we have

- (iii) $\langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathbf{c} \rangle \leq 0$.
- (iv) $\langle \mathbf{v}_D, \frac{1}{\tau} \mathbf{v}_D - \mathbf{b} \rangle \leq 0$.

Proof. (i): Combine Proposition 3.10(i), Lemma 5.1(iv) and Lemma 5.3(iii) to learn that $0 \geq \langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathcal{A}^* \mathbf{v}_D - \mathbf{r} \rangle = \langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathbf{r} \rangle - \langle \mathbf{v}_R, \mathcal{A}^* \mathbf{v}_D \rangle = \langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathbf{r} \rangle - \langle \mathcal{A} \mathbf{v}_R, \mathbf{v}_D \rangle = \langle \mathbf{v}_R, \frac{1}{\sigma} \mathbf{v}_R - \mathbf{r} \rangle$. (ii): Proceed similarly to (i) but use Proposition 3.10(ii) instead of Proposition 3.10(i). (iii): Apply (i) with \mathbf{r} replaced by \mathbf{c} . (iv): Apply (ii) with \mathbf{d} replaced by \mathbf{b} . ■

Theorem 5.7. *The following hold:*

- (i) *If $\mathbf{v}_R \neq \mathbf{0}$, then the dual (79) is infeasible, and \mathbf{v}_R is an infeasibility certificate.*
- (ii) *If $\mathbf{v}_D \neq \mathbf{0}$, then the primal (78) is infeasible, and \mathbf{v}_D is an infeasibility certificate.*

Proof. (i): Indeed, on the one hand it follows from Lemma 5.6(iii) that $\langle \mathbf{c}, \mathbf{v}_R \rangle \geq \frac{1}{\sigma} \|\mathbf{v}_R\|^2 > 0$. On the other hand Lemma 5.3(ii) implies that $\mathcal{A} \mathbf{v}_R \geq \mathbf{0}$. In addition, $H \mathbf{v}_R = \mathbf{0}$. Thus, taking the inner product of the dual constraint $H \mathbf{q} = -(\mathbf{c} + \mathcal{A}^* \mathbf{y})$ with \mathbf{v}_R yields $0 = \langle \mathbf{c}, \mathbf{v}_R \rangle + \langle \mathcal{A} \mathbf{v}_R, \mathbf{y} \rangle$, which contradicts the other constraint $\mathbf{y} \geq \mathbf{0}$. Altogether, \mathbf{v}_R is an infeasibility certificate for the dual (79). (ii): Indeed, on the one hand it follows from Lemma 5.6(iv) that $\langle \mathbf{b}, \mathbf{v}_D \rangle \geq \frac{1}{\tau} \|\mathbf{v}_D\|^2 > 0$. On the other hand Lemma 5.3(vi) implies that $\mathcal{A}^* \mathbf{v}_D = \mathbf{0}$. Finally, $\mathbf{v}_D \leq \mathbf{0}$ by Lemma 5.3(i). Taking the inner product of \mathbf{v}_D with both sides of the constraint $\mathcal{A} \mathbf{x} \leq \mathbf{b}$ yields $\langle \mathbf{v}_D, \mathcal{A} \mathbf{x} \rangle \geq \langle \mathbf{v}_D, \mathbf{b} \rangle$, i.e., $0 \geq \langle \mathbf{v}_D, \mathbf{b} \rangle$, a contradiction. Altogether, \mathbf{v}_D is an infeasibility certificate for (78). ■

5.3 PDHG for QP: Dynamic behaviour

In this section, we show that in the case of QP, $((\mathbf{x}_k, \mathbf{y}_k) + k\mathbf{v})_{k \in \mathbb{N}}$ converges as $k \rightarrow \infty$. Our result extends the analogous result for LP due to Applegate et al. Our proof technique is somewhat different from that of [1] in that it builds on our previous characterization of $\text{ran}(\text{Id} - T)$. This result strengthens both parts of Fact 3.8 for the special case of PDHG applied to QP.

We start with the following useful lemma.

Lemma 5.8. *Let C be a nonempty closed convex cone of X such that $\text{int } C \neq \emptyset$. Let $\mathbf{w} \in \text{int } C$ and let $M > 0$. Then there exists $\bar{\alpha} \geq 0$ such that $(\forall \alpha \geq \bar{\alpha}) \text{ball}(\mathbf{0}; M) + \alpha \mathbf{w} \subseteq \text{int } C$.*

Proof. Indeed, observe that $\text{int } C$ is a nonempty convex cone of X . If $\text{ball}(\mathbf{0}; M) \subseteq \text{int } C$ then $\bar{\alpha} = 0$ and the conclusion follows. Otherwise, by assumption $(\exists \epsilon > 0)$ such that $\mathbf{w} + \text{ball}(\mathbf{0}; \epsilon) \subseteq \text{int } C$. Let $\bar{\alpha} \geq \frac{M}{\epsilon} + 1$ and observe that $(\forall \alpha \geq \bar{\alpha}) \mathbf{w} + \text{ball}(\mathbf{0}; \epsilon - \frac{M}{\alpha}) \subseteq \text{int } C$. Now, $\text{ball}(\mathbf{0}; M) + \alpha \mathbf{w} \subseteq \text{ball}(\mathbf{0}; M) + \alpha \mathbf{w} + \text{ball}(\mathbf{0}; \alpha \epsilon - M) = \alpha \mathbf{w} + \text{ball}(\mathbf{0}; \alpha \epsilon) = \alpha(\mathbf{w} + \text{ball}(\mathbf{0}; \epsilon)) \subseteq \alpha(\text{int } C) = \text{int } C$. The proof is complete. ■

Corollary 5.9. *Let C be a nonempty closed convex cone of X such that $\text{int } C \neq \emptyset$. Let $\mathbf{w} \in \text{int } C$ and suppose that $(\mathbf{x}_k)_{k \in \mathbb{N}}$ is a bounded sequence in X . Then there exists $\bar{\alpha} \geq 0$ such that $(\forall \alpha \geq \bar{\alpha})$ the following hold:*

- (i) $(\mathbf{x}_k + \alpha \mathbf{w})_{k \in \mathbb{N}}$ lies in $\text{int } C$.
- (ii) $P_C(\mathbf{x}_k + \alpha \mathbf{w}) = \mathbf{x}_k + \alpha \mathbf{w}$.
- (iii) $P_{C^\ominus}(\mathbf{x}_k + \alpha \mathbf{w}) = \mathbf{0}$.

Proof. (i): Because $(\mathbf{x}_k)_{k \in \mathbb{N}}$ is bounded there exists $M > 0$ such that $(\mathbf{x}_k)_{k \in \mathbb{N}}$ lies in $\text{ball}(\mathbf{0}; M)$. Now combine this with Lemma 5.8. (ii)&(iii): This is a direct consequence of (i). ■

Because $\mathbf{v} \in \text{ran}(\text{Id} - T)$ (see (80)) we learn that

$$\text{Fix}(\mathbf{v} + T) \neq \emptyset, \quad (83)$$

where $(\forall \mathbf{z} \in \mathbb{R}^n \times \mathbb{R}^m) (\mathbf{v} + T)(\mathbf{z}) = \mathbf{v} + T\mathbf{z}$.

We are now ready for the main result in this section. We point out that [Theorem 5.10](#) below generalizes [\[1, Theorem 5\]](#) to quadratic programming.

Theorem 5.10. *Let $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^n \times \mathbb{R}^m$. Update via $(\forall k \in \mathbb{N})$*

$$\mathbf{z}_{k+1} = T\mathbf{z}_k. \quad (84)$$

Then $(\exists \alpha \geq 0)$ such that the sequence $(\mathbf{z}_k + k\mathbf{v})_{k \in \mathbb{N}}$ converges to a point in $\alpha\mathbf{v} + \text{Fix}(\mathbf{v} + T)$.

Proof. Observe that the sequence

$$(\mathbf{z}_k + k\mathbf{v})_{k \in \mathbb{N}} = \begin{pmatrix} (\mathbf{x}_k + k\mathbf{v}_R)_{k \in \mathbb{N}} \\ (\mathbf{y}_k + k\mathbf{v}_D)_{k \in \mathbb{N}} \end{pmatrix} \text{ is bounded} \quad (85)$$

by [Fact 3.6](#). Furthermore, the sequence

$$(\mathbf{z}_k - \mathbf{z}_{k+1})_{k \in \mathbb{N}} = \begin{pmatrix} (\mathbf{x}_k - \mathbf{x}_{k+1})_{k \in \mathbb{N}} \\ (\mathbf{y}_k - \mathbf{y}_{k+1})_{k \in \mathbb{N}} \end{pmatrix} \text{ is convergent, hence bounded} \quad (86)$$

by [Fact 3.8\(ii\)](#).

We set $I = \{i \in \{1, \dots, m\} \mid (\mathcal{A}\mathbf{v}_R)_i > 0 \text{ or } (\mathbf{v}_D)_i < 0\}$. We proceed by verifying the following claims.

CLAIM 1: There exists $K \geq 0$ such that $(\forall k \geq K) (\forall i \in I)$

$$((\mathbf{y}_k + \tau\mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau\mathbf{b})_i)_+ = \begin{cases} 0, & (\mathcal{A}\mathbf{v}_R)_i > 0; \\ (\mathbf{y}_k + \tau\mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau\mathbf{b})_i, & (\mathbf{v}_D)_i < 0, \end{cases} \quad (87)$$

and $(\forall k \in \mathbb{N}) (\forall i \in I)$

$$\begin{aligned} & ((\mathbf{y}_k + k\mathbf{v}_D + \tau\mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k + (k+2)\mathbf{v}_R) - \tau\mathbf{b} - K\mathcal{A}\mathbf{v}_R - K\mathbf{v}_D)_i)_+ \\ &= \begin{cases} 0, & (\mathcal{A}\mathbf{v}_R)_i > 0; \\ (\mathbf{y}_k + k\mathbf{v}_D + \tau\mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau\mathbf{b} - K\mathbf{v}_D)_i, & (\mathbf{v}_D)_i < 0. \end{cases} \end{aligned} \quad (88)$$

We proceed by verifying the following two sub-claims.

CLAIM 1-A: There exists $\bar{K}_1 \geq 0$ such that $(\forall k \geq \bar{K}_1) (\forall i \in I)$ (87) holds.

Indeed, let $i \in I$. First, suppose that $(\mathcal{A}\mathbf{v}_R)_i > 0$. In this case, [Lemma 5.3\(viii\)](#) implies that $(\mathbf{v}_D)_i = 0$. Therefore

$$\begin{aligned} & (\mathbf{y}_k + \tau\mathcal{A}(2\mathbf{x}_{k+1} - \mathbf{x}_k) - \tau\mathbf{b})_i \\ &= (\mathbf{y}_k + \tau\mathcal{A}(\mathbf{x}_{k+1} - \mathbf{x}_k + (\mathbf{x}_{k+1} + (k+1)\mathbf{v}_R)) - \tau\mathbf{b} - \tau(k+1)\mathcal{A}\mathbf{v}_R)_i. \end{aligned} \quad (89)$$

It follows from (85), the continuity of \mathcal{A} , (86) and (85) again that the sequences $((\mathbf{y}_k)_i)_{k \in \mathbb{N}} = ((\mathbf{y}_k + k\mathbf{v}_D)_i)_{k \in \mathbb{N}}$ and $((\tau\mathcal{A}(\mathbf{x}_{k+1} - \mathbf{x}_k + (\mathbf{x}_{k+1} + (k+1)\mathbf{v}_R)) - \tau\mathbf{b})_i)_{k \in \mathbb{N}}$ are bounded. Hence, their

sum is bounded. Applying [Corollary 5.9\(iii\)](#) with C replaced by $] -\infty, 0]$, $(x_k)_{k \in \mathbb{N}}$ replaced by $((y_k + kv_D + \tau\mathcal{A}(x_{k+1} - x_k + (x_{k+1} + (k+1)v_R)) - \tau\mathbf{b}))_{k \in \mathbb{N}}$ and w replaced by $(-\tau\mathcal{A}v_R)_i < 0$ we learn that there exists K_i such that

$$(\forall k \geq K_i) \quad ((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b}))_+ = ((y_k + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b}))_+ = 0. \quad (90)$$

Now, suppose that $(v_D)_i < 0$. In this case, [Lemma 5.3\(ix\)](#) implies that $(\mathcal{A}v_R)_i = 0$. Therefore,

$$(y_k + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b})_i = (y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b} - kv_D)_i. \quad (91)$$

It follows from [\(85\)](#), the continuity of \mathcal{A} , [\(86\)](#) and [\(85\)](#) again that the sequences $((y_k + kv_D)_i)_{k \in \mathbb{N}}$ and $((\tau\mathcal{A}(x_{k+1} - x_k + x_{k+1}) - \tau\mathbf{b}))_{k \in \mathbb{N}} = ((\tau\mathcal{A}(x_{k+1} - x_k + (x_{k+1} + (k+1)v_R)) - \tau\mathbf{b}))_{k \in \mathbb{N}}$ are bounded. Hence, their sum is bounded. Applying [Corollary 5.9\(ii\)](#) with C replaced by $[0, +\infty[$, $(x_k)_{k \in \mathbb{N}}$ replaced by $((y_k + kv_D + \tau\mathcal{A}(x_{k+1} - x_k + x_{k+1}) - \tau\mathbf{b}))_{k \in \mathbb{N}}$ and w replaced by $(-v_D)_i > 0$ we learn that there exists K_i such that

$$(\forall k \geq K_i) \quad ((y_k + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b}))_+ = (y_k + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b})_i. \quad (92)$$

We set

$$\bar{K}_1 = \max_{i \in I} \{K_i\}. \quad (93)$$

Then \bar{K}_1 satisfies [\(87\)](#) in view of [\(90\)](#) and [\(92\)](#).

CLAIM 1-B: There exists $\bar{K}_2 \geq 0$ such that $(\forall K \geq \bar{K}_2) (\forall k \in \mathbb{N}) (\forall i \in I)$ [\(88\)](#) holds.

Indeed, observe that $(\forall k \in \mathbb{N})$

$$\begin{aligned} & y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b} \\ &= y_k + kv_D + \tau\mathcal{A}(2(x_{k+1} + (k+1)v_R) - (x_k + kv_R)) - \tau\mathbf{b}. \end{aligned} \quad (94)$$

Hence, $(\forall i \in I)$ $(y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b})_i$ is bounded by [\(85\)](#). As before, if $(\mathcal{A}v_R)_i > 0$ then [Lemma 5.3\(viii\)](#) implies that $(v_D)_i = 0$. Applying [Corollary 5.9\(ii\)&\(iii\)](#) with C replaced by $] -\infty, 0]$, $(x_k)_{k \in \mathbb{N}}$ replaced by $((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b}))_{k \in \mathbb{N}}$ and w replaced by $(-\mathcal{A}v_R)_i < 0$ implies that there exists $\hat{K}_i \geq 0$ such that $(\forall \hat{K} \geq \hat{K}_i)$ such that $(\forall k \in \mathbb{N})$

$$\begin{aligned} & ((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b} - K\mathcal{A}v_R - Kv_D)_i)_+ \\ &= ((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b} - K\mathcal{A}v_R)_i)_+ = 0. \end{aligned} \quad (95)$$

If $(v_D)_i < 0$ then [Lemma 5.3\(viii\)](#) implies that $(\mathcal{A}v_R)_i = 0$. Applying [Corollary 5.9\(ii\)&\(iii\)](#) with C replaced by $] -\infty, 0]$, $(x_k)_{k \in \mathbb{N}}$ replaced by $((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b}))_{k \in \mathbb{N}}$ and w replaced by $(-v_D)_i > 0$ implies that there exists $\hat{K}_i \geq 0$ such that $(\forall k \in \mathbb{N}) (\forall K \geq \hat{K}_i \geq 0)$

$$\begin{aligned} & ((y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k + (k+2)v_R) - \tau\mathbf{b} - K\mathcal{A}v_R - Kv_D)_i)_+ \\ &= (y_k + kv_D + \tau\mathcal{A}(2x_{k+1} - x_k) - \tau\mathbf{b} - Kv_D)_i. \end{aligned} \quad (96)$$

We set

$$\bar{K}_2 = \max_{i \in I} \{\hat{K}_i\}. \quad (97)$$

Then \bar{K}_2 satisfies [\(87\)](#) in view of [\(90\)](#) and [\(92\)](#).

Finally, we set $K = \max\{\bar{K}_1, \bar{K}_2\}$. This verifies CLAIM 1.

CLAIM 2: We have ($\forall k \in \mathbb{N}$)

$$z_{k+K} + kv = (v + T)^k(z_K). \quad (98)$$

To simplify the notation, we set ($\forall k \in \mathbb{N}$)

$$w_k := (v + T)^k(z_K) := \begin{pmatrix} p_k \\ q_k \end{pmatrix}. \quad (99)$$

Therefore, (98) reduces to proving

$$\begin{pmatrix} (p_k)_{k \in \mathbb{N}} \\ (q_k)_{k \in \mathbb{N}} \end{pmatrix} = \begin{pmatrix} (x_{k+K} + kv_R)_{k \in \mathbb{N}} \\ (y_{k+K} + kv_D)_{k \in \mathbb{N}} \end{pmatrix}. \quad (100)$$

We use induction on k . The base case at $k = 0$ is clear. Now suppose that for some $k \geq 0$ (98) holds. We first verify that

$$(p_k)_{k \in \mathbb{N}} = (x_{k+K} + kv_R)_{k \in \mathbb{N}} = (x_{k+K} + (k+K)v_R - Kv_R)_{k \in \mathbb{N}}. \quad (101)$$

Indeed, the inductive hypothesis, the linearity of $J_H = (\text{Id} + H)^{-1}$, and Lemma 5.3(vi)&(vii) yield

$$p_{k+1} = v_R + J_{\sigma H}(p_k - \sigma \mathcal{A}^* q_k - \sigma c) \quad (102a)$$

$$= v_R + J_{\sigma H}(x_{k+K} + kv_R - \sigma \mathcal{A}^*(y_{k+K} + kv_D) - \sigma c) \quad (102b)$$

$$= v_R + J_{\sigma H}(x_{k+K} - \sigma \mathcal{A}^* y_{k+K} + \sigma k \mathcal{A}^* v_D - \sigma c) + k J_{\sigma H}(v_R) \quad (102c)$$

$$= v_R + J_{\sigma H}(x_{k+K} - \sigma \mathcal{A}^* y_{k+K} - \sigma c) + kv_R \quad (102d)$$

$$= J_{\sigma H}(x_{k+K} - \sigma \mathcal{A}^* y_{k+K} - \sigma c) + (k+1)v_R \quad (102e)$$

$$= x_{k+K+1} + (k+1)v_R. \quad (102f)$$

We now verify that

$$(q_k)_{k \in \mathbb{N}} = (y_{k+K} + kv_D)_{k \in \mathbb{N}} = (y_{k+K} + (k+K)v_D - Kv_D)_{k \in \mathbb{N}}. \quad (103)$$

It follows from the the inductive hypothesis, (88) and (101) that

$$q_{k+1} = v_D + (q_k + \tau \mathcal{A}(2p_{k+1} - p_k) - \tau b)_+ \quad (104a)$$

$$= v_D + (y_{k+K} + (k+K)v_D + \tau \mathcal{A}(2(x_{k+K+1} + (k+K+1)v_R) - (x_{k+K} + (k+K)v_R)) - \tau b - \tau K \mathcal{A} v_R - Kv_D)_+. \quad (104b)$$

Let $i \in \{1, \dots, m\}$. We examine the following cases.

CASE 1: $(\mathcal{A}v_R)_i > 0$. In this case $(v_D)_i = 0$. On the one hand, (87) implies that

$$(y_{k+K+1} + (k+1)v_D)_i = 0 + 0 = 0. \quad (105)$$

On the other hand, in view of (88) and (104b) we have

$$(q_{k+1})_i = (v_D)_i + ((q_k + \tau \mathcal{A}(2p_{k+1} - p_k) - \tau b)_i)_+ = 0 + 0 = 0 = (y_{k+K+1} + (k+1)v_D)_i. \quad (106)$$

CASE 2: $(v_D)_i < 0$. In this case $(\mathcal{A}v_R)_i = 0$. On the one hand, (87) implies that

$$(y_{k+K+1})_i = (y_{k+K} + \tau \mathcal{A}(2x_{k+K+1} - x_{k+K}) - \tau b)_i. \quad (107)$$

On the other hand, in view of (88) and (104b) we have

$$(\mathbf{q}_{k+1})_i = (\mathbf{v}_D)_i + ((\mathbf{q}_k + \tau\mathcal{A}(2\mathbf{p}_{k+1} - \mathbf{p}_k) - \tau\mathbf{b})_i)_+ \quad (108a)$$

$$= (\mathbf{v}_D)_i + (\mathbf{y}_{k+K} + (k+K)\mathbf{v}_D + \tau\mathcal{A}(2\mathbf{x}_{k+K+1} - \mathbf{x}_{k+K}) - \tau\mathbf{b} - K\mathbf{v}_D)_i \quad (108b)$$

$$= (\mathbf{y}_{k+K} + (k+1)\mathbf{v}_D + \tau\mathcal{A}(2\mathbf{x}_{k+K+1} - \mathbf{x}_{k+K}) - \tau\mathbf{b})_i \quad (108c)$$

$$= (\mathbf{y}_{k+K+1})_i + (k+1)(\mathbf{v}_D)_i. \quad (108d)$$

CASE 3: $(\mathcal{A}\mathbf{v}_R)_i = (\mathbf{v}_D)_i = 0$. In this case, it is straightforward to verify the inductive step and the conclusion is obvious.

CLAIM 3: There exists $\alpha \geq 0$ such that the sequence $(\mathbf{z}_k + k\mathbf{v})_{k \in \mathbb{N}}$ converges to a $\bar{\mathbf{z}} \in \alpha\mathbf{v} + \text{Fix}(\mathbf{v} + T)$. Indeed, (99) means that \mathbf{w}_k can be obtained by iterating $(\mathbf{v} + T)$ on $\mathbf{w}_0 \equiv \mathbf{z}_K$. Since T is firmly nonexpansive (Theorem 3.5(ii)), so is $\mathbf{v} + T$. By Example 5.18 of [6], in view of (83), we learn that the sequence $(\mathbf{w}_k)_{k \in \mathbb{N}}$ converges to a point in $\text{Fix}(\mathbf{v} + T)$. Now combine with (98) to learn that $\mathbf{z}_k + k\mathbf{v}$ converges to a point in $K\mathbf{v} + \text{Fix}(\mathbf{v} + T) = K\mathbf{v} + \text{Fix}(\mathbf{v} + T)$, and the conclusion follows by recalling that $\text{Fix}(\mathbf{v} + T) = R_- \cdot \mathbf{v} + \text{Fix}(\mathbf{v} + T)$ by Fact 3.7. \blacksquare

5.4 A numerical example

Consider the linear program (LP) (respectively the quadratic program (QP))

$$\begin{aligned} \text{minimize} \quad & x_1 - 2x_2 \\ \text{subject to} \quad & -x_1 + x_2 \leq -2 \\ & x_1 - x_2 \leq 1 \quad (\text{LP}) \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

$$\begin{aligned} \text{minimize} \quad & 0.5x_1^2 + 0.5x_2^2 + x_1 - 2x_2 \\ \text{subject to} \quad & -x_1 + x_2 \leq -2 \\ & x_1 - x_2 \leq 1 \quad (\text{QP}) \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

which was given in Example 5.5(i) (respectively Example 5.5(ii)). In this section, we provide numerical illustrations of Theorem 5.10 when applied to (LP) and (QP). Additionally, we numerically verify the conclusion of Example 5.5. For both (LP) and (QP) we set $\sigma = \tau = 0.3$, $\mathbf{x}_0 = (0, 0)$, $\mathbf{y}_0 = (0, 0, -1, -1)$, and $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0)^\top$. Finally, following the notation of Theorem 5.10 we set $(\forall k \in \mathbb{N}) \mathbf{z}_k = T^k \mathbf{z}_0$. Let $k \in \mathbb{N}$. We denote the component of $(\mathbf{z}_k + k\mathbf{v} - (\mathbf{v} + T)(\mathbf{z}_k + k\mathbf{v}))$ corresponding to \mathbf{x}_k (respectively \mathbf{y}_k) by $(\mathbf{z}_k + k\mathbf{v} - (\mathbf{v} + T)(\mathbf{z}_k + k\mathbf{v}))_x$ (respectively $(\mathbf{z}_k + k\mathbf{v} - (\mathbf{v} + T)(\mathbf{z}_k + k\mathbf{v}))_y$).

Remark 5.11. *Some comments are in order.*

- (i) Let $\mathbf{w}_0 \in Z$ and let $Q: Z \rightarrow Z$ be an affine firmly nonexpansive operator. We set $(\mathbf{w}_k)_{k \in \mathbb{N}} = (Q^k \mathbf{w}_0)_{k \in \mathbb{N}}$ and we let \mathbf{v}_Q be the minimal norm vector in $\text{ran}(\text{Id} - Q)$. The authors in [9, Theorem 3.2] proved that $\mathbf{w}_k + k\mathbf{v}_Q$ converges to a point in $\text{Fix}(\mathbf{v}_Q + Q)$.
- (ii) In view of (i) one wonders if the limit of $(\mathbf{z}_k + k\mathbf{v})$ lies $\text{Fix}(\mathbf{v} + T)$. Our numerical experiments provide a negative answer to this question, which proves the tightness of the conclusion of Theorem 5.10. Indeed, as the plots in Figure 1 and Figure 2 below show, the sequence $\mathbf{z}_k + k\mathbf{v} - (\mathbf{v} + T)(\mathbf{z}_k + k\mathbf{v}) \rightarrow \mathbf{u}^* \neq \mathbf{0}$. Recalling Theorem 5.10, this in turn implies that $\mathbf{z}_k + k\mathbf{v} \rightarrow \mathbf{z}^* \notin \text{Fix}(\mathbf{v} + T)$.

5.5 Computing the infimal displacement vector

In this section, we derive a characterization of $\text{ran}(\text{Id} - T)$ for (59) as the solution to a system of convex constraints in the case that K is polyhedral. A special case is $K = \mathbb{R}_-^m$, i.e., QP (78). This

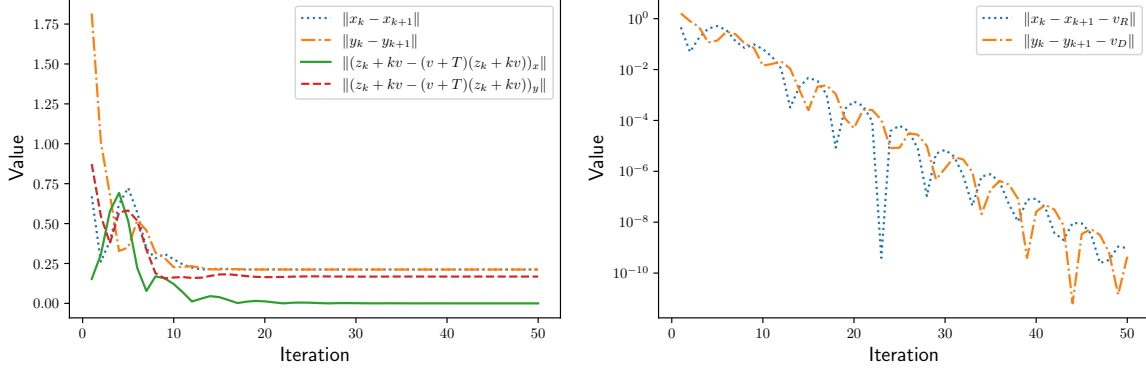


Figure 1: Python plots to illustrate the iterative behavior of PDHG when applied to solve (LP). Left: The first 50 terms of the sequences $(\|x_k - x_{k+1}\|)_{k \in \mathbb{N}}$ (the blue dotted curve) and $(\|y_k - y_{k+1}\|)_{k \in \mathbb{N}}$ (the orange dash-dotted curve) are depicted. Also, the first 50 terms of both components of the sequence $(\|z_k + kv - (v + T)(z_k + kv)\|)_{k \in \mathbb{N}}$ are depicted (the solid green curve and the dashed red curve). Right: The first 50 terms of the sequences $(\|x_k - x_{k+1} - v_R\|)_{k \in \mathbb{N}}$ (the orange dotted curve) and $(\|y_k - y_{k+1} - v_D\|)_{k \in \mathbb{N}}$ (the blue dotted curve) are depicted.

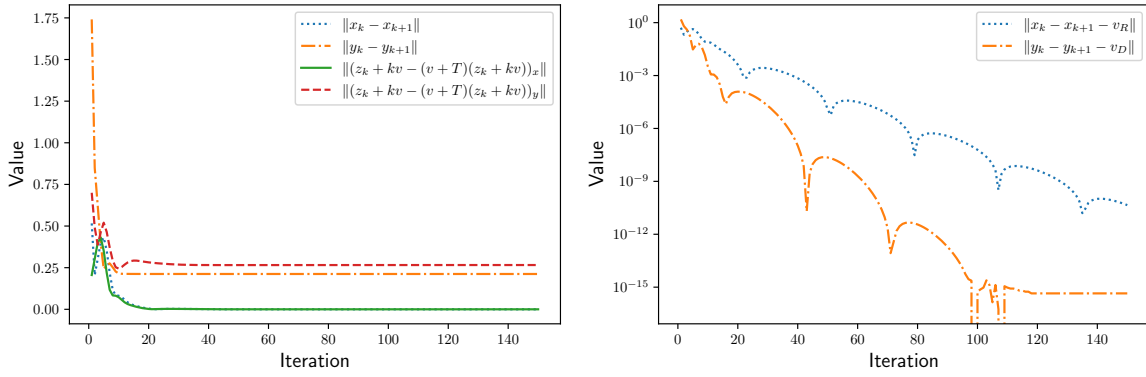


Figure 2: Python plots to illustrate the iterative behavior of PDHG when applied to solve (QP). Left: The first 150 terms of the sequences $(\|x_k - x_{k+1}\|)_{k \in \mathbb{N}}$ (the blue dotted curve) and $(\|y_k - y_{k+1}\|)_{k \in \mathbb{N}}$ (the orange dash-dotted curve) are depicted. Also, the first 150 terms of both components of the sequence $(\|z_k + kv - (v + T)(z_k + kv)\|)_{k \in \mathbb{N}}$ are depicted (the solid green curve and the dashed red curve). Right: The first 150 terms of the sequences $(\|x_k - x_{k+1} - v_R\|)_{k \in \mathbb{N}}$ (the orange dotted curve) and $(\|y_k - y_{k+1} - v_D\|)_{k \in \mathbb{N}}$ (the blue dotted curve) are depicted.

formula is useful in case one wants to compute with $\text{ran}(\text{Id} - T)$, e.g., the determination of the infimal displacement vector via an interior-point method.

We note that Lemma 5.1(vi) already yields such a characterization. However, a naive translation of (vi) to a system of constraints introduces four auxiliary vectors. The following result shows that two auxiliary vectors (denoted w and y below) suffice.

Lemma 5.12. *In the setting of Problem (59), suppose that K is a polyhedral cone. Define*

$$V := \left\{ (r, d) \in X \times Y \mid \exists (w, y) \in X \times Y \text{ such that } \begin{cases} \frac{1}{\tau}d - (Aw + b) \in K, \\ y - d \in K^\ominus, \\ \frac{1}{\sigma}r - Hr = -Hw + \mathcal{A}^*y + c \end{cases} \right\}. \quad (109)$$

Then

$$V = M^{-1}(\text{ran}(\partial F + S)) = \text{ran}(\text{Id} - T). \quad (110)$$

Proof. In view of [Remark 5.2](#), it is sufficient to verify the first identity in (110). Let $(r, d) \in V$. Then there exists $(w, y) \in X \times Y$ such that

$$\begin{aligned} \frac{1}{\sigma}r - \mathcal{A}^*d &= Hr - Hw + \mathcal{A}^*y + c - \mathcal{A}^*d \in \text{ran} H + \mathcal{A}^*(K^\ominus) + c, \\ \frac{1}{\tau}d - \mathcal{A}r &= \frac{1}{\tau}d - (\mathcal{A}w + b) + (\mathcal{A}w + b) - \mathcal{A}r \in K + \text{ran} \mathcal{A} + b. \end{aligned} \quad (111)$$

Recalling [Remark 5.2](#), for simplicity we set $R := \text{ran}(\partial F + S) = (\text{ran} H + \mathcal{A}^*(K^\ominus) + c) \times (K + \text{ran} \mathcal{A} + b)$. The inclusion $V \subseteq M^{-1}R$ follows from the definition of M given in (23) and the nonsingularity of M due to the choices of τ, σ . We now show that $M^{-1}R \subseteq V$. Indeed let $z := (z_1, z_2) \in R$. Then there exist $x \in X, s \in K^\ominus, t \in K, u \in X$ such that

$$z = (z_1, z_2) = (Hx + \mathcal{A}^*s + c, t - Au + b). \quad (112)$$

We claim that there exist $x^* \in X, s^* \in K^\ominus, t^* \in K, u^* \in X$ such that

$$z = (Hx^* + \mathcal{A}^*s^* + c, t^* - Au^* + b) \text{ and } Hx^* = Hu^*. \quad (113)$$

To this end consider the problem:

$$\begin{aligned} \underset{\bar{u} \in X}{\text{minimize}} \quad & \frac{1}{2} \langle \bar{u}, H\bar{u} \rangle - \langle z_1 - c, \bar{u} \rangle \\ \text{subject to} \quad & \mathcal{A}\bar{u} - b + z_2 \in K. \end{aligned} \quad (114)$$

Standard techniques yield that the Lagrangian dual of (114) is

$$\begin{aligned} \underset{\bar{s} \in Y, \bar{x} \in X}{\text{maximize}} \quad & -\frac{1}{2} \langle \bar{x}, H\bar{x} \rangle - \langle b - z_2, \bar{s} \rangle \\ \text{subject to} \quad & -\mathcal{A}^*\bar{s} + z_1 - c = H\bar{x} \\ & \bar{s} \in K^\ominus. \end{aligned} \quad (115)$$

It follows from (112) that u satisfies the primal constraint and the pair (x, s) satisfies the dual constraints. Therefore, because K is a polyhedral cone, strong duality holds for the primal-dual problem (114)–(115) (see, e.g., [15, Comment on Page 227]). Let $(u^*, (x^*, s^*))$ denote its primal-dual optimal solution. Then there exists $(t^*, s^*) \in K \times K^\ominus$ such that

$$z_1 = Hx^* + \mathcal{A}^*s^* + c \text{ (dual feasibility) and } z_2 = t^* - Au^* + b \text{ (primal feasibility)}. \quad (116)$$

Moreover, strong duality and KKT conditions imply that $0 = Hu^* - (z_1 - c) + \mathcal{A}^*s^* = Hu^* - Hx^*$. This proves (113). Now define $(r, d) := M^{-1}z$, i.e., $(z_1, z_2) = M(r, d) = (\frac{1}{\sigma}r - \mathcal{A}^*d, \frac{1}{\tau}d - \mathcal{A}r)$. This and (116) implies

$$\left(\frac{1}{\sigma}r, \frac{1}{\tau}d\right) = (Hx^* + \mathcal{A}^*s^* + c + \mathcal{A}^*d, t^* - Au^* + b + \mathcal{A}r). \quad (117)$$

Define $w := r - u^*$ and $y := d + s^*$. In view of (113), (117) and (116) we have

$$\begin{aligned} \frac{1}{\tau}d - (\mathcal{A}w + b) &= \mathcal{A}r + t^* - Au^* + b - (\mathcal{A}w + b) = t^* \in K \\ y - d &= s^* \in K^\ominus \\ \frac{1}{\sigma}r - Hr &= \mathcal{A}^*d + Hx^* + \mathcal{A}^*s^* + c - Hr = -Hw + \mathcal{A}^*y + c, \end{aligned}$$

which implies that $(r, d) \in V$. The inclusion $M^{-1}R \subseteq V$ follows from the construction $(r, d) = M^{-1}z$. \blacksquare

6 Application to standard conic primal form

In this section, we consider problems of the form (63) under the assumptions

$$C \text{ is a nonempty closed convex cone of } X, c \in X, \text{ and } K = \{0\}. \quad (118)$$

In other words, the problem under consideration is:

$$\begin{aligned} & \underset{x \in C}{\text{minimize}} && \langle c, x \rangle \\ & \text{subject to} && \mathcal{A}x - b = 0. \end{aligned} \quad (119)$$

Problem (119) is commonly known as standard conic primal form since it generalizes linear programming in standard equality form, which takes $C = \mathbb{R}_+^n$. However, the results in this section extend beyond LP since polyhedrality is not assumed. Specializing (66), the PDHG update to solve (119) is

$$\begin{pmatrix} x^+ \\ y^+ \end{pmatrix} = T \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} P_C(x - \sigma \mathcal{A}^* y - c) \\ y + \tau \mathcal{A}(2x^+ - x) - \tau b \end{pmatrix}. \quad (120)$$

Lemma 6.1. For (120) we have

$$\overline{\text{ran}}(\text{Id} - T) = M^{-1} \left((\overline{C^\ominus + \text{ran } \mathcal{A}^* + c}) \times (b - \overline{\mathcal{A}(C)}) \right). \quad (121)$$

Proof. Recalling (56) and (64), on the one hand we have

$$\text{dom } f = C, \text{ dom } g = \{b\} \text{ and } \text{dom } g^* = X. \quad (122)$$

On the other hand, it follows from [6, Corollary 16.39 and Corollary 16.30], [31, Theorem 3.1] and [6, Corollary 6.50] that

$$\overline{\text{dom}} f^* = \overline{\text{dom}} (\langle c, x \rangle + \iota_C)^* = \overline{\text{dom}} \partial(\langle c, x \rangle + \iota_C)^* \quad (123a)$$

$$= \overline{\text{ran}} \partial(\langle c, x \rangle + \iota_C) = \overline{\text{ran}} (c + N_C) \quad (123b)$$

$$= c + \overline{\text{ran}} N_C = c + C^\ominus. \quad (123c)$$

Now combine (122), (123) and Theorem 3.5(vi). ■

The following lemma, analogous to Lemma 5.12, presents a parsimonious description of $\overline{\text{ran}}(\text{Id} - T)$ via constraints in the context of the standard conic primal form.

Lemma 6.2. In the setting of Problem (120), define $R := ((C^\ominus + \text{ran } \mathcal{A}^* + c) \times (b - \mathcal{A}(C)))$, and

$$V := \left\{ (r, d) \in X \times Y \mid \exists (w, y) \in X \times Y \text{ such that } \begin{array}{l} \frac{1}{\tau} d = \mathcal{A}w + b, \\ r - w \in C, \\ \frac{1}{\sigma} r - (\mathcal{A}^* y + c) \in C^\ominus \end{array} \right\}. \quad (124)$$

Then

$$V = M^{-1}R. \quad (125)$$

Proof. Let $(r, d) \in V$. Then there exists w, y such that

$$\begin{aligned} \frac{1}{\sigma} r - \mathcal{A}^* d &= \frac{1}{\sigma} r - (\mathcal{A}^* y + c) + (\mathcal{A}^* y + c) - \mathcal{A}^* d \in C^\ominus + \text{ran } \mathcal{A}^* + c, \\ \frac{1}{\tau} d - \mathcal{A}r &= \mathcal{A}(w - r) + b \in b - \mathcal{A}(C). \end{aligned}$$

The inclusion $V \subseteq M^{-1}R$ follows from the definition of M (see (23)). We now show that $M^{-1}R \subseteq V$. To this end let $z \in R$ and define $(r, d) := M^{-1}z$. Then there exist $s \in C^\ominus, t \in Y, u \in C$ such that

$$z = (s + \mathcal{A}^*t + c, b - \mathcal{A}u) = \left(\frac{1}{\sigma}r - \mathcal{A}^*d, \frac{1}{\tau}d - \mathcal{A}r\right) = M(r, d). \quad (126)$$

Define $w := r - u$ and $y := d + t$. Combining this with (126) yields

$$\begin{aligned} \frac{1}{\tau}d &= \mathcal{A}r + b - \mathcal{A}u = \mathcal{A}w + b \\ r - w &= u \in C \\ \frac{1}{\sigma}r - (\mathcal{A}^*y + c) &= \mathcal{A}^*d + s + \mathcal{A}^*t + c - (\mathcal{A}^*y + c) = s + \mathcal{A}^*y + c - (\mathcal{A}^*y + c) \in C^\ominus, \end{aligned}$$

which implies that $M^{-1}z = (r, d) \in V$. This completes the proof. \blacksquare

Remark 6.3. Let R and V be defined as in Lemma 6.2. In view of (121) it is clear that $\overline{\text{ran}}(\text{Id} - T) = \overline{V}$.

6.1 Special case: $\ker \mathcal{A} \cap C = \{0\}$

In this section, we establish that $v \in \text{ran}(\text{Id} - T)$ for a subclass of (119). The general version of our result is in Proposition 6.9, and then the general result is specialized to a particular problem class in Section 6.2. We start with the following useful lemma.

Lemma 6.4. Suppose X finite-dimensional and that C_1 and C_2 are nearly convex¹⁰ subsets of X such that $\overline{C_1} = \overline{C_2}$. Then the following hold.

- (i) $\text{ri } C_1 = \text{ri } C_2$.
- (ii) Suppose that $(\exists i \in \{1, 2\}) C_i = X$. Then $C_1 = C_2 = X$.

Proof. (i): This is [8, Proposition 2.12]. (ii): Indeed, without loss of generality suppose that $C_1 = X$. Observe that (i) implies $X = \text{ri } X = \text{ri } C_1 = \text{ri } C_2 \subseteq C_2 \subseteq X$. Hence, $C_2 = X$ as claimed. \blacksquare

We now have the following corollary which will be used in the sequel.

Corollary 6.5. Suppose that X is finite-dimensional and that K_1 and K_2 are closed convex cones of X such that $K_1 \cap K_2 = \{0\}$. Then $K_1^\ominus + K_2^\ominus = X$.

Proof. Indeed, it follows from [27, Corollary 16.4.2.] that $\overline{K_1^\ominus + K_2^\ominus} = (K_1 \cap K_2)^\ominus = \{0\}^\ominus = X$. Now combine this with Lemma 6.4(ii) applied with $C_1 = K_1^\ominus + K_2^\ominus$ and $C_2 = X$. \blacksquare

Lemma 6.6. Recalling (120), for Problem (119), the following hold:

- (i) $v_R \in -C$.
- (ii) $\sigma \mathcal{A}^*v_D = P_{C^\ominus}(-v_R + \sigma \mathcal{A}^*v_D) \in C^\ominus$.
- (iii) $-v_R = P_C(-v_R + \sigma \mathcal{A}^*v_D) \in C$.
- (iv) $v_D \in (\mathcal{A}(C))^\ominus$.
- (v) Suppose that $\ker \mathcal{A} \cap C = \{0\}$. Then
 - (a) $v_R = 0$.
 - (b) $C^\ominus + \text{ran } \mathcal{A}^* = X$.

¹⁰Suppose that X is finite-dimensional. A subset E of X is *nearly convex* if there exists a convex set $C \subseteq X$ such that $C \subseteq E \subseteq \overline{C}$.

Proof. Let $z_0 = (x_0, y_0) \in X \times Y$. Update via $z_{k+1} = Tz_k$, where T is defined as in (120). Then

$$\text{the sequence } (x_k)_{k \in \mathbb{N}} \text{ lies in } C. \quad (127)$$

(i): Because C is a cone, (127) implies that $(x_k/k)_{k \geq 1}$ lies in C . It follows from Fact 3.8(i) that

$$x_k/k \rightarrow -v_R \in C, \quad (128)$$

where the inclusion follows from the closedness of C . (ii): It follows from (120) applied with x replaced by x_k and the Moreau decomposition, see, e.g., [6, Theorem 6.30], that

$$x_k - x_{k+1} - \sigma \mathcal{A}^* y_k - c = x_k - P_C(x_k - \sigma \mathcal{A}^* y_k - c) - \sigma \mathcal{A}^* y_k - c = P_{C^\ominus}(x_k - \sigma \mathcal{A}^* y_k - c). \quad (129)$$

Dividing the above equation by $k \geq 1$, using the positive homogeneity of P_{C^\ominus} (see [19, Theorem 5.6(7)]) and taking the limit as $k \rightarrow \infty$ in view of Fact 3.8(i)&(ii) and the continuity of P_{C^\ominus} we learn that

$$\sigma \mathcal{A}^* v_D \leftarrow \frac{x_k - x_{k+1}}{k} + \sigma \mathcal{A}^* \left(\frac{-y_k}{k} \right) - \left(\frac{c}{k} \right) = P_{C^\ominus} \left(\frac{x_k}{k} + \sigma \mathcal{A}^* \left(\frac{-y_k}{k} \right) - \frac{c}{k} \right) \rightarrow P_{C^\ominus}(-v_R + \sigma \mathcal{A}^* v_D). \quad (130)$$

That is, $\sigma \mathcal{A}^* v_D = P_{C^\ominus}(-v_R + \sigma \mathcal{A}^* v_D) \in C^\ominus$. (iii): It follows from Proposition 4.2(iii) that $\langle v_R, \mathcal{A}^* v_D \rangle = \langle \mathcal{A} v_R, v_D \rangle = 0$. Now combine this with (i) and (ii) in view of, e.g., [6, Proposition 6.28]. (iv): It follows from (ii) and [6, Proposition 6.37(ii)] that $v_D \in (\mathcal{A}^*)^{-1} C^\ominus = (\mathcal{A}(C))^\ominus$. (v)(a): Indeed, (i) and Proposition 4.2(v) yield $v_R \in (-C) \cap \ker \mathcal{A} = -(C \cap \ker \mathcal{A}) = \{0\}$. Hence, $v_R = 0$ as claimed. (v)(b): It follows from [6, Fact 2.25(iv) and Proposition 6.35] that $\overline{C^\ominus + \text{ran } \mathcal{A}^*} = \overline{C^\ominus + \overline{\text{ran } \mathcal{A}^*}} = \overline{C^\ominus + (\ker \mathcal{A})^\perp} = \overline{C^\ominus + (\ker \mathcal{A})^\ominus} = (C \cap \ker \mathcal{A})^\ominus = \{0\}^\ominus = X$. The proof is complete. \blacksquare

Before we proceed we recall the following useful fact.

Fact 6.7. *Suppose that X is finite-dimensional and that $\ker \mathcal{A} \cap C = \{0\}$. Then $\mathcal{A}(C)$ is closed.*

Proof. See [27, Theorem 9.1]. \blacksquare

Lemma 6.8. *Suppose that X and Y are finite-dimensional and that $\ker \mathcal{A} \cap C = \{0\}$. Then for Problem (119) we have:*

- (i) $\mathcal{A}(C)$ is a nonempty closed convex cone.
- (ii) $C^\ominus + \text{ran } \mathcal{A}^* = X$.
- (iii) $\overline{\text{ran } (\text{Id} - T)} = M^{-1} \left((X \times (\mathbf{b} - \mathcal{A}(C))) \right)$.
- (iv) Let $z \in \mathcal{A}(C)$. Then $\langle v_D, \frac{1}{\tau} v_D - (\mathbf{b} - z) \rangle \leq 0$.
- (v) $\langle v_D, \frac{1}{\tau} v_D - \mathbf{b} \rangle \leq 0$.
- (vi) $\frac{1}{\tau} v_D = P_{\mathbf{b} - \mathcal{A}(C)}(\mathbf{0}) = P_{(\mathcal{A}(C))^\ominus}(\mathbf{b})$.

Let $\bar{u} \in C$ be such that $\frac{1}{\tau} v_D = \mathbf{b} - \mathcal{A}\bar{u}$. Then we have

- (vii) $\langle \mathcal{A}\bar{u}, v_D \rangle = \langle \bar{u}, \mathcal{A}^* v_D \rangle = 0$.
- (viii) $\frac{1}{\tau} \|v_D\|^2 = \langle \mathbf{b}, v_D \rangle$.
- (ix) $\bar{u} = P_C(\bar{u} + \mathcal{A}^* v_D)$ and $\mathcal{A}^* v_D = P_{C^\ominus}(\bar{u} + \mathcal{A}^* v_D)$.

Proof. (i): Use Fact 6.7 to learn that $\mathcal{A}(C)$ is closed. The convexity is clear because \mathcal{A} is linear and C is convex. The conclusion $\mathcal{A}(C)$ is a cone is straightforward. (ii): Combine Lemma 6.6(v)(b)

and Lemma 6.4(ii) applied with $C_1 = X$ and $C_2 = \overline{C^\ominus + \text{ran } \mathcal{A}^*}$. (iii): Combine Lemma 6.1, (ii) and (i). (iv): Combine (122) and Proposition 3.10(ii) in view of Lemma 6.6(v)(b). (v): This is a direct consequence of (iv) by setting $z = \mathbf{0}$. (vi): It follows from (iii) in view of (80) that $(-\mathcal{A}^*v_D, \frac{1}{\tau}v_D) = Mv \in X \times (\mathbf{b} - \mathcal{A}(C))$. That is, $\frac{1}{\tau}v_D \in \mathbf{b} - \mathcal{A}(C)$. Now combine this with (iv) and (i) in view of [6, Theorem 3.16] to learn that $\frac{1}{\tau}v_D = P_{\mathbf{b} - \mathcal{A}(C)}(\mathbf{0})$. Finally, using, e.g., [6, Proposition 3.19] we have

$$\frac{1}{\tau}v_D = P_{\mathbf{b} - \mathcal{A}(C)}(\mathbf{0}) = \mathbf{b} + P_{-\mathcal{A}(C)}(-\mathbf{b}) = \mathbf{b} - P_{\mathcal{A}(C)}(\mathbf{b}) \quad (131a)$$

$$= (\text{Id} - P_{\mathcal{A}(C)})(\mathbf{b}) = P_{(\mathcal{A}(C))^\ominus}(\mathbf{b}). \quad (131b)$$

(vii)&(viii): Indeed, it follows from (v) and Lemma 6.6(ii) that

$$0 \leq \langle \mathbf{b} - \frac{1}{\tau}v_D, v_D \rangle = \langle \mathcal{A}\bar{u}, v_D \rangle = \langle \bar{u}, \mathcal{A}^*v_D \rangle \leq 0, \quad (132)$$

hence $\langle \mathbf{b} - \frac{1}{\tau}v_D, v_D \rangle = \langle \mathcal{A}\bar{u}, v_D \rangle = \langle \bar{u}, \mathcal{A}^*v_D \rangle = 0$ and the conclusion follows. (ix): Combine (vii) and Lemma 6.6(ii) in view of, e.g., [6, Proposition 6.28]. ■

We now show a sufficient condition to have $v \in \text{ran}(\text{Id} - T)$.

Proposition 6.9. *Suppose that X and Y are finite-dimensional, that $\ker \mathcal{A} \cap C = \{\mathbf{0}\}$ and that $c = \mathbf{0}$. Then $v \in \text{ran}(\text{Id} - T)$.*

Proof. In view of Theorem 3.5(iv) and Lemma 6.6(v)(b) we have

$$v \in \text{ran}(\text{Id} - T) \Leftrightarrow Mv = \begin{pmatrix} -\mathcal{A}^*v_D \\ \frac{1}{\tau}v_D \end{pmatrix} \in \text{ran}(\partial F + S), \quad (133)$$

where $\partial F = N_C \times \{\mathbf{b}\}$. It follows from Lemma 6.8(vi) that $(\exists \bar{u} \in C)$ such that $\frac{1}{\tau}v_D = \mathbf{b} - \mathcal{A}\bar{u}$. Now consider the point $\begin{pmatrix} \bar{u} \\ -v_D \end{pmatrix} \in C \times Y = \text{dom } \partial F = \text{dom}(\partial F + S)$. We have

$$(\partial F + S) \begin{pmatrix} \bar{u} \\ -v_D \end{pmatrix} = \begin{pmatrix} N_C(\bar{u}) - \mathcal{A}^*v_D \\ \mathbf{b} - \mathcal{A}\bar{u} \end{pmatrix} \ni \begin{pmatrix} \mathbf{0} - \mathcal{A}^*v_D \\ \mathbf{b} - \mathcal{A}\bar{u} \end{pmatrix} = \begin{pmatrix} -\mathcal{A}^*v_D \\ \frac{1}{\tau}v_D \end{pmatrix} = Mv. \quad (134)$$

This completes the proof in view of (133). ■

Proposition 6.10. *Suppose that X and Y are finite-dimensional, that $\ker \mathcal{A} \cap C = \{\mathbf{0}\}$ and that $c = \mathbf{0}$. Let $(x_0, y_0) \in X \times Y$. Update via $(\forall k \in \mathbb{N})$*

$$(x_{k+1}, y_{k+1}) = T(x_k, y_k). \quad (135)$$

Then the following hold.

(i) *The sequence $(x_k, y_k + kv_D)_{k \in \mathbb{N}}$ is bounded.*

Let \bar{x} be a cluster point of $(x_k)_{k \in \mathbb{N}}$. Then we have

(ii) $\bar{x} \in C$.

(iii) $\tau \mathcal{A}\bar{x} = \tau \mathbf{b} - v_D$.

(iv) $(\bar{x}, \mathbf{0}) \in \text{Fix}(v + T)$.

Proof. (i): Combine [Theorem 3.5\(ii\)](#), and [Fact 3.6](#) in view of [Lemma 6.6\(v\)\(a\)](#). (ii): This follows from the fact that $(\mathbf{x}_k)_{k \in \mathbb{N}}$ lies in C and C is closed. (iii): Suppose that $\mathbf{x}_{n_k} \rightarrow \bar{\mathbf{x}}$. It follows from [Fact 3.8\(i\)](#) in view of [Theorem 3.5\(ii\)](#) that $\mathbf{x}_{n_k} - \mathbf{x}_{n_k+1} \rightarrow \mathbf{0}$, hence $\mathbf{x}_{n_k+1} \rightarrow \bar{\mathbf{x}}$. Therefore, using this and (120) applied with $\mathbf{c} = \mathbf{0}$ we have $\mathbf{v}_D \leftarrow \mathbf{y}_{n_k} - \mathbf{y}_{n_k+1} = \tau \mathbf{b} - \tau \mathcal{A}(2\mathbf{x}_{n_k+1} - \mathbf{x}_{n_k}) = \tau \mathbf{b} - \tau \mathcal{A}\bar{\mathbf{x}}$. (iv): Indeed, using (ii) and (iii) we have

$$\mathbf{v} + T(\bar{\mathbf{x}}, \mathbf{0}) = (\mathbf{0}, \mathbf{v}_D) + (P_C(\bar{\mathbf{x}} - \sigma \mathcal{A}^* \mathbf{0}), \mathbf{0} + \tau \mathcal{A}(2P_C(\bar{\mathbf{x}} - \sigma \mathcal{A}^* \mathbf{0}) - \bar{\mathbf{x}}) - \tau \mathbf{b}) \quad (136a)$$

$$= (\bar{\mathbf{x}}, \mathbf{v}_D + \tau \mathcal{A}\bar{\mathbf{x}} - \tau \mathbf{b}) = (\bar{\mathbf{x}}, \mathbf{0}), \quad (136b)$$

and the conclusion follows. \blacksquare

We recall the following fact.

Fact 6.11. *Let $(\mathbf{z}_k)_{k \in \mathbb{N}}$ be Fejér monotone with respect to a nonempty closed convex subset D of X . Let \mathbf{w}_1 and \mathbf{w}_2 be two cluster points of $(\mathbf{z}_k)_{k \in \mathbb{N}}$. Then $\mathbf{w}_1 - \mathbf{w}_2 \in (D - D)^\perp$.*

Proof. See, e.g., [[7](#), Lemma 2.2] or [[5](#), Theorem 6.2.2(ii)]. \blacksquare

In the special case considered in this section, namely $\ker \mathcal{A} \cap C = \{\mathbf{0}\}$, $\mathbf{c} = \mathbf{0}$, for which we already know $\mathbf{v}_R = \mathbf{0}$ by [Lemma 6.6\(v\)\(a\)](#), we can show that the first component of the sequence $(\mathbf{x}_k, \mathbf{y}_k)_{k \in \mathbb{N}}$ converges, and furthermore, we can partly characterize the limit point as follows.

Theorem 6.12. *Suppose that X and Y are finite-dimensional, that $\ker \mathcal{A} \cap C = \{\mathbf{0}\}$ and that $\mathbf{c} = \mathbf{0}$. Let $(\mathbf{x}_0, \mathbf{y}_0) \in X \times Y$. Update via $(\forall k \in \mathbb{N})$*

$$(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = T(\mathbf{x}_k, \mathbf{y}_k). \quad (137)$$

Then there exists $\bar{\mathbf{x}} \in C$ such that the following hold.

- (i) *The sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converges to $\bar{\mathbf{x}}$.*
- (ii) *$\bar{\mathbf{x}} \in C \cap L$, where $L = \{\mathbf{x} \in X \mid \mathcal{A}\mathbf{x} = \mathbf{b} - \frac{1}{\tau} \mathbf{v}_D\}$.*

Proof. (i): In view of [Proposition 6.10\(i\)](#) it suffices to show that $(\mathbf{x}_k)_{k \in \mathbb{N}}$ has at most one cluster point. To this end suppose that $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$ are two cluster points of $(\mathbf{x}_k)_{k \in \mathbb{N}}$, say $\mathbf{x}_{n_k} \rightarrow \bar{\mathbf{x}}$ and $\mathbf{x}_{l_k} \rightarrow \hat{\mathbf{x}}$. After dropping to a subsubsequence and relabelling if needed we can and do assume that $\mathbf{z}_{n_k} + n_k \mathbf{v} \rightarrow (\bar{\mathbf{x}}, \bar{\mathbf{y}})$ and $\mathbf{z}_{l_k} + l_k \mathbf{v} \rightarrow (\hat{\mathbf{x}}, \hat{\mathbf{y}})$. On the one hand, it follows from [Proposition 6.10\(iv\)](#) that $(\bar{\mathbf{x}}, \mathbf{0})$ and $(\hat{\mathbf{x}}, \mathbf{0})$ lie in $\text{Fix}(\mathbf{v} + T)$. On the other hand, applying [Fact 6.11](#) with D replaced by $\text{Fix}(\mathbf{v} + T)$, \mathbf{w}_1 replaced by $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, and \mathbf{w}_2 replaced by $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ in view of [Proposition 6.10\(iii\)](#) applied to $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$ yields

$$0 = \langle (\bar{\mathbf{x}}, \bar{\mathbf{y}}) - (\hat{\mathbf{x}}, \hat{\mathbf{y}}), (\bar{\mathbf{x}}, \mathbf{0}) - (\hat{\mathbf{x}}, \mathbf{0}) \rangle_M \quad (138a)$$

$$= \langle (\bar{\mathbf{x}} - \hat{\mathbf{x}}, \bar{\mathbf{y}} - \hat{\mathbf{y}}), (\bar{\mathbf{x}} - \hat{\mathbf{x}}, \mathbf{0}) \rangle_M \quad (138b)$$

$$= \langle (\bar{\mathbf{x}} - \hat{\mathbf{x}}, \bar{\mathbf{y}} - \hat{\mathbf{y}}), (\frac{1}{\sigma}(\bar{\mathbf{x}} - \hat{\mathbf{x}}) - \mathcal{A}^*(\mathbf{0}), \mathbf{0} - (\mathcal{A}\bar{\mathbf{x}} - \mathcal{A}\hat{\mathbf{x}})) \rangle \quad (138c)$$

$$= \frac{1}{\sigma} \langle (\bar{\mathbf{x}} - \hat{\mathbf{x}}, \bar{\mathbf{y}} - \hat{\mathbf{y}}), (\bar{\mathbf{x}} - \hat{\mathbf{x}}, \mathbf{0}) \rangle = \frac{1}{\sigma} \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2. \quad (138d)$$

That is $\bar{\mathbf{x}} = \hat{\mathbf{x}}$ and the conclusion follows. (ii): Combine (i) and [Proposition 6.10\(ii\)&\(iii\)](#). \blacksquare

6.2 Application to the ellipsoid separation problem

An example of [Problem \(119\)](#) in which $\mathbf{c} = \mathbf{0}$ and $\ker \mathcal{A} \cap C = \{\mathbf{0}\}$ is the ellipsoid separation problem, which we describe in this section. This problem asks: given two collections of finitely many

ellipsoids, say E_1, \dots, E_k and E'_1, \dots, E'_l all lying in \mathbb{R}^d , is there a hyperplane that strictly separates E_1, \dots, E_k from E'_1, \dots, E'_l ? This problem is a robust extension of the classic binary classification problem. “Robust” in this context means that the locations of the data points are known only up to an ellipsoid, and that the separating hyperplane should be correct for all possible actual locations of the points. See, e.g., Shivaswamy et al. [29]. We start with a characterization of separators whose proof (omitted) follows directly from the standard hyperplane separation theorem.

Fact 6.13. *Suppose that X is finite-dimensional. Let E_1, \dots, E_k and E'_1, \dots, E'_l be $k + l$ nonempty convex compact bodies lying in X . Then there exists $\mathbf{a} \in X \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$ such that $\langle \mathbf{a}, \mathbf{x} \rangle < b$ for all $\mathbf{x} \in E_1 \cup \dots \cup E_k$ and $\langle \mathbf{a}, \mathbf{x} \rangle > b$ for all $\mathbf{x} \in E'_1 \cup \dots \cup E'_l$ if and only if $\text{conv}(E_1 \cup \dots \cup E_k) \cap \text{conv}(E'_1 \cup \dots \cup E'_l) = \emptyset$.*

Let us introduce further notation for the ellipsoids: say that

$$E_i := \{\mathbf{x} \in \mathbb{R}^d : \|A_i^{-1}(\mathbf{x} - \mathbf{c}_i)\| \leq 1\}, \quad i \in \{1, \dots, k\}, \quad (139a)$$

$$E'_i := \{\mathbf{x} \in \mathbb{R}^d : \|B_i^{-1}(\mathbf{x} - \mathbf{d}_i)\| \leq 1\}, \quad i \in \{1, \dots, l\}. \quad (139b)$$

Here, $A_1, \dots, A_k, B_1, \dots, B_l$ are $d \times d$ invertible matrices and $\mathbf{c}_1, \dots, \mathbf{c}_k, \mathbf{d}_1, \dots, \mathbf{d}_l$ are vectors (centers of the ellipsoids). The naive way of writing the problem “Is $\text{conv}(E_1 \cup \dots \cup E_k) \cap \text{conv}(E'_1 \cup \dots \cup E'_l)$ nonempty?” would introduce variables $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l \in \mathbb{R}^d$ constrained to lie in the respective ellipsoids, and nonnegative multipliers $\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l$ satisfying $\lambda_1 + \dots + \lambda_k = \mu_1 + \dots + \mu_l = 1$ and $\lambda_1 \mathbf{v}_1 + \dots + \lambda_k \mathbf{v}_k = \mu_1 \mathbf{w}_1 + \dots + \mu_l \mathbf{w}_l$. However, this formulation is not convex due to the products $\lambda_i \mathbf{v}_i, \mu_i \mathbf{w}_i$.

A standard rescaling trick (see., e.g., Boyd & Vandenberghe [15, Exercise 4.56] attributed to Parrilo) reformulates the problem of nonemptiness of the intersection of convex hulls as standard SOCP with variables $\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l, \mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{q}_1, \dots, \mathbf{q}_l$:

$$\begin{aligned} & \text{minimize} && 0 \\ & \text{subject to} && \lambda_1 + \dots + \lambda_k = 1, \\ & && \mu_1 + \dots + \mu_l = 1, \\ & && \lambda_1 \mathbf{c}_1 + A_1 \mathbf{p}_1 + \dots + \lambda_k \mathbf{c}_k + A_k \mathbf{p}_k - \mu_1 \mathbf{d}_1 + B_1 \mathbf{q}_1 - \dots - \mu_l \mathbf{d}_l + B_l \mathbf{q}_l = \mathbf{0}, \\ & && \|\mathbf{p}_i\| \leq \lambda_i \quad \forall i \in \{1, \dots, k\}, \\ & && \|\mathbf{q}_i\| \leq \mu_i \quad \forall i \in \{1, \dots, l\}. \end{aligned}$$

Note that the constraints $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$ are redundant in this formulation and hence are dropped. The objective “min 0” indicates that any feasible solution to the constraints yields a common point in the convex hulls. Let $X := \mathbb{R}^{d+1} \times \dots \times \mathbb{R}^{d+1}$. We further rewrite this problem in the form:

$$\text{Find } \mathbf{x} := \begin{pmatrix} \lambda_1 \\ \mathbf{p}_1 \\ \vdots \\ \lambda_k \\ \mathbf{p}_k \\ \mu_1 \\ \mathbf{q}_1 \\ \vdots \\ \mu_l \\ \mathbf{q}_l \end{pmatrix} \in X \quad \text{subject to} \quad \begin{aligned} \mathbf{A}\mathbf{x} &= \begin{pmatrix} 1 \\ 1 \\ \mathbf{0} \end{pmatrix} =: \mathbf{b}, \\ \mathbf{x} &\in \mathbb{C}_2^{d+1} \times \dots \times \mathbb{C}_2^{d+1} =: \mathbf{C}, \end{aligned} \quad (140)$$

where C_2^{d+1} refers to the second-order cone in \mathbb{R}^{d+1} , and where

$$\mathcal{A} = \begin{pmatrix} 1 & \mathbf{0}^T & \cdots & 1 & \mathbf{0}^T & 0 & \mathbf{0}^T & \cdots & 0 & \mathbf{0}^T \\ 0 & \mathbf{0}^T & \cdots & 0 & \mathbf{0}^T & 1 & \mathbf{0}^T & \cdots & 1 & \mathbf{0}^T \\ \mathbf{c}_1 & A_1 & \cdots & \mathbf{c}_k & A_k & -\mathbf{d}_1 & B_1 & \cdots & -\mathbf{d}_l & B_l \end{pmatrix}. \quad (141)$$

This is a convex feasibility problem which can be recast as:

$$\min_{\mathbf{x} \in X} \iota_{\mathbf{C}}(\mathbf{x}) + \iota_{\{\mathbf{0}\}}(\mathcal{A}\mathbf{x} - \mathbf{b}). \quad (142)$$

By setting $K = \{\mathbf{0}\}$ and $g = \iota_{\mathbf{C}}$ in (54) and recalling (120) we learn that the PDHG update for the problem becomes

$$\begin{pmatrix} \mathbf{x}^+ \\ \mathbf{y}^+ \end{pmatrix} := T \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} := \begin{pmatrix} P_{\mathbf{C}}(\mathbf{x} - \sigma \mathcal{A}^* \mathbf{y}) \\ \mathbf{y} + \tau \mathcal{A}(2\mathbf{x}^+ - \mathbf{x}) - \tau \mathbf{b} \end{pmatrix}. \quad (143)$$

Thus, the main work for PDHG in this case is multiplication by \mathcal{A} and \mathcal{A}^* and projection onto \mathbf{C} .

Lemma 6.14. *We have $\mathbf{C} \cap \ker \mathcal{A} = (-\mathbf{C}) \cap \ker \mathcal{A} = \{\mathbf{0}\}$.*

Proof. It suffices to show $\mathbf{C} \cap \ker \mathcal{A} = \{\mathbf{0}\}$. Indeed, let $\mathbf{z} = (\lambda_1, \mathbf{p}_1, \dots, \lambda_k, \mathbf{p}_k, \mu_1, \mathbf{q}_1, \dots, \mu_l, \mathbf{q}_l) \in \mathbf{C} \cap \ker \mathcal{A}$. On the one hand

$$\mathbf{z} \in \mathbf{C} \Rightarrow (\forall i \in \{1, \dots, k\}) \lambda_i \geq \|\mathbf{p}_i\| \geq 0 \text{ and } (\forall j \in \{1, \dots, l\}) \mu_j \geq \|\mathbf{q}_j\| \geq 0. \quad (144)$$

On the other hand

$$\mathbf{z} \in \ker \mathcal{A} \Rightarrow \sum_{i=1}^k \lambda_i = \sum_{j=1}^l \mu_j = 0. \quad (145)$$

We learn from (144) and (145) that $(\forall i \in \{1, \dots, k\}) \lambda_i = 0$ and $(\forall j \in \{1, \dots, l\}) \mu_j = 0$. This, together with (144) yield that $(\forall i \in \{1, \dots, k\}) \mathbf{p}_i = \mathbf{0}$ and $(\forall j \in \{1, \dots, l\}) \mathbf{q}_j = \mathbf{0}$. That is $\mathbf{z} = \mathbf{0}$ as claimed. The proof is complete. \blacksquare

We have the following two results.

Theorem 6.15. *For Problem (142) we have*

- (i) $\mathbf{v}_R = \mathbf{0}$.
- (ii) $\mathbf{v} \in \text{ran}(\text{Id} - T)$.

Proof. (i): Combine Lemma 6.14 and Lemma 6.6(v)(a). (ii): Combine Lemma 6.14 and Proposition 6.9. \blacksquare

Lemma 6.16. *Recalling Problem (142), let $\mathbf{v}_D = (s, t, \mathbf{w}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$ and set $\mathcal{A}^* \mathbf{v}_D = (\lambda_1, \mathbf{p}_1, \dots, \lambda_k, \mathbf{p}_k, \mu_1, \mathbf{q}_1, \dots, \mu_l, \mathbf{q}_l)$. Then the following hold:*

- (i) $\mathbf{w} = \mathbf{0} \Rightarrow \mathbf{v}_D = \mathbf{0}$.
- (ii) $\mathbf{v}_D \neq \mathbf{0} \Rightarrow s + t > 0$.
- (iii) $\mathbf{v}_D = \mathbf{0} \Leftrightarrow \mathcal{A}^* \mathbf{v}_D = \mathbf{0}$.
- (iv) Suppose that $\mathbf{0} \in \{\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l\}$. Then $\mathbf{v}_D = \mathbf{0}$.
- (v) Suppose that $\mathbf{0} \in \{\mathbf{p}_1, \dots, \mathbf{p}_k, \mathbf{q}_1, \dots, \mathbf{q}_l\}$. Then $\mathbf{v}_D = \mathbf{0}$.

Proof. It follows from [Lemma 6.8\(viii\)](#) that

$$s + t = \frac{1}{\tau}(s^2 + t^2 + \|\mathbf{w}\|^2) \geq 0. \quad (146)$$

(i): Indeed, [Lemma 6.6\(ii\)](#) implies that $\|\mathbf{p}_1\| = \|A_1^* \mathbf{w}\| \leq -\lambda_1 = -s - \langle \mathbf{c}, \mathbf{w} \rangle$ and similarly $\|B_1^* \mathbf{w}\| \leq -t + \langle \mathbf{d}, \mathbf{w} \rangle$. Consequently, $\mathbf{w} = \mathbf{0} \Rightarrow s \leq 0$ and $t \leq 0$. In view of (146) we conclude that $s + t = 0$ and hence $s^2 + t^2 = 0$, equivalently, $s = t = 0$. That is $\mathbf{v}_D = \mathbf{0}$. (ii): Combine (i) and (146). (iii): “ \Rightarrow ”: This is clear. “ \Leftarrow ”: Observe that, because A_1 is invertible and $A_1^* \mathbf{w} = \mathbf{0}$ we must have $\mathbf{w} = \mathbf{0}$. Now combine this with (i). (iv): Without loss of generality, we may and do assume that $\lambda_1 = 0$. Then $\mathbf{p}_1 = A_1^* \mathbf{w} = \mathbf{0}$, hence $\mathbf{w} = \mathbf{0}$. Now combine this with (i). (v): Without loss of generality, we may and do assume that $\mathbf{p}_1 = A_1^* \mathbf{w} = \mathbf{0}$. Then $\mathbf{w} = \mathbf{0}$ and the conclusion follows in view of (i). ■

Theorem 6.17. *Recalling Problem (142), let T be defined as in (143). Let $\mathbf{x}_0 \in \mathbb{R}^{d+1} \times \dots \times \mathbb{R}^{d+1}$ and let $\mathbf{y}_0 \in \mathbb{R}^{d+2}$. Update via ($\forall k \in \mathbb{N}$)*

$$(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = T(\mathbf{x}_k, \mathbf{y}_k). \quad (147)$$

Then there exists $\bar{\mathbf{x}} \in C$ such that the following hold.

- (i) *The sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converges to $\bar{\mathbf{x}}$.*
- (ii) *$\bar{\mathbf{x}} \in C \cap L$, where $L = \{\mathbf{x} \in \mathbb{R}^{d+1} \times \dots \times \mathbb{R}^{d+1} \mid \mathcal{A}\mathbf{x} = \mathbf{b} - \frac{\mathbf{v}_D}{\tau}\}$.*

Proof. (i)–(ii): Combine [Lemma 6.14](#) and [Theorem 6.12\(i\)&\(ii\)](#). ■

As indicated by [Fact 6.13](#), disjointness of the convex hulls, i.e., primal infeasibility of (140), is certified by a separating hyperplane. Furthermore, we know from [Theorem 6.15](#) that $\mathbf{v}_D \neq \mathbf{0}$ in the infeasible case. We now argue a nonzero \mathbf{v}_D encodes a separating hyperplane. We first characterize such a hyperplane with the following lemma.

Lemma 6.18. *Given invertible $A \in \mathbb{R}^{n \times n}$, $\mathbf{c} \in \mathbb{R}^n$, $s \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, consider the ellipsoid $E := \{\mathbf{x} \in \mathbb{R}^n \mid \|A^{-1}(\mathbf{x} - \mathbf{c})\| \leq 1\}$ and the halfspace $H := \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{w}, \mathbf{x} \rangle \leq s\}$. The following hold.*

- (i) $E \subseteq H \Leftrightarrow s \geq \|A^* \mathbf{w}\| + \langle \mathbf{c}, \mathbf{w} \rangle$.
- (ii) $E \subseteq \text{int } H \Leftrightarrow s > \|A^* \mathbf{w}\| + \langle \mathbf{c}, \mathbf{w} \rangle$.

Proof. Let $\mathbf{x} \in \mathbb{R}^n$. Then

$$\langle \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{w}, \mathbf{x} - \mathbf{c} \rangle + \langle \mathbf{w}, \mathbf{c} \rangle = \langle A^* \mathbf{w}, A^{-1}(\mathbf{x} - \mathbf{c}) \rangle + \langle \mathbf{w}, \mathbf{c} \rangle. \quad (148)$$

(i): “ \Leftarrow ”: Let $\mathbf{x} \in E$. Using (148) and Cauchy–Schwarz we have $\langle \mathbf{w}, \mathbf{x} \rangle \leq \|A^* \mathbf{w}\| \cdot \|A^{-1}(\mathbf{x} - \mathbf{c})\| + \langle \mathbf{w}, \mathbf{c} \rangle \leq \|A^* \mathbf{w}\| + \langle \mathbf{w}, \mathbf{c} \rangle \leq s$. “ \Rightarrow ”: Let $\mathbf{x} = A(A^* \mathbf{w} / \|A^* \mathbf{w}\|) + \mathbf{c}$. Then $\mathbf{x} \in E$, hence $\mathbf{x} \in H$ and (148) implies $s \geq \langle \mathbf{w}, \mathbf{x} \rangle = \langle A^* \mathbf{w}, A^{-1}(\mathbf{x} - \mathbf{c}) \rangle + \langle \mathbf{w}, \mathbf{c} \rangle = \|A^* \mathbf{w}\| + \langle \mathbf{w}, \mathbf{c} \rangle$. (ii): The proof proceeds similar to the proof of (i). ■

We now state and prove our main result for the ellipsoid separation problem, which states that, because $\mathbf{v} \in \text{ran}(\text{Id} - T)$ ([Theorem 6.15\(ii\)](#)), a nonzero \mathbf{v} indicates inconsistency, and furthermore, a nonzero \mathbf{v} encodes a strict separating hyperplane.

Theorem 6.19. *Given $k + 1$ ellipsoids specified by (139), let $\mathbf{v}_D =: (s, t, \mathbf{w})$ and recall (143). Then the following are equivalent.*

- (i) $\text{conv}(E_1 \cup \dots \cup E_k) \cap \text{conv}(E'_1 \cup \dots \cup E'_l) = \emptyset$,
- (ii) SOCP problem (140) is infeasible,
- (iii) $\mathbf{0} \notin \text{ran}(\text{Id} - T)$, where T is the PDHG operator given by (143),

- (iv) $\mathbf{v} \neq \mathbf{0}$,
- (v) $\mathbf{v}_D \neq \mathbf{0}$,
- (vi) $s > -t$.

Any one of these statements implies:

- (vii) The hyperplane $\{\mathbf{x} \mid \langle \mathbf{w}, \mathbf{x} \rangle = s'\}$ strictly separates E_1, \dots, E_k from E'_1, \dots, E'_l , where s' is chosen arbitrarily in $] -t, s[$.

Conversely, the existence of (\mathbf{w}, s') as in (vii) implies all of (i)–(vi).

Proof. (i) \Leftrightarrow (ii): This was explained earlier in the formulation of (140). (ii) \Leftrightarrow (iii): We show the contrapositives. If (140) has a solution say \mathbf{x}^* , then $(\mathbf{x}^*, \mathbf{0})$ is a fixed point of T defined in (143). Equivalently, $\mathbf{0} \in \text{ran}(\text{Id} - T)$. Conversely, suppose that $(\mathbf{x}, \mathbf{y}) \in \text{Fix } T$. Then $\mathbf{x} \in \mathbf{C}$ and $\mathcal{A}\mathbf{x} = \mathbf{b}$, i.e., \mathbf{x} solves (140). (iii) \Leftrightarrow (iv): This follows from Theorem 6.15(ii). (iv) \Leftrightarrow (v): This follows from Theorem 6.15(i). (v) \Leftrightarrow (vi): The forward direction is established by Lemma 6.16(ii), while the reverse direction is trivial. (vi) \Rightarrow (vii): Recalling the form of \mathcal{A} in (141), we have

$$\mathcal{A}^* \mathbf{v}_D = \begin{pmatrix} s + \langle \mathbf{c}_1, \mathbf{w} \rangle \\ A_1^* \mathbf{w} \\ \vdots \\ s + \langle \mathbf{c}_k, \mathbf{w} \rangle \\ A_k^* \mathbf{w} \\ t - \langle \mathbf{d}_1, \mathbf{w} \rangle \\ B_1^* \mathbf{w} \\ \vdots \\ t - \langle \mathbf{d}_l, \mathbf{w} \rangle \\ B_l^* \mathbf{w} \end{pmatrix}.$$

By Lemma 6.6(ii), we know that $\mathcal{A}^* \mathbf{v}_D \in \mathbf{C}^\ominus$, in other words,

$$\begin{aligned} -s - \langle \mathbf{c}_i, \mathbf{w} \rangle &\geq \|A_i^* \mathbf{w}\|, \quad i \in \{1, \dots, k\}, \\ -t + \langle \mathbf{d}_i, \mathbf{w} \rangle &\geq \|B_i^* \mathbf{w}\|, \quad i \in \{1, \dots, l\}. \end{aligned}$$

Since $s > -t$, select an arbitrary s' satisfying $s > -s' > -t$. Then we obtain the inequalities

$$\begin{aligned} s' &> \|A_i^* \mathbf{w}\| + \langle \mathbf{c}_i, \mathbf{w} \rangle, \quad i \in \{1, \dots, k\}, \\ -s' &> \|B_i^* \mathbf{w}\| + \langle \mathbf{d}_i, -\mathbf{w} \rangle, \quad i \in \{1, \dots, l\}. \end{aligned}$$

In view of Lemma 6.18, these inequalities show that E_1, \dots, E_k are strictly on one side of the hyperplane $\{\mathbf{x} \mid \langle \mathbf{w}, \mathbf{x} \rangle = s'\}$ while E'_1, \dots, E'_l are strictly on the other side, thus establishing (vii). Finally, the converse statement at the end of the theorem follows from Fact 6.13. \blacksquare

7 Conclusion

We have developed a new formula for $\overline{\text{ran}}(\text{Id} - T)$ when T is the PDHG operator. We applied this formula to quadratic programming and the ellipsoid separation problem to show that in both cases, PDHG can diagnose inconsistency by checking the limiting value of $\mathbf{z}_k - \mathbf{z}_{k+1}$ as per Fact 3.8(ii). Both results used the conclusion that $\mathbf{v} \in \text{ran}(\text{Id} - T)$, where \mathbf{v} is the infimal displacement vector.

We provided new results on the convergence of PDHG iterates for both problems. Many issues remain in understanding the landscape of PDHG for infeasible conic optimization problems. Lest the reader suspect that [Fact 3.8\(ii\)](#) can always diagnose inconsistency, we point out that it is relatively easy to construct small contrived inconsistent problems such that $\mathbf{0} \in \overline{\text{ran}}(\text{Id} - T) \setminus \text{ran}(\text{Id} - T)$, meaning that the test based on [Fact 3.8\(ii\)](#) will fail to detect inconsistency. There are also realistic examples when this occurs, for example, the unbounded case of the min-volume-ellipsoid problem (see, e.g., formulation (12a) in [\[30\]](#)), which arises when the data points lie in a low-dimensional affine space.

References

- [1] David Applegate, Mateo Díaz, Haihao Lu, and Miles Lubin. Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. <https://arxiv.org/abs/2102.04592>, 2021.
- [2] J. B. Baillon, R. E. Bruck, and S. Reich. On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. *Houston Journal of Mathematics*, 4:1–9, 1978.
- [3] Goran Banjac. On the minimal displacement vector of the Douglas–Rachford operator. *Operations Research Letters*, 49(2):197–200, 2021.
- [4] Sedi Bartz, Heinz H. Bauschke, Jonathan M. Borwein, Simeon Reich, and Xianfu Wang. Fitzpatrick functions, cyclic monotonicity and Rockafellar’s antiderivative. *Nonlinear Analysis: Theory, Methods & Applications*, 66(5):1198–1223, 2007.
- [5] Heinz H. Bauschke. *Projection Algorithms and Monotone Operators*. PhD thesis, Simon Fraser University, Burnaby, B.C., Canada, 1996.
- [6] Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2 edition, 2017.
- [7] Heinz H. Bauschke, Minh N. Dao, and Walaa M. Moursi. On Fejér monotone sequences and nonexpansive mappings. *Linear and Nonlinear Analysis*, 1:287–295, 2015.
- [8] Heinz H. Bauschke, Sarah M. Moffat, and Xianfu Wang. Near equality, near convexity, sums of maximally monotone operators, and averages of firmly nonexpansive mappings. *Mathematical Programming Series B*, 139:55–70, 2013.
- [9] Heinz H. Bauschke and Walaa M. Moursi. The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces. *SIAM Journal on Optimization*, 26(2):968–985, 2016.
- [10] Heinz H. Bauschke and Walaa M. Moursi. On the Douglas–Rachford algorithm. *Mathematical Programming Series A*, 164:263–284, 2017.
- [11] Heinz H. Bauschke and Walaa M. Moursi. On the Douglas–Rachford algorithm for solving possibly inconsistent optimization problems. *Mathematics of Operations Research*, 2023. <https://doi.org/10.1287/moor.2022.1347>.
- [12] Radu Ioan Bot. *Conjugate duality in convex optimization*, volume 637 of *Lecture Notes in Economics and Mathematical Systems*. Springer Science & Business Media, 2009.
- [13] Radu Ioan Bot, Sorin-Mihai Grad, and Gert Wanka. *Duality in vector optimization*. Springer Science & Business Media, 2009.
- [14] Radu Ioan Bot and Christopher Hendrich. A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. *SIAM Journal on Optimization*, 23(4):2541–2565, 2013.
- [15] Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2004.
- [16] Luis M. Briceño Arias and Patrick L. Combettes. A monotone+skew splitting model for composite monotone inclusions in duality. *SIAM Journal on Optimization*, 21(4):1230–1250, 2011.

- [17] R.E. Bruck and S. Reich. Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston Journal of Mathematics*, 3:459–470, 1977.
- [18] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
- [19] Frank Deutsch. *Best Approximation in Inner Product Spaces*. Springer, 2001.
- [20] Yanli Liu, Ernest K Ryu, and Wotao Yin. A new use of Douglas–Rachford splitting for identifying infeasible, unbounded, and pathological conic programs. *Mathematical Programming Series A*, 177(1-2):225–253, 2019.
- [21] Haihao Lu. First-order methods for linear programming: Theory and computation. Cornell University Operations Research and Information Engineering Colloquium, <https://www.youtube.com/watch?v=15kR6I6zesA>, 2022.
- [22] Walaa M. Moursi. *The Douglas–Rachford operator in the possibly inconsistent case: static properties and dynamic behaviour*. PhD thesis, University of British Columbia, 2016. <https://doi.org/10.14288/1.0340501>.
- [23] Walaa M. Moursi. The forward–backward algorithm and the normal problem. *Journal of Optimization Theory and Applications*, 176:605–624, 2018.
- [24] Daniel O’Connor and Lieven Vandenbergh. On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. *Mathematical Programming*, 179(1-2):85–108, 2020.
- [25] A Pazy. Asymptotic behavior of contractions in Hilbert space. *Israel Journal of Mathematics*, 9:235–240, 1971.
- [26] Stephen M. Robinson. Some continuity properties of polyhedral multifunctions. In H. König, B. Korte, and K. Ritter, editors, *Mathematical Programming at Oberwolfach*, pages 206–214. Springer, 1980.
- [27] R Tyrrell Rockafellar. *Convex analysis*, volume 11. Princeton University Press, 1970.
- [28] Ernest K Ryu, Yanli Liu, and Wotao Yin. Douglas–Rachford splitting and ADMM for pathological convex optimization. *Computational Optimization and Applications*, 74(3):747–778, 2019.
- [29] Pannagadatta K Shivaswamy, Chiranjib Bhattacharyya, and Alexander J Smola. Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research*, pages 1283–1314, 2006.
- [30] Michael J. Todd and E. Alper Yıldırım. On Khachiyan’s algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744, 2007.
- [31] Eduardo H Zarantonello. Projections on convex sets in Hilbert space and spectral theory: Part I. projections on convex sets: Part II. spectral theory. In Eduardo H Zarantonello, editor, *Contributions to nonlinear functional analysis*, pages 237–424. Elsevier, 1971.