Math. Program., Ser. A (2019) 177:225–253
https://doi.org/10.1007/s10107-018-1265-5

FULL LENGTH PAPER

A new use of Douglas–Rachford splitting for identifying infeasible, unbounded, and pathological conic programs

Yanli Liu[1] · **Ernest K. Ryu**[1] · **Wotao Yin**[1]

Received: 15 June 2017 / Accepted: 21 March 2018 / Published online: 2 April 2018

Abstract In this paper, we present a method for identifying infeasible, unbounded, and pathological conic programs based on Douglas–Rachford splitting. When an optimization program is infeasible, unbounded, or pathological, the iterates of Douglas–Rachford splitting diverge. Somewhat surprisingly, such divergent iterates still provide useful information, which our method uses for identification. In addition, for strongly infeasible problems the method produces a separating hyperplane and informs the user on how to minimally modify the given problem to achieve strong feasibility. As a first-order method, the proposed algorithm relies on simple subroutines, and therefore is simple to implement and has low per-iteration cost.

Keywords Douglas–Rachford splitting · Infeasible · Unbounded · Pathological · Conic programs

Mathematics Subject Classification 47H05 · 65K05 · 65K15 · 90C25

This work is supported in part by NSF Grant DMS-1720237 and ONR Grant N000141712162.

✉ Wotao Yin
wotaoyin@math.ucla.edu

Yanli Liu
yanli@math.ucla.edu

Ernest K. Ryu
eryu@math.ucla.edu

[1] Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA

1 Introduction

Many convex optimization algorithms have strong theoretical guarantees and empirical performance, but they are often limited to non-pathological, feasible problems; under pathologies often the theory breaks down and the empirical performance degrades significantly. In fact, the behavior of convex optimization algorithms under pathologies has been studied much less, and many existing solvers often simply report "failure" without informing the users of what went wrong upon encountering infeasibility, unboundedness, or pathology. Pathological problem are numerically challenging, but they are not impossible to deal with. As infeasibility, unboundedness, and pathology can arise in practice (see, for example [12,17,18,36,38]), designing a robust algorithm that behaves well in all cases is important to the completion of a robust solver.

In this paper, we propose a method based on Douglas–Rachford splitting (DRS) that identifies infeasible, unbounded, and pathological conic programs. First-order methods such as DRS are simple and can quickly provide a solution with low or moderate accuracy. It is well known, for example by combining Theorem 1 of [32] and Proposition 4.4 of [13], that the iterates of DRS converge to a fixed point if there is one (a fixed point z^* of an operator T satisfies $z^* = Tz^*$), and when there is no fixed point, the iterates diverge unboundedly. However, the precise manner in which they diverge has been studied much less. Somewhat surprisingly, when iterates of DRS diverge, the divergent iterates still provide useful information, which we use to classify the conic program. For example, a separating hyperplane can be found when the conic program is strongly infeasible, and an improving direction can be obtained when there is one. When the problem is infeasible or weakly feasible, we can get information of how to minimally modify the problem data to achieve strong feasibility.

Facial reduction is one approach to handle infeasible or pathological conic programs. Facial reduction reduces an infeasible or pathological problem into a reduced problem that is strongly feasible, strongly infeasible, or unbounded with an improving direction, which are the easier cases [9,10,26,37]. This reduced problem can then be solved with, say, interior point methods [25]. However, facial reduction introduces a new set of computational issues. After completing the facial reduction step, which has its own the computational challenge and cost, the reduced problem must be solved. The reduced problem involves a cone expressed as an intersection of the original cone with an linear subspace, and in general such cones neither are self-dual nor have a simple formula for projection. This makes applying an interior point method or a first-order method difficult, and existing work on facial reduction do not provide an efficient way to address this issue.

Homogeneous self-dual embedding is a transformation that embeds a conic program and its dual into a single larger conic program. In conjunction with interior point methods, one can use the homogeneous self-dual embedding to identify and solve some pathologies [14,19,28,39,40].

In contrast, our proposed method directly addresses infeasibility, unboundedness, and pathology without transforming to a larger problem. Some cases are always identified, and some are identifiable under certain conditions. Being a first-order method, the proposed algorithm relies on simple subroutines; each iteration performs projections onto the cone and the affine space of the conic program and elementary operations such as vector addition. Consequently, the method is simple to implement and has a lower per-iteration cost than interior point methods.

1.1 Basic definitions

Cones A set $K \subseteq \mathbb{R}^n$ is a cone if $K = \lambda K$ for any $\lambda > 0$. We write and define the dual cone of K as

$$K^* = \left\{u \in \mathbb{R}^n |\ u^T v \geq 0, \quad \text{for all } v \in K\right\}.$$

Throughout this paper, we focus on nonempty closed convex cones that we can efficiently project onto. In particular, we do *not* require that the cone be self-dual. Example of such cones include:

– The positive orthant:

$$\mathbb{R}^k_+ = \left\{x \in \mathbb{R}^k \mid x_i \geq 0, \quad i = 1, \ldots, n\right\}$$

– Second order cone:

$$Q^{k+1} = \left\{(x_1, \ldots, x_k, x_{k+1}) \in \mathbb{R}^k \times \mathbb{R}_+ \mid x_{k+1} \geq \sqrt{x_1^2 + \cdots + x_k^2}\right\}$$

– Rotated second order cone:

$$Q_r^{k+2} = \left\{(x_1, \ldots, x_k, x_{k+1}, x_{k+2}) \in \mathbb{R}^k \times \mathbb{R}^2_+ \mid 2x_{k+1}x_{k+2} \geq x_1^2 + \cdots + x_k^2\right\}.$$

– Positive semidefinite cone:

$$S^k_+ = \left\{M = M^T \in \mathbb{R}^{k\times k} |\ x^T M x \geq 0 \text{ for any } x \in \mathbb{R}^k\right\}$$

Conic programs Consider the conic program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \in K, \end{array} \tag{P}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$ are problem data, and $K \subseteq \mathbb{R}^n$ is a nonempty closed convex cone. We write $p^\star = \inf\{c^T x \mid Ax = b,\ x \in K\}$ to denote the optimal value of (P). For simplicity, we assume $m \leq n$ and A is full rank.

The dual problem of (P) is

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y + s = c \\ & s \in K^*, \end{array} \tag{D}$$

where $y \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$ are the optimization variables. We write $d^\star = \sup\{b^T y \mid A^T y + s = c,\ s \in K^*\}$ to denote the optimal value of (D).

The optimization problem (P) is either feasible or infeasible; (P) is feasible if there is an $x \in K \cap \{x \mid Ax = b\}$ and infeasible if there is not. When (P) is feasible, it is strongly feasible if there is an $x \in \mathbf{relint} K \cap \{x \mid Ax = b\}$ and weakly feasible if there is not, where **relint** denotes the relative interior. When (P) is infeasible, it is strongly infeasible if there is a non-zero distance between K and $\{x \mid Ax = b\}$, i.e., $d(K, \{x \mid Ax = b\}) > 0$, and weakly infeasible if $d(K, \{x \mid Ax = b\}) = 0$, where

$$d(C_1, C_2) = \inf\{\|x - y\| \mid x \in C_1,\ y \in C_2\},$$

and $\|\cdot\|$ denotes the Euclidean norm. Note that $d(C_1, C_2) = 0$ does not necessarily imply C_1 and C_2 intersect. When (P) is infeasible we say $p^\star = \infty$ and when feasible $p^\star \in \mathbb{R} \cup \{-\infty\}$. Likewise, when (D) is infeasible we say $d^\star = -\infty$ and when feasible $d^\star \in \mathbb{R} \cup \{\infty\}$.

As special cases, (P) is called a linear program when K is the positive orthant, a second-order cone program when K is the second-order cone, and a semidefinite program when K is the positive semidefinite cone.

1.2 Classification of conic programs

Every conic program of the form (P) falls under exactly one of the following 7 cases (some of the following examples are taken from [19–22]). Discussions on most of these cases exist in the literature. Some of these cases have a corresponding dual characterization, but we skip this discussion as it is not directly relevant to our method. We report the results of SDPT3 [35], SeDuMi [33], and MOSEK [23] using their default settings. In Sect. 2, we discuss how to identify most of these 7 cases.

Case (a) $p^\star$ is finite, both (P) and (D) have solutions, and $d^\star = p^\star$, which is the most common case. For example, the problem

$$\begin{array}{ll} \text{minimize} & x_3 \\ \text{subject to} & x_1 = 1 \\ & x_3 \geq \sqrt{x_1^2 + x_2^2} \end{array}$$

has the solution $x^\star = (1, 0, 1)$ and $p^\star = 1$. (The inequality constraint corresponds to $x \in Q^3$.) SDPT3, SeDuMi and MOSEK can solve this example.

The dual problem, after some simplification, is

$$\begin{array}{ll} \text{maximize} & y \\ \text{subject to} & 1 \geq y^2, \end{array}$$

which has the solution $y^\star = 1$ and $d^\star = 1$.

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

$$\begin{array}{ll} \text{minimize } x_2 \\ \text{subject to } x_1 = x_3 = 1 \\ \qquad\qquad x_3 \geq \sqrt{x_1^2 + x_2^2} \end{array}$$

has the solution $x^\star = (1, 0, 1)$ and optimal value $p^\star = 0$. (The inequality constraint corresponds to $x \in Q^3$.)

The dual problem, after some simplification, is

$$\text{maximize } y_1 - \sqrt{1 + y_1^2}.$$

By taking $y_1 \to \infty$, we achieve the dual optimal value $d^\star = 0$, but no finite y_1 achieves it.

In this example, SDPT3 reports "Inaccurate/Solved" and -2.99305×10^{-5} as the optimal value; SeDuMi reports "Solved" and -1.54566×10^{-4} as the optimal value; MOSEK reports "Solved" and -2.71919×10^{-8} as the optimal value.

As another example, the problem

$$\begin{array}{ll} \text{minimize} & 2x_{12} \\ \text{subject to} & X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & 0 & x_{23} \\ x_{13} & x_{23} & x_{12}+1 \end{bmatrix} \in S_+^3, \end{array}$$

has the solution

$$X^\star = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and optimal value $p^\star = 0$.

The dual problem, after some simplification, is

$$\begin{array}{ll} \text{maximize} & 2y_2 \\ \text{subject to} & \begin{bmatrix} 0 & y_2+1 & 0 \\ y_2+1 & -y_1 & 0 \\ 0 & 0 & -2y_2 \end{bmatrix} \in S_+^3, \end{array}$$

which has the solution $y^\star = (0, -1)$ and optimal value $d^\star = -2$.

In this example, SDPT3 reports "Solved" and -2 as the optimal value; SeDuMi reports "Solved" and -0.602351 as the optimal value; MOSEK reports "Failed" and does not report an optimal value.

Note that case (b) can happen only when (P) is weakly feasible, by standard convex duality [31].

Case (c) (P) is feasible, $p^\star$ is finite, but there is no solution. For example, the problem

$$\begin{aligned} \text{minimize } & x_3 \\ \text{subject to } & x_1 = \sqrt{2} \\ & 2x_2x_3 \geq x_1^2 \\ & x_2, x_3 \geq 0 \end{aligned}$$

has an optimal value $p^\star = 0$ but has no solution since any feasible x satisfies $x_3 > 0$. (The inequality constraints correspond to $x \in Q_r^3$.)

In this example, SDPT3 reports "Inaccurate/Solved" and 7.9509×10^{-5} as the optimal value; SeDuMi reports "Solved" and 8.75436×10^{-5} as the optimal value; MOSEK reports "Solved" and 4.07385×10^{-8} as the optimal value.

Case (d) (P) is feasible, $p^\star = -\infty$, and there is an improving direction, i.e., there is a $u \in \mathcal{N}(A) \cap K$ satisfying $c^T u < 0$. For example, the problem

$$\begin{aligned} \text{minimize } & x_1 \\ \text{subject to } & x_2 = 0 \\ & x_3 \geq \sqrt{x_1^2 + x_2^2} \end{aligned}$$

has an improving direction $u = (-1, 0, 1)$. If x is any feasible point, $x + tu$ is feasible for $t \geq 0$, and the objective value goes to $-\infty$ as $t \to \infty$. (The inequality constraint corresponds to $x \in Q^3$.)

In this example, SDPT3 reports "Failed" and does not report an optimal value; SeDuMi reports "Unbounded" and $-\infty$ as the optimal value; MOSEK reports "Unbounded" and $-\infty$ as the optimal value.

Case (e) (P) is feasible, $p^\star = -\infty$, but there is no improving direction, i.e., there is no $u \in \mathcal{N}(A) \cap K$ satisfying $c^T u < 0$. For example, consider the problem

$$\begin{aligned} \text{minimize } & x_1 \\ \text{subject to } & x_2 = 1 \\ & 2x_2x_3 \geq x_1^2 \\ & x_2, x_3 \geq 0. \end{aligned}$$

(The inequality constraints correspond to $x \in Q_r^3$.) Any improving direction $u = (u_1, u_2, u_3)$ would satisfy $u_2 = 0$, and this in turn, with the cone constraint, implies $u_1 = 0$ and $c^T u = 0$. However, even though there is no improving direction, we can eliminate the variables x_1 and x_2 to verify that

$$p^\star = \inf \left\{ -\sqrt{2x_3} \mid x_3 \geq 0 \right\} = -\infty.$$

In this example, SDPT3 reports "Failed" and does not report an optimal value; SeDuMi reports "Inaccurate/Solved" and -175514 as the optimal value; MOSEK reports "Inaccurate/Unbounded" and $-\infty$ as the optimal value.

Case (f) Strongly infeasible, where $p^\star = \infty$ and $d(K, \{x \mid Ax = b\}) > 0$. For example, the problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & x_3 = -1 \\ & x_3 \geq \sqrt{x_1^2 + x_2^2} \end{array}$$

satisfies $d(K, \{x \mid Ax = b\}) = 1$. (The inequality constraint corresponds to $x \in Q^3$.)

In this example, SDPT3 reports "Failed" and does not report an optimal value; SeDuMi reports "Infeasible" and ∞ as the optimal value; MOSEK reports "Infeasible" and ∞ as the optimal value.

Case (g) Weakly infeasible, where $p^\star = \infty$ but $d(K, \{x \mid Ax = b\}) = 0$. For example, the problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \begin{bmatrix} 0, & 1, & 1 \\ 1, & 0, & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ & x_3 \geq \sqrt{x_1^2 + x_2^2} \end{array}$$

satisfies $d(K, \{x \mid Ax = b\}) = 0$, since

$$d(K, \{x \mid Ax = b\}) \leq \|(1, -y, y) - (1, -y, \sqrt{y^2 + 1})\| \to 0$$

as $y \to \infty$. (The inequality constraint corresponds to $x \in Q^3$.)

In this example, SDPT3 reports "Infeasible" and ∞ as the optimal value; SeDuMi reports "Solved" and 0 as the optimal value; MOSEK reports "Failed" and does not report an optimal value.

Remark In the case of linear programming, i.e., when K in (P) is the positive orthant, there are only three possible cases: (a), (d), and (f).

1.3 Classification method overview

At a high level, our proposed method for classifying the 7 cases is quite simple. Given an operator T and a starting point z^0, we call $z^{k+1} = T\left(z^k\right)$ the *fixed-point iteration* of T. Our proposed method runs three similar but distinct fixed-point iterations with the operators

$$\begin{aligned} T_1(z) &= \tilde{T}(z) + x_0 - \gamma Dc \\ T_2(z) &= \tilde{T}(z) + x_0 \\ T_3(z) &= \tilde{T}(z) - \gamma Dc, \end{aligned} \qquad \text{(Operators)}$$

where $\tilde{T}(z) = (1/2)(I + R_{\mathcal{N}(A)} R_K)(z)$, $D = I - A^T(AA^T)^{-1}A$, $x_0 = A^T(AA^T)^{-1}b$, and $\gamma > 0$. We explain the notation in more detail in Sect. 2.

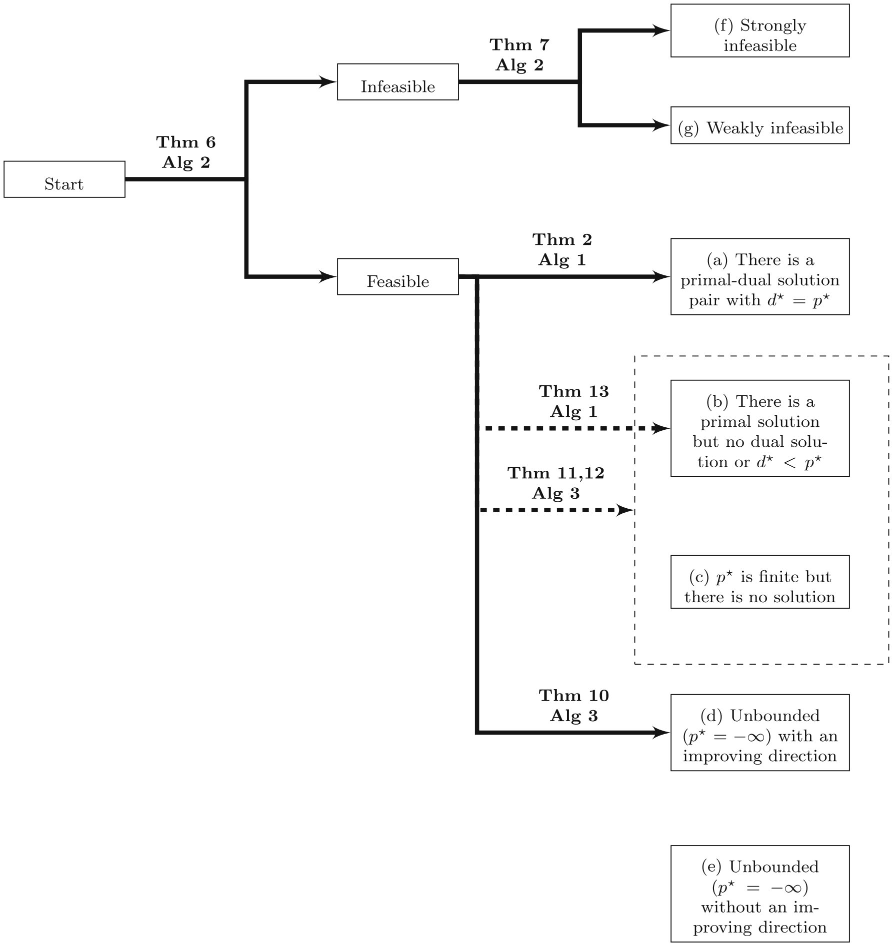

Fig. 1 The flowchart for identifying cases (**a**–**g**). A solid arrow means the cases are always identifiable, a dashed arrow means the cases sometimes identifiable

We can view T_1 as the DRS operator of (P), T_2 as the DRS operator with c set to $\mathbf{0}$ in (P), and T_3 as the DRS operator with b set to $\mathbf{0}$ in (P). We use the information provided by the iterates of these fixed-point iterations to solve (P) and classify the cases. As outlined in Sect. 2.8, this is based on the theory of Sect. 2 and the flowchart shown in Fig. 1.

1.4 Previous work

Previously, Bauschke, Combettes, Hare, Luke, and Moursi have analyzed Douglas–Rachford splitting in other pathological problems such as: feasibility problems between 2 affine sets [7], feasibility problems between 2 convex sets [4,8], and general setups [2,5,6,24]. Our work builds on these past results.

2 Obtaining certificates from Douglas–Rachford Splitting

The primal problem (P) is equivalent to

$$\text{minimize} \;\; f(x) + g(x), \tag{1}$$

where

$$\begin{aligned} f(x) &= c^T x + \delta_{\{x \mid Ax=b\}}(x) \\ g(x) &= \delta_K(x), \end{aligned} \tag{2}$$

and $\delta_C(x)$ is the indicator function of a set C defined as

$$\delta_C(x) = \begin{cases} 0 & \text{if } \; x \in C \\ \infty & \text{if } \; x \notin C. \end{cases}$$

Douglas–Rachford splitting (DRS) [15] applied to (1) is

$$\begin{aligned} x^{k+1/2} &= \mathrm{Prox}_{\gamma g}\left(z^k\right) \\ x^{k+1} &= \mathrm{Prox}_{\gamma f}\left(2x^{k+1/2} - z^k\right) \\ z^{k+1} &= z^k + x^{k+1} - x^{k+1/2}, \end{aligned} \tag{3}$$

which updates z^k to z^{k+1} for $k = 0, 1, \ldots$. Given $\gamma > 0$ and function h,

$$\mathrm{Prox}_{\gamma h}(x) = \mathrm{argmin}_{z\in\mathbb{R}^n} \left\{ h(z) + (1/2\gamma)\|z - x\|^2 \right\}$$

denotes the proximal operator with respect to γh.

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

$$P_C(x) = \mathrm{argmin}_{y\in C} \|y - x\|^2$$

and the reflection with respect to C as

$$R_C(x) = 2P_C(x) - x.$$

Write I to denote both the $n \times n$ identity matrix and the identity map from $\mathbb{R}^n \to \mathbb{R}^n$. Write $\mathbf{0}$ to denote the origin point in $\mathbb{R}^n$. Define

$$\begin{aligned} D &= I - A^T(AA^T)^{-1}A \\ x_0 &= A^T(AA^T)^{-1}b = P_{\{x \mid Ax=b\}}(\mathbf{0}). \end{aligned} \tag{4}$$

Write $\mathcal{N}(A)$ for the null space of A and $\mathcal{R}(A^T)$ for the range of A^T. Then

$$\begin{aligned} P_{\{x \mid Ax=b\}}(x) &= Dx + x_0, \\ P_{\mathcal{N}(A)}(x) &= Dx. \end{aligned}$$

Finally, define

$$\tilde{T}(z) = \frac{1}{2}(I + R_{\mathcal{N}(A)} R_K)(z).$$

Now we can rewrite the DRS iteration (3) as

$$\begin{aligned} x^{k+1/2} &= P_K\left(z^k\right) \\ x^{k+1} &= D\left(2x^{k+1/2} - z^k\right) + x_0 - \gamma Dc \\ z^{k+1} &= z^k + x^{k+1} - x^{k+1/2}. \end{aligned} \tag{5}$$

Equivalently and more compactly, we can write

$$z^{k+1} = \tilde{T}(z^k) + x_0 - \gamma Dc, \tag{6}$$

which is also $z^{k+1} = T_1\left(z^k\right)$ with T_1 definied in (Operators).

Remark Instead of (2), we could have considered the more general form

$$\begin{aligned} f(x) &= (1-\alpha)c^T x + \delta_{\{x \mid Ax=b\}}(x), \\ g(x) &= \alpha c^T x + \delta_K(x) \end{aligned}$$

with $\alpha \in \mathbb{R}$. By simplifying the resulting DRS iteration, one can verify that the iterates are equivalent to the $\alpha = 0$ case. Since the choice of α does not affect the DRS iteration at all, we will only work with the case $\alpha = 0$.

2.1 Convergence of DRS

A point $x^\star \in \mathbb{R}^n$ is a solution of (1) if and only if

$$\mathbf{0} \in \partial(f+g)(x^\star).$$

DRS, however, converges if and only if there is a point $x^\star$ such that

$$\mathbf{0} \in \partial f(x^\star) + \partial g(x^\star).$$

In general,

$$\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x)$$

for all $x \in \mathbb{R}^n$, but the two are not necessarily equal.

We summarize the convergence of DRS in the theorem below. Its main part is a direct result of Theorem 1 of [32] and Propositions 4.4 and 4.8 of [13]. The convergence of $x^{k+1/2}$ and x^{k+1} is due to [34]. Therefore, we do not prove it.

Theorem 1 *Consider the iteration* (6) *with any starting point* z^0. *If there is an* x *such that*

$$\mathbf{0} \in \partial f(x) + \partial g(x),$$

then z^k *converges to a limit* $z^\star$, $x^{k+1/2} \to x^\star = \mathrm{Prox}_{\gamma g}(z^\star)$, $x^{k+1} \to x^\star = \mathrm{Prox}_{\gamma g}(z^\star)$, *and*

$$\mathbf{0} \in \partial f(x^\star) + \partial g(x^\star).$$

If there is no x *such that*

$$\mathbf{0} \in \partial f(x) + \partial g(x),$$

then z^k *diverges in that* $\|z^k\| \to \infty$.

DRS can fail to find a solution to (P) even when one exists. Slater's constraint qualification is a sufficient condition that prevents such pathologies: if (P) is strongly feasible, then

$$\mathbf{0} \in \partial f(x^\star) + \partial g(x^\star)$$

for all solutions $x^\star$ [30, Theorem 23.8]. This fact and Theorem 1 tell us that under Slater's constraint qualifications DRS finds a solution of (P) if one exists.

The following theorem, however, provides a stronger, necessary and sufficient characterization of when the DRS iteration converges.

Theorem 2 ([31]) *There is an* $x^\star$ *such that*

$$\mathbf{0} \in \partial f(x^\star) + \partial g(x^\star)$$

if and only if $x^\star$ *is a solution to* (P), (D) *has a solution, and* $d^\star = p^\star$.

Based on Theorems 1 and 2 we can determine whether we have case (a) with the iteration (6)

with any starting point z^0 and $\gamma > 0$.

- If $\lim_{k\to\infty} \|z^k\| < \infty$, we have case (a), and vice versa.
- If $\lim_{k\to\infty} \|z^k\| = \infty$, we do not have case (a), and vice versa.

With a finite number of iterations, we test $\|z^k\| \geq M$ for some large $M > 0$. However, distinguishing the two cases can be numerically difficult as the rate of $\|z^k\| \to \infty$ can be very slow.

2.2 Fixed-point iterations without fixed points

We say an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive if

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

for all $x, y \in \mathbb{R}^n$. We say T is firmly nonexpansive (FNE) if

$$\|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \|(I - T)(x) - (I - T)(y)\|^2$$

for all $x, y \in \mathbb{R}^n$. (FNE operators are nonexpansive.) In particular, all three operators defined in (Operators) are FNE, as they are DRS operators [3]. It is well known [11] that if a FNE operator T has a fixed point, its fixed-point iteration $z^{k+1} = T(z^k)$ converges to one with rate

$$\|z^{k+1} - z^k\| = o(1/\sqrt{k}).$$

Now consider the case where a FNE operator T has no fixed point, which has been studied to a lesser extent. In this case, the fixed-point iteration $z^{k+1} = T(z^k)$ diverges in that $\|z^k\| \to \infty$ [32, Theorem 1]. Precisely in what manner z^k diverges is characterized by the *infimal displacement vector* [27]. Given a FNE operator T, we call

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

the *infimal displacement vector* of T. To clarify, $\overline{\mathbf{ran}(I - T)}$ denotes the closure of the set

$$\mathbf{ran}(I - T) = \{x - T(x) \mid x \in \mathbb{R}^n\}.$$

Because T is nonexpansive, the closed set $\overline{\mathbf{ran}(I - T)}$ is convex [27], so v is uniquely defined. We can interpret the infimal displacement vector v as the asymptotic output of $I - T$ corresponding to the best effort to find a fixed point.

Lemma 1 (Corollary 2.3 of [1]) *Let T be FNE, and consider its fixed-point iteration $z^{k+1} = T(z^k)$ with any starting point z^0. Then*

$$z^k - z^{k+1} \to v = P_{\overline{\mathbf{ran}(I-T)}}(\mathbf{0}).$$

In [1], Lemma 1 is proved in generality for nonexpansive operators, but we provide a simpler proof in our setting in Theorem 3.

When T has a fixed point $v = \mathbf{0}$, but $v = \mathbf{0}$ is possible even when T has no fixed point. In the following sections, we use Lemma 1 to determine the status of a conic program, but, in general, $z^k - z^{k+1} \to v$ has no rate. However, we only need to determine whether $\lim_{k\to\infty}(z^{k+1} - z^k) = \mathbf{0}$ or $\lim_{k\to\infty}(z^{k+1} - z^k) \neq \mathbf{0}$, and we do so by checking whether $\|z^{k+1} - z^k\| \geq \varepsilon$ for some tolerance $\varepsilon > 0$. For this purpose, the following rate of approximate convergence is good enough.

Theorem 3 *Let T be FNE, and consider its fixed point iteration*

$$z^{k+1} = T(z^k),$$

with any starting point z^0, then

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

And for any $\varepsilon > 0$, there is an $M_\varepsilon > 0$ (which depends on T, z^0, and ε) such that

$$\|v\| \le \min_{0\le j\le k} \|z^j - z^{j+1}\| \le \|v\| + \frac{M_\varepsilon}{\sqrt{k+1}} + \frac{\varepsilon}{2}.$$

Proof (Proof of Theorem 3) For simplicity, we prove the result for $0 < \varepsilon \le 1$. The result for $\varepsilon = 1$ applies to the $\varepsilon > 1$ case.

Given any x_ε, we use the triangle inequality to get

$$\begin{aligned} \|z^k - z^{k+1} - v\| &= \|T^k(z^0) - T^{k+1}(z^0) - v\| && (7)\\ &\le \|(T^k(z^0) - T^{k+1}(z^0)) - (T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon))\| \\ &\quad + \|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon) - v\|. && (8)\end{aligned}$$

To bound the second term, pick an x_ε such that

$$\|x_\varepsilon - T(x_\varepsilon) - v\| \le \frac{\varepsilon^2}{4(2\|v\|+1)},$$

which we can do since $v = P_{\overline{\mathbf{ran}(I-T)}}(\mathbf{0}) \in \overline{\mathbf{ran}(I-T)}$. Since T is nonexpansive, we have

$$\|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\| - \|v\| \le \|x_\varepsilon - T(x_\varepsilon)\| - \|v\| \le \|x_\varepsilon - T(x_\varepsilon) - v\|.$$

Since $v = \operatorname{argmin}_{\overline{\mathbf{ran}(I-T)}} \|x\|$, we have $\|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\| - \|v\| \ge 0$. Putting this together we get

$$0 \le \|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\| - \|v\| \le \frac{\epsilon^2}{4(2\|v\|+1)}.$$

Since $v = P_{\overline{\mathbf{ran}(I-T)}}(\mathbf{0})$,

$$\|v\|^2 \le y^T v$$

for any $y \in \overline{\mathbf{ran}(I-T)}$. Putting these together we get

$$\begin{aligned}
\|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon) - v\|^2 &= \|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\|^2 + \|v\|^2 \\
&\quad - 2(T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon))^T v \\
&\leq \|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\|^2 + \|v\|^2 - 2\|v\|^2 \\
&= \big(\|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\| + \|v\|\big)\big(\|T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon)\| - \|v\|\big) \\
&\leq (2\|v\| + \frac{\varepsilon^2}{4(2\|v\|+1)})\frac{\varepsilon^2}{4(2\|v\|+1)} \\
&\leq (2\|v\|+1)\frac{\varepsilon^2}{4(2\|v\|+1)} = \frac{\varepsilon^2}{4}
\end{aligned} \tag{9}$$

for $0 < \varepsilon \leq 1$.

Now let us bound the first term $\|(T^k(z^0) - T^{k+1}(z^0)) - (T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon))\|$ on the righthand side of (8). Since T is FNE, we have

$$\begin{aligned}
&\|(T^k(z^0) - T^{k+1}(z^0)) - (T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon))\|^2 = \|T^k(z^0) - T^k(x_\varepsilon)\|^2 \\
&\quad - \|T^{k+1}(z^0) - T^{k+1}(x_\varepsilon)\|^2.
\end{aligned}$$

Summing this inequality we have

$$\sum_{j=0}^{k} \|(T^k(z^0) - T^{k+1}(z^0)) - (T^k(x_\varepsilon) - T^{k+1}(x_\varepsilon))\|^2 \leq \|z^0 - x_\varepsilon\|^2. \tag{10}$$

(8), (9), and (10) imply that

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

Furthermore,

$$\min_{0\leq j\leq k} \|z^j - z^{j+1} - v\| \leq \frac{M_\varepsilon}{\sqrt{k+1}} + \frac{\varepsilon}{2},$$

where $M_\varepsilon = \|z^0 - x_\varepsilon\|$. As a result,

$$\|v\| \leq \min_{0\leq j\leq k} \|z^j - z^{j+1}\| \leq \|v\| + \frac{M_\varepsilon}{\sqrt{k+1}} + \frac{\varepsilon}{2}.$$

□

2.3 Feasibility and infeasibility

We now return to conic programs. Consider the operator T_2 defined by $T_2(z) = \tilde{T}(z) + x_0$. As mentioned, we can view T_2 as the DRS operator with c set to $\mathbf{0}$ in (P).

The infimal displacement vector of T_2 has a nice geometric interpretation: it is the best approximation displacement between the sets K and $\{x \mid Ax = b\}$, and $\|v\| = d(K, \{x \mid Ax = b\})$. Define the set

$$K - \{x \mid Ax = b\} = \{y - x \mid y \in K,\ Ax = b\}.$$

Theorem 4 (Theorem 3.4 of [4], Proposition 11.22 of [24]) *The operator T_2 has the infimal displacement vector $v = P_{\overline{K-\{x \mid Ax=b\}}}(\mathbf{0})$.*

We can further understand v in terms of the projection $P_{\overline{P_{\mathcal{R}(A^T)}(K)}}$. Note that $P_{\mathcal{R}(A^T)}(K)$ is a cone because K is. $P_{\mathcal{R}(A^T)}(K)$ is not always closed, but its closure $\overline{P_{\mathcal{R}(A^T)}(K)}$ is. We prove the following result at the end of this subsection.

Lemma 2 (Interpretation of v) *The infimal displacement vector v of T_2 satisfies*

$$v = P_{\overline{K-\{x \mid Ax=b\}}}(\mathbf{0}) = P_{\overline{P_{\mathcal{R}(A^T)}(K)-x_0}}(\mathbf{0}) = P_{\overline{P_{\mathcal{R}(A^T)}(K)}}(x_0) - x_0,$$

where x_0 is given in (4) and K is any nonempty set.

Combining the discussion of Sect. 2.2 with Theorem 4 gives us Theorems 5 and 6.

Theorem 5 (Certificate of feasibility) *Consider the iteration $z^{k+1} = T_2(z^k)$ with any starting point $z^0 \in \mathbb{R}^n$, then*

1. *(P) is feasible if and only if z^k converges, and in this case $x^{k+1/2}$ converges to a feasible point of (P).*
2. *(P) is infeasible if and only if z^k diverges in that $\|z^k\| \to \infty$.*

Theorem 6 (Certificate of strong infeasibility) *Consider the iteration $z^{k+1} = T_2(z^k)$ with any starting point z^0. We have $z^k - z^{k+1} \to v$ and*

1. *(P) is strongly infeasible if and only if $v \neq \mathbf{0}$.*
2. *(P) is weakly infeasible or feasible if and only if $v = \mathbf{0}$.*

When (P) is strongly infeasible, we can obtain a separating hyperplane from v. We prove the following result at the end of this subsection.

Theorem 7 (Separating hyperplane) *Consider the iteration $z^{k+1} = T_2(z^k)$ with any starting point z^0. When (P) is strongly infeasible, $z^k - z^{k+1} \to v \neq \mathbf{0}$, and the hyperplane*

$$\{x \mid h^T x = \beta\},$$

where $h = -v \in K^* \cap \mathcal{R}(A^T)$ and $\beta = -(v^T x_0)/2 > 0$, strictly separates K and $\{x \mid Ax = b\}$. More precisely, for any $y_1 \in K$ and $y_2 \in \{x \mid Ax = b\}$ we have

$$h^T y_1 < \beta < h^T y_2.$$

Based on Theorems 5, 6, and 7, we can determine feasibility, weak infeasiblity, and strong infeasibility and obtain a strictly separating hyperplane if one exists with the iteration $z^{k+1} = T_2(z^k)$ with any starting point z^0.

- $\lim_{k\to\infty} \|z^k\| < \infty$ if and only if (P) is feasible.
- $\lim_{k\to\infty} \|z^k - z^{k+1}\| > 0$ if and only if (P) is strongly infeasible, and Theorem 7 provides a strictly separating hyperplane.
- $\lim_{k\to\infty} \|z^k\| = \infty$ and $\lim_{k\to\infty} \|z^k - z^{k+1}\| = 0$ if and only if (P) is weakly infeasible.

With a finite number of iterations, we distinguish the three cases by testing $\|z^{k+1} - z^k\| \leq \varepsilon$ and $\|z^k\| \geq M$ for some small $\varepsilon > 0$ and large $M > 0$. By Theorem 3, we can distinguish strong infeasibility from weak infeasibility or feasibility at a rate of $O(1/\sqrt{k})$. However, distinguishing feasibility from weak infeasibility can be numerically difficult as the rate of $\|z^k\| \to \infty$ can be very slow when (P) is weakly infeasible.

Proof (Proof of Lemma 2) Remember that by definition (4), we have $x_0 \in \mathcal{R}(A^T)$ and

$$\{x \mid Ax = b\} = x_0 + \mathcal{N}(A) = x_0 - \mathcal{N}(A).$$

Also note that for any $y \in \mathbb{R}^n$, we have

$$y + \mathcal{N}(A) = P_{\mathcal{R}(A^T)}(y) + \mathcal{N}(A).$$

So

$$K - \{x \mid Ax = b\} = K + \mathcal{N}(A) - x_0 = P_{\mathcal{R}(A^T)}(K) - x_0 + \mathcal{N}(A),$$

and

$$\overline{K - \{x \mid Ax = b\}} = \overline{P_{\mathcal{R}(A^T)}(K) + \mathcal{N}(A)} - x_0 = \overline{P_{\mathcal{R}(A^T)}(K)} - x_0 + \mathcal{N}(A). \quad (11)$$

Since $x_0 \in \mathcal{R}(A^T)$, we have $\overline{P_{\mathcal{R}(A^T)}(K)} - x_0 \subseteq \mathcal{R}(A^T)$, and, in particular, $\overline{P_{\mathcal{R}(A^T)}(K)} - x_0$ is orthogonal to the subspace $\mathcal{N}(A)$. Recall

$$v = P_{\overline{P_{\mathcal{R}(A^T)}(K)} - x_0 + \mathcal{N}(A)}(\mathbf{0}).$$

So $v \in \overline{P_{\mathcal{R}(A^T)}(K)} - x_0 \subseteq \mathcal{R}(A^T)$ and

$$v = P_{\overline{P_{\mathcal{R}(A^T)}(K)} - x_0}(\mathbf{0}).$$

Finally,

$$v = \operatorname{argmin}_{x \in \overline{P_{\mathcal{R}(A^T)}(K) - x_0}} \left\{ \|x\|_2^2 \right\} = \operatorname{argmin}_{y \in \overline{P_{\mathcal{R}(A^T)}(K)}} \left\{ \|y - x_0\|_2^2 \right\} - x_0$$
$$= P_{\overline{P_{\mathcal{R}(A^T)}(K)}}(x_0) - x_0$$

□

Proof (Proof of Theorem 7) Note that

$$v = P_{\overline{K - \{x \mid Ax = b\}}}(\mathbf{0}) = P_{\overline{K + \mathcal{N}(A) - x_0}}(\mathbf{0}) = P_{\overline{K + \mathcal{N}(A)}}(x_0) - x_0$$

Using $I = P_{K^* \cap \mathcal{R}(A^T)} + P_{-(K^* \cap \mathcal{R}(A^T))^*}$ and $(K^* \cap \mathcal{R}(A^T))^* = \overline{K + \mathcal{N}(A)}$ [3], we have

$$v = P_{\overline{K + \mathcal{N}(A)}}(x_0) - x_0 = -P_{-(K^* \cap \mathcal{R}(A^T))}(x_0) = P_{K^* \cap \mathcal{R}(A^T)}(-x_0).$$

Since the projection operator is FNE, we have

$$-v^T x_0 = (v - \mathbf{0})^T (-x_0 - \mathbf{0}) \geq \|P_{K^* \cap \mathcal{R}(A^T)}(-x_0)\|^2 = \|v\|^2 > 0$$

and therefore $v^T x_0 < 0$, $\beta = -v^T x_0/2 > 0$.

So for any $y_1 \in K$ and $y_2 \in \{x \mid Ax = b\}$, we have

$$h^T y_1 = -v^T y_1 \leq 0 < -(v^T x_0)/2 = \beta < -v^T x_0 = h^T y_2,$$

where we have used $h = -v = -P_{K^* \cap \mathcal{R}(A^T)}(-x_0) \in -K^*$ in the first inequality. □

2.4 Modifying affine constraints to achieve strong feasibility

Loosely speaking, strongly feasible problems are the good cases that are easier to solve, compared to weakly feasible or infeasible problems. Given a problem that is not strongly feasible, how to minimally modify the problem to achieve strong feasibility is often useful to know.

The limit $z^k - z^{k+1} \to v$ informs us of how to do this. When $d(K, \{x \mid Ax = b\}) = \|v\| > 0$, the constraint $K \cap \{x \mid A(x - y) = b\}$ is infeasible for any y such that $\|y\| < \|v\|$. In general, the constraint $K \cap \{x \mid A(x - v) = b\}$ can be feasible or weakly infeasible, but is not strongly feasible. The constraint $K \cap \{x \mid A(x - v - d) = b\}$ is strongly feasible for an arbitrarily small $d \in \mathbf{relint} K$. In other words, $K \cap \{x \mid A(x - v - d) = b\}$ achieves strong feasibility with the minimal modification (measured by the Euclidean norm $\|\cdot\|$) to the original constraint $K \cap \{x \mid Ax = b\}$.

Theorem 8 (Achieving strong feasibility) *Let $v = P_{\overline{K - \{x \mid Ax = b\}}}(\mathbf{0})$, and let d be any vector satisfying $d \in \mathbf{relint} K$. Then the constraint $K \cap \{x \mid A(x - v - d) = b\}$ is strongly feasible, i.e., there is an x such that $x \in \mathbf{relint} K \cap \{x \mid A(x - v - d) = b\}$.*

Proof (Proof of Theorem 8) By Lemma 2 we have

$$v + x_0 \in \overline{P_{\mathcal{R}(A^T)}(K)}. \tag{12}$$

Because $P_{\mathcal{R}(A^T)}$ is a linear transformation, by Lemma 3 below

$$P_{\mathcal{R}(A^T)}(\textbf{relint}\, K) = \textbf{relint}\, P_{\mathcal{R}(A^T)}(K).$$

Since $d \in \textbf{relint}\, K$,

$$P_{\mathcal{R}(A^T)}(d) \in P_{\mathcal{R}(A^T)}(\textbf{relint}\, K) = \textbf{relint}\, P_{\mathcal{R}(A^T)}(K). \tag{13}$$

Applying Lemma 4 to (12) and (13), we have

$$v + x_0 + P_{\mathcal{R}(A^T)}(d) \in \textbf{relint}\, P_{\mathcal{R}(A^T)}(K) = P_{\mathcal{R}(A^T)}(\textbf{relint}\, K).$$

Finally we have

$$\mathbf{0} \in P_{\mathcal{R}(A^T)}(\textbf{relint}\, K) - x_0 - v - d + \mathcal{N}(A) = \textbf{relint}\, K - \{x \mid A(x - v - d) = b\}.$$

□

Lemma 3 (Theorem 6.6 of [30]) *If $A(\cdot)$ is a linear transformation and C is a convex set, then $A(\textbf{relint}\, C) = \textbf{relint}\, A(C)$.*

Lemma 4 (Theorem 6.1 [30]) *Let K be a convex cone. If $x \in K$ and $y \in \textbf{relint}\, K$, then $x + y \in \textbf{relint}\, K$.*

2.5 Improving direction

(P) has an improving direction if and only if the dual problem (D) is strongly infeasible:

$$0 < d(0, K^\star + \mathcal{R}(A^T) - c) = d(\{(y,s) \mid A^T y + s = c\}, \{(y,s) \mid s \in K^*\}).$$

Theorem 9 (Certificate of improving direction) *Exactly one of the following is true:*

1. (P) *has an improving direction,* (D) *is strongly infeasible, and $P_{\mathcal{N}(A)\cap K}(-c) \neq \mathbf{0}$ is an improving direction.*
2. (P) *has no improving direction,* (D) *is feasible or weakly infeasible, and $P_{\mathcal{N}(A)\cap K}(-c) = \mathbf{0}$.*

Furthermore,

$$P_{\mathcal{N}(A)\cap K}(-c) = P_{\overline{K^* + \mathcal{R}(A^T) - c}}(\mathbf{0}).$$

Theorem 10 *Consider the iteration* $z^{k+1} = T_3\big(z^k\big) = \tilde{T}\big(z^k\big) - \gamma Dc$ *with any starting point* z^0 *and* $\gamma > 0$. *If* (P) *has an improving direction, then*

$$d = \lim_{k\to\infty} z^{k+1} - z^k = P_{\overline{K^*+\mathcal{R}(A^T)-c}}(\mathbf{0}) \neq \mathbf{0}$$

gives one. If (P) *has no improving direction, then*

$$\lim_{k\to\infty} z^{k+1} - z^k = \mathbf{0}.$$

Based on Theorems 9 and 10 we can determine whether there is an improving direction and find one if one exists with the iteration $z^{k+1} = \tilde{T}\big(z^k\big) - \gamma Dc$ with any starting point z^0 and $\gamma > 0$.

- $\lim_{k\to\infty} z^{k+1} - z^k = \mathbf{0}$ if and only if there is no improving direction.
- $\lim_{k\to\infty} z^{k+1} - z^k = d \neq \mathbf{0}$ if and only if d is an improving direction.

With a finite number of iterations, we test $\|z^{k+1} - z^k\| \le \varepsilon$ for some small $\varepsilon > 0$. By Theorem 3, we can distinguish whether there is an improving direction or not at a rate of $O(1/\sqrt{k})$.

We need the following theorem for Sect. 2.7, it is proved similarly to Theorem 5 below.

Theorem 11 *Consider the iteration*

$$z^{k+1} = \tilde{T}\big(z^k\big) - \gamma Dc$$

with any starting point z^0 *and* $\gamma > 0$. *If* (D) *is feasible, then* z^k *converges. If* (D) *is infeasible, then* z^k *diverges in that* $\|z^k\| \to \infty$.

Proof (Proof of Theorem 9) The qualitative aspect of this theorem (duality between existence of improving directions and strong infeasibility) is known [21]. To the best of our knowledge, the quantitative aspect of this theorem (the meaning and characterization of $P_{\mathcal{N}(A)\cap K}(-c)$) has not been explicitly addressed before. The following proof slightly extends the argument of [21] to show both the qualitative and the quantitative parts.

(P) has no improving direction if and only if

$$\{x \in \mathbb{R}^n | x \in \mathcal{N}(A) \cap K, c^T x < 0\} = \emptyset,$$

which is equivalent to $c^T x \ge 0$ for all $x \in \mathcal{N}(A) \cap K$. This is in turn equivalent to $c \in (\mathcal{N}(A) \cap K)^*$. So

$$-c = P_{-(\mathcal{N}(A)\cap K)^*}(-c).$$

if and only if there is no improving direction, which holds if and only if

$$\mathbf{0} = P_{\mathcal{N}(A)\cap K}(-c).$$

Assume there is an improving direction. Since the projection operator is firmly nonexpansive, we have

$$0 < \|P_{\mathcal{N}(A)\cap K}(-c)\|^2 \leq (P_{\mathcal{N}(A)\cap K}(-c))^T(-c).$$

This simplifies to

$$(P_{\mathcal{N}(A)\cap K}(-c))^T c < 0,$$

and we conclude $P_{\mathcal{N}(A)\cap K}(-c)$ is an improving direction.

Using the fact that $(\mathcal{N}(A)\cap K)^* = \overline{K^* + \mathcal{R}(A^T)}$, we have

$$P_{\mathcal{N}(A)\cap K}(-c) = -P_{\mathcal{N}(A)\cap K}(c) = (P_{\overline{K^*+\mathcal{R}(A^T)}} - I)(c) = P_{\overline{K^*+\mathcal{R}(A^T)-c}}(\mathbf{0}),$$

where we have used the identity $I = P_{\mathcal{N}(A)\cap K} + P_{\overline{K^*+\mathcal{R}(A^T)}}$ in the second equality. □

Proof (Proof of Theorems 10 and 11) Using the identities $I = P_{\mathcal{N}(A)} + P_{\mathcal{R}(A^T)}$, $I = P_K + P_{-K^*}$, and $R_{\mathcal{R}(A^T)-\gamma c}(z) = R_{\mathcal{R}(A^T)}(z) - 2\gamma Dc$, we have

$$T_3(z) = \tilde{T}(z) - \gamma Dc = \frac{1}{2}(I + R_{\mathcal{R}(A^T)-\gamma c}R_{-K^*})(z).$$

In other words, we can interpret the fixed point iteration

$$z^{k+1} = \tilde{T}\left(z^k\right) - \gamma Dc$$

as the DRS iteration on

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & x \in \mathcal{R}(A^T) - \gamma c \\ & x \in -K^*. \end{array}$$

This proves Theorem 11.

Using Lemma 1, applying Theorem 3.4 of [4] as we did for Theorem 4, and applying Theorem 9, we get

$$\begin{aligned} z^k - z^{k+1} &\to P_{\overline{\mathbf{ran}(I-T_3)}}(\mathbf{0}) \\ &= P_{\overline{-K^*-\mathcal{R}(A^T)+\gamma c}}(\mathbf{0}) \\ &= -\gamma P_{\overline{K^*+\mathcal{R}(A^T)-c}}(\mathbf{0}) \\ &= -\gamma P_{\mathcal{N}(A)\cap K}(-c). \end{aligned}$$

□

2.6 Modifying the objective to achieve finite optimal value

Similar to Theorem 8, we can achieve strong feasibility of (D) by modifying c, and (P) will have a finite optimal value.

Theorem 12 (Achieving finite $p^\star$) *Let $w = P_{\overline{K^* + \mathcal{R}(A^T) - c}}(\mathbf{0})$, and let s be any vector satisfying $s \in \mathbf{relint} K^*$. If (P) is feasible and has an unbounded direction, then by replacing c with $c' = c + w + s$, (P) will have a finite optimal value.*

Proof (Proof of Theorem 12) Similar to Lemma 2, we have

$$w = P_{\overline{P_{\mathcal{N}(A)}(K^*) - P_{\mathcal{N}(A)(c)}}}(\mathbf{0}).$$

And similar to Theorem 8, the new constraint of (D)

$$K^* \cap \{c + w + s - A^T y\}$$

is strongly feasible. The constraint of (P) is still $K \cap \{x \mid Ax = b\}$, which is feasible. By weak duality of we conclude that the optimal value of (P) becomes finite. □

2.7 Other cases

So far, we have discussed how to identify and certify cases (a), (d), (f), and (g). We now discuss sufficient conditions to certify the remaining cases.

The following theorem follows from weak duality.

Theorem 13 ([31] Certificate of finite $p^\star$) *If (P) and (D) are feasible, then $p^\star$ is finite.*

Based on Theorem 11, we can determine whether (D) is feasible with the iteration $z^{k+1} = T_3(z^k) = \tilde{T}(z^k) - \gamma Dc$,

with any starting point z^0 and $\gamma > 0$.

- $\lim_{k\to\infty} \|z^k\| < \infty$ if and only if (D) is feasible.
- $\lim_{k\to\infty} \|z^k\| = \infty$ if and only if (D) is infeasible.

With a finite number of iterations, we test $\|z^k\| \geq M$ for some large $M > 0$. However, distinguishing the two cases can be numerically difficult as the rate of $\|z^k\| \to \infty$ can be very slow.

Theorem 14 (Primal iterate convergence) *Consider the DRS iteration as defined in (5) with any starting point z^0. Assume (P) is feasible, if $x^{k+1/2} \to x^\infty$ and $x^{k+1} \to x^\infty$, then x^∞ is primal optimal, even if z^k doesn't converge.*

When running the fixed-point iteration with $T_1(z) = \tilde{T}(z) + x_0 - \gamma Dc$, if $\|z^k\| \to \infty$ but $x^{k+1/2} \to x^\infty$ and $x^{k+1} \to x^\infty$, then we have case (b), but the converse is not necessarily true.

Proof (Proof of Theorem 14) Define

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

as in (5) Define

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

It's simple to verify that

$$\begin{aligned} \tilde{\nabla} g\big(x^{k+1/2}\big) &\in \partial g\big(x^{k+1/2}\big) \\ \tilde{\nabla} f\big(x^{k+1}\big) &\in \partial f\big(x^{k+1}\big). \end{aligned}$$

Clearly,

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

We also have

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

Consider any $x \in K \cap \{x \mid Ax = b\}$. Then, by convexity of f and g,

$$\begin{aligned} g\big(x^{k+1/2}\big) - g(x) + f\big(x^{k+1}\big) - f(x) &\le \tilde{\nabla} g\big(x^{k+1/2}\big)^T\big(x^{k+1/2} - x\big) \\ &\quad + \tilde{\nabla} f\big(x^{k+1}\big)^T\big(x^{k+1} - x\big) \\ &= \Big(\tilde{\nabla} g\big(x^{k+1/2}\big) + \tilde{\nabla} f\big(x^{k+1}\big)\Big)^T\big(x^{k+1/2} - x\big) \\ &\quad + \tilde{\nabla} f\big(x^{k+1}\big)^T\big(x^{k+1} - x^{k+1/2}\big) \\ &= \big(x^{k+1} - x^{k+1/2}\big)^T\Big(\tilde{\nabla} f\big(x^{k+1}\big) - (1/\gamma)\big(x^{k+1/2} - x\big)\Big) \\ &= (1/\gamma)\big(x^{k+1} - x^{k+1/2}\big)^T\big(x - z^{k+1}\big) \end{aligned}$$

We take the liminf on both sides and use Lemma 5 below to get

$$g(x^{\infty}) + f(x^{\infty}) \le g(x) + f(x).$$

Since this holds for any $x \in K \cap \{x \mid Ax = b\}$, x^{∞} is optimal. □

Lemma 5 *Let $\Delta^1, \Delta^2, \ldots$ be a sequence in $\mathbb{R}^n$. Then*

$$\liminf_{k\to\infty} (\Delta^k)^T \sum_{i=1}^{k} (-\Delta^i) \le 0.$$

Proof Assume for contradiction that

$$\liminf_{k\to\infty} (\Delta^k)^T \sum_{i=1}^{k} (-\Delta^i) > 2\varepsilon$$

for some $\varepsilon > 0$. Since the initial part of the sequence is irrelevant, assume without loss of generality that

$$(\Delta^j)^T \sum_{i=1}^{j} \Delta^i < -\varepsilon$$

for $j = 1, 2, \ldots$, summing both sides gives us, for all $k = 1, 2, \ldots$

$$\sum_{j=1}^{k} (\Delta^j)^T \sum_{i=1}^{j} \Delta^i < -\varepsilon k.$$

Define

$$\mathbb{1}\{i \le j\} = \begin{cases} 1, & \text{if } i \le j, \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$\sum_{j=1}^{k} \sum_{i=1}^{k} (\Delta^j)^T \Delta^i \mathbb{1}\{i \le j\} < -\varepsilon k,$$

$$0 \le \frac{1}{2} \left\| \sum_{i=1}^{k} \Delta^i \right\|^2 + \frac{1}{2} \sum_{i=1}^{k} \left\| \Delta^i \right\|^2 < -\varepsilon k,$$

which is a contradiction. □

2.8 The algorithms

We now collect the discussed classification results as three algorithms. The full algorithm is simply running Algorithms 1, 2, and 3, and applying flowchart of Fig. 1. In theory, the algorithms work with any value of $\gamma > 0$, although the empirical performance can vary with γ.

The algorithms rely on detecting whether certain quantities converge to 0 or ∞. This can be numerically challenging in certain cases. However, certain pathologies are inherently challenging, and we observe through the examples of Sect. 3 that our method is competitive with other approaches.

Algorithm 1 Finding a solution

Parameters: $\gamma, M, \varepsilon, z^0$
for $k = 1, \ldots$ **do**
  $x^{k+1/2} = P_K(z^k)$
  $x^{k+1} = D(2x^{k+1/2} - z^k) + x_0 - \gamma Dc$
  $z^{k+1} = z^k + x^{k+1} - x^{k+1/2}$
end for
if $\|z^k\| < M$ **then**
  Case (a)
  $x^{k+1/2}$ and x^{k+1} solution
else if $x^{k+1/2} \to x^\infty$ and $x^{k+1} \to x^\infty$ **then**
  Case (b)
  $x^{k+1/2}$ and x^{k+1} solution
else
  Case (b), (c), (d), (e), (f), or (g).
end if

Algorithm 2 Feasibility test

Parameters: M, ε, z^0
for $k = 1, \ldots$ **do**
  $x^{k+1/2} = P_K(z^k)$
  $x^{k+1} = D(2x^{k+1/2} - z^k) + x_0$
  $z^{k+1} = z^k + x^{k+1} - x^{k+1/2}$
end for
if $\|z^k\| \geq M$ and $\|z^{k+1} - z^k\| > \varepsilon$ **then**
  Case (f)
  Strictly separating hyperplane defined by $(z^{k+1} - z^k, ((z^{k+1} - z^k)^T x_0)/2)$
else if $\|z^k\| \geq M$ and $\|z^{k+1} - z^k\| \leq \varepsilon$ **then**
  Case (g)
else $\|z^k\| < M$
  Case (a), (b), (c), (d), or (e)
end if

2.9 Case-by-case illustration

In this section, we present a case-by-case illustration of the algorithms. We describe the empirical behavior of the algorithms on cases (b), (c), (d), and (e) and demonstrate how the classification works.

We skip the discussion of case (a), as it is the standard non-pathological case. Algorithm 1 determines whether or not we have case (a). Case (f) and (g) are the

Algorithm 3 Boundedness test

Prerequisite: (P) is feasible.
Parameters: γ, M, ε, z^0
for $k = 1, \dots$ **do**
  $x^{k+1/2} = P_K(z^k)$
  $x^{k+1} = D(2x^{k+1/2} - z^k) - \gamma Dc$
  $z^{k+1} = z^k + x^{k+1} - x^{k+1/2}$
end for
if $\|z^k\| \geq M$ and $\|z^{k+1} - z^k\| \geq \varepsilon$ **then**
  Case (d)
  Improving direction $z^{k+1} - z^k$
else if $\|z^k\| < M$ **then**
  Case (a), (b), or (c)
else
  Case (a), (b), (c), or (e)
end if

infeasible cases, and Algorithm 2 determines whether or not we have case (f) or (g). We skip the discussion of these cases, as we present more thorough experiments of them in Sect. 3.

Case (b), (P) *has a solution but* (D) *has no solution.* Consider the example problem of this case discussed in Sect. 1.2. When we run Algorithm 1, we empirically observe that $\|z^k\| \to \infty$ and $x^{k+1/2}, x^{k+1} \to x^\star$, for $\gamma = 0.1$. This tells us we have case (b).

Case (b), $-\infty < d^\star < p^\star < \infty$. Consider the example problem of this case discussed in Sect. 1.2. When we run Algorithm 1, we empirically observe that $\|z^k\| \to \infty$, $x^{k+1/2}$ and x^{k+1} do not converge, and $\lim_{k\to\infty} 2x_{12}^{k+1} = -0.2$ for $\gamma = 0.1$. When we run Algorithm 2, we empirically observe that z^k converges to a limit. When we run Algorithm 3, we empirically observe that z^k converges to a limit. From this, we can conclude we have case (b) or (c).

Case (b), $-\infty = d^\star < p^\star < \infty$ Consider the problem

$$\begin{array}{ll} \text{minimize} & x_1 \\ \text{subject to} & x_2 - x_3 = 0 \\ & x_3 \geq \sqrt{x_1^2 + x_2^2}, \end{array}$$

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

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & y \geq \sqrt{y^2 + 1}, \end{array}$$

which is infeasible. This immediately tells us that $p^\star > -\infty$ is possible even when $d^\star = -\infty$.

We can in fact analyze this example analytically. When we run Algorithm 1 with starting point $z^0 = (z_1^0, z_2^0, 0)$, the iterates $z^{k+1} = \left(z_1^{k+1}, z_2^{k+2}, z_3^{k+1}\right)$ are:

$$\begin{aligned} z_1^{k+1} &= \frac{1}{2} z_1^k - \gamma \\ z_2^{k+1} &= \frac{1}{2} z_2^k + \frac{1}{2}\sqrt{\left(z_1^k\right)^2 + \left(z_2^k\right)^2} \\ z_3^{k+1} &= 0. \end{aligned}$$

So $\|z^k\| \to \infty$. Furthermore, $x^{k+1/2} = P_K\left(z^k\right)$ satisfies $x_1^{k+1/2} \to -2\gamma$, $x_2^{k+1/2} \to \infty$ and $x_3^{k+1/2} \to \infty$, so $x^{k+1/2}$ does not converge to the solution set. When we run Algorithm 2, z^k converges to a limit. When we run Algorithm 3, $\|z^k\| \to \infty$ and $z^{k+1} - z^k \to \mathbf{0}$. From such observations, we could conclude we have case (b), (c), or (e).

This example demonstrates that the converses of Theorems 13 and 14 are not true.

Case (c) In this case, $|p^\star| < \infty$ but there is no solution. Consider the example problem of this case discussed in Sect. 1.2. When we run Algorithm 1, we empirically observe that $\|z^k\| \to \infty$, $x^{k+1/2}$ and x^{k+1} do not converge, and $\lim_{k\to\infty} 2x_3^{k+1} = p^\star$ for $\gamma = 0.1$. When we run Algorithm 2, we empirically observe that z^k converges to a limit. When we run Algorithm 3, we empirically observe that z^k converges to a limit. From this, we can conclude we have case (b) or (c).

Case (d) In this case, there is an improving direction. Consider the example problem of this case discussed in Sect. 1.2. When we run Algorithm 1, we empirically observe that $\|z^k\| \to \infty$ and $x^{k+1/2}$, x^{k+1} do not converge for $\gamma = 0.1$. When we run Algorithm 2, we empirically observe that z^k converges to a limit. When we run Algorithm 3, we empirically observe that $\|z^k\| \to \infty$ and $\lim_{k\to\infty} \|z^{k+1} - z^k\| > 0$. From this, we can conclude we have case (d).

Case (e) In this case, $p^\star = -\infty$, but there is no improving direction. Consider the example problem of this case discussed in Sect. 1.2. When we run Algorithm 1, we empirically observe that $\|z^k\| \to \infty$ and $x^{k+1/2}$, x^{k+1} do not converge for $\gamma = 0.1$. When we run Algorithm 2, we empirically observe that z^k converges to a limit. When we run Algorithm 3, we empirically observe that $\|z^k\| \to \infty$ and $z^{k+1} - z^k \to \mathbf{0}$. From this, we can conclude we have case (b), (c), or (e).

3 Numerical experiments

We test our algorithm on a library of weakly infeasible SDPs generated by [16]. These semidefinite programs are in the form:

$$\begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i,\ i = 1, \ldots, m \\ & X \in S_+^n, \end{array}$$

Table 1 Percentage of infeasibility detection in [16]

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
SeDuMi	0	0	1	0
SDPT3	0	0	0	0
MOSEK	0	0	11	0
PP+SeDuMi	100	0	100	0

Table 2 Percentage of infeasibility detection success

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
Proposed method	100	21	100	99

Table 3 Percentage of success determination that problems are not strongly infeasible

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
Proposed method	100	100	100	100

where $n = 10$, $m = 10$ or 20, and $A \bullet B = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij}$ denotes the inner product between two $n \times n$ matrices A and B.

The library provides "clean" and "messy" instances. Given a clean instance, a messy instance is created with

$$A_i \leftarrow U^T \left(\sum_{j=1}^{m} T_{ij} A_j \right) U \text{ for } i = 1, \ldots, m$$

$$b_i \leftarrow \sum_{j=1}^{m} T_{ij} b_j \text{ for } i = 1, \ldots, m,$$

where $T \in \mathbb{Z}^{m \times m}$ and $U \in \mathbb{Z}^{n \times n}$ are random invertible matrices with entries in $[-2, 2]$.

In [16], four solvers are tested, specifically, SeDuMi, SDPT3 and MOSEK from the YALMIP environment, and the preprocessing algorithm of Permenter and Parrilo [29] interfaced with SeDuMi. Table 1 reports the numbers of instances determined infeasible out of 100 weakly infeasible instances. The four solvers have varying success in detecting infeasibility of the clean instances, but none of them succeed in the messy instances.

Our proposed method performs better. However, it does require many iterations and does fail with some of the messy instances. We run the algorithm with $N = 10^7$ iterations and label an instance infeasible if $1/\|z^N\| \leq 8 \times 10^{-2}$ (cf. Theorems 5 and 6). Table 2 reports the numbers of instances determined infeasible out of 100 weakly infeasible instances. Curiously, our method and other existing methods perform better with the larger instances of $m = 20$. This behavior is also reported and discussed in [16], the paper that provides the library of pathological instances. We suspect this phenomenon is inherent to the data set, not our algorithm.

We would like to note that detecting whether or not a problem is strongly infeasible is easier than detecting whether a problem is infeasible. With $N = 5 \times 10^4$ and a tolerance of $\|z^N - z^{N+1}\| < 10^{-3}$ (c.f Theorem 6) our proposed method correctly determined that all test instances are not strongly infeasible. Table 3 reports the numbers of instances determined not strongly infeasible out of 100 weakly infeasible instances.

Acknowledgements W. Yin would like to thank Professor Yinyu Ye for his question regarding ADMM applied to infeasible linear programs during the 2014 Workshop on Optimization for Modern Computation held at Peking University.

References

1. Baillon, J.B., Bruck, R.E., Reich, S.: On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. Houst. J. Math. **4**(1), 1–9 (1978)
2. Bauschke, H.H., Boţ, R.I., Hare, W.L., Moursi, W.M.: Attouch–Théra duality revisited: paramonotonicity and operator splitting. J. Approx. Theory **164**(8), 1065–1084 (2012)
3. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics. Springer, New York (2011)
4. Bauschke, H.H., Combettes, P.L., Luke, D.: Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. J. Approx. Theory **127**(2), 178–192 (2004)
5. Bauschke, H.H., Hare, W.L., Moursi, W.M.: Generalized solutions for the sum of two maximally monotone operators. SIAM J. Control Optim. **52**(2), 1034–1047 (2014)
6. Bauschke, H.H., Hare, W.L., Moursi, W.M.: On the range of the Douglas–Rachford operator. Math. Oper. Res. **41**(3), 884–897 (2016)
7. Bauschke, H.H., Moursi, W.M.: The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces. SIAM J. Optim. **26**(2), 968–985 (2016)
8. Bauschke, H.H., Moursi, W.M.: On the Douglas–Rachford algorithm. Math. Program. **164**(1), 263–284 (2017)
9. Borwein, J., Wolkowicz, H.: Regularizing the abstract convex program. J. Math. Anal. Appl. **83**(2), 495–530 (1981)
10. Borwein, J.M., Wolkowicz, H.: Facial reduction for a cone-convex programming problem. J. Aust. Math. Soc. **30**(3), 369–380 (1981)
11. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: Glowinski, R., Osher, S., Yin, W. (eds.) Splitting Methods in Communication, Imaging, Science and Engineering, vol. 4. Springer, Berlin (2016)
12. Drusvyatskiy, D., Wolkowicz, H.: The many faces of degeneracy in conic optimization. Found. Trends Optim. **3**(2), 77–170 (2017)
13. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, MIT (1989)
14. de Klerk, E., Terlaky, T., Roos, K.: Self-dual embeddings. In: Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) Handbook of Semidefinite Programming: Theory, Algorithms, and Applications, pp. 111–138. Springer, New York (2000)
15. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
16. Liu, M., Pataki, G.: Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. Math. Program. **167**(2), 435–480 (2017)
17. Loera, J.A.D., Malkin, P.N., Parrilo, P.A.: Computation with polynomial equations and inequalities arising in combinatorial optimization. In: Lee, J., Leyffer, S. (eds.) Mixed Integer Nonlinear Programming, pp. 447–481. Springer, New York (2012)

18. Lofberg, J.: Pre- and post-processing sum-of-squares programs in practice. IEEE Trans. Autom. Control **54**(5), 1007–1011 (2009)
19. Lourenço, B.F., Muramatsu, M., Tsuchiya, T.: Solving SDP completely with an interior point oracle. arXiv:1507.08065 [math] (2015)
20. Luenberger, D.G., Ye, Y.: Linear and Nonlinear Programming, 4th edn. Springer (2016)
21. Luo, Z.Q., Sturm, J.F., Zhang, S.: Duality results for conic convex programming. Econometric Institute, Erasmus University Rotterdam, The Netherlands, Technical Report 9719/A (1997)
22. Luo, Z.Q., Sturm, J.F., Zhang, S.: Conic convex programming and self-dual embedding. Optim. Methods Softw. **14**(3), 169–218 (2000)
23. Mosek, A.: The MOSEK optimization toolbox for matlab manual. Version 7.1 (Revision 28) (2015)
24. Moursi, W.M.: The Douglas–Rachford operator in the possibly inconsistent case: static properties and dynamic behaviour. Ph.D. thesis, University of British Columbia (2017)
25. Nesterov, Y., Todd, M.J., Ye, Y.: Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. Math. Program. **84**(2), 227–267 (1999)
26. Pataki, G.: A simple derivation of a facial reduction algorithm and extended dual systems. Columbia University, Technical report (2000)
27. Pazy, A.: Asymptotic behavior of contractions in Hilbert space. Israel J. Math. **9**(2), 235–240 (1971)
28. Permenter, F., Friberg, H.A., Andersen, E.D.: Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach. SIAM J. Optim. **27**(3), 1257–1282 (2017)
29. Permenter, F., Parrilo, P.: Partial facial reduction: simplified, equivalent SDPs via approximations of the PSD cone. arXiv:1408.4685 (2014)
30. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
31. Rockafellar, R.T.: Conjugate Duality and Optimization. Society for Industrial and Applied Mathematics, Philadelphia (1974)
32. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976)
33. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim. Methods Softw. **11**(1–4), 625–653 (1999)
34. Svaiter, B.F.: On weak convergence of the Douglas–Rachford method. SIAM J. Control Optim. **49**(1), 280–287 (2011)
35. Tütüncü, R.H., Toh, K.C., Todd, M.J.: Solving semidefinite-quadratic-linear programs using SDPT3. Math. Program. **95**(2), 189–217 (2003)
36. Waki, H.: How to generate weakly infeasible semidefinite programs via Lasserre's relaxations for polynomial optimization. Optim. Lett. **6**(8), 1883–1896 (2012)
37. Waki, H., Muramatsu, M.: Facial reduction algorithms for conic optimization problems. J. Optim. Theory Appl. **158**(1), 188–215 (2013)
38. Waki, H., Nakata, M., Muramatsu, M.: Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization. Comput. Optim. Appl. **53**(3), 823–844 (2012)
39. Ye, Y., Todd, M.J., Mizuno, S.: An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. Math. Oper. Res. **19**(1), 53–67 (1994)
40. Yoshise, A.: Complementarity problems over symmetric cones: a survey of recent developments in several aspects. In: Anjos, M.F., Lasserre, J.B. (eds.) Handbook on Semidefinite, Conic and Polynomial Optimization, no. 166 in International Series in Operations Research and Management Science, pp. 339–375. Springer, New York (2012)