Douglas-Rachford Splitting and ADMM for Pathological Convex Optimization

Ernest K. Ryu
UCLA, eryu@math.ucla.edu
Yanli Liu
UCLA, yanli@math.ucla.edu
Wotao Yin
UCLA, wotaoyin@math.ucla.edu

Despite the vast literature on DRS and ADMM, there had 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. We then translate the pathological analyses for DRS to pathological analyses for ADMM.

Key words: Douglas-Rachford splitting, ADMM, Strong Duality, Pathological convex programs
MSC2000 subject classification: 90C46, 49N15, 90C25

1. Introduction

Douglas-Rachford splitting (DRS) and alternating directions method of multipliers (ADMM) are classical methods originally presented in [75, 38, 54, 52] and [48, 51], respectively. DRS and ADMM are closely related. Over the last decade, these methods have enjoyed a resurgence of popularity, as the demand to solve ever larger problems grew.

DRS and ADMM have 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 had been very little work analyzing DRS and ADMM under pathologies, despite the vast literature on these methods. There has been some recent exciting progress in this area, which we review in Section 1.2.

In this paper, we analyze the asymptotic behavior of DRS and ADMM 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 \( p^* = d^* \in [-\infty, \infty] \), 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. We then translate the pathological analyses for DRS to pathological analyses for ADMM.

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.

1.1. Summary of results, contribution, and organization

Sections 4 and 5 present what we consider the fruits of this work, the convergence analyses of DRS and ADMM under various pathologies. In fact, we suggest readers read Sections 4 and 5 before reading the theory of Section 3, as doing so will give a sense of direction.
We quickly illustrate, through examples, the kinds of results we show. Precise definitions and statements are presented later. We want DRS and ADMM to find a point that is approximately feasible and, when applicable, approximately optimal. For example, if the primal problem is weakly infeasible, we want the DRS iterates to satisfy

\[ x^{k+1} - x^{k+1/2} \to 0 \]

and we show this as Theorem 10. As another example, if the primal problem is feasible but has no solution and \( d^* = p^* > -\infty \), we want the DRS iterates to satisfy

\[ x^{k+1} - x^{k+1/2} \to 0, \quad f(x^{k+1/2}) + g(x^{k+1}) \to p^* \]

and we show related results as Theorems 7 and 8. We can say something for all the pathological cases, so long as \( d^* = p^* \).

Section 3 presents the main theoretical contribution of this work. To show that DRS and ADMM successfully achieve the 2 goals of approximate feasibility and approximate optimality, we need 2 separate major theoretical components.

Section 3.1 presents the first component, which analyzes the “fixed-point iteration” without a fixed point with tools from operator theory. With this machinery, we show things like \( x^{k+1} - x^k \to 0 \) or \( x^{k+1} - x^k \to v \), where \( v \) is a certificate of (primal or dual) infeasibility. Our contribution is defining the notion of improving directions via recession functions and fully characterizing the infimal displacement vector with this notion.

Section 3.2 presents the second component, the function-value analysis, which is based on ideas from convex optimization and subgradient inequalities. With these techniques, we show things like \( f(x^{k+1/2}) + g(x^{k+1}) \to p^* \). This part requires the \( d^* = p^* \) assumption. Our function-value analysis uses, but does not immediately follow from, the results of Section 3.1. To the best of our knowledge, analyzing the convergence of objective values for DRS or ADMM applied to pathological problems has not been done before.

Section 3.3 presents a third, relatively minor theoretical component, which we use later in Section 5 to translate analyses for DRS to analyses for ADMM.

As the goal of this work is to prove several theorems, one each for the many pathological cases, we build up our theory in a series of lemmas and corollaries. Some of these lemmas are rather simple extensions of known results while some are novel. All results of Section 3 are eventually used in proving the 5 theorems of Section 4 and the 3 theorems of Section 5.

The paper is organized 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 analyzes ADMM under pathologies with the theory of Sections 4 and 3. Section 6 presents counterexamples to make additional observations. Section 7 concludes the paper.

1.2. Prior work
Duality for linear programming (LP) is classic subject with very thorough and complete theory [32, 31]. For example, the classic theory states that the primal problem has an “improving direction” if and only if the dual problem is strongly infeasible, and therefore the improving direction certifies dual infeasibility [43]. A consequence of our theory is that DRS and ADMM applied to pathological LPs produce a certificate of infeasibility as an improving direction.

Duality for general conic programs, which is more complicated, has also been thoroughly studied [61, 81, 67, 63, 91, 26, 69, 59, 60, 71, 77, 55], mostly under the explicit or implicit context of second order methods. Some of the definitions and ideas of our work were directly influenced by this line of work. We believe further applying ideas of this classical literature to analyze pathological DRS and ADMM is a direction with much potential, especially in understanding the conjecture of Section 6.
Although negative results are often not published, there has been some reporting of pathologies happening in practice \cite{58, 36, 39, 92, 95}. Therefore, there is practical value in studying how well-behaved and robust an algorithm is when faced with pathological problems.

Despite their huge popularity, there had been surprisingly little work investigating the behavior of DRS and ADMM under pathologies. The understanding is still incomplete, but there has been some recent progress: \cite{10, 15, 16, 56} analyze DRS under specific pathological setups, \cite{12, 13, 16} analyze DRS under general setups, and \cite{80, 89, 6} analyze ADMM under specific pathological setups for conic programs. These studies, however, are limited to more specific setups and pathologies where an improving direction exists or the primal problem is strongly feasible.

The convex feasibility problem of finding an \(x \in A \cap B\), where \(A\) and \(B\) are nonempty closed convex sets, is a subclass of problems with practical importance. This problem can be recast into an equivalent optimization problem. Under our definitions, the convex feasibility problem is pathological when \(A \cap B = \emptyset\), and strong duality fails when \(A \cap B = \emptyset\) and the sets are tangent. Our analysis of Theorem 10 can be extended in a straightforward manner to handle convex feasibility problems without strong duality. There has been some past work studying DRS and conditions that allow the shadow iterates to converge in this setup \cite{10, 15, 16}. The results specific to convex feasibility problems presented in these papers are stronger than what Theorem 10 or its immediate extension can show.

DRS has strong primal-dual symmetry, in the sense of Fenchel duality for convex optimization \cite{45, 86} and, more generally, Attouch-Théra duality for monotone operators \cite[p. 40]{65} and \cite{4}. See \cite[Lemma 3.6 p. 133]{40} or \cite[7, 14, 16]{4} for in-depth studies on this subject. Naturally, our results also exhibit a degree of primal-dual symmetry, although we do not explicitly address it 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.

In operator theory, and especially in infinite dimensional problems arising from physics and PDEs, the sum of two maximal monotone operators may not maximal, and one can consider this a pathology. One remedy to such pathologies is to generalize the notion of the sum of two operators by regularizing the operators and then considering the limit as the regularization is reduced to zero \cite{3, 83, 84}. This notion of pathology and the remedy is quite different from what we consider. For some of the pathologies we consider, \(\partial f + \partial g\) is a perfectly well-defined maximal monotone operator. Moreover, we do not remedy the pathology but rather simply analyze how DRS and ADMM behave under the pathology. Finally, we work in finite dimensions and thereby avoid certain complications exclusive to infinite dimensional problems.

When a problem is known to be pathological a priori, one can first modify or regularize the problem and then solve the non-pathological problem. One such approach is facial reduction, a pre-processing step that rids a pathological conic program of difficult pathologies \cite{19, 21, 22, 82, 68, 27, 93, 94, 79, 59, 77, 78, 71}. In contrast, the goal of this work is to analyze DRS and ADMM when directly applied to pathological convex programs. To put in differently, we do not assume users of DRS or ADMM have a priori knowledge of whether the problem is pathological or not.

The standard analysis for DRS proves the iterates converge using ideas from operator theory and fixed point iterations \cite{54, 40, 41, 28, 30, 42, 29}. The standard analyses of ADMM prove the iterates converge by reducing ADMM to DRS \cite{47, 41, 42} or with a direct analysis via a Lyapunov function \cite{46, 50, 18, 37, 25}. These analyses rely on the existence of a primal-dual saddle point, which only exists under the non-pathological case, and therefore do not immediately generalize to pathological setups.

The first part of our analysis relies on a classical result by Pazy \cite{74} and Baillon et al. \cite{5} from the 1970s, which characterize the asymptotic behavior of fixed-point iterations without fixed-points. There has been some recent work that analyze algorithms that can be interpreted as fixed-point iterations without fixed-points \cite{8, 10, 20, 1, 12, 11, 15, 2, 66, 16, 87, 56}. The analysis of Section 3.1 was inspired by these works.
Another recent line of analysis for DRS and ADMM is function-value analysis, which establishes the objective values, rather than the iterates, converge \([33, 34, 35]\). These analyses, however, also rely on the existence of a primal-dual saddle point and do not immediately generalize to the pathological setups. The function-value analysis of Section 3.2 was inspired by these works.

Yet another recent line of work analyzes DRS with the so-called Douglas-Rachford envelope \([72, 73, 88, 49, 57]\). These analyses reinterpret DRS as a (variable-metric) gradient method on a certain non-convex envelope function, which minimizers coincide with the original problem under additional assumptions. This line of reasoning could be useful in the analysis of DRS applied to pathological convex programs.

There has been a rich line of work analyzing convex programs for which strong duality fails \([85, 82, 64, 76, 90, 91, 39, 70]\). In this work, we do not analyze this case, aside from the discussion and conjecture of Section 6.

### 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 \([86, 85, 9]\).

Throughout this section, \(f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}\) is a function and \(A, B, C \subseteq \mathbb{R}^n\) are nonempty sets.

A set \(C\) is convex, if \(x, y \in C\) and \(\theta \in [0, 1]\) implies \(\theta x + (1 - \theta)y \in C\). Write \(\overline{C}\) for the closure of \(C\). If \(C\) is convex \(\overline{C}\) is convex. The Minkowski sum and differences of \(A\) and \(B\) are

\[
A + B = \{a + b | a \in A, b \in B\}, \quad A - B = \{a - b | a \in A, b \in B\},
\]

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.

For the distance between \(x \in \mathbb{R}^n\) and the set \(A\), write

\[
\text{dist}(x, A) = \inf \{\|x - a\| | a \in A\}.
\]

For the distance between \(A\) and \(B\), write

\[
\text{dist}(A, B) = \inf \{\|a - b\| | a \in A, b \in B\}.
\]

Note that \(\text{dist}(A, B) = 0\) if and only if \(0 \in \overline{A - B}\).

A function \(f\) 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 \(\theta \in [0, 1]\). A function \(f\) is closed if its epigraph \(\{(x, \alpha) \in \mathbb{R}^{n+1} | f(x) \leq \alpha\}\) 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, proper, and convex (CPC) functions. If \(f\) and \(g\) are CPC functions, then \(f + g\) is CPC or \(f + g = \infty\) everywhere. If \(\gamma > 0\), then \(\gamma f\) is CPC.

Define the (effective) domain of \(f\) as \(\text{dom} f = \{x \in \mathbb{R}^n | f(x) < \infty\}\). If \(f\) is a convex function, then \(\text{dom} f\) is a convex set. However, \(\text{dom} f\) may not be closed even when \(f\) is CPC. For any \(\gamma > 0\), we have \(\text{dom} \gamma f = \text{dom} f\).

Define the conjugate of \(f\) as \(f^* (y) = \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\}\). If \(f\) is convex and proper, then \(f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}\) is CPC. If \(f\) is CPC, then \((f^*)^* = f\). For any \(\gamma > 0\), we have \((\gamma f)^* (x) = f^* (x/\gamma)\) and \(\text{dom}(\gamma f)^* = \gamma \text{dom} f^*\). If \(h(x) = g(-x)\), then \(h^*(y) = g^*(-y)\).

Define the projection onto \(C\) as \(\Pi_C (x_0) = \arg \min_{x \in C} \{\|x - x_0\|\}\). When \(C\) is closed and convex, \(\Pi_C : \mathbb{R}^n \to \mathbb{R}^n\) is well-defined, i.e., the minimizer uniquely exists.

Define the indicator function with respect to \(C\) as

\[
\delta_C (x) = \begin{cases} 
0 & \text{if } x \in C \\
\infty & \text{otherwise}
\end{cases}
\]
When $C$ is closed convex, $\delta_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is CPC.

Define the support function of $C$ as

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

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

Define the recession function of $f$ as

$$\text{rec}(f)(d) = \lim_{\alpha \to \infty} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

for any $x \in \text{dom}\ f$. Loosely speaking, the recession function characterizes the asymptotic change of $f$ as we go in direction $d$. In fact,

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

as $\alpha \to \infty$ for any $x \in \text{dom}\ f$. The recession function $\text{rec} f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a positively homogeneous CPC function. If $h(x) = g(-x)$, then $\text{rec}(h^*)(d) = \text{rec}(g^*)(-d)$. When $f$ and $g$ are CPC, either $f(x) + g(x) = \infty$ for all $x \in \mathbb{R}^n$ or $\text{rec}(f + g) = \text{rec} f + \text{rec} g$. If $f$ is CPC, then $\sigma_{\text{dom}\ f^*} = \text{rec} f$.

Define the proximal operator $\text{Prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\text{Prox}_f(z) = \arg\min_{x \in \mathbb{R}^n} \{f(x) + (1/2)\|x - z\|^2\}.$$ 

When $f$ is CPC, the arg min uniquely exists, and therefore $\text{Prox}_f$ is well-defined. When $C$ is closed and convex, $\text{Prox}_{\delta_C} = \Pi_C$. When $f$ is CPC, $\text{Prox}_f + \text{Prox}_{f^*} = I$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity operator.

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. $T : \mathbb{R}^n \to \mathbb{R}^n$ is firmly-nonexpansive if

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

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

$$\min_{x \in \mathbb{R}^n} f(x) + g(x),$$

the primal problem. We call the optimization problem

$$\max_{\nu \in \mathbb{R}^n} -f^*(\nu) - g^*(-\nu),$$

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

$(P)$ is feasible if $0 \in \text{dom} f - \text{dom} g$, strongly infeasible if $0 \notin \Pi_{\text{dom} f - \text{dom} g}(0)$, and weakly infeasible otherwise. $(P)$ falls under exactly one of the three cases. $(P)$ is infeasible if it is not feasible. $(D)$ is feasible if $0 \in \text{dom}(f^*) + \text{dom}(g^*)$, strongly infeasible if $0 \notin \Pi_{\text{dom}(f^*) + \text{dom}(g^*)}(0)$, and weakly infeasible otherwise.

We call $p^* = \inf \{f(x) + g(x) | x \in \mathbb{R}^n\}$ the primal optimal value and $d^* = \sup\{-f^*(\nu) - g^*(-\nu) | \nu \in \mathbb{R}^n\}$ the dual optimal value. We let $p^* = \infty$ if $(P)$ is infeasible and $d^* = -\infty$ if $(D)$ is infeasible. Weak duality, which always holds, states $d^* \leq p^*$. We say strong duality holds between
(P) and (D), if \( d^* = p^* \in [-\infty, \infty] \). We say total duality holds between (P) and (D), if (P) has a solution, (D) has a solution, and strong duality holds.

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 [53, 96, 62, 64]. Here, we generalize it to general convex programs. The following theorem is well known [85], although we have not seen it stated exactly in this form.

**Theorem 1.** If \( f \) and \( g \) are CPC, then \( d^* = p^- \leq p^* \).

With Theorem 1, 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-infinitesimal improvement to the function value, we can consider (P) ill-posed.

### 2.2. Douglas-Rachford operator

Douglas-Rachford splitting (DRS) applied to (P) is
\[
x^{k+1/2} = \text{Prox}_{\gamma f}(z^k) \\
x^{k+1} = \text{Prox}_{\gamma g}(2x^{k+1/2} - z^k) \\
z^{k+1} = z^k + x^{k+1} - x^{k+1/2}
\]

with a starting point \( z^0 \in \mathbb{R}^n \) and a parameter \( \gamma > 0 \). We also express this iteration 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. Write \( T_1 \) for \( T_\gamma \) with \( \gamma = 1 \).

The standard analysis of DRS assumes total duality, which, again, means (P) has a solution, (D) has a solution, and \( d^* = p^* \).

**Theorem 2 (Theorem 7.1 and 8.1 of [7] and Proposition 4.8 of [40]).** Total duality holds between (P) and (D) if and only if \( T_\gamma \) has a fixed point for some \( \gamma > 0 \). If total duality holds between (P) and (D), then DRS converges in that \( z^k \to z^* \), where \( x^* = \text{Prox}_{\gamma f}(z^*) \) 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 is 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.

### 2.3. Fixed-point iterations without fixed points

Theorem 2 states the DRS iteration has no fixed points under pathologies. Analyzing fixed-point iterations without fixed points is the first part of our pathological analysis.

Let \( T : \mathbb{R}^n \to \mathbb{R}^n \) be a firmly-nonexpansive operator. Write
\[
\text{range}(I - T) = \{ z - T(z) \mid z \in \mathbb{R}^n \}.
\]

Note that \( T \) has a fixed point if and only if \( 0 \in \text{range}(I - T) \). The closure of this set, \( \overline{\text{range}(I - T)} \), is closed and convex [74]. We call
\[
v = \Pi_{\overline{\text{range}(I - T)}}(0)
\]
the infimal displacement vector of \( T \). (The term was coined in [12].) If \( T \) has a fixed point, then \( v = 0 \), but \( v = 0 \) is possible even when \( T \) has no fixed point.

The following classical result by Pazy and Baillon et al. elegantly characterizes the asymptotic behavior of fixed-point iterations with respect to \( T \).
Theorem 3 (Theorem 2 of [74] and Corollary 2.3 of [5]). If $T$ is firmly-nonexpansive and $v$ is its infimal displacement vector, the iteration $z^{k+1} = T(z^k)$ satisfies

$$z^k = -kv + o(k), \quad z^{k+1} - z^k \to -v.$$  

Theorem 3 is especially powerful when we can concretely characterize $v$. Recently, Bauschke, Hare, and Moursi published the following elegant formula.

Theorem 4 ([13]). The infimal displacement vector $v$ of $T_1$, the DRS operator, satisfies

$$v = \min \{ \|z\| : z \in \dom f - \dom g \cap \dom f^* + \dom g^* \}.$$  

The original result in [13] is more general as it applies to the DRS operator of monotone operators. In Section 3.1, we use Theorem 4 and the notion of improving directions provide a further concrete characterization of $v$.

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 Sections 4 and 5 we combine these results to analyze the asymptotic behavior of DRS and ADMM applied to pathological convex programs.

While the formula of Theorem 4 is known, the use of improving directions to concretely characterize the infimal displacement vector $v$ is new. An improving direction may or may not exist, and we analyze both cases. Our analysis shows that existence of an improving direction is a key deciding factor in how DRS behaves.

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

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

Note $\text{rec}(f)(d) + \text{rec}(g)(d) = \text{rec}(f + g)(d)$ when $(P)$ is feasible. For simplicity, we only consider primal improving directions when $(P)$ is feasible. The notion of (primal) improving direction is standard in conic programming [62, 67, 64]. Here, we extend it to general convex programs of the form $(P)$.

If $(P)$ is feasible and there is a primal improving direction, then $p^* = -\infty$. To see why, let $d$ be a primal improving direction. Then

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

for any $x \in \dom f \cap \dom g$ as $\alpha \to \infty$, and therefore $p^* = -\infty$. However, $p^* = -\infty$ 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

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

If $(D)$ is feasible and there is a dual improving direction, then $d^* = \infty$.
3.1. Infimal displacement vector of the DRS operator

In this section, we provide a further concrete characterization of the infimal displacement vector $v$. When (P) or (D) is strongly infeasible, Theorem 4 states $v \neq 0$. Our contribution is to show $v$ is an improving direction in this case. For the sake of simplicity, we first analyze $T_1$ and then translate the results to $T_\gamma$ for $\gamma > 0$.

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

**Lemma 1.** (P) has an improving direction if and only if (D) is strongly infeasible. Write

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

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

**Proof.** We first show

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

Let $A$ and $B$ be nonempty convex sets. The identities of Section 2 tell us

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

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

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

Based on the identities of Section 2, we have

$$-\Pi_{\text{dom } f^* + \text{dom } g^*}(0) = \text{Prox}_{\text{dom } f^* + \text{dom } g^*}(0)
= (I - \text{Prox}_{\text{dom } f^* + \text{dom } g^*}) (0)
= - \text{Prox}_{\sigma_{\text{dom } f^* + \text{dom } g^*}}(0)
= - \text{Prox}_{\sigma_{\text{dom } f^*}}(0)
= - \text{Prox}_{\text{rec } f + \text{rec } g}(0).$$

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

$$0 = \arg \min \{ \text{rec } f(x) + \text{rec } g(x) + (1/2)||x||^2 \} = \text{Prox}_{\text{rec } f + \text{rec } g}(0)$$

if and only if $\text{rec } f(x) + \text{rec } g(x) \geq 0$ for all $x \in \mathbb{R}^n$. By our definition of an improving direction, $\text{rec } f(x) + \text{rec } g(x) \geq 0$ for all $x \in \mathbb{R}^n$ if and only if there is no improving direction. By definition, $0 = -\Pi_{\text{dom } f^* + \text{dom } g^*}(0)$ if and only if (D) is not strongly infeasible. So with (3), we conclude (P) has an improving direction if and only if (D) is strongly infeasible.

It remains to show that

$$d = \arg \min \{ \text{rec } f(x) + \text{rec } g(x) + (1/2)||x||^2 \}$$

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

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

This implies $\text{rec } f(d) + \text{rec } g(d) \leq -(1/2)||d||^2 < 0$, i.e., $d$ is an improving direction.
Lemma 2. Assume (P) is feasible. Then

\[ v = -d = \Pi_{\text{dom } f^* + \text{dom } g^*}(0) \]

is the infimal displacement vector of \( T_1 \).

Proof. Let \( x_0 \) be a feasible point of (P). Since \( \text{rec } f(d) + \text{rec } g(d) \leq 0 < \infty \) by Lemma 1 and the definition of an improving direction, we have \( x_0 \in \text{dom } f \), \( x_0 + d \in \text{dom } g \), and thus \( -d \in \text{dom } f - \text{dom } g \subseteq \text{dom } f^* - \text{dom } g^* \). Since \( -d \) is the minimum-norm element of \( \text{dom } f^* - \text{dom } g^* \), Theorem 4 tells us that \( -d \) is the infimal displacement vector of \( T_1 \). \( \square \)

Corollary 1. Assume (P) is feasible, and (D) is feasible. Then \( v = 0 \) is the infimal displacement vector of \( T_\gamma \) for any \( \gamma > 0 \).

Corollary 2. Assume (P) is feasible, and (D) is weakly infeasible. Then \( v = 0 \) is the infimal displacement vector of \( T_\gamma \) for any \( \gamma > 0 \).

Corollary 3. Assume (P) is feasible, and (D) is strongly infeasible. Then

\[ v = -\gamma d = \gamma \Pi_{\text{dom } f^* + \text{dom } g^*}(0) \neq 0 \]

is the infimal displacement vector of \( T_\gamma \) 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.

Lemma 3. Assume (D) is feasible. Then

\[ v = -d' = \Pi_{\text{dom } f^* - \text{dom } g}(0) \]

is the infimal displacement vector of \( T_1 \).

Proof. Following the same logic as in the proof of Lemma 1, we have

\[ \Pi_{\text{dom } f^* - \text{dom } g}(0) = - \arg \min \nu \{ \text{rec } (f^*(\nu)) + \text{rec } (g^*(-\nu)) + (1/2)\|\nu\|^2 \}, \]

and

\[ d' = -\Pi_{\text{dom } f^* - \text{dom } g}(0) \]

is a dual improving direction, if \( d' \neq 0 \).

Let \( \nu_0 \) be any feasible point of (D). Then \( \nu_0 \in \text{dom } f^* \) and \( -\nu_0 - d' \in \text{dom } g^* \). Therefore, \( -d' \in \text{dom } f^* + \text{dom } g^* \subseteq \text{dom } f^* - \text{dom } g^* \). Since \( -d' \) is defined to be the minimum-norm element of \( \text{dom } f^* - \text{dom } g^* \), we conclude the statement with Theorem 4. \( \square \)

Corollary 4. Assume (D) is feasible, and (P) is weakly infeasible. Then \( v = 0 \) is the infimal displacement vector of \( T_\gamma \) for any \( \gamma > 0 \).

Corollary 5. Assume (D) is feasible, and (P) is strongly infeasible. Then

\[ v = -d' = \Pi_{\text{dom } f^* - \text{dom } g}(0) \neq 0, \]

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

Note that for Corollary 5, the infimal displacement vector is independent of the value of \( \gamma \).
3.2. Function-value analysis

In this section, we present the second major theoretical component to our analysis. Section 3.1 analyzed the infimal displacement vector of $T_n$. This, however, is not sufficient for characterizing the asymptotic behavior of DRS in relation to the original optimization problem (P).

Let us briefly discuss why function-value analysis is necessary. Consider the convex function $h(x, y) = x^2/y$ defined for $y > 0$. Note that $h$ has minimizers, $(0, y)$ for any $y > 0$, and the operator $I - \nabla h$ has fixed points. However, even though, $h(\sqrt{y}, y) = 1$ we have $\nabla h(\sqrt{y}, y) = (2/\sqrt{y}, -1/y) \to 0$ as $y \to \infty$. This means the point $(\sqrt{y}, y)$ for large $y$ is not an approximate minimizer for $h$ but does approximate satisfy the fixed point condition for $I - \nabla h$. It is possible to construct a similar example with the DRS operator. If we let $f = h$ and $g = 0$, then DRS reduces to the proximal point method on $h$. This operator exhibits the same exact issue.

This means **approximate fixed points do not always correspond to approximate solutions of the original problem**. This is why we need a separate and distinct function-value analysis to accompany the fixed-point theory.

We now present function-value analysis. Throughout this section, write $x^{k+1/2}$ and $x^{k+1}$ to denote the DRS iterates of (2).

**Lemma 4.** For all $k = 0, 1, \ldots$ and any $x \in \mathbb{R}^n$

\[
\begin{align*}
    f(x^{k+1/2}) + g(x^{k+1}) - f(x) - g(x) \\
    \leq (1/\gamma)(x^{k+1} - x^{k+1/2})^T(x - z^{k+1}).
\end{align*}
\]

An inequality similar to that of Lemma 4 has been presented as Proposition 2 of [34]. We nevertheless quickly show a direct proof.

**Proof.** Write

\[
\begin{align*}
    \hat{\nabla} f(x^{k+1/2}) &= (1/\gamma)(z^k - x^{k+1/2}) \\
    \hat{\nabla} g(x^{k+1}) &= (1/\gamma)(2x^{k+1/2} - z^k - x^{k+1}).
\end{align*}
\]

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

\[
\hat{\nabla} f(x^{k+1/2}) \in \partial f(x^{k+1/2}), \quad \hat{\nabla} g(x^{k+1}) \in \partial g(x^{k+1})
\]

and that

\[
\hat{\nabla} f(x^{k+1/2}) + \hat{\nabla} g(x^{k+1}) = (1/\gamma)(x^{k+1/2} - x^{k+1}).
\]

We also have

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

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,

\[
\begin{align*}
    f(x^{k+1/2}) - f(x) + g(x^{k+1}) - g(x) \\
    \leq \hat{\nabla} f(x^{k+1/2})^T(x^{k+1/2} - x) + \hat{\nabla} g(x^{k+1})^T(x^{k+1} - x) \\
    = (\hat{\nabla} f(x^{k+1/2}) + \hat{\nabla} g(x^{k+1}))^T(x^{k+1/2} - x) + \hat{\nabla} g(x^{k+1})^T(x^{k+1} - x^{k+1/2}) \\
    = (x^{k+1} - x^{k+1/2})^T(\hat{\nabla} g(x^{k+1}) - (1/\gamma)(x^{k+1/2} - x)) \\
    = (1/\gamma)(x^{k+1} - x^{k+1/2})^T(x - z^{k+1}).
\end{align*}
\]

\[
\square
\]

The following result, which is well known for non-pathological setups, also holds under pathologies, so long as $d^* = p^*$. 
LEMMA 5. Assume \( p^* = d^* \in [-\infty, \infty] \). Assume the infimal displacement vector \( v \) of \( T_\gamma \) satisfies \( v = 0 \). Then
\[
\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^*
\]
and
\[
\liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^*.
\]
Proof. If \( \Delta^0, \Delta^1, \ldots \) is any sequence in \( \mathbb{R}^n \), then
\[
\sum_{j=0}^{k} (\Delta^j)^T \sum_{i=0}^{j} \Delta^i = \sum_{j=0}^{k} \sum_{i=0}^{j} 1 \{i \leq j\} (\Delta^j)^T (\Delta^i)
\]
\[
= \frac{1}{2} \left\| \sum_{i=0}^{k} \Delta^i \right\|^2 + \frac{1}{2} \sum_{i=0}^{k} \| \Delta^i \|^2.
\]
Let \( \Delta^k = z^{k+1} - z^k = x^{k+1} - x^{k+1/2} \) and sum the inequality of Lemma 4 to get
\[
\gamma \sum_{i=0}^{k} f(x^{i+1/2}) - f(x) + g(x^{i+1}) - g(x) \leq \sum_{j=0}^{k} (\Delta^j)^T (x - z^0) - \sum_{j=0}^{k} (\Delta^j)^T \sum_{i=0}^{j} \Delta^i
\]
\[
= (z^{k+1} - z^0)^T (x - z^0) - \frac{1}{2} \| z^{k+1} - z^0 \|^2 - \frac{1}{2} \sum_{i=0}^{k} \| z^{i+1} - z^i \|^2
\]
\[
= - \frac{1}{2} \| z^{k+1} \|^2 + \frac{1}{2} \| z^0 \|^2 + (z^{k+1} - z^0)^T x - \frac{1}{2} \sum_{i=0}^{k} \| z^{i+1} - z^i \|^2.
\]
Dividing both sides by \( (k+1)/2 \) to get
\[
\frac{2\gamma}{k+1} \sum_{i=0}^{k} \left( f(x^{i+1/2}) - f(x) + g(x^{i+1}) - g(x) \right) \leq - \frac{1}{k+1} \| z^{k+1} \|^2 + \frac{1}{k+1} \| z^0 \|^2 - \frac{1}{k+1} \sum_{i=0}^{k} \| z^{i+1} - z^i \|^2 + \frac{2}{k+1} (z^{k+1} - z^0)^T x.
\]
for all \( k = 0, 1, \ldots \) and any \( x \in \mathbb{R}^n \).

We now show
\[
\limsup_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} (f(x^{i+1/2}) + g(x^{i+1})) \leq p^*.
\]
Assume \( p^* < \infty \), as otherwise there is nothing to prove. Let \( x \) be any \( x \in \text{dom} f \cap \text{dom} g \). By Theorem 3, \( z^k = -kv + o(k) \). If \( v \neq 0 \), then the first (negative) term on the right-hand side of (4) dominates the positive terms. If \( v = 0 \), then both nonnegative terms on the right-hand side of (4) converge to 0. In both cases, we have
\[
\limsup_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) \leq f(x) + g(x)
\]
for all \( x \in \text{dom} f \cap \text{dom} g \). We minimize the right-hand side to obtain \( p^* \).

By Theorem 3, \( v = 0 \) implies \( x^{k+1/2} - x^{k+1} \to 0 \). In turn, by Theorem 1, we have
\[
\liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \geq p^*.
\]
Combining this with (5) gives us the first stated result.

It is straightforward to verify that if a real-valued sequence \( a^k \) satisfies

\[
\liminf_{k \to \infty} a^k \geq a, \quad \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} a^i = a,
\]

then

\[
\liminf_{k \to \infty} a^k = a.
\]

The second stated result follows from this argument. \( \square \)

Lemma 5 provides the function-value analysis when \( v = 0 \), and the first part of Lemma 6 provides the analysis when \( v \neq 0 \). The later parts of Lemma 6 is used in translating the analyses for DRS to analyses for ADMM in Section 5.

**Lemma 6.** Assume \( (P) \) is feasible and \( v \neq 0 \), i.e., \( (P) \) has an improving direction. Then

\[
f(x^{k+1}/2) + g(x^{k+1}) \to p^* = -\infty.
\]

Moreover, \( |f(x^{k+1}/2)| \leq O(k) \) and \( |g(x^{k+1})| \leq O(k) \) as \( k \to \infty \). Assume \( (P) \) is feasible and \( v = 0 \). Then \( |f(x^{k+1}/2)| \leq o(k) \) and \( |g(x^{k+1})| \leq o(k) \) as \( k \to \infty \).

**Proof.** When \( (P) \) has an improving direction, Corollary 3 and Theorem 3 tells us

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

Then Lemma 4 tells us that

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

which tells us

\[
\limsup_{k \to \infty}(1/k)(f(x^{k+1}/2) + g(x^{k+1})) \leq -\gamma \|d\|^2.
\] (6)

This proves the first statement.

Assume \( v = 0 \). With the same reasoning as for (6) we get

\[
\limsup_{k \to \infty}(1/k)(f(x^{k+1}/2) + g(x^{k+1})) \leq 0.
\]

Assume \( (P) \) feasible, without making any assumptions on \( v \). Write \( \tilde{\nabla}f(x^{1/2}) \) for any subgradient of \( f \) at \( x^{1/2} \). Then

\[
f(x^{k+1/2}) \geq f(x^{1/2}) + (\tilde{\nabla}f(x^{1/2}))^T(x^{k+1/2} - x^{1/2})
\]

\[
\geq f(x^{1/2}) - \|\tilde{\nabla}f(x^{1/2})\|\|x^{k+1/2} - x^{1/2}\| = k\gamma\|d\|\|\tilde{\nabla}f(x^{1/2})\| + o(k),
\]

and we conclude

\[
\liminf_{k \to \infty}(1/k)f(x^{k+1/2}) \geq -\gamma \|d\|\|\tilde{\nabla}f(x^{1/2})\|.
\]

With a similar argument, we get

\[
\liminf_{k \to \infty}(1/k)g(x^{k+1}) \geq -\gamma \|d\|\|\tilde{\nabla}g(x^{1})\|
\]

where \( \tilde{\nabla}g(x^1) \) is any subgradient of \( g \) at \( x^1 \). Combining these with (6) gives us the remaining statements. \( \square \)

**Lemma 7.** Assume \( p^* = d^* \). Assume \( x^{k+1/2} \) and \( x^{k+1} \) are the DRS iterates as defined in (2). If \( x^{k+1/2} \to x^* \) and \( x^{k+1} \to x^* \) for some \( x^* \), then \( x^* \) 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 Lemma 5 we have

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

and we conclude \( f(x^*) + g(x^*) = p^* \). \( \square \)
3.3. Evolution of shadow iterates

Section 3.1 characterized the evolution of the $z^k$-iterates, which we could all the main iterates. The $x^{k+1/2}$ and $x^{k+1}$-iterates of DRS are called the shadow iterates. Here, we analyze the evolution of the shadow iterates.

Although the results of this section are not as fundamental or important as the results of Sections 3.1 and 3.2, we do need these results later, especially when translating the analyses for DRS to analyses for ADMM.

**Lemma 8.** If $v = 0$, then $x^{k+3/2} - x^{k+1/2} \to 0$ and $x^{k+2} - x^{k+1} \to 0$.

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

**Lemma 9.** If (P) is strongly infeasible and (D) is feasible, then $x^{k+3/2} - x^{k+1/2} \to 0$ and $x^{k+2} - x^{k+1} \to 0$.

**Proof.** Write $-d'$ for the infimal displacement vector as given by Corollary 5. By Theorem 3, we have

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

The projection inequality states

$$\langle v - \Pi_C x, \Pi_C x - x \rangle \geq 0 \quad (7)$$

for any nonempty closed convex set $C$, $v \in C$, and $x \in \mathbb{R}^n$. Since $-d' = \Pi_{\text{dom } f - \text{dom } g}(0)$, (7) tells us that

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

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

$$\|\text{Prox}_{\gamma f}(z^k + d') - x^{k+1/2}\|^2 \leq \langle d', \text{Prox}_{\gamma f}(z^k + d') - x^{k+1/2} \rangle = \langle d', \text{Prox}_{\gamma f}(z^k + d') - x^{k+1} \rangle + \langle d', x^{k+1} - x^{k+1/2} \rangle \to 0$$

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

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

Putting everything together we conclude

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

Since

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

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

**Lemma 10.** 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 $-\gamma d = \gamma \Pi_{\text{dom } f + \text{dom } g}(0)$ is the infimal displacement vector as given in Corollary 3.
Proof. For simplicity, assume $\gamma = 1$. For $\gamma \neq 1$, we scale $f$ and $g$ to get the stated result. Rewrite the DRS iteration as

$$x^{k+1/2} = \text{Prox}_f(z^k)$$

$$\nu^{k+1/2} = z^k - x^{k+1/2} = \text{Prox}_{f^*}(z^k)$$

$$x^{k+1} = \text{Prox}_g(2x^{k+1/2} - z^k)$$

$$\nu^{k+1} = 2x^{k+1/2} - z^k - x^{k+1} = \text{Prox}_{g^*}(2x^{k+1/2} - z^k)$$

$$z^{k+1} = z^k - (\nu^{k+1} + \nu^{k+1/2}).$$

By Theorem 3, we have

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

By the same reasoning as in Lemma 9, we can use the (7) and firm-nonexpansiveness to show that

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

Since

$$z^{k+1} - z^k = \frac{\nu^{k+3/2} - \nu^{k+1/2}}{2} + \nu^{k+3/2} - x^{k+1/2} \to d,$$

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

Since

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

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

4. Pathological convergence: DRS

In this section, we use the theory of Section 3 to analyze DRS under pathologies. We classify the status of (P) and (D) into 7 cases and provide convergence analyses for the first 6 cases, the ones that assume strong duality.

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).**

In other words, (P) and (D) have solutions, and $d^* = p^*$. For example, the primal problem

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

and its dual problem

$$\text{maximize } 1 + \log(y)$$

subject to $y = 1$

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

**Case (b) $d^* = p^*$ is finite, (P) has a solution, (D) has no solution.**

For example, the primal problem

$$\text{minimize } \delta_{\{(x_1,x_2) | x_1^2 + x_2^2 \leq 1\}}(x_1, x_2) + x_2 + \delta_{\{(x_1,x_2) | x_1 = 1\}}(x_1, x_2)$$

has a solution but its dual problem

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

does not. Nevertheless, $d^* = p^* = 0$. 
Case (c) $d^* = p^*$ 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) $d^* = p^* = -\infty$, (P) is feasible, but there is no improving direction.
This implies (D) is weakly infeasible. For example, the primal problem
\[
\text{minimize } \delta_{\{x : x \geq 1\}}(x) - \log x
\]
has no solution and has optimal value $p^* = -\infty$. Since the asymptotic derivative of the objective $x \to \infty$ is 0, the primal problem has no improving direction. The dual problem
\[
\text{maximize } y + 1 + \log(y) \\
\text{subject to } y \leq 0
\]
is weakly infeasible.

Case (e) $d^* = p^* = -\infty$, (P) is feasible, and there is an improving direction.
This implies (D) is strongly infeasible. For example, the primal problem
\[
\text{minimize } x + x
\]
has an improving direction, namely $d = -1$, and the dual problem
\[
\text{maximize } \delta_{\{1\}}(x) + \delta_{\{1\}}(-x)
\]
is strongly infeasible.

Case (f) $d^* = p^* = \infty$ and (P) is infeasible.
For example, the problem
\[
\text{minimize } 1/\sqrt{-x} - \log(x)
\]
is infeasible, and its dual
\[
\text{maximize } (3/2^{2/3})y^{1/3} + 1 + \log(y) \\
\text{subject to } y \geq 0
\]
has optimal value $d^* = \infty$.

Case (g) $d^* < p^*$, i.e. strong duality fails.

4.2. Convergence results

Theorem 5. [54, 34] In case (a), $x^{k+1/2}, x^{k+1} \to x^*$, where $x^*$ is a solution of (P) and
\[
\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^*.
\]

Theorem 6. In case (b), $x^{k+1} - x^{k+1/2} \to 0$ and
\[
\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^*, \quad \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^*.
\]
Furthermore, if $x^{k+1/2} \to x^*$ (or equivalently if $x^{k+1} \to x^*$) then $x^*$ is a solution.

Theorem 7. In case (c), $x^{k+1} - x^{k+1/2} \to 0$,
\[
\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = p^*, \quad \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) = p^*,
\]
and $(x^{k+1/2}, x^{k+1})$ do not converge.
Theorem 8. In case (d), (D) is weakly infeasible, \( x^{k+1} - x^{k+1/2} \to 0 \),
\[
\lim_{k \to \infty} \frac{1}{k+1} \sum_{i=0}^{k} f(x^{i+1/2}) + g(x^{i+1}) = -\infty,
\]
and \((x^{k+1/2}, x^{k+1})\) do not converge.

Theorem 9. 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, \( \text{dist}(x^{k+1/2}, \text{dom } g) \to 0 \) and \( \text{dist}(x^{k+1}, \text{dom } f) \to 0 \).

Theorem 10. In case (f), \( ||x^{k+1} - x^{k+1/2}|| \to \text{dist}(\text{dom } f, \text{dom } g) \).

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. Finally, DRS provides a certificate of infeasibility in cases (e) and (f).

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

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

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.

In cases (e) and (f), the limit \( z^{k+1} - z^{k} \to -v \neq 0 \) provides a certificate of dual and primal strong infeasibility, respectively. These may be computationally useful when verifying the validity of a certificate is easy, which is the case for conic programs.

4.4. Proofs

Theorem 6 follows from Theorem 3, Corollary 1, Lemma 5, and Lemma 7. Theorem 7 follows from Theorem 3, Corollary 1, Lemma 5, and the contrapositive of Lemma 7. Theorem 8 follows from Theorem 3, Lemma 1, Corollary 2, Lemma 5, and the contrapositive of Lemma 7. Theorem 10 follows from Theorem 3 and Corollaries 4 and 5.

Finally, we prove Theorem 9. All but the last assertions follows from Theorem 3, Lemma 1, Corollary 3, Lemma 6, and the contrapositive of Lemma 7. By Lemma 10 \( x^{k+1/2} - x^{k-1/2} \to \gamma d \) and by Theorem 3 and Corollary 3 \( x^k - x^{k-1/2} \to \gamma d \). So \( x^{k+1/2} - x^k \to 0 \). Since \( x^k \in \text{dom } g \), we have
\[
\text{dist}(x^{k+1/2}, \text{dom } g) \leq \text{dist}(x^{k+1/2}, x^k) \to 0.
\]
Since \( x^{k+1/2} \in \text{dom } f \), we have
\[
\text{dist}(x^k, \text{dom } f) \leq \text{dist}(x^k, x^{k+1/2}) \to 0.
\]
5. Pathological convergence: ADMM

We now analyze ADMM under pathologies. Consider the primal problem

\[
\begin{align*}
\text{minimize} & \quad f(x) + g(y) \\
\text{subject to} & \quad Ax + By = c,
\end{align*}
\]

where \( f : \mathbb{R}^p \to \mathbb{R} \cup \{\infty\} \) and \( g : \mathbb{R}^q \to \mathbb{R} \cup \{\infty\} \) are CPC, \( A \in \mathbb{R}^{n \times p} \), \( B \in \mathbb{R}^{n \times q} \), and \( c \in \mathbb{R}^n \), and its dual problem

\[
\begin{align*}
\text{maximize} & \quad -f^*(-A^T \nu) - g^*(-B^T \nu) - c^T \nu.
\end{align*}
\]

Write \( p^* \) and \( d^* \) for the primal and dual optimal values. ADMM applied to this primal-dual problem pair is

\[
\begin{align*}
x^{k+1} & \in \arg \min_{x \in \mathbb{R}^p} \left\{ f(x) + (\nu^k)^T (Ax + By^k - c) + \frac{1}{2\gamma} \|Ax + By^k - c\|^2 \right\} \\
y^{k+1} & \in \arg \min_{y \in \mathbb{R}^q} \left\{ g(y) + (\nu^k)^T (Ax^{k+1} + By - c) + \frac{1}{2\gamma} \|Ax^{k+1} + By - c\|^2 \right\} \\
\nu^{k+1} & = \nu^k + \frac{1}{\gamma} (Ax^{k+1} + By^{k+1} - c).
\end{align*}
\]

(8)

For ADMM to be well-defined, the argmins of (8) must exist. Throughout this section, we assume the regularity conditions

\[
\begin{align*}
\text{(range} A^T \cap \text{ri dom}(f^*) \neq & \emptyset \quad (9) \\
\text{(range} B^T \cap \text{ri dom}(g^*) \neq & \emptyset \quad (10)
\end{align*}
\]

in addition to requiring \( f \) and \( g \) to be CPC. Here, ri denotes the relative interior of a set. These conditions ensure the subproblems are solvable [86, Theorem 16.3].

Without these regularity conditions, the subproblems of (8) may not have solutions. This is often overlooked and sometimes even misunderstood throughout the ADMM literature. (The highly influential paper [23] mistakenly claimed it is enough for \( f \) and \( g \) to be CPC. Chen, Sun, and Toh [25] pointed out that additional assumptions are needed.)

5.1. Classification and convergence results

Under (9) and (10), the status of (P-ADMM) and (D-ADMM) falls under exactly one of the following 5 distinct cases.

Case (a) \( d^* = p^* \), both (P-ADMM) and (D-ADMM) have solutions.

**Theorem 11 ([51, 23, 34]).** In case (a), \( Ax^k + By^k - c \to 0 \) and

\[
\lim_{k \to \infty} f(x^k) + g(y^k) \to p^*.
\]

Case (b) \( d^* = p^* \), (P-ADMM) has a solution, (D-ADMM) has no solution.

**Theorem 12.** In case (b), \( Ax^k + By^k - c \to 0 \) and

\[
\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} f(x^i) + g(y^i) = p^* , \quad \liminf_{k \to \infty} f(x^k) + g(y^k) = p^*.
\]

Furthermore, if \((x^k, y^k) \to (x^*, y^*)\), then \((x^*, y^*)\) is a solution.
Case (c)  $d^* = p^* \in [-\infty, \infty)$, (P-ADMM) is feasible but has no solution.

**Theorem 13.** In case (c), $Ax^k + By^k - c \to 0$ and

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} f(x^i) + g(y^i) = p^*, \quad \liminf_{k \to \infty} f(x^k) + g(y^k) = p^*,$$

and the sequence $(x^k, y^k)$ does not converge.

Case (d)  $d^* = p^* = \infty$, (P-ADMM) is infeasible.

**Theorem 14.** In case (d),

$$\|Ax^k + By^k - c\| \to \inf_{x \in \text{dom } f} \inf_{y \in \text{dom } g} \|Ax + By - c\|.$$

Case (e)  $d^* < p^*$, i.e. strong duality fails.

### 5.2. Interpretation

With some caveats, ADMM succeeds at achieving feasibility and optimality. In cases (a), (b), and (c) the iterates become approximately feasible in that $Ax^k + By^k - c \to 0$, and the function values on average converge to the solution. In case (d), feasibility is impossible, but ADMM does its best to achieve feasibility.

### 5.3. Proofs

ADMM is often analyzed as DRS applied to (D-ADMM) [47]. In this proof, however, we take the less common approach shown in [40, 97], which derives ADMM directly from the primal problem. We do so as the function-value analysis of Section 3.2 translate nicely with this primal approach.

Consider the equivalent primal optimization problem

$$\text{minimize } \tilde{f}(z) + \tilde{g}(z)$$

with

$$\tilde{f}(z) = \inf \{ f(x) \mid Ax + z = 0 \}, \quad \tilde{g}(z) = \inf \{ g(y) \mid By - c = z \},$$

which are CPC functions, as we assume (9) and (10) [86, Theorem 16.3]. We apply DRS to this form to get

$$\tilde{x}^{k+1/2} = \arg \min_{\tilde{x}} \left\{ \gamma \tilde{g}(\tilde{x}) + (1/2)\|\tilde{x} - z^k\|^2 \right\}$$

$$\tilde{x}^{k+1} = \arg \min_{\tilde{x}} \left\{ \gamma \tilde{f}(\tilde{x}) + (1/2)\|\tilde{x} - 2\tilde{x}^{k+1/2} + z^k\|^2 \right\}$$

$$z^{k+1} = z^k + \tilde{x}^{k+1} - \tilde{x}^{k+1/2},$$

where we perform the $\tilde{g}$-update before the $\tilde{f}$-update. We introduce and substitute the variables $x^k$, $y^k$, and $\nu^k$ defined implicitly by $\tilde{x}^{k+1/2} = By^{k+1} - c$, $\tilde{x}^{k+1} = -Ax^{k+2}$, and $z^k = -\gamma \nu^k - Ax^{k+1}$ to get

$$y^{k+1} = \arg \min_{y} \left\{ \gamma g(y) + \gamma (\nu^k)^T (Ax^{k+1} + By - c) + (1/2)\|Ax^{k+1} + By - c\|^2 \right\}$$

$$x^{k+2} = \arg \min_{x} \left\{ \gamma f(x) + \gamma (\nu^{k+1})^T (Ax + By^{k+1} - c) + (1/2)\|Ax + By^{k+1} - c\|^2 \right\}$$

$$\nu^{k+1} = \nu^k + (1/\gamma)(Ax^{k+1} + By^{k+1} - c).$$
Reordering the updates to get the dependency right, we get
\[ y^{k+1} = \arg\min_{y} \left\{ \gamma g(y) + (\nu^{k})^{T}(Ax^{k+1} + By - c) + (1/2)\left\| Ax^{k+1} + By - c \right\|^2 \right\} \]
\[ \nu^{k+1} = \nu^{k} + (1/\gamma)(Ax^{k+1} + By^{k+1} - c) \]
\[ x^{k+2} = \arg\min_{x} \left\{ \gamma f(x) + (\nu^{k+1})^{T}(Ax + By^{k+1} - c) + (1/2)\left\| Ax + By^{k+1} - c \right\|^2 \right\} . \]

Finally, redefine the start and end of an iteration so that it updates \( x^{k+1}, y^{k+1} \), and \( \nu^{k+1} \) instead of \( y^{k+1}, \nu^{k+1} \), and \( x^{k+2} \). With this, we get (8).

The last step, where we redefine the start and end of an iteration, introduces a subtlety when translating the results of Section 4.2. In particular, the results of Section 3.3 are necessary because of this.

Theorem 12 follows from Theorem 6 and Lemmas 2, 6, and 8. Theorem 14 follows from Theorem 10 and Lemma 9.

Case (c) of this section corresponds to cases (c), (d), and (e) of Section 4.2. For the three cases, we use Theorem 7 and Lemmas 2, 6, and 8, Theorem 8 and Lemmas 2, 6, and 8, and Theorem 9 and Lemma 6, and 10. Combining the three results into one gives us Theorem 13. □

6. When strong duality fails
In the analyses of DRS, we assumed strong duality holds. When strong duality fails, i.e., when \( d^{*} < p^{*} \), we conjecture that DRS fails.

**Conjecture 1** When strong duality fails, DRS necessarily fails in that
\[ \liminf_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^{*} . \]

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 \( p^{*} \) while achieving an infinitesimal infeasibility. We support the conjecture with examples.

We first present an analytical counter example. Consider the problem taken from [56]
\[
\text{minimize } \delta_{\{(x_1,x_2,x_3) \mid x_3 \geq (x_1^2 + x_2^2)^{1/2}\}}(x) + \frac{1}{2}\left\| (x_1,x_2,x_3) \right\|_{2}^2
\]
which has the solution set \( \{(0,t,t) \mid t \in \mathbb{R} \} \) and optimal value \( p^{*} = 0 \). Its dual problem
\[
\text{maximize } -\delta_{\{(\nu_1,\nu_2,\nu_3) \mid \nu_3 \geq (\nu_1^2 + \nu_2^2)^{1/2}\}}(-\nu) - \delta_{\{(\nu_1,\nu_2,\nu_3) \mid \nu_1 = 1, \nu_2 = -\nu_3\}}(-\nu)
\]
is infeasible. Given \( z^0 = (z_1^0, z_2^0, 0) \), the DRS iterates have the form
\[
\begin{align*}
  z_1^{k+1} &= \frac{1}{2}z_1^k - \gamma \\
  z_2^{k+1} &= \frac{1}{2}z_2^k + \frac{1}{2}(z_1^k)^2 + (z_2^k)^2 \\
  z_3^{k+1} &= 0.
\end{align*}
\]

With this, it is relatively straightforward to show \( x^{k+1/2} - x^{k+1} \to 0 \), \( x_1^{k+1/2} \to -2\gamma \), \( x_2^{k+1/2} \to \infty \), \( x_3^{k+1/2} \to \infty \), and \( f(x^{k+1/2}) + g(x^{k+1}) \to -2\gamma \). Also, \( x^{k+1/2} \to \text{dom } f \cap \text{dom } g \) even though \( x^{k+1/2} - x^{k+1} \to 0 \).
Note that
\[ d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^*. \]

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

Next, we present more experimental counter examples that support the conjecture. We run DRS on these problems report the experimental results.

The problem, taken from [17],
\[
\min_{x \in \mathbb{R}^2} \exp\left(-\sqrt{x_1 x_2}\right) + \delta_{\{(x_1, x_2) | x_1 = 0\}}(x)
\]
has \( p^* = 1 \) but \( d^* = 0 \). Experimentally, for all \( \gamma > 0 \) and choice of \( z^0 \) we observe \( d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^* \).

The problem, taken from [39],
\[
\min_{X \in \mathbb{S}^3} \delta_{\mathbb{S}^3_+}(X) + \frac{1}{2}X_{12} + \delta_{\{X \in \mathbb{S}^3 | X_{33} = 0, X_{22} + 2X_{13} = 1\}}(X)
\]
where \( \mathbb{S}^3 \) and \( \mathbb{S}^3_+ \) respectively denote the set of symmetric and positive semidefinite \( 3 \times 3 \) matrices, has \( p^* = 1 \) but \( d^* = 0 \). Experimentally, we observe \( d^* = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \) for \( \gamma \geq 0.5 \), and \( d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^* \) for \( 0 < \gamma < 0.5 \). This behaviour does not depend on \( z^0 \).

The problem, taken from [98],
\[
\min_{X \in \mathbb{S}^3} \delta_{\mathbb{S}^3_+}(X) + 2X_{12} + \delta_{\{X \in \mathbb{S}^3 | X_{22} = 0, -2X_{12} + 2X_{33} = 2\}}(X)
\]
has \( p^* = 0 \) but \( d^* = -2 \). Experimentally, we observe \( d^* = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \) for \( \gamma \geq 1 \), and \( d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^* \) for \( 0 < \gamma < 1 \). This behaviour does not depend on \( z^0 \).

The problem, taken from [91],
\[
\min_{X \in \mathbb{S}^5} \delta_{\mathbb{S}^5_+}(X) + X_{14} + X_{55} + \delta_{\{X \in \mathbb{S}^5 | X_{11} = 0, X_{22} = 1, X_{34} = 1, 2X_{13} + 2X_{45} + X_{55} = 1\}}(X)
\]
has \( p^* = (\sqrt{5} - 1)/2 \) but \( d^* = 0 \). Experimentally, we observe \( d^* = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \) for \( \gamma \geq 0.8 \), and \( d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^* \) 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 \( \gamma_{\text{min}} \) such that \( d^* < \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) < p^* \) when \( 0 < \gamma < \gamma_{\text{min}} \) and \( d^* = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1}) \) when \( \gamma_{\text{min}} \leq \gamma \). We do not have an explanation for this phenomenon.

7. Conclusion

In this paper, we analyzed DRS and ADMM under pathologies. With some caveats, DRS and ADMM 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 6, DRS exhibits an interesting behavior in the absence of strong duality, and we do not have an explanation for it. Analyzing this phenomenon and addressing the conjecture is an interesting direction of future research.
One conclusion of this paper is that we can solve pathological convex programs with 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 direction of future research.

For non-pathological problems, DRS can be generalized with an over-under relaxation parameter between 0 and 2. The pathological DRS analysis of this paper immediately extends to this generalized setup. For non-pathological problems, ADMM can be generalized with an over-under relaxation parameter between 0 and \((1 + \sqrt{5})/2\). This generalization arises when ADMM is analyzed directly through a Lyapunov function, and not through DRS \([46, 50, 18, 37, 44, 25, 24]\). The pathological ADMM analysis of this paper does not immediately extend to this generalized setup. Analyzing this form of ADMM applied to pathological problems is an interesting direction of future research.

References

[1] Artacho FJA, Borwein JM, Tam MK (2014) Douglas-Rachford feasibility methods for matrix completion problems. *The ANZIAM J.* 55(4):299–326.

[2] Artacho FJA, Borwein JM, Tam MK (2016) Global behavior of the Douglas-Rachford method for a nonconvex feasibility problem. *J. of Global Optim.* 65(2):309–327.

[3] Attouch H, Baillon JB, Théra M (????) Variational sum of monotone operators. *Journal of Convex Analysis* 1(1):001–030.

[4] Attouch H, Théra M (1996) A general duality principle for the sum of two operators. *Journal of Convex Analysis* 3(1):1–24.

[5] Baillon JB, Bruck RE, Reich S (1978) On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. *Houston J. Math.* 4(1):1–9.

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

[7] Bauschke HH, Boţ RI, Hare WL, Moursi WM (2012) Attouch-Théra duality revisited: Paramonotonicity and operator splitting. *J. Approx. Theory* 164(8):1065–1084.

[8] Bauschke HH, Borwein JM, Lewis AS (1997) The method of cyclic projections for closed convex sets in Hilbert space. *Contemp. Math.* 204:1–38.

[9] Bauschke HH, Combettes PL (2017) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (Springer New York), 2nd edition.

[10] Bauschke HH, Combettes PL, Luke D (2004) Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. *J. Approx. Theory* 127(2):178–192.

[11] Bauschke HH, Douglas GR, Moursi WM (2016) On a result of Pazy concerning the asymptotic behaviour of nonexpansive mappings. *J. Fixed Point Theory Appl.* 18(2):297–307.

[12] Bauschke HH, Hare WL, Moursi WM (2014) Generalized solutions for the sum of two maximally monotone operators. *SIAM J. Control Optim.* 52(2):1034–1047.

[13] Bauschke HH, Hare WL, Moursi WM (2016) On the range of the Douglas-Rachford operator. *Math. Oper. Res.* 41(3):884–897.

[14] Bauschke HH, Luikens B, Moursi WM (2017) Affine nonexpansive operators, attouch–thera duality and the douglas–rachford algorithm. *Set-Valued and Variational Analysis* 25(3):481–505.

[15] Bauschke HH, Moursi WM (2016) The Douglas-Rachford algorithm for two (not necessarily intersecting) affine subspaces. *SIAM J. Optim.* 26(2):968–985.

[16] Bauschke HH, Moursi WM (2017) On the Douglas–Rachford algorithm. *Math. Program.* 164(1–2):263–284.

[17] Bertsekas D (2009) *Convex Optimization Theory* (Athena Scientific).
[18] Bertsekas D, Tsitsiklis J (1989) Parallel and Distributed Computation: Numerical Methods (Prentice-Hall, Englewood Cliffs).

[19] Borwein J, Wolkowicz H (1981) Regularizing the abstract convex program. J. Math. Anal. Appl. 83(2):495–530.

[20] Borwein JM, Tam MK (2015) The cyclic Douglas-Rachford method for inconsistent feasibility problems. J. Nonlinear Convex Anal. 16(4):537–584.

[21] Borwein JM, Wolkowicz H (1981) Facial reduction for a cone-convex programming problem. J. Austral. Math. Soc. 30(3):369–380.

[22] Borwein JM, Wolkowicz H (1982) Characterizations of optimality without constraint qualification for the abstract convex program. Guignard M, ed., Optimality and Stability in Mathematical Programming, 77–100 (Berlin, Heidelberg: Springer Berlin Heidelberg).

[23] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1):1–122.

[24] Chen L, Li X, Sun D, Toh KC (2018) On the equivalence of inexact proximal alm and admm for a class of convex composite programming. arXiv.

[25] Chen L, Sun D, Toh KC (2017) A note on the convergence of ADMM for linearly constrained convex optimization problems. Computational Optimization and Applications 66(2):327–343.

[26] Cheung YL, Schurr S, Wolkowicz H (2013) Preprocessing and regularization for degenerate semidefinite programs. Computational and analytical mathematics, 251–303 (Springer).

[27] Cheung YL, Schurr S, Wolkowicz H (2013) Preprocessing and regularization for degenerate semidefinite programs. Bailey DH, Bauschke HH, Borwein P, Garvan F, Théra M, Vanderwerff JD, Wolkowicz H, eds., Computational and Analytical Mathematics, 251–303 (New York, NY: Springer New York).

[28] Combettes PL (2004) Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization 53(5–6):475–504.

[29] Combettes PL (2018) Monotone operator theory in convex optimization. Mathematical Programming 170(1):177–206.

[30] Combettes PL, Pesquet JC (2011) Proximal splitting methods in signal processing. Bauschke HH, Burachik RS, Combettes PL, Elser V, Luke DR, Wolkowicz H, eds., Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 185–212 (New York, NY: Springer New York).

[31] Dantzig G (2016) Linear Programming and Extensions (Princeton university press).

[32] Dantzig GB, Duffin RJ, Fan K, Mather AW (1956) Linear Inequalities and Related systems. Number 38 (Princeton university press).

[33] Davis D (2015) Convergence rate analysis of primal-dual splitting schemes. SIAM Journal on Optimization 25(3):1912–1943.

[34] Davis D, Yin W (2016) Convergence rate analysis of several splitting schemes. Glowinski R, Osher SJ, Yin W, eds., Splitting Methods in Communication, Imaging, Science, and Engineering, 115–163 (Springer International Publishing).

[35] Davis D, Yin W (2017) Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. Math. Oper. Res. 42(3):783–805.

[36] De Loera JA, Malkin PN, Parrilo PA (2012) Computation with polynomial equations and inequalities arising in combinatorial optimization. Lee J, Leyffer S, eds., Mixed Integer Nonlinear Programming, 447–481 (New York, NY: Springer New York).

[37] Deng W, Yin W (2016) On the global and linear convergence of the generalized alternating direction method of multipliers. Journal of Scientific Computing 66(3):889–916.

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

[39] Drusvyatskiy D, Wolkowicz H (2017) The many faces of degeneracy in conic optimization. Found. Trends Optim. 3(2):77–170.
[40] Eckstein J (1989) Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, MIT.
[41] Eckstein J, Bertsekas DP (1992) On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* 55(1–3).
[42] Eckstein J, Yao W (2015) Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pacific Journal on Optimization* 11(4):619–644.
[43] Farlas J (1902) Theorie der einfachen ungleichungen. *Journal für die reine und angewandte Mathematik* 124:1–27.
[44] Fazel M, Pong T, Sun D, Tseng P (2013) Hankel matrix rank minimization with applications to system identification and realization. *SIAM Journal on Matrix Analysis and Applications* 34(3):946–977.
[45] Fenchel W (1953) Convex cones, sets, and functions. Mimeographed lecture notes.
[46] Fortin M, Glowinski R (1983) On decomposition-coordination methods using an augmented lagrangian. Fortin M, Glowinski R, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, 97–146.
[47] Gabay D (1983) Applications of the method of multipliers to variational inequalities. Fortin M, Glowinski R, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*.
[48] Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* 2(1):17–40.
[49] Giselsson P, Fält M (2018) Envelope functions: Unifications and further properties. *Journal of Optimization Theory and Applications*.
[50] Glowinski R (1984) *Numerical methods for nonlinear variational problems* (Springer).
[51] Glowinski R, Marroco A (1975) Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *Revue Française d’Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique* 9(2):41–76.
[52] Kellogg RB (1969) A nonlinear alternating direction method. *Math. Comput.* 23(105):23–27.
[53] Kretschmer KS (1961) Programmes in paired spaces. *Canad. J. Math.* 13:221–238.
[54] Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* 16(6):964–979.
[55] Liu M, Pataki G (2018) Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. *Mathematical Programming* 167(2):435–480.
[56] Liu Y, Ryu EK, Yin W (2018) A new use of Douglas-Rachford splitting for identifying infeasible, unbounded, and pathological conic programs. *Math. Program.*.
[57] Liu Y, Yin W (2018) An envelope for Davis-Yin splitting and strict saddle point avoidance. *arXiv*.
[58] Lofberg J (2009) Pre- and post-processing sum-of-squares programs in practice. *IEEE Trans. Autom. Control* 54(5):1007–1011.
[59] Lourenço BF, Muramatsu M, Tsuchiya T (2015) Solving SDP completely with an interior point oracle. *arXiv preprint arXiv:1507.08065*.
[60] Lourenço BF, Muramatsu M, Tsuchiya T (2016) A structural geometrical analysis of weakly infeasible sdpss. *J. Oper. Res. Soc. Japan* 59(3):241–257.
[61] Luo ZQ, Sturm JF, Zhang S (1997) Duality results for conic convex programming. *Econometric Institute, Erasmus University Rotterdam*, The Netherlands, Technical Report 9719/A.
[62] Luo ZQ, Sturm JF, Zhang S (1997) Duality results for conic convex programming. Technical report, Erasmus University Rotterdam, Econometric Institute.
[63] Luo ZQ, Sturm JF, Zhang S (2000) Conic convex programming and self-dual embedding. *Optimization Methods and Software* 14(3):169–218.
[64] Luo ZQ, Sturm JF, Zhang S (2000) Conic convex programming and self-dual embedding. *Optim. Methods Softw.* 14(3):169–218.
[65] Mercier B (1980) *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d’Orsay, no. 80.01) (Orsay, France: Université de Paris-XI).

[66] Moursi WM (2018) The forward-backward algorithm and the normal problem. *Journal of Optimization Theory and Application* 176(3):605–624.

[67] Nesterov Y, Todd MJ, Ye Y (1999) Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Program.* 84(2):227–267.

[68] Pataki G (2000) A simple derivation of a facial reduction algorithm and extended dual systems. Technical report, Columbia University.

[69] Pataki G (2013) Strong duality in conic linear programming: facial reduction and extended duals. *Computational and analytical mathematics*, 613–634 (Springer).

[70] Pataki G (2017) Bad semidefinite programs: They all look the same. *SIAM J. Optim.* 27(1):146–172.

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

[72] Patrinos P, Stella L, Bemporad A (2014) Douglas-rachford splitting: Complexity estimates and accelerated variants. *IEEE Conference on Decision and Control (CDC)* 4234–4239.

[73] Patrinos P, Stella L, Bemporad A (2014) Forward-backward truncated newton methods for convex composite optimization. *arXiv*.

[74] Pazy A (1971) Asymptotic behavior of contractions in Hilbert space. *Israel J. Math.* 9(2):235–240.

[75] Peaceman DW, Rachford HH (1955) The numerical solution of parabolic and elliptic differential equations. *J. Soc. Ind. Appl. Math.* 3(1).

[76] Peña J, Renegar J (2000) Computing approximate solutions for convex conic systems of constraints. *Math. Program.* 87(3):351–383.

[77] Permenter F, Friberg HA, Andersen ED (2017) Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach. *SIAM J. Optim.* 27(3):1257–1282.

[78] Permenter F, Parrilo P (2017) Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. *Mathematical Programming*.

[79] Permenter F, Parrilo PA (2014) Basis selection for SOS programs via facial reduction and polyhedral approximations. *53rd IEEE Conference on Decision and Control (CDC)*, 6615–6620.

[80] Raghunathan AU, Cairano SD (2014) Infeasibility detection in alternating direction method of multipliers for convex quadratic programs. *2014 IEEE 53rd Annual Conference on Decision and Control (CDC)*, 5819–5824.

[81] Ramana MV (1997) An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming* 77(1):129–162.

[82] Ramana MV, Tunçel L, Wolkowicz H (1997) Strong duality for semidefinite programming. *SIAM J. Optim.* 7(3):641–662.

[83] Revalski JP, Théra M (1999) Generalized sums of monotone operators. *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics* 329(11):979–984.

[84] Revalski JP, Théra M (2002) Enlargements and sums of monotone operators. *Nonlinear Analysis: Theory, Methods & Applications* 48(4):505–519.

[85] Rockafellar R (1974) *Conjugate Duality and Optimization* (Society for Industrial and Applied Mathematics).

[86] Rockafellar RT (1970) *Convex Analysis*.

[87] Ryu EK (2018) Cosmic divergence, weak cosmic convergence, and fixed points at infinity. *Journal of Fixed Point Theory and Applications*.

[88] Stella L, Themelis A, Patrinos P (2017) Forward–backward quasi-newton methods for nonsmooth optimization problems. *Computational Optimization and Applications* 67(3):443–487.

[89] Stellato B, Banjac G, Goulart P, Bemporad A, Boyd S (2017) OSQP: An operator splitting solver for quadratic programs. *arXiv preprint arXiv:1711.08013*.
[90] Tseng P (2009) Some convex programs without a duality gap. Math. Program. 116(1–2):553–578.

[91] Tunçel L, Wolkowicz H (2012) Strong duality and minimal representations for cone optimization. Comput. Optim. Appl. 53(2):619–648.

[92] Waki H (2012) How to generate weakly infeasible semidefinite programs via Lasserre’s relaxations for polynomial optimization. Optim. Lett. 6(8):1883–1896.

[93] Waki H, Muramatsu M (2010) A facial reduction algorithm for finding sparse sos representations. Operations Research Letters 38(5):361 – 365.

[94] Waki H, Muramatsu M (2013) Facial reduction algorithms for conic optimization problems. J. Optim. Theory Appl. 158(1):188–215.

[95] Waki H, Nakata M, Muramatsu M (2012) Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization. Comput. Optim. Appl. 53(3):823–844.

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

[97] Yan M, Yin W (2016) Self equivalence of the alternating direction method of multipliers. Glowinski R, Osher SJ, Yin W, eds., Splitting Methods in Communication, Imaging, Science, and Engineering, 165–194 (Springer International Publishing).

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