On the Asymptotic Behavior of the Douglas-Rachford and Proximal-Point Algorithms for Convex Optimization

Goran Banjac and John Lygeros

May 1, 2020

Abstract

The authors in [BGSB19] recently showed that the Douglas-Rachford algorithm provides certificates of infeasibility for a class of convex optimization problems. In particular, they showed that the difference between consecutive iterates generated by the algorithm converges to certificates of primal and dual strong infeasibility. Their result was shown in a finite dimensional Euclidean setting and for a particular structure of the constraint set. In this paper, we extend the result to Hilbert spaces and a general nonempty closed convex set. Moreover, we show that the proximal-point algorithm applied to the set of optimality conditions of the problem generates similar infeasibility certificates.

1 Introduction

Due to its very good practical performance and ability to handle nonsmooth functions, the Douglas-Rachford algorithm has attracted a lot of interest for solving convex optimization problems. Provided that a problem is solvable and satisfies certain constraint qualification, the algorithm is known to converge to an optimal solution [BC17, Cor. 27.3]. When the problem is infeasible, then some of its iterates diverge [EB92].

Results on the asymptotic behavior of the Douglas-Rachford algorithm for infeasible problems are very scarce, and most of them study some specific cases such as feasibility problems involving two convex sets that do not intersect [BDM16, BM16, BM17]. Although there have been some recent results studying a more general setting [RLY19], they do not provide practical conditions to detect infeasibility of a problem. Instead, the asymptotic behavior is characterized via the so called infimal displacement vector, which is not known prior to solving the problem. The authors in [BGSB19] consider a problem of minimizing a convex quadratic function over a particular constraint set, and show that the iterates of the
Douglas-Rachford algorithm generate an infeasibility certificate when the problem is primal or dual strongly infeasible. A similar analysis was applied in [LMK20] to show that the proximal-point algorithm used for solving a convex quadratic program can also detect infeasibility.

The constraint set of the problem studied in [BGSB19] is represented in the form $Ax \in C$, where $A$ is a real matrix and $C$ the Cartesian product of a convex compact set and a translated closed convex cone. This paper extends the result of [BGSB19] to Hilbert spaces and a general nonempty closed convex set $C$. Moreover, we show that a similar analysis can be used to show that the proximal-point algorithm for solving the same class of problems generates similar infeasibility certificates.

The paper is organized as follows. We introduce some definitions and notation in the sequel of Section 1, and the problem under consideration in Section 2. Section 3 presents some supporting results that are essential for generalizing the results in [BGSB19]. Finally, Section 4 and Section 5 analyze the asymptotic behavior of the Douglas-Rachford and proximal-point algorithms, respectively, and show that they provide infeasibility certificates for the considered problem.

1.1 Notation

Let $\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$ be real Hilbert spaces with inner products $\langle \cdot \mid \cdot \rangle$, induced norms $\| \cdot \|$, and identity operators $\text{Id}$. The power set of $\mathcal{H}$ is denoted by $2^\mathcal{H}$. Let $\mathbb{N}$ denote the set of positive integers, $\mathbb{R}$ the set of real numbers, $\mathbb{R}^n$ the $n$-dimensional Euclidean space, and $\mathbb{R}^{m \times n}$ the space of real $m$-by-$n$ matrices. For a sequence $(s_n)_{n \in \mathbb{N}}$, we define $\delta_{s_{n+1}} = s_{n+1} - s_n$.

Let $D$ be a nonempty subset of $\mathcal{H}$. Then $T : D \to \mathcal{H}$ is nonexpansive if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq \|x - y\|,$$

and it is $\alpha$-averaged with $\alpha \in [0,1]$ if there exists a nonexpansive operator $R : D \to \mathcal{H}$ such that $T = (1 - \alpha)\text{Id} + \alpha R$. We denote the range of $T$ by $\text{ran} T$. A set-valued operator $B : \mathcal{H} \to 2^\mathcal{H}$, characterized by its graph

$$\text{gra} B = \{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Bx\},$$

is monotone if

$$(\forall (x,u) \in \text{gra} B)(\forall (y,v) \in \text{gra} B) \quad \langle x - y \mid u - v \rangle \geq 0.$$ 

For a proper lower semicontinuous convex function $f : \mathcal{H} \to (-\infty, +\infty]$, we define its:

- **Fenchel conjugate**: $f^* : \mathcal{H} \to (-\infty, +\infty] : u \mapsto \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - f(x)$,
- **proximity operator**: $\text{Prox}_f : \mathcal{H} \to \mathcal{H} : x \mapsto \arg\min_{y \in \mathcal{H}} \left( f(y) + \frac{1}{2} \|y - x\|^2 \right)$,
- **subdifferential**: $\partial f : \mathcal{H} \to 2^\mathcal{H} : x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \langle y - x \mid u \rangle + f(x) \leq f(y)\}$. 


For a nonempty closed convex set $C \subseteq \mathcal{H}$, we denote its closure by $\overline{C}$ and define its:

- **polar cone**: $C^\circ = \left\{ u \in \mathcal{H} \mid \sup_{x \in C} \langle x \mid u \rangle \leq 0 \right\}$,
- **recession cone**: $\text{rec} \, C = \left\{ x \in \mathcal{H} \mid (\forall y \in C) \, x + y \in C \right\}$,
- **indicator function**: $\iota_C : \mathcal{H} \to [0, +\infty] : x \mapsto \begin{cases} 0 & x \in C \\ +\infty & \text{otherwise} \end{cases}$,
- **support function**: $\sigma_C : \mathcal{H} \to ]-\infty, +\infty] : u \mapsto \sup_{x \in C} \langle x \mid u \rangle$,
- **projection operator**: $P_C : \mathcal{H} \to \mathcal{H} : x \mapsto \arg\min_{y \in C} \|y - x\|,$
- **normal cone operator**: $N_C : \mathcal{H} \to 2^\mathcal{H} : x \mapsto \begin{cases} \left\{ u \in \mathcal{H} \mid \sup_{y \in C} \langle y - x \mid u \rangle \leq 0 \right\} & x \in C \\ \emptyset & x \notin C. \end{cases}$

## 2 Problem of Interest

Consider the following convex optimization problem:

$$\begin{align*}
\text{minimize} \quad & \frac{1}{2} \langle Qx \mid x \rangle + \langle q \mid x \rangle + \iota_C(\mathcal{A}x) \\
\text{subject to} \quad & \mathcal{A}x \in C,
\end{align*}$$

with $Q : \mathcal{H}_1 \to \mathcal{H}_1$ a monotone self-adjoint bounded linear operator, $q \in \mathcal{H}_1$, $\mathcal{A} : \mathcal{H}_1 \to \mathcal{H}_2$ a bounded linear operator with a closed range, and $C$ a nonempty closed convex subset of $\mathcal{H}_2$. The objective function of the problem is convex, continuous, and Fréchet differentiable [BC17, Prop. 17.36].

When $\mathcal{H}_1 = \mathbb{R}^n$ and $\mathcal{H}_2 = \mathbb{R}^m$, problem (1) reduces to the one considered in [BGSB19], where the Douglas-Rachford algorithm (which is equivalent to the alternating direction method of multipliers) was shown to generate certificates of primal and dual strong infeasibility. Moreover, the authors proposed termination criteria for infeasibility detection, which are easy to implement and are used in several numerical solvers; see, e.g., [SBG+20, GCG19, HTP19].

To prove the main results, they used the assumption that $C$ can be represented as the Cartesian product of a convex compact set and a translated closed convex cone, which was exploited heavily in their proofs. In this paper we extend these results to the case where $\mathcal{H}_1$ and $\mathcal{H}_2$ are real Hilbert spaces, and $C$ is a general nonempty closed convex set.

### 2.1 Optimality Conditions

We can rewrite problem (1) in the form

$$\begin{align*}
\text{minimize} \quad & \frac{1}{2} \langle Qx \mid x \rangle + \langle q \mid x \rangle + \iota_C(\mathcal{A}x).
\end{align*}$$


Provided that a certain constraint qualification holds, we can characterize its solution by [BC17, Thm. 27.2]

\[ 0 \in Qx + q + A^*\partial\iota_C(Ax), \]

and introducing a dual variable \( y \in \partial\iota_C(Ax) \), we can rewrite the inclusion as

\[
0 \in \left( Qx + q + A^*y \right) - y + \partial\iota_C(Ax). 
\]

Introducing an auxiliary variable \( z \in C \) and using \( \partial\iota = N_C \), we can write the optimality conditions for problem \((1)\) as

\[
\begin{align*}
Ax - z &= 0 \quad &\text{(3a)} \\
Qx + q + A^*y &= 0 \quad &\text{(3b)} \\
z \in C, \quad y \in N_C &z. \quad &\text{(3c)}
\end{align*}
\]

### 2.2 Infeasibility Certificates

The authors in [BGSB19] derived the following conditions for characterizing strong infeasibility of problem \((1)\) and its dual:

**Proposition 2.1.**

(i) If there exists \( \bar{y} \in (\text{rec } C)^\oplus \) such that

\[ A^*\bar{y} = 0 \quad \text{and} \quad \sigma_C(\bar{y}) < 0, \]

then problem \((1)\) is strongly infeasible.

(ii) If there exists \( \bar{x} \in H_1 \) such that

\[ Q\bar{x} = 0, \quad A\bar{x} \in \text{rec } C, \quad \text{and} \quad \langle q | \bar{x} \rangle < 0, \]

then the dual of problem \((1)\) is strongly infeasible.

**Proof.** See [BGSB19, Prop. 3.1].

### 3 Auxiliary Results

**Lemma 3.1.** Suppose that \( T: H \to H \) is an averaged operator and let \( s_0 \in H, s_n = T^ns_0, \) and \( \delta s := P_{\text{ran}(T - \text{Id})}(0). \) Then

(i) \( \frac{1}{n}s_n \to \delta s. \)

(ii) \( \delta s_n \to \delta s. \)
Proof. The first result is [Paz71, Cor. 3] and the second is [BBR78, Cor. 2.3]. □

The following proposition provides essential ingredients for generalizing the results in [BGSB19, §5].

**Proposition 3.2.** Let \((s_n)_{n \in \mathbb{N}}\) be a sequence in \(\mathcal{H}\) and \(D \subseteq \mathcal{H}\) a nonempty closed convex set. Define sequences \((p_n)_{n \in \mathbb{N}}\) and \((r_n)_{n \in \mathbb{N}}\) by

\[
p_n := P_D s_n, \quad r_n := (\text{Id} - P_D) s_n,
\]

and suppose that the limits \(\delta s := \lim_{n \to \infty} \frac{1}{n} s_n\) and \(\lim_{n \to \infty} \frac{1}{n} p_n\) exist. Then

(i) \(r_n \in (\text{rec } D)^\ominus\).

(ii) \(\frac{1}{n} r_n \to \delta r := P_{(\text{rec } D)^\ominus}(\delta s)\).

(iii) \(\frac{1}{n} p_n \to \delta p := P_{\text{rec } D}(\delta s)\).

(iv) \(\lim_{n \to \infty} \frac{1}{n} \langle p_n \mid r_n \rangle = \sigma_D(\delta r)\).

Proof. (i): Due to [BC17, Thm. 3.16 & Def. 6.48],

\[
(\forall d \in \text{rec } D) \quad \langle d \mid (\text{Id} - P_D) s_n \rangle \leq 0,
\]

which implies

\[
r_n = (\text{Id} - P_D) s_n \in (\text{rec } D)^\ominus.
\]

(ii)&(iii): Due to [BC17, Prop. 6.51], we have

\[
\lim_{n \to \infty} \frac{1}{n} p_n = \lim_{n \to \infty} \frac{1}{n} P_D s_n \in \text{rec } D. \tag{4}
\]

Since \((\text{rec } D)^\ominus\) is a cone, part (i) implies

\[
\lim_{n \to \infty} \frac{1}{n} r_n \in (\text{rec } D)^\ominus. \tag{5}
\]

From [BC17, Prop. 6.47], we have

\[
r_n = s_n - p_n \in N_D P_n,
\]

which, due to [BC17, Thm. 16.29] and the facts that \(\iota_D^* = \sigma_D\) and \(\partial \iota_D = N_D\), is equivalent to

\[
\frac{1}{n} \langle p_n \mid r_n \rangle = \sigma_D \left(\frac{1}{n} r_n\right), \tag{6}
\]

which implies

\[
\langle p_n \mid \frac{1}{n} r_n \rangle = \sup_{p \in D} \langle p \mid \frac{1}{n} r_n \rangle \geq \langle \hat{p} \mid \frac{1}{n} r_n \rangle,
\]

for any fixed \(\hat{p} \in D\). Dividing by \(n\) and taking the limit, we obtain

\[
\lim_{n \to \infty} \frac{1}{n} p_n \mid \frac{1}{n} r_n \rangle \geq \lim_{n \to \infty} \frac{1}{n} \hat{p} \mid \frac{1}{n} r_n \rangle = 0.
\]
Due to (4) and (5), the left-hand side of the inequality is the inner product of terms in rec $D$ and $(\text{rec } D)^\ominus$, and is thus always nonpositive. Therefore, it must be
\[
\left\langle \lim_{n \to \infty} \frac{1}{n} p_n, \lim_{n \to \infty} \frac{1}{n} r_n \right\rangle = 0.
\]
The result then follows from [BC17, Cor. 6.31].

(iv): Taking the limit of the inequality
\[
(\forall n \in \mathbb{N})(\forall p \in D) \quad \left\langle \hat{p} \mid \frac{1}{n} r_n \right\rangle \leq \sup_{p \in D} \left\langle p \mid \frac{1}{n} r_n \right\rangle,
\]
we obtain
\[
(\forall \hat{p} \in D) \quad \lim_{n \to \infty} \left\langle \hat{p} \mid \frac{1}{n} r_n \right\rangle \leq \lim_{n \to \infty} \sup_{p \in D} \left\langle p \mid \frac{1}{n} r_n \right\rangle,
\]
and taking the supremum of the left-hand side over $D$, we get
\[
\sup_{p \in D} \lim_{n \to \infty} \left\langle p \mid \frac{1}{n} r_n \right\rangle \leq \lim_{n \to \infty} \sup_{p \in D} \left\langle p \mid \frac{1}{n} r_n \right\rangle.
\]
Taking the limit of (6) and using the inequality above, we obtain
\[
\lim_{n \to \infty} \frac{1}{n} \langle p_n \mid r_n \rangle = \lim_{n \to \infty} \sup_{p \in D} \left\langle p \mid \frac{1}{n} r_n \right\rangle \geq \sup_{p \in D} \lim_{n \to \infty} \left\langle p \mid \frac{1}{n} r_n \right\rangle = \sigma_D(\delta r).
\]
Since $p_n \in D$, we also have
\[
\lim_{n \to \infty} \frac{1}{n} \langle p_n \mid r_n \rangle \leq \sup_{p \in D} \lim_{n \to \infty} \left\langle p \mid \frac{1}{n} r_n \right\rangle = \sigma_D(\delta r).
\]
The result follows by combining the two inequalities above. \(\square\)

The results of Prop. 3.2 are straightforward under the additional assumption that $D$ is compact, since then $\text{rec } D = \{0\}$ and $(\text{rec } D)^\ominus = H_2$, and thus
\[
\lim_{n \to \infty} \frac{1}{n} p_n = \lim_{n \to \infty} \frac{1}{n} P_D s_n = 0 = P_{\text{rec } D}(\delta s)
\]
\[
\lim_{n \to \infty} \frac{1}{n} r_n = \lim_{n \to \infty} \frac{1}{n} (s_n - p_n) = \delta s = P_{(\text{rec } D)^\ominus}(\delta s).
\]
Moreover, due to the continuity of $\sigma_D$ [BC17, Ex. 11.2], taking the limit of (6) implies
\[
\lim_{n \to \infty} \frac{1}{n} \langle p_n \mid r_n \rangle = \lim_{n \to \infty} \sigma_D \left( \frac{1}{n} r_n \right) = \sigma_D \left( \lim_{n \to \infty} \frac{1}{n} r_n \right) = \sigma_D(\delta r).
\]
When $D$ is a (translated) closed convex cone, the recession cone is the cone itself, and the results of Prop. 3.2 can be shown using the Moreau decomposition and some basic properties of the projection operator; see [BGSB19, Lem. A.3 & Lem. A.4] for details.
4 Douglas-Rachford Algorithm

The Douglas-Rachford algorithm is an operator splitting method, which can be used to solve composite minimization problems of the form

$$\min_{w \in H} f(w) + g(w),$$

where $f$ and $g$ are proper lower semicontinuous convex functions. An iteration of the algorithm in application to problem (7) can be written as

$$w_n = \text{Prox}_g s_n$$
$$\tilde{w}_n = \text{Prox}_f (2w_n - s_n)$$
$$s_{n+1} = s_n + \alpha (\tilde{w}_n - w_n).$$

where $\alpha \in ]0, 2[$ is the relaxation parameter.

If we rewrite problem (1) as

$$f(x, z) = \frac{1}{2} \langle Qx \mid x \rangle + \langle q \mid x \rangle + \langle Ax = z \mid x, z \rangle$$
$$g(x, z) = \iota_C(z),$$

then an iteration of the Douglas-Rachford algorithm takes the following form [BGSB19, SBG+20]:

$$\tilde{x}_n = \arg\min_{x \in H_1} \left( \frac{1}{2} \langle Qx \mid x \rangle + \langle q \mid x \rangle + \frac{1}{2} \|x - x_n\|^2 + \frac{1}{2} \|Ax - (2P_C - \text{Id})v_n\|^2 \right)$$

$$x_{n+1} = x_n + \alpha (\tilde{x}_n - x_n)$$
$$v_{n+1} = v_n + \alpha (A\tilde{x}_n - P_C v_n)$$

We will exploit the following well-known result to analyze the asymptotic behavior of the algorithm [LM79]:

**Fact 4.1.** Iteration (8) amounts to

$$(x_{n+1}, v_{n+1}) = T_{\text{DR}}(x_n, v_n),$$

where $T_{\text{DR}}: (H_1 \times H_2) \to (H_1 \times H_2)$ is an $(\alpha/2)$-averaged operator.

The solution to the subproblem in (8a) satisfies the optimality condition

$$Q\tilde{x}_n + q + (\tilde{x}_n - x_n) + A^* (A\tilde{x}_n - (2P_C - \text{Id})v_n) = 0.$$

If we rearrange (8b) to isolate $\tilde{x}_n$,

$$\tilde{x}_n = x_n + \alpha^{-1} \delta x_{n+1},$$
and substitute it into (8c) and (9), we obtain the following relations between the iterates:

\[
Ax_n - P_Cv_n = -\alpha^{-1}(A\delta x_{n+1} - \delta v_{n+1})
\]

(10a)

\[
Qx_n + q + A^*(\text{Id} - P_C)v_n = -\alpha^{-1}((Q + \text{Id})\delta x_{n+1} + A^*\delta v_{n+1}).
\]

(10b)

Let us define the following auxiliary iterates of iteration (8):

\[
z_n := P_Cv_n
\]

(11a)

\[
y_n := (\text{Id} - P_C)v_n.
\]

(11b)

Observe that the pair \((z_n, y_n)\) satisfies optimality condition (3c) for all \(n \in \mathbb{N}\) [BC17, Prop. 6.47], and that the right-hand terms in (10) indicate how far the iterates \((x_n, z_n, y_n)\) are from satisfying (3a) and (3b).

The following corollary follows directly from Lem. 3.1, Prop. 3.2, Fact 4.1, and the Moreau decomposition [BC17, Thm. 6.30]:

**Corollary 4.2.** Let the sequences \((x_n)_{n \in \mathbb{N}}, (v_n)_{n \in \mathbb{N}}, (z_n)_{n \in \mathbb{N}}, \text{ and } (y_n)_{n \in \mathbb{N}\) be given by (8) and (11), and \((\delta x, \delta v) := P_{\text{ran}(T_{\text{DR}} - \text{Id})}(0). Then

(i) \(\frac{1}{n}(x_n, v_n) \rightarrow (\delta x, \delta v)\).

(ii) \((\delta x_n, \delta v_n) \rightarrow (\delta x, \delta v)\).

(iii) \(y_n \in (\text{rec } C)\).

(iv) \(\frac{1}{n}y_n \rightarrow \delta y := P_{(\text{rec } C)}(\delta v)\).

(v) \(\frac{1}{n}z_n \rightarrow \delta z := P_{\text{rec } C}(\delta v)\).

(vi) \(\lim_{n \to \infty} \frac{1}{n} \langle z_n | y_n \rangle = \sigma_C(\delta y)\).

(vii) \(\delta z + \delta y = \delta v\).

(viii) \(\langle \delta z | \delta y \rangle = 0\).

(ix) \(\|\delta z\|^2 + \|\delta y\|^2 = \|\delta v\|^2\).

The following two propositions generalize [BGSB19, Prop. 5.1 & Prop. 5.2], though the proofs follow very similar arguments.

**Proposition 4.3.** The following relations hold between \(\delta x, \delta z, \text{ and } \delta y, \text{ which are defined in Cor. 4.2:}

(i) \(A\delta x = \delta z\).

(ii) \(Q\delta x = 0\).

(iii) \(A^*\delta y = 0\).

(iv) \(\delta z_n \rightarrow \delta z\).

(v) \(\delta y_n \rightarrow \delta y\).

Proof. (i): Divide (10a) by \(n\), take the limit, and use Cor. 4.2(v) to get

\[
A\delta x = \lim_{n \to \infty} \frac{1}{n} P_Cv_n = \delta z.
\]

(12)
(ii): Divide (10b) by $n$, take the inner product of both sides with $\delta x$ and take the limit to obtain

$$
\langle Q\delta x \mid \delta x \rangle = -\lim_{n \to \infty} \langle A\delta x, \frac{1}{n}(\text{Id} - P_C)v_n \rangle = -\langle \delta z \mid \delta y \rangle = 0,
$$

where we used (12) and Cor. 4.2(iv) in the second equality, and Cor. 4.2(viii) in the third. Due to [BC17, Cor. 18.18], the equality above implies

$$
Q\delta x = 0. \quad (13)
$$

(iii): Divide (10b) by $n$, take the limit, and use (13) to obtain

$$
0 = \lim_{n \to \infty} \frac{1}{n}A^*(\text{Id} - P_C)v_n = A^*\delta y,
$$

where we used Cor. 4.2(iv) in the second equality.

(iv): Subtracting (10a) at iterations $n + 1$ and $n$, and taking the limit yield

$$
\lim_{n \to \infty} \delta z_n = A\delta x = \delta z,
$$

where the second equality follows from (12).

(v): From (11) we have

$$
\lim_{n \to \infty} \delta y_n = \lim_{n \to \infty} (\delta v_n - \delta z_n) = \delta v - \delta z = \delta y,
$$

where the last equality follows from Cor. 4.2(vii).

Proposition 4.4. The following identities hold for $\delta x$ and $\delta y$, which are defined in Cor. 4.2:

(i) $\langle q \mid \delta x \rangle = -\alpha^{-1}\|\delta x\|^2 - \alpha^{-1}\|A\delta x\|^2$.

(ii) $\sigma_C(\delta y) = -\alpha^{-1}\|\delta y\|^2$.

Proof. Take the inner product of both sides of (10b) with $\delta x$ and use (13) to obtain

$$
\langle q \mid \delta x \rangle + \langle A\delta x \mid y_n \rangle = -\alpha^{-1}\langle \delta x \mid \delta x_{n+1} \rangle - \alpha^{-1}\langle A\delta x \mid \delta v_{n+1} \rangle.
$$

Taking the limit and using Prop. 4.3(i) and Cor. 4.2(vii)&(viii) give

$$
\langle q \mid \delta x \rangle + \alpha^{-1}\|\delta x\|^2 + \alpha^{-1}\|\delta z\|^2 = -\lim_{n \to \infty} \langle \delta z \mid y_n \rangle \geq 0, \quad (14)
$$

where the inequality follows from Cor. 4.2(iii)&(v) as the inner product of terms in rec $C$ and (rec $C)^\ominus$ is nonpositive. Now take the inner product of both sides of (10a) with $\delta y$ to obtain

$$
\langle A^*\delta y \mid x_n + \alpha^{-1}\delta x_{n+1} \rangle - \langle \delta y \mid P_Cv_n \rangle = \alpha^{-1}\langle \delta y \mid \delta v_{n+1} \rangle.
$$

Due to Prop. 4.3(iii), the first inner product on the left-hand side is zero. Taking the limit and using Cor. 4.2(vii)&(viii), we obtain

$$
-\alpha^{-1}\|\delta y\|^2 = \lim_{n \to \infty} \langle \delta y \mid P_Cv_n \rangle \leq \sup_{z \in C} \langle \delta y \mid z \rangle = \sigma_C(\delta y),
$$

where we used (12) and Cor. 4.2(iv) in the second equality, and Cor. 4.2(viii) in the third. Due to [BC17, Cor. 18.18], the equality above implies

$$
Q\delta x = 0. \quad (13)
$$

(iii): Divide (10b) by $n$, take the limit, and use (13) to obtain

$$
0 = \lim_{n \to \infty} \frac{1}{n}A^*(\text{Id} - P_C)v_n = A^*\delta y,
$$

where we used Cor. 4.2(iv) in the second equality.

(iv): Subtracting (10a) at iterations $n + 1$ and $n$, and taking the limit yield

$$
\lim_{n \to \infty} \delta z_n = A\delta x = \delta z,
$$

where the second equality follows from (12).

(v): From (11) we have

$$
\lim_{n \to \infty} \delta y_n = \lim_{n \to \infty} (\delta v_n - \delta z_n) = \delta v - \delta z = \delta y,
$$

where the last equality follows from Cor. 4.2(vii).
or equivalently,
\[ \sigma_C(\delta y) + \alpha^{-1}\|\delta y\|^2 \geq 0. \] (15)

Summing (14) and (15) and using Cor. 4.2(ix), we obtain
\[ \langle \delta x \mid q \rangle + \sigma_C(\delta y) + \alpha^{-1}\|\delta x\|^2 + \alpha^{-1}\|\delta v\|^2 \geq 0. \] (16)

Now take the inner product of both sides of (10b) with \( x_n \) to obtain
\[ \langle Qx_n \mid x_n \rangle + \langle q \mid x_n \rangle + \langle Ax_n \mid y_n \rangle = -\alpha^{-1}\|\delta x\|^2 - \alpha^{-1}\|\delta v\|^2. \]

We can write the last term on the left-hand side as
\[ \lim_{n \to \infty} \frac{1}{n} \langle Qx_n \mid x_n \rangle + \langle q \mid \delta x \rangle + \lim_{n \to \infty} \frac{1}{n} \langle Ax_n \mid y_n \rangle = -\alpha^{-1}\|\delta x\|^2 - \alpha^{-1}\|\delta z\|^2. \]

Due to Prop. 2.1, Prop. 4.3 and Prop. 4.4 imply that, if the limit \( \delta y \) is nonzero, then problem (1) is infeasible, and similarly, if \( \delta x \) is nonzero, then its dual is infeasible. Thanks to the fact that \( (\delta y_n, \delta x_n) \to (\delta y, \delta x) \), we can now extend the termination criteria proposed in [BGB19, §5.2] for the more general case where \( C \) is a general nonempty closed convex set. The criteria in [BGB19, §5.2] evaluate conditions given in Prop. 2.1 at \( \delta y_n \) and \( \delta x_n \), and have already formed the basis for stable numerical implementations [SBG+20, GCG19]. Our results pave the way for similar developments in the more general setting considered here.

5 Proximal-Point Algorithm

The proximal-point algorithm is a method for finding a vector \( w \in H \) that solves the following inclusion problem:
\[ 0 \in B(w), \] (18)
where $B: \mathcal{H} \to 2^\mathcal{H}$ is a maximally monotone operator. An iteration of the algorithm in application to problem (18) can be written as

$$w_{n+1} = (\text{Id} + \gamma B)^{-1} w_n,$$

where $\gamma > 0$ is the regularization parameter. 

Due to [BC17, Cor. 16.30], we can rewrite (2) as

$$0 \in M(x, y) := \left( Qx + q + A^* y, -Ax + \partial i_C^*(y) \right),$$

where $M: (\mathcal{H}_1 \times \mathcal{H}_2) \to 2^{(\mathcal{H}_1 \times \mathcal{H}_2)}$ is a maximally monotone operator [Roc76]. An iteration of the proximal-point algorithm in application to the inclusion above is then

$$(x_{n+1}, y_{n+1}) = (\text{Id} + \gamma M)^{-1} (x_n, y_n),$$

which was also analyzed in [HTP19]. We will exploit the following result [BC17, Prop. 23.8] to analyze the algorithm:

**Fact 5.1.** Operator $T_{PP} := (\text{Id} + \gamma M)^{-1}$ is the resolvent of a maximally monotone operator and is thus $(1/2)$-averaged.

Iteration (19) reads

$$0 = x_{n+1} - x_n + \gamma (Qx_{n+1} + q + A^* y_{n+1})$$

$$0 \in y_{n+1} - y_n + \gamma (-Ax_{n+1} + \partial i_C^*(y_{n+1})).$$

Inclusion (20b) can be written as

$$\gamma Ax_{n+1} + y_n \in (\text{Id} + \gamma \partial i_C^*) y_{n+1},$$

which is equivalent to [BC17, Prop. 16.44]

$$y_{n+1} = \text{Prox}_{\gamma \partial i_C^*} (\gamma Ax_{n+1} + y_n) = \gamma Ax_{n+1} + y_n - \gamma P_C(Ax_{n+1} + \gamma^{-1} y_n),$$

where the second equality follows from [BC17, Thm. 14.3]. Let us define the following auxiliary iterates of iteration (19):

$$v_{n+1} := Ax_{n+1} + \gamma^{-1} y_n$$

$$z_{n+1} := P_C v_{n+1},$$

and observe from (21) that

$$y_{n+1} = \gamma (\text{Id} - P_C) v_{n+1}.$$

Using (20a) and (21), we now obtain the following relations between the iterates:

$$Ax_{n+1} - P_C v_{n+1} = \gamma^{-1} \delta y_{n+1}$$

$$Qx_{n+1} + q + \gamma A^* (\text{Id} - P_C) v_{n+1} = -\gamma^{-1} \delta x_{n+1}.$$
Similarly as for the Douglas-Rachford algorithm, the pair \((z_{n+1}, y_{n+1})\) satisfies optimality condition (3c) for all \(n \in \mathbb{N}\). Observe that the optimality residuals, given by the norms of the left-hand terms in (23), can be computed by evaluating the norms of \(\delta y_{n+1}\) and \(\delta x_{n+1}\).

The following corollary follows directly from Lem. 3.1, Prop. 3.2, and Fact 5.1:

**Corollary 5.2.** Let the sequences \((x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}, (v_n)_{n\in\mathbb{N}},\) and \((z_n)_{n\in\mathbb{N}}\) be given by (19) and (22), and \((\delta x, \delta y) := P_{\text{ran}(T_{\text{proj}})}(0)\). Then

1. \(\frac{1}{n}(x_n, y_n, v_n) \rightarrow (\delta x, \delta y, A\delta x + \gamma^{-1}\delta y)\).
2. \((\delta x_n, \delta y_n, \delta v_n) \rightarrow (\delta x, \delta y, A\delta x + \gamma^{-1}\delta y)\).
3. \(y_{n+1} \in (\text{rec} C)^\ominus\).
4. \(\delta y = \gamma P_{(\text{rec} C)^\ominus}(\delta v)\).
5. \(\frac{1}{n}z_n \rightarrow \delta z := P_{\text{rec} C}(\delta v)\).
6. \(\lim_{n \to \infty} \frac{1}{n} \langle z_n \mid y_n \rangle = \sigma_C(\delta y)\).

The proofs of the following two propositions follow similar arguments as those in Section 4, and are thus omitted.

**Proposition 5.3.** The following relations hold between \(\delta x, \delta z,\) and \(\delta y,\) which are defined in Cor. 5.2:

1. \(A\delta x = \delta z\).
2. \(Q\delta x = 0\).
3. \(A^*\delta y = 0\).

**Proposition 5.4.** The following identities hold for \(\delta x\) and \(\delta y\), which are defined in Cor. 5.2:

1. \(\langle q \mid \delta x \rangle = -\gamma^{-1}\|\delta x\|^2\).
2. \(\sigma_C(\delta y) = -\gamma^{-1}\|\delta y\|^2\).

The authors in [HTP19] use similar termination criteria to those given in [BGSB19, §5.2] to detect infeasibility of convex quadratic programs using the algorithm given by iteration (19), though they do not prove that \(\delta y\) and \(\delta x\) are indeed infeasibility certificates whenever the problem is strongly infeasible. Identities in (23) show that when \((\delta y, \delta x) = (0,0)\), the optimality conditions (3) are satisfied in the limit. Otherwise, Prop. 2.1, Prop. 5.3, and Prop. 5.4 imply that problem (1) and/or its dual is strongly infeasible.

**References**

[BBR78] 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):1–9, 1978.

[BC17] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer International Publishing, 2nd edition, 2017.
[BDM16] H. H. Bauschke, M. N. Dao, and W. M. Moursi. The Douglas-Rachford algorithm in the affine-convex case. *Operations Research Letters*, 44(3):379–382, 2016.

[BGSB19] G. Banjac, P. Goulart, B. Stellato, and S. Boyd. Infeasibility detection in the alternating direction method of multipliers for convex optimization. *Journal of Optimization Theory and Applications*, 183(2):490–519, 2019.

[BM16] H. H. Bauschke and W. M. Moursi. The Douglas-Rachford algorithm for two (not necessarily intersecting) affine subspaces. *SIAM Journal on Optimization*, 26(2):968–985, 2016.

[BM17] H. H. Bauschke and W. M. Moursi. On the Douglas-Rachford algorithm. *Mathematical Programming*, 164(1):263–284, 2017.

[EB92] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[GCG19] M. Garstka, M. Cannon, and P. Goulart. COSMO: a conic operator splitting method for large convex problems. In *European Control Conference (ECC)*, 2019.

[HTP19] B. Hermans, A. Themelis, and P. Patrinos. QPALM: a Newton-type proximal augmented Lagrangian method for quadratic programs. In *IEEE Conference on Decision and Control (CDC)*, 2019.

[LM79] P. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[LMK20] D. Liao-McPherson and I. Kolmanovsky. FBstab: a proximally stabilized semismooth algorithm for convex quadratic programming. *Automatica*, 113, 2020.

[Paz71] A. Pazy. Asymptotic behavior of contractions in Hilbert space. *Israel Journal of Mathematics*, 9(2):235–240, 1971.

[RLY19] E. Ryu, Y. Liu, and W. Yin. Douglas-Rachford splitting and ADMM for pathological convex optimization. *Computational Optimization and Applications*, 74:747–778, 2019.

[Roc76] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976.

[SBG+20] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd. OSQP: an operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 2020.