Douglas-Rachford Splitting for Pathological Convex Optimization

Ernest K. Ryu*   Yanli Liu*   Wotao Yin*

January 19, 2018

Abstract

Despite the vast literature on DRS, there has been very little work analyzing their behavior under pathologies. Most analyses assume a primal solution exists, a dual solution exists, and strong duality holds. When these assumptions are not met, i.e., under pathologies, the theory often breaks down and the empirical performance may degrade significantly. In this paper, we establish that DRS only requires strong duality to work, in the sense that asymptotically iterates are approximately feasible and approximately optimal.

1   Introduction

Douglas-Rachford splitting (DRS) is a classical method originally presented in [45, 24, 34, 32]. Over the last decade, DRS has enjoyed a resurgence of popularity, as the demand to solve ever larger problems grew.

DRS has strong theoretical guarantees and empirical performance, but such results are often limited to non-pathological problems; in particular, most analyses assume a primal solution exists, a dual solution exists, and strong duality holds. When these assumptions are not met, i.e., under pathologies, the theory often breaks down and the empirical performance may degrade significantly. Surprisingly, there has been very little work analyzing DRS under pathologies, despite the vast DRS literature.

In this paper, we analyze the asymptotic behavior of DRS under pathologies. While it is well-known that the iterates "diverge" in such cases, the precise manner in which they do so was not known. We establish that when strong duality holds, i.e., when $p^\star = d^\star \in [-\infty, \infty]$, DRS works, in the sense that asymptotically the divergent iterates are approximately feasible and approximately optimal. The assumption that primal and dual solutions exist is not necessary.

Furthermore, we conjecture that DRS necessarily fails when strong duality fails, and we present empirical evidence that supports (but does not prove) this conjecture. In other words, we believe strong duality is the necessary and sufficient condition for DRS to work.

We organize the paper as follows. Section 2 reviews standard notions of convex analysis, states several known results, and sets up the notation. Section 3 presents the main theoretical contribution of this paper. Section 4 analyzes DRS under pathologies with the theory of Section 3. Section 5 presents counter examples to make additional observations. Section 6 concludes the paper.

Notes Alternating directions method of multipliers (ADMM) is a method closely related to DRS [28, 29, 27]. There also has not been much work analyzing ADMM under pathologies. In this paper, we focus on DRS.

There is a degree of primal-dual symmetry to our results that we do not explicitly address in the interest of space. Rather, we take the viewpoint that the primal problem is the problem of interest and the dual problem is an auxiliary conceptual and computational tool.

Pathological conic programs can be solved with facial reduction [15, 17, 18, 51, 41, 19, 59, 60, 48, 36, 47, 43]. Facial reduction is a pre-processing step that rids the problem of difficult pathologies, and the simplified problem can be solved with other methods. This approach, however, does not immediately generalize to non-conic setups, and facial reduction is numerically challenging. In any case, this fact is besides the point, as our goal is to analyze DRS when it is directly applied to pathological convex programs.

We reference and discuss other work that are related or that we build upon later as we discuss the relevant concepts.

---

*UCLA, Mathematics

2   Preliminaries

In this section, we review standard notions of convex analysis, state several known results, and set up the notation. For the sake of brevity, we omit proofs or direct references of the standard results and refer interested readers to standard references such as [53, 52, 7].

Throughout this paper, we distinguish the scalar and vector zero by writing 0 for $0 \in \mathbb{R}$, and $\mathbf{0}$ for $\mathbf{0} \in \mathbb{R}^n$.

Convex set A set $C \subseteq \mathbb{R}^n$ is convex, if $x, y \in C$ and $\theta \in [0, 1]$ implies

$$\theta x + (1 - \theta)y \in C.$$

Given a set $A \subseteq \mathbb{R}^n$, write $\overline{A}$ for its closure. If C is convex $\overline{C}$ is convex.

Minkowski sum Let $A, B \subseteq \mathbb{R}^n$ be two nonempty sets. The Minkowski sum and differences of A and B are

$$A + B = \{a + b \,|\, a \in A,\, b \in B\}, \qquad A - B = \{a - b \,|\, a \in A,\, b \in B\},$$

respectively. If A and B are convex, then $A + B$ and $A - B$ are convex. However, neither $A + B$ nor $A - B$ is guaranteed to be closed, even when A and B are nonempty closed convex sets.

Distance between sets Let $x \in \mathbb{R}^n$ be a point and $A, B \subseteq \mathbb{R}^n$ be nomempty sets. Write

$$\mathrm{dist}(x, A) = \inf\{\|x - a\| \,|\, a \in A\}$$

for the distance between the point x and the set A. Write

$$\mathrm{dist}(A, B) = \inf\{\|a - b\| \,|\, a \in A,\, b \in B\}$$

for the distance between the sets A and B.

If A is a nonempty closed set, then $\mathrm{dist}(x, A) = 0$ if and only if $x \in A$. However, $\mathrm{dist}(A, B) = 0$ is possible even when $A \cap B = \emptyset$, regardless of whether A and B are closed. By definition $\mathrm{dist}(A, B) = \mathrm{dist}(\mathbf{0}, A - B)$, so $\mathrm{dist}(A, B) = 0$ if and only if $\mathbf{0} \in \overline{A - B}$.

Convex function A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex if

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

for all $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$.

Closed convex proper functions We say $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is closed if its epigraph

$$\big\{(x, \alpha) \in \mathbb{R}^{n+1} \mid f(x) \leq \alpha\big\}$$

is a closed subset of $\mathbb{R}^{n+1}$. We say $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper if $f(x) < \infty$ for some x. In this paper, we focus our attention on closed, convex, and proper (CCP) functions. Closedness and properness are mild assumptions.

If $A \in \mathbb{R}^{n \times m}$ and f is CCP, then $f(Ax)$ is CCP or $f(Ax) = \infty$ everywhere. If f and g are CCP functions, then $f + g$ is CCP or $f + g = \infty$ everywhere. If $\gamma > 0$, then γf is CCP.

Domain Given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, define its (effective) domain as

$$\mathrm{dom}\, f = \{x \in \mathbb{R}^n \,|\, f(x) < \infty\}.$$

If f is a convex function, then $\mathrm{dom}\, f$ is a convex set. However, $\mathrm{dom}\, f$ may not be closed even when f is CCP. For any $\gamma > 0$, we have $\mathrm{dom}\, \gamma f = \mathrm{dom}\, f$.

Conjugate function Given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, define its conjugate as

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\} = -\inf_{x \in \mathbb{R}^n} \{f(x) - y^T x\}.$$

If f is convex and proper (but not necessarily closed), then $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is CCP. If f is CCP, then $(f^*)^* = f$. For any $\gamma > 0$, we have $(\gamma f)^*(x) = \gamma f^*(x/\gamma)$ and $\mathrm{dom}(\gamma f)^* = \gamma \,\mathrm{dom}\, f^*$. If $h(x) = g(-x)$, then $h^*(y) = g^*(-y)$.

Projection Given a nonempty closed convex set $C \subseteq \mathbb{R}^n$, define the projection onto C as

$$\Pi_C(x_0) = \underset{x \in C}{\arg\min} \|x - x_0\|^2.$$

The operator $\Pi_C : \mathbb{R}^n \to \mathbb{R}^n$ is well-defined, i.e., the minimizer uniquely exists. The projection inequality states that

$$\langle v - \Pi_C x, \Pi_C x - x \rangle \geq 0$$

for any $v \in C$ and $x \in \mathbb{R}^n$.

Indicator function Given a nonempty set $C \subseteq \mathbb{R}^n$, define its indicator function as

$$\delta_C(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \in C \\ \infty & \text{otherwise.} \end{array} \right.$$

When C is nonempty closed convex, $\delta_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is CCP.

Support function Given a nonempty set $C \subseteq \mathbb{R}^n$, define its support function as

$$\sigma_C(y) = \sup_{x \in C} \{x^T y\}.$$

$\sigma_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is CCP. When C is nonempty and convex, we have $\sigma_C = \sigma_{\overline{C}}$. If $A, B \in \mathbb{R}^n$ are nonempty convex sets, then $\sigma_{A+B} = \sigma_A + \sigma_B$. If C is a nonempty closed convex set then $(\sigma_C)^* = \delta_C$.

Recession function Given a CCP function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, define its recession function as

$$\operatorname{rec} f(d) = \lim_{\alpha \to \infty} \frac{f(x + \alpha d) - f(x)}{\alpha}. \tag{1}$$

The value $\operatorname{rec} f(d)$ is independent of the choice $x \in \operatorname{dom} f$. An alternate equivalent definition is

$$\operatorname{rec} f(d) = \sup_{x \in \operatorname{dom} f} \{f(x + d) - f(x)\}.$$

$\operatorname{rec} f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a positively homogeneous CCP function. By positively homogeneous, we mean $\operatorname{rec} f(\alpha d) = \alpha \operatorname{rec} f(d)$ for any $\alpha \geq 0$. If $h(x) = g(-x)$, then $\operatorname{rec}(h^*)(d) = \operatorname{rec}(g^*)(-d)$. When f and g are CCP functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{\infty\}$ either $f(x) + g(x) = \infty$ for all $x \in \mathbb{R}^n$ or $\operatorname{rec}(f + g) = \operatorname{rec} f + \operatorname{rec} g$.

Recession functions and support functions are related through the following duality relationship. If f is CCP, then

$$\sigma_{\operatorname{dom} f^*} = \operatorname{rec} f, \quad \sigma_{\operatorname{dom} f} = \operatorname{rec}(f^*).$$

Proximal operator Given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, define its proximal operator $\operatorname{Prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\operatorname{Prox}_f(z) = \underset{x \in \mathbb{R}^n}{\arg\min} \left\{ f(x) + (1/2)\|x - z\|_2^2 \right\}.$$

When f is CCP, the arg min uniquely exists, and therefore Prox_f is well-defined.

When C is a nonempty closed convex set, $\operatorname{Prox}_{\delta_C} = \Pi_C$. When f is CCP, $\operatorname{Prox}_f + \operatorname{Prox}_{f^*} = I$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity operator.

Nonexpansive and firmly-nonexpansive operators A mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive if

$$\|T(x) - T(y)\| \leq \|x - y\|$$

for all $x, y \in \mathbb{R}^n$. Nonexpansive mappings are, by definition, Lipschitz continuous with Lipschitz constant 1. T is firmly-nonexpansive if $T = (1/2)I + (1/2)S$ where $S : \mathbb{R}^n \to \mathbb{R}^n$ is another nonexpansive operator. Alternatively, $T : \mathbb{R}^n \to \mathbb{R}^n$ is firmly-nonexpansive if and only if

$$\|T(x) - T(y)\|^2 \leq \langle x - y, T(x) - T(y) \rangle$$

holds for all $x, y \in \mathbb{R}^n$. Proximal and projection operators are firmly-nonexpansive.

2.1 Duality and primal subvalue

We call the optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + g(x), \tag{P}$$

the primal problem. We call the optimization problem

$$\underset{\nu \in \mathbb{R}^n}{\text{maximize}} \quad -f^*(\nu) - g^*(-\nu), \tag{D}$$

the dual problem. Throughout this paper, we assume f and g are CCP.

Feasibility The primal problem (P) is feasible, weakly infeasible, and strongly infeasible if

$$\mathbf{0} \in \text{dom}\, f - \text{dom}\, g \tag{feasible}$$

$$\mathbf{0} \in \overline{\text{dom}\, f - \text{dom}\, g} \backslash (\text{dom}\, f - \text{dom}\, g) \tag{weakly infeasible}$$

$$\mathbf{0} \neq \Pi_{\overline{\text{dom}\, f - \text{dom}\, g}}(\mathbf{0}), \tag{strongly infeasible}$$

respectively. This characterization is complete; (P) always falls under exactly one of the three cases. (P) is infeasible if it is not feasible.

The dual problem (D) is feasible, weakly infeasible, and strongly infeasible if

$$\mathbf{0} \in \text{dom}(f^*) + \text{dom}(g^*) \tag{feasible}$$

$$\mathbf{0} \in \overline{\text{dom}(f^*) + \text{dom}(g^*)} \backslash (\text{dom}(f^*) + \text{dom}(g^*)) \tag{weakly infeasible}$$

$$\mathbf{0} \neq \Pi_{\overline{\text{dom}(f^*) + \text{dom}(g^*)}}(\mathbf{0}), \tag{strongly infeasible}$$

respectively. This characterization is complete; (D) always falls under exactly one of the three cases. (D) is infeasible if it is not feasible.

These terse definitions are equivalent to the usual definitions of feasibility and infeasibility: (P) is feasible if there is an $x \in \mathbb{R}^n$ such that $f(x) + g(x) < \infty$, and infeasible otherwise; an infeasible problem (P) is weakly infeasible if $\text{dist}(\text{dom}\, f, \text{dom}\, g) = 0$ and strongly infeasible if $\text{dist}(\text{dom}\, f, \text{dom}\, g) > 0$. The same can be said with the usual analogous definitions for (D).

Optimal values and weak duality We call $p^\star = \inf\{f(x) + g(x) \mid x \in \mathbb{R}^n\}$ the primal optimal value and $d^\star = \sup\{-f^*(\nu) - g^*(-\nu) \mid \nu \in \mathbb{R}^n\}$ the dual optimal value. We let $p^\star = \infty$ if (P) is infeasible and $d^\star = -\infty$ if (D) is infeasible. Weak duality, which always holds, states that $d^\star \leq p^\star$.

Strong duality If $d^\star = p^\star$, we say strong duality holds between the primal-dual problem pair (P) and (D). This includes the cases $d^\star = p^\star = -\infty$ and $d^\star = p^\star = +\infty$. When f and g are convex, strong duality holds often, but not always. This paper focuses on the case when strong duality holds.

Total duality Solutions to problems (P) and (D) may or may not exist. If (P) has a solution, (D) has a solution, and strong duality holds, we say *total duality* holds between the primal-dual problem pair (P) and (D).

Primal subvalue Define the *primal subvalue* of (P) as

$$p^- = \lim_{\varepsilon \to 0^+} \inf_{x, y \in \mathbb{R}^n} \{f(x) + g(y) \mid \|x - y\| \leq \varepsilon\}.$$

The notion of primal subvalue is standard in conic programming [33, 61, 37, 38]. Here, we generalize it to general convex programs.

Lemma 1. *If f and g are CCP, then $d^\star = p^- \leq p^\star$.*

Although we have not seen Lemma 1 stated exactly in this form, the idea is well known [52].

Strong duality is well-posedness Lemma 1 tells us that strong duality holds when $p^- = p^\star$, and we can interpret strong duality as well-posedness of (P). The primal subvalue p^- is the optimal value of (P) achieved with infinitesimal infeasibilities. When the infinitesimal infeasibilities provide a non-infinitestimal improvement to the function value, we can consider (P) ill-posed [52, 38, 46, 25, 26].

4

2.2 Douglas-Rachford operator and fixed-point encoding

Douglas-Rachford splitting (DRS) applied to (P) is

$$\begin{aligned}
x^{k+1/2} &= \text{Prox}_{\gamma f}(z^k) \\
x^{k+1} &= \text{Prox}_{\gamma g}(2x^{k+1/2} - z^k) \\
z^{k+1} &= z^k + x^{k+1} - x^{k+1/2}
\end{aligned} \tag{2}$$

with a starting point $z^0 \in \mathbb{R}^n$ and a parameter $\gamma > 0$. This iteration can be expressed more concisely as

$$z^{k+1} = T_\gamma(z^k)$$

where

$$T_\gamma = \frac{1}{2}I + \frac{1}{2}(2\,\text{Prox}_{\gamma g} - I)(2\,\text{Prox}_{\gamma f} - I).$$

$T_\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is a firmly-nonexpansive operator, and we interpret DRS as a fixed-point iteration.

Total duality and convergence of DRS The standard analysis of DRS assumes total duality.

Theorem 2. *[5] Total duality holds between* (P) *and* (D) *if and only if the DRS operator T_γ has a fixed point for some $\gamma > 0$.*

Theorem 3. *[5] If total duality holds between* (P) *and* (D), *then DRS converges in that $z^k \to z^\star$, where $x^\star = \text{Prox}_{\gamma f}(z^\star)$ is a solution of* (P). *If total duality does not hold between* (P) *and* (D), *then DRS diverges in that $\|z^k\| \to \infty$.*

Theorem 2 and 3 are well known, although the term "total duality" is not always used. More often, total duality is assumed by instead assuming a saddle point exists for an appropriate Lagrangian.

Despite the huge popularity of DRS, there has been surprisingly little work investigating the behavior of DRS when total duality does not hold. The understanding is still incomplete, but there has been some recent progress: [8, 12, 13, 35] present results in specific pathological setups, [10, 11, 13] present results on general setups, and [49, 56, 4] analyze ADMM, which is closely related to DRS, under specific pathological setups for conic programs.

Fixed-point encoding We can view DRS as an instance of a more general approach called *fixed-point encoding*. The idea of fixed-point encoding is to transform a problem into an equivalent fixed-point problem such that solutions of the original problem are encoded as fixed points. See, for example, the surveys [20, 22, 21, 55] or the book [7] for more information on this subject. Another example of fixed-point encoding is

$$\underset{x}{\text{minimize}} \ f(x) \quad \Longleftrightarrow \quad \underset{x}{\text{solve}} \ (I - \nabla f)(x) = x \tag{3}$$

where f is a differentiable convex function.

A problem arises when the original problem has no solution or when the transformation fails due to pathologies. For example, even when f has no minimizer in (3) the optimization problem can still be meaningful, but it's harder to ascribe meaning to the fixed-point problem without fixed points.

As stated in Theorem 2, the DRS operator T_γ has no fixed point when total duality fails. As we see in Section 4, this can happen even when the primal problem (P) has a solution. In this case, the DRS iteration is a fixed-point iteration without a fixed point.

2.3 Fixed-point iterations without fixed points

Analyzing fixed-point iterations without fixed points is a large part of our analysis of DRS in the absence of total duality.

Infimal displacement vector Let T be a firmly-nonexpansive operator. Write

$$\text{range}(I - T) = \{z - T(z) \,|\, z \in \mathbb{R}^n\}$$

for the range of the operator $I - T$. Note that T has a fixed point if and only if $\mathbf{0} \in \text{range}(I - T)$. The closure of this set, $\overline{\text{range}(I - T)}$, is closed and convex [44].

We call

$$v = P_{\overline{\text{range}(I-T)}}(\mathbf{0})$$

the *infimal displacement vector* of T. If T has a fixed point, then $v = \mathbf{0}$, but $v = \mathbf{0}$ is possible even when T has no fixed point. Remarkably, v elegantly characterizes the asymptotic behavior of fixed-point iterations with respect to T.

Theorem 4 ([44, 3]). *Assume T is firmly-nonexpansive. The fixed-point iteration $z^{k+1} = T(z^k)$ satisfies*

$$z^{k+1} - z^k \to -v.$$

Some recent work used Theorem 4 to analyze algorithms that can be interpreted as fixed-point iterations without fixed-points [6, 8, 16, 1, 10, 9, 12, 2, 39, 13, 54]. In fact, [10] coined the term infimal displacement vector.

Finding approximate fixed points In a loose sense, we say z is an approximate fixed point if $T(z) \approx z$. Theorem 4 states the fixed-point iteration with respect to T finds an approximate fixed point if T is firmly-nonexpansive and $v = \mathbf{0}$.

However, an approximate fixed point of T_γ does not always correspond to an approximate solution of the (P), and analyzing the DRS iteration via approximate fixed points is not sufficient. We return to this point in Section 3.2.

2.4 Infimal displacement vector of the DRS operator

Since DRS is a fixed-point iteration with respect to the firmly-nonexpansive operator T_γ, characterizing the infimal displacement vector of T_γ is central to analyzing asymptotic behavior of DRS.

Recently, Bauschke et al. published the following elegant result.

Theorem 5 ([11]). *The infimal displacement vector v of T_1 is characterized by*

$$v = \arg\min_{z \in S} \|z\|^2$$

where

$$S = \overline{\text{dom } f - \text{dom } g} \cap \overline{\text{dom } f^* + \text{dom } g^*}.$$

The original result stated in [11] is more general as it applies to the DRS operator of monotone operators. In Section 3.1, we further refine this characterization of the infimal displacement vector for our setup.

3 Theoretical results

In this section, we present the main theoretical contribution of this paper. Our analysis requires a generalized notion of improving directions, so we define it first. Section 3.1 analyzes DRS as a fixed-point iteration without fixed points. Section 3.2 analyzes DRS as an optimization method that reduces function values. Section 3.3 directly analyzes the evolution of the $x^{k+1/2}$ and x^{k+1}-iterates of DRS. Later in Section 4 we combine these results to analyze the asymptotic behavior of DRS. The definition of improving directions via recession functions is new, and identifying its value in the analysis is a key contribution of this work.

Interpretation of recession function Recession functions and directional derivatives are related. Given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, its directional derivative is

$$f'(x; d) = \lim_{\lambda \searrow 0} \frac{f(x + \lambda d) - f(x)}{\lambda}.$$

When f is CCP, $f'(x; d) \in \mathbb{R} \cup \{\pm\infty\}$ is well-defined for all $x \in \operatorname{dom} f$ and $d \in \mathbb{R}^n$.

Lemma 6. *If f is CCP, $\operatorname{rec} f(d) < \infty$, and $x \in \operatorname{dom} f$, then $x + \alpha d \in \operatorname{dom} f$ for all $\alpha \geq 0$ and*

$$f'(x + \alpha d; d) \nearrow \operatorname{rec} f(d)$$

as $\alpha \to \infty$. In other words, $f'(x + \alpha d; d)$ monotonically increases to $\operatorname{rec} f(d)$.

Proof. That $x + \alpha d \in \operatorname{dom} f$ for all $\alpha \geq 0$ when $\operatorname{rec} f(d) < \infty$ follows from the definition of the recession function.

Write $g(\alpha) = f(x + \alpha d)$, which is a convex function on $\mathbb{R}$.

$$\frac{f(x + \alpha d) - f(x)}{\alpha} = \frac{g(\alpha) - g(\mathbf{0})}{\alpha} = \frac{1}{\alpha} \int_0^\alpha g'(a)\, da$$

where g' is the right-derivative of g. Note that $g'(x) = f'(x + \alpha d; d)$. Since g' is a non-decreasing function on $\mathbb{R}$,

$$\operatorname{rec} f(d) = \lim_{\alpha \to \infty} \frac{f(x + \alpha d) - f(x)}{\alpha} = \lim_{\alpha \to \infty} \frac{1}{\alpha} \int_0^\alpha g'(a)\, da.$$

This implies $f'(x + \alpha d; d) = g'(\alpha) \to \operatorname{rec} f(d)$. $\qquad\square$

Lemma 6 allows us to interpret $\operatorname{rec} f(d)$ as the asymptotic directional derivative of f as we move in direction d. In light of this interpretation, the following result makes more sense.

Lemma 7. *If f is CCP, then*

$$(1/\alpha) f(x + \alpha d) = \operatorname{rec} f(d) + o(1)$$

as $\alpha \to \infty$.

Proof. From the definition (1), $f(x + \alpha d) - f(x) \sim \alpha \operatorname{rec} f(d)$ as $\alpha \to \infty$. $\qquad\square$

Improving direction We say $d \in \mathbb{R}^n$ is a *primal improving direction* for the primal problem (P) if

$$\operatorname{rec} f(d) + \operatorname{rec} g(d) < 0.$$

Note $\operatorname{rec} f(d) + \operatorname{rec} g(d) = \operatorname{rec}(f + g)(d)$ when (P) is feasible. While this definition still makes sense when (P) is infeasible, we do not consider the case where an infeasible (P) has an improving direction in this paper.

The notion of (primal) improving direction is standard in conic programming [37, 40, 38]. Here, we extend it to general convex programs of the form (P).

Corollary 8. *Assume* (P) *has a feasible point $x \in \operatorname{dom} f \cap \operatorname{dom} g$ and a primal improving direction d. Then*

$$f(x + \alpha d) + g(x + \alpha d) = \alpha \operatorname{rec}(f + g)(d) + o(\alpha)$$

as $\alpha \to \infty$, and therefore $p^\star = -\infty$.

So existence of a primal improving direction implies $p^\star = -\infty$, but $p^\star = -\infty$ is possible even when (P) has no improving direction. We discuss such an example in Section 4.

Likewise, we say $d \in \mathbb{R}^n$ is a *dual improving direction* if

$$\operatorname{rec}(f^*)(d) + \operatorname{rec}(g^*)(-d) < 0.$$

Note (D) is a maximization problem. So decreasing $f^*(\nu) + g^*(-\nu)$ is the same as increasing $-f^*(\nu) - g^*(-\nu)$, which is an improvement.

Corollary 9. *Assume* (D) *is feasible and there is a dual improving direction. Then $d^\star = \infty$.*

3.1 Infimal displacement vector of the DRS operator

In this section, we refine the characterization of the infimal displacement vector discussed in Section 2.4 using the generalized notion of improving directions. For the sake of simplicity, we first analyze

$$T_1 = \frac{1}{2}I + \frac{1}{2}(2\operatorname{Prox}_g - I)(2\operatorname{Prox}_f - I)$$

and then translate the results to T_γ.

We first consider the case where (P) is feasible and characterize the infimal displacement vector based on the primal improving direction or the absence of it.

Lemma 10. *For the primal-dual problem pair* (P) *and* (D), *we have*

$$-\Pi_{\overline{(\operatorname{dom} f^* + \operatorname{dom} g^*)}}(\mathbf{0}) = \operatorname{Prox}_{\operatorname{rec} f + \operatorname{rec} g}(\mathbf{0}).$$

Proof. Let A and B be nonempty convex sets. The properties of the support function discussed in Section 2 tell us

$$(\delta_{\overline{A+B}})^*(x) = \sigma_{\overline{A+B}}(x) = \sigma_{A+B}(x) = \sigma_A(x) + \sigma_B(x).$$

Setting $A = \operatorname{dom} f^*$ and $B = \operatorname{dom} g^*$ gives us

$$(\delta_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}})^*(x) = \sigma_{\operatorname{dom} f^*}(x) + \sigma_{\operatorname{dom} g^*}(x).$$

Based on the previous identities we've discussed, we have

$$\begin{aligned}
\Pi_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0}) &= \operatorname{Prox}_{\delta_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}}(\mathbf{0}) \\
&= (I - \operatorname{Prox}_{\sigma_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}})(\mathbf{0}) \\
&= -\operatorname{Prox}_{\sigma_{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0}) \\
&= -\operatorname{Prox}_{\sigma_{\operatorname{dom} f^*} + \sigma_{\operatorname{dom} g^*}}(\mathbf{0}) \\
&= -\operatorname{Prox}_{\operatorname{rec} f + \operatorname{rec} g}(\mathbf{0})
\end{aligned}$$

$\square$

Theorem 11. (P) *has an improving direction if and only if* (D) *is strongly infeasible. Write*

$$d = -\Pi_{\overline{(\operatorname{dom} f^* + \operatorname{dom} g^*)}}(\mathbf{0}).$$

If (P) *has an improving direction, then* $d \neq \mathbf{0}$ *and* d *is an improving direction. If* (P) *has no improving direction, then* $d = \mathbf{0}$.

Proof. Remember that $\operatorname{rec} f + \operatorname{rec} g$ is a convex positively homogeneous function. Since $\operatorname{rec} f(\mathbf{0}) + \operatorname{rec} g(\mathbf{0}) = 0$,

$$\mathbf{0} = \arg\min_x \left\{ \operatorname{rec} f(x) + \operatorname{rec} g(x) + (1/2)\|x\|^2 \right\} = \operatorname{Prox}_{\operatorname{rec} f + \operatorname{rec} g}(\mathbf{0})$$

if and only if $\operatorname{rec} f(x) + \operatorname{rec} g(x) \geq 0$ for all $x \in \mathbb{R}^n$. By our definition of an improving direction, $\operatorname{rec} f(x) + \operatorname{rec} g(x) \geq 0$ for all $x \in \mathbb{R}^n$ if and only if there is no improving direction.

By definition, $\mathbf{0} = \Pi_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0})$ if and only if (D) is not strongly infeasible. So with Lemma 10, we conclude (P) has an improving direction if and only if (D) is strongly infeasible.

It remains to show that

$$d = \arg\min_x \left\{ \operatorname{rec} f(x) + \operatorname{rec} g(x) + (1/2)\|x\|^2 \right\}$$

is an improving direction, if $d \neq \mathbf{0}$. Since d is defined as a minimizer, we have

$$\operatorname{rec} f(d) + \operatorname{rec} g(d) + (1/2)\|d\|^2 \leq \operatorname{rec} f(\mathbf{0}) + \operatorname{rec} g(\mathbf{0}) + (1/2)\|\mathbf{0}\|^2 = 0.$$

This implies $\operatorname{rec} f(d) + \operatorname{rec} g(d) \leq -(1/2)\|d\|^2 < 0$, i.e., d is an improving direction. $\square$

Theorem 12. *Assume* (P) *is feasible. Then*

$$v = -d = \Pi_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0})$$

is the infimal displacement vector of T_γ for $\gamma = 1$.

Proof. Let x_0 be a feasible point of (P). Since $\operatorname{rec} f(d) + \operatorname{rec} g(d) \leq 0 < \infty$ by Theorem 11 and the definition of an improving direction, we have $x_0 \in \operatorname{dom} f$, $x_0 + d \in \operatorname{dom} g$, and thus $-d \in \operatorname{dom} f - \operatorname{dom} g \subseteq \overline{\operatorname{dom} f - \operatorname{dom} g}$. Since $-d$ is the minimum-norm element of $\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}$, Theorem 5 tells us that $-d$ is the infimal displacement vector of T_1. $\square$

Corollary 13. *Assume* (P) *is feasible, and* (D) *is feasible. Then $v = \mathbf{0}$ is the infimal displacement vector of T_γ for any $\gamma > 0$.*

Corollary 14. *Assume* (P) *is feasible, and* (D) *is weakly infeasible. Then $v = \mathbf{0}$ is the infimal displacement vector of T_γ for any $\gamma > 0$.*

Corollary 15. *Assume* (P) *is feasible, and* (D) *is strongly infeasible. Then*

$$v = -\gamma d = \gamma \Pi_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0}) \neq \mathbf{0}$$

is the infimal displacement vector of T_γ for any $\gamma > 0$. Furthermore, d is an improving direction of (P).

Next, we consider the case where (D) is feasible and characterize the infimal displacement vector based on the dual improving direction or the absence of it.

Theorem 16. *Assume* (D) *is feasible. Then*

$$v = -d' = \Pi_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(\mathbf{0})$$

is the infimal displacement vector of T_1.

Proof. Following the same logic as before, we have

$$\Pi_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(\mathbf{0}) = -\arg\min_{\nu} \left\{ \operatorname{rec}(f^*)(\nu) + \operatorname{rec}(g^*)(-\nu) + (1/2)\|\nu\|^2 \right\},$$

and

$$d' = -\Pi_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(\mathbf{0})$$

is a dual improving direction, if $d' \neq \mathbf{0}$.

Let ν_0 be any feasible point of (D). Then $\nu_0 \in \operatorname{dom} f^*$ and $-\nu_0 - d' \in \operatorname{dom} g^*$. Therefore, $-d' \in \operatorname{dom} f^* + \operatorname{dom} g^* \subseteq \overline{\operatorname{dom} f^* + \operatorname{dom} g^*}$. Since $-d'$ is defined to be the minimum-norm element of $\overline{\operatorname{dom} f - \operatorname{dom} g}$ we conclude the statement with Theorem 5. $\square$

Corollary 17. *Assume* (D) *is feasible, and* (P) *is weakly infeasible. Then $v = \mathbf{0}$ is the infimal displacement vector of T_γ for any $\gamma > 0$.*

Corollary 18. *Assume* (D) *is feasible, and* (P) *is strongly infeasible. Then*

$$v = -d' = \Pi_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(\mathbf{0}) \neq \mathbf{0},$$

is the infimal displacement vector of T_γ for any $\gamma > 0$. Furthermore, d' is a dual improving direction.

Note that for Corollary 18, the infimal displacement vector is independent of the value of γ.

3.2 Function-value analysis

In Section 3.1, we analyzed the infimal displacement vector of the fixed-point iteration $z^{k+1} = T_\gamma(z^k)$. This, however, is not sufficient for characterizing the asymptotic behavior of DRS in relation to the original optimization problem (P).

Why function-value analysis is necessary Consider the convex function

$$f(x, y) = x^2/y$$

defined for $y > 0$. Note that f has minimizers, $(0, y)$ for any $y > 0$, and the operator $I - \nabla f$ has fixed points.

However, even though, $f(\sqrt{y}, y) = 1$ we have $\nabla f(\sqrt{y}, y) = (2/\sqrt{y}, -1/y) \to 0$ as $y \to \infty$. This means the point $(\sqrt{y}, y)$ for large y is not an approximate minimizer for f but is an approximate fixed point for the operator $I - \nabla f$. It is possible to construct a similar example with a DRS operator.

So approximate fixed points do not always correspond to approximate solutions of the original problem. This is why we need the function-value analysis to accompany the fixed-point theory in order to conclude the DRS iteration provides approximate solutions to the original optimization problem.

The function-value analysis We now present function-value analysis of DRS that is distinct from the fixed-point analysis of Section 3.1.

Lemma 19. *Assume $x^{k+1/2}$ and x^{k+1} are the DRS iterates as defined in (2). Then*

$$f(x^{k+1/2}) + g(x^{k+1}) - f(x) - g(x)$$
$$\leq (1/\gamma)(x^{k+1} - x^{k+1/2})^T(x - z^{k+1})$$

for all $k = 0, 1, \dots$ and any $x \in \mathbb{R}^n$.

Proof. Write

$$\tilde{\nabla} f(x^{k+1/2}) = (1/\gamma)(z^k - x^{k+1/2})$$
$$\tilde{\nabla} g(x^{k+1}) = (1/\gamma)(2x^{k+1/2} - z^k - x^{k+1}).$$

From the definition of the DRS iteration (2), we can verify that

$$\tilde{\nabla} f(x^{k+1/2}) \in \partial f(x^{k+1/2}), \qquad \tilde{\nabla} g(x^{k+1}) \in \partial g(x^{k+1})$$

and that

$$\tilde{\nabla} f(x^{k+1/2}) + \tilde{\nabla} g(x^{k+1}) = (1/\gamma)(x^{k+1/2} - x^{k+1}).$$

We also have

$$z^{k+1} = z^k - \gamma \tilde{\nabla} f(x^{k+1/2}) - \gamma \tilde{\nabla} g(x^{k+1}) = x^{k+1/2} - \gamma \tilde{\nabla} g(x^{k+1}).$$

If $x \notin \text{dom} f \cap \text{dom} g$, then $f(x) + g(x) = \infty$ for all $x \in \mathbb{R}^n$, and there is nothing to prove. Now, consider any $x \in \text{dom} f \cap \text{dom} g$. Then, by definition of subdifferentials,

$$f(x^{k+1/2}) - f(x) + g(x^{k+1}) - g(x)$$
$$\leq \tilde{\nabla} f(x^{k+1/2})^T(x^{k+1/2} - x) + \tilde{\nabla} g(x^{k+1})^T(x^{k+1} - x)$$
$$= (\tilde{\nabla} f(x^{k+1/2}) + \tilde{\nabla} g(x^{k+1}))^T(x^{k+1/2} - x) + \tilde{\nabla} g(x^{k+1})^T(x^{k+1} - x^{k+1/2})$$
$$= (x^{k+1} - x^{k+1/2})^T(\tilde{\nabla} g(x^{k+1}) - (1/\gamma)(x^{k+1/2} - x))$$
$$= (1/\gamma)(x^{k+1} - x^{k+1/2})^T(x - z^{k+1}).$$

$\square$

Corollary 20. *Assume $x^{k+1/2}$ and x^{k+1} are the DRS iterates as defined in (2). Then*

$$\frac{2\gamma}{k+1} \sum_{i=0}^{k} \left(f(x^{i+1/2}) - f(x) + g(x^{i+1}) - g(x) \right)$$

$$\leq -\frac{1}{k+1}\|z^{k+1}\|^2 + \frac{1}{k+1}\|z^0\|^2 - \frac{1}{k+1}\sum_{i=0}^{k}\|z^{k+1} - z^k\|^2 + \frac{2}{k+1}(z^{k+1} - z^0)^T x.$$

for all $k = 0, 1, \dots$ and any $x \in \mathbb{R}^n$.

Proof. If $\Delta^0, \Delta^1, \ldots$ is any sequence in $\mathbb{R}^n$, then

$$\sum_{j=0}^{k} (\Delta^j)^T \sum_{i=0}^{j} \Delta^i = \sum_{j=0}^{k} \sum_{i=0}^{k} \mathbb{1}\{i \leq j\} (\Delta^j)^T (\Delta^i)$$

$$= \frac{1}{2} \left\| \sum_{i=0}^{k} \Delta^i \right\|^2 + \frac{1}{2} \sum_{i=0}^{k} \|\Delta^i\|^2.$$

More specifically, we let $\Delta^k = z^{k+1} - z^k = x^{k+1} - x^{k+1/2}$. Now summing inequality of Lemma 19 gives us

$$\gamma \sum_{i=0}^{k} f(x^{i+1/2}) - f(x) + g(x^{i+1}) - g(x) \leq \sum_{j=0}^{k} (\Delta^j)^T (x - z^0) - \sum_{j=0}^{k} (\Delta^j)^T \sum_{i=0}^{j} \Delta^i$$

$$= (z^{k+1} - z^0)^T (x - z^0) - \frac{1}{2} \|z^{k+1} - z^0\|^2 - \frac{1}{2} \sum_{i=0}^{k} \|z^{i+1} - z^i\|^2$$

$$= -\frac{1}{2} \|z^{k+1}\|^2 + \frac{1}{2} \|z^0\|^2 + (z^{k+1} - z^0)^T x - \frac{1}{2} \sum_{i=0}^{k} \|z^{i+1} - z^i\|^2.$$

Finally, dividing both sides by $(k+1)/2$ produces the stated result. $\qquad\square$

Lemma 21. *Assume $x^{k+1/2}$ and x^{k+1} are the DRS iterates as defined in (2). Then*

$$\limsup_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \left(f(x^{i+1/2}) + g(x^{i+1}) \right) \leq p^\star$$

Proof. Assume $p^\star < \infty$, as otherwise there is nothing to prove. Let x be any $x \in \operatorname{dom} f \cap \operatorname{dom} g$.

By Theorem 4, $z^k \sim kv$. If $v \neq \mathbf{0}$, then the first (negative) term on the right-hand side of Corollary 20 dominates the positive terms. If $v = \mathbf{0}$, then $z^{k+1} - z^k = o(1)$, and both positive terms on the right-hand side of Corollary 20 converge to 0. In both cases, we have

$$\limsup_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) - f(x) + g(x^{i+1}) - g(x) \leq 0.$$

Since

$$\limsup_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) \leq f(x) + g(x)$$

for all $x \in \operatorname{dom} f \cap \operatorname{dom} g$, we minimize the right-hand side to obtain $p^\star$. $\qquad\square$

The following result is well known for the non-pathological setup where (P) has a solution, (D) has a solution, and $p^\star = d^\star \in (-\infty, \infty)$. It is still true under certain pathologies.

Theorem 22. *Assume $p^\star = d^\star \in [-\infty, \infty]$. Assume $x^{k+1/2}$ and x^{k+1} are the DRS iterates as defined in (2). Assume the infimal displacement vector v of T_γ satisfies $v = \mathbf{0}$. Then*

$$\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^\star$$

and

$$\liminf_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star.$$

11

Proof. By Theorem 4, $v = \mathbf{0}$ implies $x^{k+1/2} - x^{k+1} \to \mathbf{0}$. In turn, by Lemma 1, we have

$$\liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \geq p^\star.$$

Combining this with Lemma 21 gives us the first stated result.

If a real-valued sequence a^k satisfies

$$\liminf_{k \to \infty} a^k \geq a, \qquad \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} a^i = a,$$

then

$$\liminf_{k \to \infty} a^k = a.$$

The second stated result follows from this argument. $\square$

Corollary 23. *Assume $p^\star = d^\star$. Assume $x^{k+1/2}$ and x^{k+1} are the DRS iterates as defined in (2). If $x^{k+1/2}, x^{k+1} \to x^\star$ for some $x^\star$, then $x^\star$ is a solution.*

Proof. We first note that closed functions are by definition lower semi-continuous, and that f and g are assumed to be closed. By Theorem 22 we have

$$f(x^\star) + g(x^\star) \leq \liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star,$$

and we conclude $f(x^\star) + g(x^\star) = p^\star$. $\square$

Corollary 24. *Assume (P) is feasible and has an improving direction. Then*

$$f(x^{k+1/2}) + g(x^{k+1}) \to -\infty.$$

Proof. By Corollary 15 and Theorem 4, we have

$$z^{k+1} - z^k = x^{k+1} - x^{k+1/2} \to \gamma d.$$

Then Lemma 19 tells us that

$$(1/k)(f(x^{k+1/2}) + g(x^{k+1})) \leq -(1/\gamma)(x^{k+1} - x^{k+1/2})^T((1/k)z^{k+1}) + O(1/k)$$
$$\to -\gamma \|d\|^2,$$

which proves the statement. $\square$

3.3 Evolution of shadow iterates

Since DRS is a fixed-point iteration with respect to the z^k-iterates, the $x^{k+1/2}$ and x^{k+1}-iterates are sometimes called the *shadow iterates*. Section 3.1 characterized the evolution of the z^k-iterates. Here, we analyze the evolution of the shadow iterates.

Theorem 25. *If $v = \mathbf{0}$, then $x^{k+3/2} - x^{k+1/2} \to \mathbf{0}$ and $x^{k+2} - x^{k+1} \to \mathbf{0}$.*

Proof. Since $v = \mathbf{0}$, we have $z^{k+1} - z^k \to \mathbf{0}$. Since the map the defines $z^k \mapsto x^{k+1/2}$ and $z^{k+1} \mapsto x^{k+3/2}$ is Lipschitz continuous, $x^{k+3/2} - x^{k+1/2} \to \mathbf{0}$. Finally, $z^{k+1} - z^k \to \mathbf{0}$ and $x^{k+3/2} - x^{k+1/2} \to \mathbf{0}$ implies $x^{k+2} - x^{k+1} \to \mathbf{0}$. $\square$

Theorem 26. *If (P) is strongly infeasible and (D) is feasible, then $x^{k+3/2} - x^{k+1/2} \to \mathbf{0}$ and $x^{k+2} - x^{k+1} \to \mathbf{0}$.*

Proof. Write $-d'$ for the infimal displacement vector as given by Corollary 18. By Theorem 4, we have

$$z^{k+1} - z^k = x^{k+1} - x^{k+1/2} \to d'.$$

Since $-d' = \Pi_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(\mathbf{0})$ the projection inequality tells us that

$$\langle d', x - x^{k+1} \rangle + \|d'\|^2 \le 0$$

for any $x \in \operatorname{dom} f$. Using the fact that $x^{k+1/2} = \operatorname{Prox}_{\gamma f}(z^k)$ and firm-nonexpansiveness of Prox, we get

$$\begin{aligned}
\| \operatorname{Prox}_{\gamma f}(z^k + d') - x^{k+1/2} \|^2 &\le \langle d', \operatorname{Prox}_{\gamma f}(z^k + d') - x^{k+1/2} \rangle \\
&= \langle d', \operatorname{Prox}_{\gamma f}(z^k + d') - x^{k+1} \rangle + \langle d', x^{k+1} - x^{k+1/2} \rangle \\
&\to 0
\end{aligned}$$

since $\langle d', x^{k+1} - x^{k+1/2} \rangle \to \|d'\|^2$. So $\operatorname{Prox}_{\gamma f}(z^k + d') - x^{k+1/2} \to \mathbf{0}$. Since Prox is Lipschitz continuous, $z^{k+1} - z^k - d \to \mathbf{0}$ implies

$$\operatorname{Prox}_{\gamma f}(z^k + d) - \operatorname{Prox}_{\gamma f}(z^{k+1}) \to \mathbf{0}.$$

Putting everything together we conclude

$$x^{k+3/2} - x^{k+1/2} = \operatorname{Prox}_{\gamma f}(z^{k+1}) - \operatorname{Prox}_{\gamma f}(z^k) \to \mathbf{0}.$$

Since

$$z^{k+2} - z^{k+1} = \underbrace{z^{k+1} - z^k}_{\to d'} + x^{k+2} - x^{k+1} - \underbrace{(x^{k+3/2} - x^{k+1/2})}_{\to \mathbf{0}} \to d'$$

we also conclude that $x^{k+2} - x^{k+1} \to \mathbf{0}$. $\qquad\square$

Theorem 27. *If (P) has an improving direction, and (P) is feasible, then $x^{k+3/2} - x^{k+1/2} \to \gamma d$ and $x^{k+2} - x^{k+1} \to \gamma d$, where $-\gamma d = \gamma \Pi_{\overline{\operatorname{dom} f^* + \operatorname{dom} g^*}}(\mathbf{0})$ is the infimal displacement vector as given in Corollary 15.*

Proof. For simplicity, assume $\gamma = 1$. For $\gamma \ne 1$, we scale f and g to get the stated result.

Rewrite the DRS iteration as

$$\begin{aligned}
x^{k+1/2} &= \operatorname{Prox}_f(z^k) \\
\nu^{k+1/2} &= z^k - x^{k+1/2} = \operatorname{Prox}_{f^*}(z^k) \\
x^{k+1} &= \operatorname{Prox}_g(2x^{k+1/2} - z^k) \\
\nu^{k+1} &= 2x^{k+1/2} - z^k - x^{k+1} = \operatorname{Prox}_{g^*}(2x^{k+1/2} - z^k) \\
z^{k+1} &= z^k - (\nu^{k+1} + \nu^{k+1/2}).
\end{aligned}$$

By Theorem 4, we have

$$z^{k+1} - z^k = x^{k+1} - x^{k+1/2} \to d.$$

By the same reasoning as in Theorem 26, we can use the projection inequality and firm-nonexpansiveness to show that

$$\nu^{k+3/2} - \nu^{k+1/2} = \operatorname{Prox}_{f^*}(z^{k+1}) - \operatorname{Prox}_{f^*}(z^k) \to \mathbf{0}.$$

Since

$$z^{k+1} - z^k = \underbrace{\nu^{k+3/2} - \nu^{k+1/2}}_{\to \mathbf{0}} + x^{k+3/2} - x^{k+1/2} \to d,$$

we have $x^{k+3/2} - x^{k+1/2} \to d$.

Since

$$z^{k+2} - z^{k+1} = \underbrace{z^{k+1} - z^k}_{\to d} + x^{k+2} - x^{k+1} - \underbrace{(x^{k+3/2} - x^{k+1/2})}_{\to d} \to d$$

we also conclude that $x^{k+2} - x^{k+1} \to d$. $\qquad\square$

4 Application to DRS

In this section, we use the theory of Section 3 to analyze DRS under pathologies. We classify the primal-dual problem pair (P) and (D) into 7 cases and provide convergence analyses for the first 6 cases, which are the ones that assume strong duality. A rough summary is that DRS essentially works when strong duality holds.

4.1 Classification

The primal-dual problem pair, (P) and (D), falls under exactly one of the following 7 distinct cases.

Case (a) Total duality holds between (P) and (D), i.e., $d^\star = p^\star$ is finite, both (P) and (D) have solutions, and $d^\star = p^\star$. For example, the primal problem

$$\text{minimize} \quad x - \log x$$

and its dual problem

$$\text{maximize} \quad 1 + \log(y)$$
$$\text{subject to} \quad y = 1$$

both have solutions, and $d^\star = p^\star = 1$.

Case (b) $d^\star = p^\star$ is finite, (P) has a solution, but (D) has no solution. For example, the primal problem

$$\text{minimize} \quad \underbrace{\delta_{\{(x_1,x_2) \,|\, x_1^2 + x_2^2 \le 1\}}(x_1, x_2)}_{f(x)} + \underbrace{x_2 + \delta_{\{(x_1,x_2) \,|\, x_1 = 1\}}(x_1, x_2)}_{g(x)}$$

has a solution but its dual problem

$$\text{maximize} \quad -\sqrt{\nu_1^2 + \nu_2^2} + \nu_1 - \delta_{\{\nu_2 = 1\}}(-\nu_2)$$

does not. Nevertheless, $d^\star = p^\star = 0$.

Case (c) $d^\star = p^\star$ is finite, (P) is feasible, but (P) has no solution. To get such an example, swap the role of the primal and the dual in the example for case (b).

Case (d) $d^\star = p^\star = -\infty$, (P) is feasible, but there is no improving direction. This implies (D) is weakly infeasible. For example, the primal problem

$$\text{minimize} \quad \delta_{\{x \,|\, x \ge 1\}}(x) - \log x$$

has no solution and has optimal value $p^\star = -\infty$. Since the asymptotic derivative of the objective $x \to \infty$ is 0, the primal problem has no improving direction. The dual problem

$$\text{maximize} \quad y + 1 + \log(y)$$
$$\text{subject to} \quad y \le 0$$

is weakly infeasible.

Case (e) $d^\star = p^\star = -\infty$, (P) is feasible, and there is an improving direction. This implies (D) is strongly infeasible. For example, the primal problem

$$\text{minimize} \quad x + x$$

has an improving direction, namely $d = -1$, and the dual problem

$$\text{maximize} \quad \delta_{\{1\}}(x) + \delta_{\{1\}}(-x)$$

is strongly infeasible.

Case (f) $d^\star = p^\star = \infty$ and (P) is infeasible. For example, the problem

$$\text{minimize} \quad 1/\sqrt{-x} - \log(x)$$

is infeasible, and its dual

$$\begin{aligned} \text{maximize} \quad & (3/2^{2/3})y^{1/3} + 1 + \log(y) \\ \text{subject to} \quad & y \geq 0 \end{aligned}$$

has optimal value $d^\star = \infty$.

Case (g) $d^\star < p^\star$, i.e. strong duality fails. We do not consider this case.

4.2 Convergence results

We now provide convergence analyses for cases (a) through (f).

Theorem 28. *[34, 23] In case (a), $x^{k+1/2}, x^{k+1} \to x^\star$, where $x^\star$ is a solution of (P) and*

$$\lim_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star.$$

Theorem 29. *In case (b), $x^{k+1} - x^{k+1/2} \to \mathbf{0}$ and*

$$\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^\star, \qquad \liminf_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star.$$

Furthermore, if $x^{k+1/2} \to x^\star$ (or equivalently if $x^{k+1} \to x^\star$) then $x^\star$ is a solution.

Proof. This follows from Theorem 4, Corollary 13, Theorem 22, and Corollary 23. $\qquad\square$

Theorem 30. *In case (c), $x^{k+1} - x^{k+1/2} \to \mathbf{0}$,*

$$\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^\star, \qquad \liminf_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star,$$

and $(x^{k+1/2}, x^{k+1})$ do not converge.

Proof. This follows from Theorem 4, Corollary 13, Theorem 22, and the contrapositive of Corollary 23. $\quad\square$

Theorem 31. *In case (d), (D) is weakly infeasible, $x^{k+1} - x^{k+1/2} \to \mathbf{0}$,*

$$\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = -\infty, \qquad \liminf_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = -\infty,$$

and $(x^{k+1/2}, x^{k+1})$ do not converge.

Proof. This follows from Theorem 4, Theorem 11, Corollary 14, Theorem 22, and the contrapositive of Corollary 23. $\qquad\square$

Theorem 32. *In case (e), (D) is strongly infeasible, $x^{k+1} - x^{k+1/2} \to \gamma d$, where d is an improving direction,*

$$\lim_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = -\infty,$$

and $(x^{k+1/2}, x^{k+1})$ do not converge. Furthermore, $\operatorname{dist}(x^{k+1/2}, \operatorname{dom} g) \to 0$ and $\operatorname{dist}(x^{k+1}, \operatorname{dom} f) \to 0$.

Proof. All but the last assertions follows from Theorem 4, Theorem 11, Corollary 15, Corollary 24, and the contrapositive of Corollary 23.

By Theorem 27 $x^{k+1/2} - x^{k-1/2} \to \gamma d$ and by Theorem 4 and Corollary 15 $x^k - x^{k-1/2} \to \gamma d$. So $x^{k+1/2} - x^k \to \mathbf{0}$. Since $x^k \in \operatorname{dom} g$, we have

$$\operatorname{dist}(x^{k+1/2}, \operatorname{dom} g) \leq \operatorname{dist}(x^{k+1/2}, x^k) \to 0.$$

Since $x^{k+1/2} \in \operatorname{dom} f$, we have

$$\operatorname{dist}(x^k, \operatorname{dom} f) \leq \operatorname{dist}(x^k, x^{k+1/2}) \to 0.$$

$\square$

Theorem 33. *In case (f),* $\|x^{k+1} - x^{k+1/2}\| \to \operatorname{dist}(\operatorname{dom} f, \operatorname{dom} g)$.

Proof. This follows from Theorem 4 and Corollary 17 and 18. $\square$

4.3 Interpretation

We can view the DRS as an algorithm with two major goals: make the iterates feasible and optimal. With some caveats, DRS succeeds at both. As an auxiliary goal, we want the shadow iterates of DRS to converge to a solution if one exists. With some caveats, DRS succeeds at this as well.

Feasibility In cases (a), (b), (c), and (d) the iterates become approximately feasible in that $x^{k+1} - x^{k+1/2} \to \mathbf{0}$. In case (e) the iterates become approximately feasible in that $\operatorname{dist}(x^{k+1/2}, \operatorname{dom} g) \to 0$ and $\operatorname{dist}(x^{k+1}, \operatorname{dom} f) \to 0$. In case (f), feasibility is impossible, but DRS does its best to achieve feasibility.

Optimality In cases (a), (b), (c), (d), and (e), the function values on average converge to the solution. In other words, DRS finds the correct optimal value in these cases.

Shadow iterate convergence In case (a), the shadow iterates, the $x^{k+1/2}$ and x^{k+1} iterates, converge to a solution. In case (b), we do not know whether the shadow iterates converge to a solution. However, if they converge the limit is a solution. In cases (c), (d), and (e), the shadow iterates do not converge, which is good since there is no solution to converge to.

4.4 Notes

For in-depth studies problems of case (g), i.e., convex programs for which strong duality fails, see [50, 57, 58, 26, 42].

The papers [8, 12, 13] study DRS applied to infeasible convex feasibility problems and conditions that allow the shadow iterates to converge. In other words, these work study setups of case (f) for which the shadow iterates converge.

Finally, we clarify how some aspects of the results of Section 4.2 are, in subtle ways, not as strong as the standard convergence results for case (a).

In case (a) and (e), we have

$$\lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star. \tag{4}$$

For cases (b), (c), and (d), we are only able to show

$$\liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star$$

and

$$\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^\star. \tag{5}$$

The stronger result (4) says the function values are near optimal for large k. The weaker result (5) says the function values are *usually* near optimal for large k, since the average function value is near optimal for large k. Results of the form (5) are sometimes called *ergodic convergence*.

In case (a), the standard results state the shadow iterates converge to a single solution. In case (b), we are only able to show that the iterates achieve approximate optimality. When there are many solutions in case (b), we do not know whether or not the shadow iterates have multiple optimal limit points.

Often, $\operatorname{dom} f \cap \operatorname{dom} g$ is called the feasible set. In case (a), the standard results state $\operatorname{dist}(x^{k+1/2}, \operatorname{dom} f \cap \operatorname{dom} g) \to 0$ and $\operatorname{dist}(x^{k+1}, \operatorname{dom} f \cap \operatorname{dom} g) \to 0$, i.e., $x^{k+1/2}$ and x^{k+1} converge to the feasible set. In cases (b), (c), (d), and (e), we show $x^{k+1/2} - x^k \to \mathbf{0}$, $x^{k+1/2} \in \operatorname{dom} f$, and $x^{k+1} \in \operatorname{dom} g$. This, however, does not imply $x^{k+1/2}$ or x^{k+1} converge to the feasible set. To see why the implication fails, see the example of Section 5.2. So the iterates approach approximate feasibility, but we have not shown whether the iterates converge to the feasible set.

5   Counter examples

Throughout this paper, we analyzed and proved convergence without asking how fast the convergence is, and we assumed strong duality without asking what happens without it. In this section, we address these issues through counter examples.

The claims of this section are based largely on experimental observations rather than analytical results. For the sake of scientific reproducibility, we release the code used for the experiments.

5.1   Rate of convergence

When total duality holds between the primal-dual problem pair (P) and (D), i.e., when we have case (a), the convergence rate
$$\|z^{k+1} - z^k\|^2 \le o(1/k)$$
is guaranteed for DRS [27, 23].

Under pathologies, however, the rate can be much slower. Consider the problems

$$\text{minimize} \quad \underbrace{\delta_{\{(x_1, x_2) \,|\, x_1^q + x_2^q \le 1\}}(x_1, x_2)}_{f(x)} + \underbrace{x_2 + \delta_{\{(x_1, x_2) \,|\, (x_1 - 2)^2 + x_2^2 \le 1\}}(x_1, x_2)}_{g(x)},$$

where $q > 1$. These problems are in case (b), and has the unique solution $x^\star = (1, 0)$. We run DRS with $\gamma = 1$ and experimentally observe a rate of

$$\|z^{k+1} - z^k\|^2 \le \mathcal{O}(1/k^{4/(q+1)})$$

as shown in Figure 1. For $q > 3$, the rate is slower than $\mathcal{O}(1/k)$. To solve the projections onto the q-norm ball, we used CVX [31, 30].

Although the results are purely experimental, it does make the following point: There is no reason to expect the usual $o(1/k)$ rate under pathologies, and in some cases we do observe a slower rate.

5.2   When strong duality fails

In the analyses of DRS in Sections 4 we assumed strong duality. When strong duality fails, i.e., when $d^\star < p^\star$, we conjecture that DRS fails.

Conjecture. *When strong duality fails, DRS necessarily fails in that*

$$\liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star.$$

In other words, DRS finds the wrong objective value.

As discussed in Section 4.2, DRS tries to achieve feasibility and optimality. As discussed in Section 2.1, strong duality is well-posedness. Therefore, when the problem is ill-posed, we expect DRS to reduce the function value below $p^\star$ while achieving an infinitesimal infeasibility. We support the conjecture with examples.
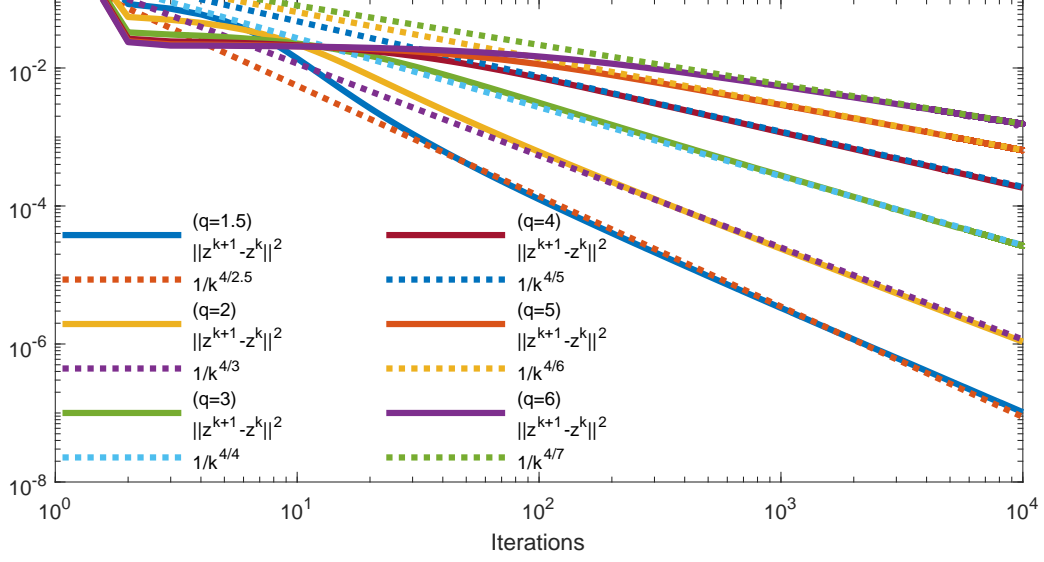
Figure 1: Plot of $\|z^{k+1} - z^k\|^2$ and $\mathcal{O}(1/k^{4/(q+1)})$.

Analytical counter example As our first example, taken from [35], consider the primal problem

$$\text{minimize} \quad \underbrace{\delta_{\{(x_1,x_2,x_3)\,|\,x_3 \geq (x_1^2+x_2^2)^{1/2}\}}(x)}_{f(x)} + \underbrace{x_1 + \delta_{\{(x_1,x_2,x_3)\,|\,x_2=x_3\}}(x)}_{g(x)}$$

which has the solution set $\{(0,t,t)\,|\,t \in \mathbb{R}\}$ and optimal value $p^\star = 0$. Its dual problem

$$\text{maximize} \quad -\delta_{\{(\nu_1,\nu_2,\nu_3)\,|\,-\nu_3 \geq (\nu_1^2+\nu_2^2)^{1/2}\}}(\nu) - \delta_{\{(\nu_1,\nu_2,\nu_3)\,|\,\nu_1=1,\,\nu_2=-\nu_3\}}(-\nu)$$

is infeasible. Given $z^0 = (z_1^0, z_2^0, 0)$, the DRS iterates have the form

$$z_1^{k+1} = \frac{1}{2}z_1^k - \gamma$$
$$z_2^{k+1} = \frac{1}{2}z_2^k + \frac{1}{2}\sqrt{(z_1^k)^2 + (z_2^k)^2}$$
$$z_3^{k+1} = 0.$$

With this, it is relatively straightforward to show $x^{k+1/2} - x^{k+1} \to \mathbf{0}$, $x_1^{k+1/2} \to -2\gamma$, $x_2^{k+1/2} \to \infty$, $x_3^{k+1/2} \to \infty$, and $f(x^{k+1/2}) + g(x^{k+1}) \to -2\gamma$. Also, $x^{k+1/2} \not\to \text{dom } f \cap \text{dom } g$ even though $x^{k+1/2} - x^{k+1} \to \mathbf{0}$.

Note that

$$d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star.$$

So this counter example proves, at least in some cases, that DRS solves neither the primal nor the dual problem in the absence of strong duality.

Experimental counter examples To further support the conjecture, we run DRS on several problems for which strong duality fails and report the experimental results.

The problem, taken from [14],

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \underbrace{\exp(-\sqrt{x_1 x_2})}_{f(x)} + \underbrace{\delta_{\{(x_1,x_2)\,|\,x_1=0\}}(x)}_{g(x)}$$

has optimal value $p^\star = 1$ but $d^\star = 0$. Experimentally, for all $\gamma > 0$ and choice of z^0 we observe $d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star$.

The problem, taken from [26],

$$\underset{X \in \mathbf{S}^3}{\text{minimize}} \quad \underbrace{\delta_{\mathbf{S}^3_+}(X)}_{f(X)} + \underbrace{X_{22} + \delta_{\{X \in \mathbf{S}^3 \mid X_{33}=0, X_{22}+2X_{13}=1\}}(X)}_{g(X)},$$

where $\mathbf{S}^3$ and $\mathbf{S}^3_+$ respectively denote the set of symmetric and positive semidefinite 3×3 matrices, has optimal value $p^\star = 1$ but $d^\star = 0$. Experimentally, we observe $d^\star = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1})$ for $\gamma \geq 0.5$, and $d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star$ for $0 < \gamma < 0.5$. This behaviour does not depend on z^0.

The problem, taken from [62],

$$\underset{X \in \mathbf{S}^3}{\text{minimize}} \quad \underbrace{\delta_{\mathbf{S}^3_+}(X)}_{f(X)} + \underbrace{2X_{12} + \delta_{\{X \in \mathbf{S}^3 \mid X_{22}=0, -2X_{12}+2X_{33}=2\}}(X)}_{g(X)}$$

has optimal value $p^\star = 0$ but $d^\star = -2$. Experimentally, we observe $d^\star = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1})$ for $\gamma \geq 1$, and $d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star$ for $0 < \gamma < 1$. This behaviour does not depend on z^0.

The problem, taken from [58],

$$\underset{X \in \mathbf{S}^5}{\text{minimize}} \quad \underbrace{\delta_{\mathbf{S}^5_+}(X)}_{f(X)} + \underbrace{X_{44} + X_{55} + \delta_{\{X \in \mathbf{S}^3 \mid X_{11}=0, X_{22}=1, X_{34}=1, 2X_{13}+2X_{45}+X_{55}=1\}}(X)}_{g(X)}$$

has optimal value $p^\star = (\sqrt{5}-1)/2$ but $d^\star = 0$. Experimentally, we observe $d^\star = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1})$ for $\gamma \geq 0.8$, and $d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star$ for $0 < \gamma < 0.8$. This behaviour does not depend on z^0.

The conjecture holds for all examples. Interestingly, for some examples, there is a threshold $\gamma_{\min}$ such that $d^\star < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star$ when $0 < \gamma < \gamma_{\min}$ and $d^\star = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1})$ when $\gamma_{\min} \leq \gamma$. Investigating this behavior is an interesting direction of future work.

6   Conclusion

In this paper, we analyzed DRS under pathologies. With some caveats, DRS works when strong duality holds. Furthermore, we conjectured that DRS necessarily fails when strong duality fails, and we provided empirical evidence supporting this conjecture.

As discussed in Section 4.4 and 5, there are some theoretical questions remaining. Strengthening the convergence results, analyzing the rate of convergence, and analyzing DRS in the absence of strong duality are all interesting future directions of research.

One conclusion of this paper is that we can solve pathological convex programs DRS, provided strong duality holds. However, should we? Studying how effective DRS is compared to other approaches that can solve pathological convex programs is another interesting future direction of research.

As mentioned in Section 1, ADMM is a closely related method, which behavior under pathologies also has not been studied extensively. Analyzing ADMM under pathologies is another direction worth pursuing.

References

[1] F. J. A. Artacho, J. M. Borwein, and M. K. Tam, *Douglas-Rachford feasibility methods for matrix completion problems*, The ANZIAM J., 55 (2014), pp. 299–326.

[2] F. J. A. Artacho, J. M. Borwein, and M. K. Tam, *Global behavior of the Douglas-Rachford method for a nonconvex feasibility problem*, J. of Global Optim., 65 (2016), pp. 309–327.

[3] J.-B. Baillon, R. E. Bruck, and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston J. Math., 4 (1978), pp. 1–9.

[4] G. Banjac, P. Goulart, B. Stellato, and S. Boyd, *Infeasibility detection in the alternating direction method of multipliers for convex optimization*, Optimization-online.org, (2017).

[5] H. H. Bauschke, R. I. Boţ, W. L. Hare, and W. M. Moursi, *Attouch-Théra duality revisited: Paramonotonicity and operator splitting*, J. Approx. Theory, 164 (2012), pp. 1065–1084.

[6] H. H. Bauschke, J. M. Borwein, and A. S. Lewis, *The method of cyclic projections for closed convex sets in Hilbert space*, Contemp. Math., 204 (1997), pp. 1–38.

[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer New York, 2nd ed., 2017.

[8] H. H. Bauschke, P. L. Combettes, and D. Luke, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*, J. Approx. Theory, 127 (2004), pp. 178–192.

[9] H. H. Bauschke, G. R. Douglas, and W. M. Moursi, *On a result of Pazy concerning the asymptotic behaviour of nonexpansive mappings*, J. Fixed Point Theory Appl., 18 (2016), pp. 297–307.

[10] H. H. Bauschke, W. L. Hare, and W. M. Moursi, *Generalized solutions for the sum of two maximally monotone operators*, SIAM J. Control Optim., 52 (2014), pp. 1034–1047.

[11] H. H. Bauschke, W. L. Hare, and W. M. Moursi, *On the range of the Douglas-Rachford operator*, Math. Oper. Res., 41 (2016), pp. 884–897.

[12] H. H. Bauschke and W. M. Moursi, *The Douglas-Rachford algorithm for two (not necessarily intersecting) affine subspaces*, SIAM J. Optim., 26 (2016), pp. 968–985.

[13] H. H. Bauschke and W. M. Moursi, *On the Douglas–Rachford algorithm*, Math. Program., 164 (2017), pp. 263–284.

[14] D. Bertsekas, *Convex Optimization Theory*, Athena Scientific, 2009.

[15] J. Borwein and H. Wolkowicz, *Regularizing the abstract convex program*, J. Math. Anal. Appl., 83 (1981), pp. 495–530.

[16] J. M. Borwein and M. K. Tam, *The cyclic Douglas-Rachford method for inconsistent feasibility problems*, J. Nonlinear Convex Anal., 16 (2015), pp. 537–584.

[17] J. M. Borwein and H. Wolkowicz, *Facial reduction for a cone-convex programming problem*, J. Austral. Math. Soc., 30 (1981), pp. 369–380.

[18] J. M. Borwein and H. Wolkowicz, *Characterizations of optimality without constraint qualification for the abstract convex program*, in Optimality and Stability in Mathematical Programming, M. Guignard, ed., Springer Berlin Heidelberg, Berlin, Heidelberg, 1982, pp. 77–100.

[19] Y.-L. Cheung, S. Schurr, and H. Wolkowicz, *Preprocessing and regularization for degenerate semidefinite programs*, in Computational and Analytical Mathematics, D. H. Bailey, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. D. Vanderwerff, and H. Wolkowicz, eds., Springer New York, New York, NY, 2013, pp. 251–303.

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

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

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

[23] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, in Splitting Methods in Communication, Imaging, Science, and Engineering, R. Glowinski, S. J. Osher, and W. Yin, eds., Springer International Publishing, 2016, pp. 115–163.

[24] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[25] D. Drusvyatskiy, G. Li, and H. Wolkowicz, *A note on alternating projections for ill-posed semidefinite feasibility problems*, Math. Program., 162 (2017), pp. 537–548.

[26] D. Drusvyatskiy and H. Wolkowicz, *The many faces of degeneracy in conic optimization*, Found. Trends Optim., 3 (2017), pp. 77–170.

[27] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., 1983.

[28] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[29] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problmes de Dirichlet non linéaires*, Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique, 9 (1975), pp. 41–76.

[30] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, eds., Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110. `http://stanford.edu/~boyd/graph_dcp.html`.

[31] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*. `http://cvxr.com/cvx`, Mar. 2014.

[32] R. B. Kellogg, *A nonlinear alternating direction method*, Math. Comput., 23 (1969), pp. 23–27.

[33] K. S. Kretschmer, *Programmes in paired spaces*, Canad. J. Math., 13 (1961), pp. 221–238.

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

[35] Y. Liu, E. K. Ryu, and W. Yin, *A new use of douglas-rachford splitting and admm for identifying infeasible, unbounded, and pathological conic programs*, arXiv preprint arXiv:1706.02374, (2017).

[36] B. F. Lourenço, M. Muramatsu, and T. Tsuchiya, *Solving SDP completely with an interior point oracle*, arXiv preprint arXiv:1507.08065, (2015).

[37] Z.-Q. Luo, J. F. Sturm, and S. Zhang, *Duality results for conic convex programming*, tech. report, Erasmus University Rotterdam, Econometric Institute, 1997.

[38] Z.-Q. Luo, J. F. Sturm, and S. Zhang, *Conic convex programming and self-dual embedding*, Optim. Methods Softw., 14 (2000), pp. 169–218.

[39] W. M. Moursi, *The forward-backward algorithm and the normal problem*, arXiv preprint arXiv:1608.02240, (2016).

[40] Y. Nesterov, M. J. Todd, and Y. Ye, *Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems*, Math. Program., 84 (1999), pp. 227–267.

[41] G. Pataki, *A simple derivation of a facial reduction algorithm and extended dual systems*, tech. report, Columbia University, 2000.

[42] G. Pataki, *Bad semidefinite programs: They all look the same*, SIAM J. Optim., 27 (2017), pp. 146–172.

[43] G. Pataki, Q. Tran-Dinh, et al., *Sieve-SDP: a simple facial reduction algorithm to preprocess semidefinite programs*, arXiv preprint arXiv:1710.08954, (2017).

[44] A. Pazy, *Asymptotic behavior of contractions in Hilbert space*, Israel J. Math., 9 (1971), pp. 235–240.

[45] D. W. Peaceman and H. H. Rachford, *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math., 3 (1955).

[46] J. Peña and J. Renegar, *Computing approximate solutions for convex conic systems of constraints*, Math. Program., 87 (2000), pp. 351–383.

[47] F. Permenter, H. A. Friberg, and E. D. Andersen, *Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach*, SIAM J. Optim., 27 (2017), pp. 1257–1282.

[48] F. Permenter and P. A. Parrilo, *Basis selection for SOS programs via facial reduction and polyhedral approximations*, in 53rd IEEE Conference on Decision and Control, CDC 2014, 2014, pp. 6615–6620.

[49] A. U. Raghunathan and S. D. Cairano, *Infeasibility detection in alternating direction method of multipliers for convex quadratic programs*, in 2014 IEEE 53rd Annual Conference on Decision and Control (CDC), 2014, pp. 5819–5824.

[50] M. V. Ramana, L. Tunçel, and H. Wolkowicz, *Strong duality for semidefinite programming*, SIAM J. Optim., 7 (1997), pp. 641–662.

[51] M. V. Ramana, L. Tunel, and H. Wolkowicz, *Strong duality for semidefinite programming*, SIAM J. Optim., 7 (1997), pp. 641–662.

[52] R. Rockafellar, *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics, 1974.

[53] R. T. Rockafellar, *Convex Analysis*, 1970.

[54] E. K. Ryu, *Cosmic divergence, weak cosmic convergence, and fixed points at infinity*, arXiv preprint arXiv:1709.02838, (2017).

[55] E. K. Ryu and S. Boyd, *Primer on monotone operator methods*, Appl. Comput. Math., 15 (2016), pp. 3–43.

[56] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, *OSQP: An operator splitting solver for quadratic programs*, arXiv preprint arXiv:1711.08013, (2017).

[57] P. Tseng, *Some convex programs without a duality gap*, Math. Program., 116 (2009), pp. 553–578.

[58] L. Tunçel and H. Wolkowicz, *Strong duality and minimal representations for cone optimization*, Comput. Optim. Appl., 53 (2012), pp. 619–648.

[59] H. Waki and M. Muramatsu, *A facial reduction algorithm for finding sparse sos representations*, Operations Research Letters, 38 (2010), pp. 361 – 365.

[60] H. Waki and M. Muramatsu, *Facial reduction algorithms for conic optimization problems*, J. Optim. Theory Appl., 158 (2013), pp. 188–215.

[61] M. Yamasaki, *Some generalizations of duality theorems in math. program. problems*, Math. J. Okayama Univ., 14 (1969), pp. 69–81.

[62] Y. Ye, *Linear conic programming*, Manuscript. Stanford University, Stanford, CA, (2004).