On Convergence of Heuristics Based on Douglas-Rachford Splitting and ADMM to Minimize Convex Functions over Nonconvex Sets

Shuvomoy Das Gupta

Abstract—**Recently, heuristics based on the Douglas-Rachford splitting algorithm and the alternating direction method of multipliers (ADMM) have found empirical success in minimizing convex functions over nonconvex sets, but not much has been done to improve the theoretical understanding of them. In this paper, we investigate convergence of these heuristics. First, we characterize optimal solutions of minimization problems involving convex cost functions over nonconvex constraint sets. We show that these optimal solutions are related to the fixed point set of the underlying nonconvex Douglas-Rachford operator. Next, we establish sufficient conditions under which the Douglas-Rachford splitting heuristic either converges to a point or its cluster points form a nonempty compact connected set. In the case where the heuristic converges to a point, we establish sufficient conditions for that point to be an optimal solution. Then, we discuss how the ADMM heuristic can be constructed from the Douglas-Rachford splitting algorithm. We show that, unlike in the convex case, the algorithms in our nonconvex setup are not equivalent to each other and have a rather involved relationship between them. Finally, we comment on convergence of the ADMM heuristic and compare it with the Douglas-Rachford splitting heuristic.**

Index Terms—**Alternating direction method of multipliers (ADMM), Douglas-Rachford splitting, optimization algorithms, nonconvex optimization problems.**

I. Introduction

In this paper, we study convergence of heuristics based on the Douglas-Rachford splitting algorithm and the alternating direction method of multipliers (ADMM) for minimization of convex functions over nonconvex sets. Such optimization problems can be described as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C}, \end{array} \tag{OPT}$$

where $x \in \mathbf{R}^n$ is the decision variable. The constraint set $\mathcal{C} \subseteq \mathbf{R}^n$ is nonempty and compact (closed and bounded), but it is not necessarily convex. The cost function $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ is CPC. This means that f is (i) convex, (ii) proper, *i.e.,* its domain $\mathbf{dom}\, f = \{x \in \mathbf{R}^n \mid f(x) < \infty\}$ is nonempty, and (iii) closed (or lower-semicontinuous), *i.e.,* its epigraph $\mathbf{epi}\, f = \{(x, \xi) \in \mathbf{R}^n \times \mathbf{R} \mid f(x) \leq \xi\}$ is a closed set. The constraint set $\mathcal{C}$ is assumed to be closed so that projection onto $\mathcal{C}$ is well-defined, and it is assumed to be bounded to avoid the possibility of an unbounded optimal solution.

S. Das Gupta is with the Research & Technology Department, Thales Canada, 105 Moatfield Drive, Toronto, ON, Canada. Email: shuvomoy.dasgupta@thalesgroup.com

We consider the following **heuristic based on the Douglas-Rachford splitting algorithm** [1, §27.2] to solve (OPT):

$$\begin{aligned} x_{n+1} &= \mathbf{prox}_{\gamma f}(z_n) \\ y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(2x_{n+1} - z_n) \\ z_{n+1} &= z_n + y_{n+1} - x_{n+1}, \end{aligned} \tag{NC-DRS}$$

where $n \in \mathbf{N}$ is an iteration counter, $\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x)$ is a **Euclidean projection** of x onto $\mathcal{C}$ (as $\mathcal{C}$ is not necessarily convex, there can be multiple projections onto it from a point outside $\mathcal{C}$), and

$$\mathbf{prox}_{\gamma f}(x) = \mathrm{argmin}_{y \in \mathbf{dom}\, f}\big(f(y) + \frac{1}{2\gamma}\|y - x\|^2\big)$$

is the **proximal operator** of f at x with parameter $\gamma > 0$. We also investigate the following **heuristic based on ADMM** (also known as NC-ADMM [2, §3.1]) to solve (OPT):

$$\begin{aligned} x_{n+1} &= \mathbf{prox}_{\gamma f}(y_n - z_n) \\ y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(x_{n+1} + z_n) \\ z_{n+1} &= z_n - y_{n+1} + x_{n+1}, \end{aligned} \tag{NC-ADMM}$$

where $n \in \mathbf{N}$ is an iteration counter, and $\gamma > 0$. Note that both heuristics consist of the same subroutines, but different inputs are fed into them. So, any software package that implements one of the heuristics can be easily modified to implement the other.

If the set $\mathcal{C}$ is convex, then the iterates x_n, y_n in both (NC-DRS) and (NC-ADMM) converge to an optimal solution for any initial point [1, Corollary 27.4], [3]. The convergence conditions for the nonconvex case, studied in this paper, are far more complicated.

Motivation. This paper is motivated by the recent success of ADMM in solving nonconvex problems. ADMM, which is a special case of the Douglas-Rachford splitting algorithm in a convex setup, was originally designed to solve convex optimization problems [3]. However, since the idea of implementing this algorithm as a general purpose heuristics to solve nonconvex optimization problems was introduced in [3], (NC-ADMM) has been applied successfully to minimization of convex functions over nonconvex sets [2], [4], [5], [6], and it has been implemented recently in the Python package NCVX—an extension of CVXPY—to formulate and solve problems of the form (OPT) [7]. In these works, the nonconvex projection step of (NC-ADMM), if computationally too costly, is replaced with a tractable "approximate" projection onto the nonconvex set, *e.g.,* rounding for Boolean variables; yet it finds approximate solutions to a wide variety of nonconvex problems effectively. In spite of the empirical success, not much has been done to improve the theoretical understanding of such heuristics. Some recent progress has been made on understanding convergence of ADMM for specialized nonconvex setups, such as (i) minimizing a nonconvex function over an affine set [8], and (ii) minimizing the sum of a smooth function with a bounded Hessian and a nonsmooth function with an easy to compute proximal mapping [9]. However, these works are not applicable to (NC-DRS) and (NC-ADMM), which has motivated us to investigate the convergence properties of these heuristics.

Contributions. Our contributions in this paper are as follows. First, we characterize global minimizers of (OPT) and show that they are related to the fixed point set of the underlying Douglas-Rachford operator constructed from (NC-DRS). Then, we establish conditions under which (NC-DRS) either converges to a point (not necessarily an optimal solution) or its cluster points form a nonempty compact connected set. In the case where the heuristic converges to a point, we provide sufficient conditions for that point to be an optimal solution. Then, we investigate the relationship between (NC-DRS) and (NC-ADMM). For a convex optimization problem, ADMM is the Douglas-Rachford algorithm splitting applied to the dual problem [10], but their relationship is more involved in our nonconvex setup. Applying the Douglas-Rachford splitting algorithm to the convex dual of (OPT) results in a relaxed version of (NC-ADMM), where the projection is onto the convex hull of $\mathcal{C}$. We show that this relaxed algorithm finds a minimizer of f over the convex hull of $\mathcal{C}$, and by restricting its projection step onto the original constraint set $\mathcal{C}$, we arrive at (NC-ADMM). The construction procedure also explains why, when compared with exact solvers, (NC-ADMM) often achieves lower objective values in many numerical experiments performed in [2], [4], [5], [6]. We comment on the convergence properties of (NC-ADMM) and compare it with (NC-DRS). To the best of our knowledge, we are not aware of similar results in the existing literature.

Notation and notions. We denote the sets of real numbers and natural numbers by $\mathbf{R}$ and $\mathbf{N}$, respectively. Furthermore, $\overline{\mathbf{R}} = \mathbf{R} \cup \{\infty\}$ denotes the extended real line. The set of real column vectors of length n is denoted by $\mathbf{R}^n$. Depending on the context, 0 may be a scalar or a column vector of zeros. The $n \times n$ identity matrix is denoted by I_n. The standard Euclidean norm is denoted by $\|\cdot\|$. We use $\langle \cdot \mid \cdot \rangle$ as the inner product in the Euclidean space. Let $\mathcal{X}, \mathcal{Y}$ be two nonempty subsets of $\mathbf{R}^n$, and let $z \in \mathbf{R}^n$. Then, $\mathcal{X} + \mathcal{Y} = \{x + y \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$, $\mathcal{X} - \mathcal{Y} = \{x - y \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$, $z + \mathcal{X} = \{z\} + \mathcal{X}$, and $\mathcal{X} - z = \mathcal{X} - \{z\}$. If one of the sets is empty, then the resultant addition or subtraction is an empty set, *i.e.,* $\mathcal{X} + \emptyset = \emptyset$. Finally, the **indicator function** of a nonempty set $\mathcal{X} \subseteq \mathbf{R}^n$, denoted by $\delta_{\mathcal{X}}$, is defined as

$$\delta_{\mathcal{X}}(x) = \begin{cases} 0, & \text{if } x \in \mathcal{X} \\ \infty, & \text{if } x \notin \mathcal{X}. \end{cases}$$

Using indicator function, (OPT) can be expressed as minimize $f(x) + \delta_{\mathcal{C}}(x)$.

II. Background on monotone operator theory

In this section, we present some definitions and preliminary results on monotone operator theory and relate them to our setup. First, in §II-A, we briefly review the essential operator theoretic notions and provide examples that relate these concepts to (OPT). In §II-B, we review nonexpansiveness and its several variants for an operator. These concepts are essential for characterizing (i) the operators $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ and $\mathbf{prox}_{\gamma f}$ and (ii) the fixed point sets of the underlying operators of (NC-DRS) and (NC-ADMM). Finally, in §II-C, we introduce resolvent and reflected resolvent of an operator to provide additional characterizations of $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ and $\mathbf{prox}_{\gamma f}$.

A. *Operator theoretic notions*

A **set-valued operator** $T : \mathbf{R}^n \rightrightarrows \mathbf{R}^n$ maps each element in $\mathbf{R}^n$ to a set in $\mathbf{R}^n$; its **domain** is defined as $\mathbf{dom}\, T = \{x \in \mathbf{R}^n \mid T(x) \neq \emptyset\}$, its **range** is defined as $\mathbf{ran}\, T = \bigcup_{x \in \mathbf{R}^n} T(x)$, and it is completely completely characterized by its **graph**: $\mathbf{gra}\, T = \{(x, u) \mid u \in Tx\}$. Furthermore, the **zero set** of T is defined as $\mathbf{zer}\, T = \{x \mid 0 \in A(x)\}$, and the **fixed point set** of T is defined as $\mathbf{fix}\, T = \{x \mid Tx \ni x\}$. The operator $2T - I_n$ is called the **reflection of** T. **Inverse** of T, denoted by T^{-1}, is defined through its graph: $\mathbf{gra}\, T^{-1} = \{(u, x) \mid (x, u) \in \mathbf{gra}\, T\}$, so $x \in T(u) \Leftrightarrow u \in T^{-1}(x)$. For every x, **addition of two operators** $T_1, T_2 : \mathbf{R}^n \rightrightarrows \mathbf{R}^n$, denoted by $T_1 + T_2$, is defined as $(T_1 + T_2)(x) = T_1(x) + T_2(x)$ (subtraction is defined analogously), and **composition** of these operators, denoted by $T_1 T_2$, is defined as $T_1 T_2(x) = T_1(T_2(x))$; note that order matters for composition. Also, if $\mathcal{X} \subseteq \mathbf{R}^n$ is a nonempty set, then $T(\mathcal{X}) = \bigcup_{x \in \mathcal{X}} T(x)$.

A set-valued operator $T : \mathbf{R}^n \rightrightarrows \mathbf{R}^n$ is **monotone** if, for every $(x, u), (y, v) \in \mathbf{gra}\, T$, it satisfies $\langle x - y \mid u - v \rangle \geq 0$. A monotone operator T is **maximally monotone** if $\mathbf{gra}\, T$ is not properly contained by the graph of any other monotone operator.

Finally, a **single-valued operator** $T : \mathcal{D} \to \mathbf{R}^n$ is a special type of set-valued operator, which maps every x in its domain $\mathcal{D} \subseteq \mathbf{R}^n$ to a singleton $T(x)$ in $\mathbf{R}^n$.

Example 1 (**projection operator**)**.** Recall that $\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x)$ is a Euclidean projection of x onto the constraint set $\mathcal{C}$. The operator $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ is single-valued. The set of all projections onto $\mathcal{C}$, denoted by $\mathbf{\Pi}_{\mathcal{C}}$, is the **set-valued projection operator** onto $\mathcal{C}$, and it is defined as $\mathbf{\Pi}_{\mathcal{C}}(x) = \mathrm{argmin}_{y \in \mathcal{C}} \|x - y\|^2$. Clearly, $\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x) \subseteq \mathbf{\Pi}_{\mathcal{C}}(x)$ for every x. Both $\mathbf{\Pi}_{\mathcal{C}}$ and $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ are monotone operators, but not necessarily maximally monotone [1, Example 20.12]. The projection operator onto a nonempty closed convex set, however, is maximally monotone [1, Example 20.12, Corollary 20.27, and Proposition 4.8].

Example 2 (**subdifferential operator**). For every proper function $g : \mathbf{R}^n \to \overline{\mathbf{R}}$, its subdifferential operator is the set-valued operator $\partial g : \mathbf{R}^n \rightrightarrows \mathbf{R}^n$, which is defined as

$$\partial g\,(x) = \{u \in \mathbf{R}^n \mid (\forall y \in \mathbf{R}^n)\ g\,(y) \geq g\,(x) + \langle u \mid y - x\rangle\}\,. \tag{1}$$

A vector $u \in \partial g\,(x)$ is called a **subgradient** of g at x. The subdifferential operator of a proper function is monotone, hence $\partial \delta_{\mathcal{C}}$ is monotone [1, Example 20.3]. On the other hand, the subdifferential operator of a CPC function is maximally monotone, thus ∂f is maximally monotone [1, Theorem 20.40]. The following result regarding the subdifferential operator plays a key role in characterizing global minimizers of (OPT) in §III.

Theorem 1 (**Fermat's rule** [1, page 223], [11, §2.3]). *The set of all global minimizers of a proper function* $g : \mathbf{R}^n \to \overline{\mathbf{R}}$*, denoted by* $\operatorname{argmin} g$*, is equal to the zero set of its subdifferential operator* ∂g*, i.e.,* $\operatorname{argmin} g = \mathbf{zer}\,\partial g = \{x \in \mathbf{R}^n \mid 0 \in \partial g(x)\}\,.$

Proof: Take $x \in \operatorname{argmin} g$ which is equivalent to the statement $(\forall y \in \mathbf{R}^n)\ \ g(y) \geq g(x) + \langle 0 \mid y - x\rangle \Leftrightarrow \partial g(x) \ni 0 \Leftrightarrow x \in \mathbf{zer}\,\partial g$. ■

While this simple characterization of optimality via the subdifferential holds for every nonconvex functions, it may not be particularly useful in practice if we cannot compute the subdifferential in an algorithmic manner [11, page 4].

We now present a lemma regarding the subdifferential operator of the sum of two proper functions, which is used later in §III. Recall that $\partial\,(g + h)\,(x) = \partial\,(g(x) + h(x))$ according to our notation.

Lemma 1 (**subdifferential of sum of proper functions**). *Let* $g : \mathbf{R}^n \to \overline{\mathbf{R}}$ *and* $h : \mathbf{R}^n \to \overline{\mathbf{R}}$ *be proper functions such that* $\mathbf{dom}\,g \cap \mathbf{dom}\,h \neq \emptyset$*. Then,*

(i) *the function* $g + h$ *is proper,*

(ii) *for every* x *in* $\mathbf{R}^n$*, we have* $\partial g\,(x) + \partial h\,(x) \subseteq \partial(g + h)(x)$*, and*

(iii) *both* $\partial(g + h)$ *and* $\partial g + \partial h$ *are monotone operators.*

Proof: (i): By definition, $\mathbf{dom}\,(g + h) = \{x \mid g(x) + h(x) < \infty\} = \mathbf{dom}\,g \cap \mathbf{dom}\,h \neq \emptyset$. Thus, $g + h$ is proper.

(ii): Take $x \in \mathbf{R}^n$, and denote $u \in \partial g\,(x)$ and $v \in \partial h\,(x)$. We want to prove that $u + v \in \partial\,(g + h)\,(x) = \partial\,(g(x) + h(x))$. Using (1), we have $g\,(y) \geq g\,(x) + \langle u \mid y - x\rangle$ and $h\,(y) \geq h\,(x) + \langle v \mid y - x\rangle$ for every $y \in \mathbf{R}^n$. Adding the last two inequalities we get $(g\,(y) + h\,(y)) \geq (g\,(x) + h\,(x)) + \langle u + v \mid y - x\rangle$ for every $y \in \mathbf{R}^n$, *i.e.,* $u + v \in \partial\,(g\,(x) + h\,(x))$.

(iii): Denote $\phi := g + h$, which is proper due to (i). Now take $(x, u), (y, v)$ in $\mathbf{gra}\,\partial\phi$, so we have $\phi\,(y) \geq \phi\,(x) + \langle u \mid y - x\rangle$ and $\phi(x) \geq \phi(y) + \langle v \mid x - y\rangle$ using (1); adding these inequalities we have $0 \geq \langle u \mid y - x\rangle + \langle v \mid x - y\rangle$ *i.e.,* $\langle u - v \mid x - y\rangle \geq 0$, so $\partial\phi = \partial\,(g + h)$ is a monotone operator by definition. Furthermore, both ∂g and ∂h are monotone, as the subdifferential operator of a proper function is monotone [1, Example 20.3]. Using also the fact that sum of two monotone operators is a monotone operator [1, page 351], we conclude that $\partial g + \partial h$ is monotone. ■

B. Nonexpansive and firmly nonexpansive operator

Let $T : \mathcal{D} \to \mathbf{R}^n$ be a single-valued operator, where $\mathcal{D} \subseteq \mathbf{R}^n$ is nonempty. Then, T is

1) **nonexpansive** on $\mathcal{D}$ if for every $x, y \in \mathcal{D}$ it satisfies $\|T\,(x) - T\,(y)\| \leq \|x - y\|$, and
2) **firmly nonexpansive** on $\mathcal{D}$ if for every $x, y \in \mathcal{D}$ it satisfies $\|T\,(x) - T\,(y)\|^2 + \|(I_n - T)(x) - (I_n - T)(y)\|^2 \leq \|x - y\|^2$.

An operator $T : \mathcal{D} \to \mathbf{R}^n$ is firmly nonexpansive on $\mathcal{D}$ if and only if its reflection operator $2T - I_n$ is nonexpansive [1, Proposition 4.2]. Furthermore, a firmly nonexpansive operator is also nonexpansive [1, page 59].

Example 3 (**proximal operator**). The proximal operator of a CPC function is both firmly nonexpansive and nonexpansive [1, Proposition 12.27, Example 23.3], hence $\mathbf{prox}_{\gamma f}$ in (NC-ADMM) and (NC-DRS) is both firmly nonexpansive and nonexpansive. Furthermore, its reflection $2\,\mathbf{prox}_{\gamma f} - I_n$ is nonexpansive [1, Proposition 4.2].

Example 4 (**projection operator**). We remind the reader that, a set is called **proximinal** if every point has at least one projection onto it, whereas it is called a **Chebyshev** set if every point has exactly one projection onto it. A nonempty subset in $\mathbf{R}^n$ is Chebyshev if and only if it is closed and convex [1, Remark 3.15], and the projection operator onto such a set is single-valued and firmly nonexpansive on $\mathbf{R}^n$ [1, Proposition 4.8]. However, for the constraint set $\mathcal{C}$ in (OPT), which is possibly nonconvex, the projection operator $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ is not, in general, nonexpansive, hence not firmly nonexpansive. For example, consider the set $\{0, 1\}$; the projections of 0.4 and 0.6 onto this set are 0 and 1, respectively, so $|0.6 - 0.4| = 0.2 < 1$, which violates the definition of nonexpansiveness. In such a case, $2\tilde{\mathbf{\Pi}}_{\mathcal{C}} - I_n$ is also not nonexpansive, because an operator is firmly nonexpansive if and only if its reflection operator is nonexpansive [1, Proposition 4.2].

We now introduce the following definitions to (i) characterize an operator that is not necessarily nonexpansive (*e.g.,* $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ and $2\tilde{\mathbf{\Pi}}_{\mathcal{C}} - I_n$) and (ii) measure the deviation of such an operator from being nonexpansive.

Expansiveness of an operator. Let $T : \mathcal{D} \to \mathbf{R}^n$ be a single-valued operator. The **expansiveness** of T at x, y in $\mathcal{D}$, denoted by $\varepsilon_{xy}^{(T)}$, is defined as

$$\varepsilon_{xy}^{(T)} = \begin{cases} \|T(x) - T(y)\| - \|x - y\|, & \\ & \text{if } \|x - y\| < \|T(x) - T(y)\| \\ 0, & \text{else.} \end{cases}$$

where it is nonnegative and symmetric, *i.e.,* $\varepsilon_{xy}^{(T)} = \varepsilon_{yx}^{(T)} \geq 0$. It follows that for every x, y in $\mathcal{D}$,

$$\|T(x) - T(y)\| \leq \|x - y\| + \varepsilon_{xy}^{(T)}. \tag{2}$$

Furthermore, define, **squared expansiveness** of T at x, y in $\mathcal{D}$ as

$$\sigma_{xy}^{(T)} = \begin{cases} \sqrt{\|T(x) - T(y)\|^2 - \|x - y\|^2}, & \\ & \text{if } \|x - y\| < \|T(x) - T(y)\| \\ 0, & \text{else.} \end{cases}$$

Clearly, $\sigma_{xy}^{(T)}$ can be defined through $\varepsilon_{xy}^{(T)}$ as

$$\sigma_{xy}^{(T)} = \sqrt{\varepsilon_{xy}^{(T)}} \sqrt{\|T(x) - T(y)\| + \|x - y\|}.$$

It follows that for every x, y in $\mathcal{D}$,

$$\|T(x) - T(y)\|^2 \leq \|x - y\|^2 + \left(\sigma_{xy}^{(T)}\right)^2. \tag{3}$$

Remark 1 (**further characterization of nonexpansive operators**). An operator T is nonexpansive on $\mathbf{R}^n$ if and only if $\varepsilon_{xy}^{(T)} = \sigma_{xy}^{(T)} = 0$ for every x, y in $\mathbf{R}^n$. On the other hand, an operator T is not nonexpansive if and only if there exist x, y in its domain such that $\varepsilon_{xy}^{(T)}$ is positive. Thus, $\varepsilon_{xy}^{(T)}$ and $\sigma_{xy}^{(T)}$ measure the deviation of T from being a nonexpansive operator at x, y.

C. Resolvent and reflected resolvent of an operator

Let $T : \mathbf{R}^n \rightrightarrows \mathbf{R}^n$ be a set-valued operator and let $\gamma > 0$. The **resolvent** of T, denoted by $J_{\gamma T}$, is defined as $J_{\gamma T} = (I_n + \gamma T)^{-1}$, and its **reflected resolvent,** denoted by $R_{\gamma T}$, is defined as $R_{\gamma T} = 2J_{\gamma T} - I_n$. The proximal operator of a function is intimately connected to the resolvent of that function's subdifferential operator as follows.

Lemma 2 (**resolvent characterization of proximal operator**). *Let $g : \mathbf{R}^n \to \overline{\mathbf{R}}$ be proper, let $x \in \mathbf{R}^n$, and let $\gamma > 0$. Then, both $\mathbf{prox}_{\gamma g}$ and $J_{\gamma \partial g}$ are set-valued, and $\mathbf{prox}_{\gamma g}(x) \subseteq J_{\gamma \partial g}(x)$. Moreover, if g is CPC, then both $\mathbf{prox}_{\gamma g}$ and $J_{\gamma \partial g}$ are single-valued, firmly nonexpansive and continuous on $\mathbf{R}^n$, and $\mathbf{prox}_{\gamma g}(x) = J_{\gamma \partial g}(x)$.*

Proof: When g is proper, the claim follows from [12, Example 10.2]. When, g is CPC, the claim follows from [1, Proposition 12.27], [1, pages 59-60], and [1, Example 23.3]. ■

The following corollary applies Lemma 2 to the constraint set $\mathcal{C}$ in (OPT).

Corollary 1 (**resolvent characterization of projection**). *For the constraint set $\mathcal{C}$ in* (OPT), *$\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x) \subseteq \mathbf{prox}_{\gamma \delta_{\mathcal{C}}}(x) = \mathbf{\Pi}_{\mathcal{C}}(x) \subseteq J_{\gamma \partial \delta_C}(x)$ for every $x \in \mathbf{R}^n$. For a convex set, all these operators are single-valued, firmly nonexpansive, and equal to each other.*

Proof: Follows directly from Lemma 2 and the definitions of the proximal operator and the projection operator. ■

III. Convergence of (NC-DRS)

This section is organized as follows. First, in §III-A, we present some supporting lemmas on convergence of sequences. Then, in §III-B, we describe three interrelated operators to develop the machinery for the convergence analysis of (NC-DRS), and in §III-C, we characterize global minimizers of (OPT) using these operators. In §III-D, we present our main convergence result.

A. Supporting lemmas on sequences

In this subsection, we present some supporting lemmas on sequences to be used later; the first three results concern convergence of a sequence of scalars, and the fourth result is about convergence of a sequence of vectors in a compact set.

First, we briefly review the definitions and basic properties of limit inferior and limit superior of a sequence. **Limit inferior** and **limit superior** of a scalar sequence $(\alpha_n)_{n \in \mathbf{N}}$ are defined as

$$\begin{aligned} \varliminf_{n \to \infty} \alpha_n &= \lim_{n \to \infty} \big(\inf_{m \geq n} \alpha_m \big), \text{and} \\ \varlimsup_{n \to \infty} \alpha_n &= \lim_{n \to \infty} \big(\sup_{m \geq n} \alpha_m \big), \end{aligned}$$

respectively, where they can be extended real-valued. For a bounded sequence, both $\varliminf_{n \to \infty} \alpha_n$ and $\varlimsup_{n \to \infty} \alpha_n$ exist, and they are finite. Clearly, $\varliminf_{n \to \infty} \alpha_n \leq \varlimsup_{n \to \infty} \alpha_n$. The sequence converges if and only if $\varliminf_{n \to \infty} \alpha_n = \varlimsup_{n \to \infty} \alpha_n = \lim_{n \to \infty} \alpha_n \in \mathbf{R}$. Furthermore, limit inferior satisfies **superadditivity**, *i.e.,* for every two sequences of real numbers, $(\alpha_n)_{n \in \mathbf{N}}, (\beta_n)_{n \in \mathbf{N}}$ we have $\varliminf_{n \to \infty} (\alpha_n + \beta_n) \geq \varliminf_{n \to \infty} \alpha_n + \varliminf_{n \to \infty} \beta_n$.

Lemma 3 (**limit of a nonnegative scalar sequence**). *Let $(\alpha_n)_{n \in \mathbf{N}}$ be a sequence of nonnegative scalars such that $\sum_{n \in \mathbf{N}} \alpha_n$ is bounded above. Then, $\lim_{n \to \infty} \alpha_n = 0$.*

Proof: Directly follows from [13, Proposition 3.2.1] and [13, Theorem 3.1.4]. ■

Lemma 4 (**convergence of a nonnegative scalar sequence** [14, page 44, Lemma 2]). *Let $(u_n)_{n \in \mathbf{N}}, (\alpha_n)_{n \in \mathbf{N}}$, and $(\beta_n)_{n \in \mathbf{N}}$ be sequences of nonnegative scalars such that for every $n \in \mathbf{N}$, we have $u_{n+1} \leq (1+\alpha_n)u_n + \beta_n$, $\sum_{n \in \mathbf{N}} \alpha_n < \infty$, and $\sum_{n \in \mathbf{N}} \beta_n < \infty$. Then, there is a nonnegative scalar u such that u_n converges to u.*

Lemma 5 (**limit inferior of addition of two sequences** [15, Proposition 2.3]). *Let $(\alpha_n)_{n \in \mathbf{N}}$ and $(\beta_n)_{n \in \mathbf{N}}$ be two bounded scalar sequences. If $\lim_{n \to \infty} \alpha_n = \alpha$, then $\varliminf_{n \to \infty} (\alpha_n + \beta_n) = \alpha + \varliminf_{n \to \infty} \beta_n$.*

Now we record a result about convergence of a sequence of vectors in a compact set. We remind the reader that, a set is **connected** if it is not the union of two disjoint nonempty closed sets. A compact and connected set is called a **continuum**. Moreover, a set is called a **nontrivial continuum**, if it is a continuum, and it does not reduce to $\emptyset$ or a singleton

[16]. Finally, x is a **cluster point** of a sequence $(x_n)_{n\in\mathbf{N}}$ if the sequence has a subsequence that converges to x.

Lemma 6 (**convergence of a sequence of vectors in a compact set** [16, Theorem 4.2])**.** *Let $(x_n)_{n\in\mathbf{N}}$ be a sequence of vectors in a compact set $\mathcal{S}\subseteq\mathbf{R}^n$ such that $\|x_{n+1}-x_n\|$ converges to zero. Then, either $(x_n)_{n\in\mathbf{N}}$ converges to a point in $\mathcal{S}$, or its set of cluster points is a nontrivial continuum in $\mathcal{S}$.*

B. Nonconvex Douglas-Rachford, Cayley and Peaceman-Rachford operators

To facilitate our convergence analysis, we define the following operators for (OPT).

- The **nonconvex Douglas-Rachford operator** with parameter $\gamma>0$, denoted by $\tilde{T}$, is defined as
$$\tilde{T}=\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}-I_n\right)+I_n-\mathbf{prox}_{\gamma f}. \tag{4}$$
- The **nonconvex Cayley operator** of $\tilde{T}$ (also known as the reflection operator of $\tilde{T}$) with parameter $\gamma>0$, denoted by $\tilde{R}$, is defined as
$$\tilde{R}=2\tilde{T}-I_n. \tag{5}$$
- The **nonconvex Peaceman-Rachford operator** with parameter $\gamma>0$, denoted by $\tilde{S}$, is defined as
$$\tilde{S}=(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n)(2\mathbf{prox}_{\gamma f}-I_n). \tag{6}$$

Remark 2 (**nonconvex Peaceman-Rachford operator $\tilde{S}$ is not nonexpansive**). Note that $\tilde{S}$ is a composition of $2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n$ and $2\mathbf{prox}_{\gamma f}-I_n$, where the latter is nonexpansive (see Example 3), but the former is not nonexpansive in general (see Example 4). Hence $\tilde{S}$ is not a nonexpansive operator in general.

These operators allow us to write (NC-DRS) in the following compact form:
$$z_{n+1}=\tilde{T}z_n=\frac{1}{2}\tilde{R}z_n+\frac{1}{2}z_n. \tag{Compact-NC-DRS}$$

The following lemma will be used later to characterize global minimizers of (OPT).

Lemma 7 (**characterization of nonconvex Peaceman-Rachford operator**)**.** *For* (OPT)*, let $\tilde{S}$ be the nonconvex Peaceman-Rachford operator with parameter $\gamma>0$ defined in* (6)*. Then, $\tilde{S}(x)\subseteq R_{\gamma\partial\delta_{\mathcal{C}}}R_{\gamma\partial f}(x)$ for every $x\in\mathbf{R}^n$.*

Proof: As f is CPC, we have
$$R_{\gamma\partial f}=2\mathbf{prox}_{\gamma f}-I_n, \tag{7}$$
using Lemma 2 and the definition of the reflected resolvent in §II-C. Now for every $x\in\mathbf{R}^n$,
$$\begin{aligned}\left(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n\right)(x)&=2\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x)-x\\&\overset{a)}{\subseteq}2\mathbf{\Pi}_{\mathcal{C}}(x)-x\\&\overset{b)}{\subseteq}2J_{\gamma\delta_{\mathcal{C}}}(x)-x\\&=\left(2J_{\gamma\delta_{\mathcal{C}}}-I_n\right)x\\&\overset{c)}{=}R_{\gamma\delta_{\mathcal{C}}}(x),\end{aligned} \tag{8}$$
where $a)$ follows from $\tilde{\mathbf{\Pi}}_{\mathcal{C}}(x)\subseteq\mathbf{\Pi}_{\mathcal{C}}(x)$ for every x in $\mathbf{R}^n$ (Example 1), $b)$ follows from Corollary 1, and $c)$ follows from the definition of reflected resolvent in §II-C. Thus, for every $x\in\mathbf{R}^n$,
$$\begin{aligned}\tilde{S}(x)&=\left(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n\right)\left(2\mathbf{prox}_{\gamma f}-I_n\right)(x)\\&\overset{a)}{=}\left(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n\right)R_{\gamma\partial f}(x)\\&\overset{b)}{\subseteq}R_{\gamma\partial\delta_{\mathcal{C}}}R_{\gamma\partial f}(x),\end{aligned}$$
where $a)$ and $b)$ use (7) and (8), respectively. ∎

Proposition 1 (**relationship between $\tilde{T}$, $\tilde{R}$, and $\tilde{S}$**)**.** *For* (OPT)*, let $\tilde{T}$, $\tilde{R}$, and $\tilde{S}$ be the operators with parameter $\gamma>0$ defined in* (4)*,* (5)*, and* (6)*, respectively. Then,*

(i) *the operators $\tilde{R}$ and $\tilde{S}$ are equal, i.e., $\tilde{R}(x)=\tilde{S}(x)$ for every $x\in\mathbf{R}^n$, and*

(ii) *the fixed point sets of $\tilde{T}$, $\tilde{R}$, and $\tilde{S}$ are equal, i.e., $\mathbf{fix}\,\tilde{R}=\mathbf{fix}\,\tilde{S}=\mathbf{fix}\,\tilde{T}$.*

Proof: (i): For every $x\in\mathbf{R}^n$,
$$\begin{aligned}\tilde{T}(x)&=\left(\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}-I_n\right)+I_n-\mathbf{prox}_{\gamma f}\right)(x)\\&=\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}-I_n\right)(x)+x-\mathbf{prox}_{\gamma f}(x)\\&=\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}(x)-x\right)+x-\mathbf{prox}_{\gamma f}(x).\end{aligned} \tag{9}$$

Furthermore, for every $x\in\mathbf{R}^n$,
$$\begin{aligned}\tilde{R}(x)&=\left(2\tilde{T}-I_n\right)(x)\\&=2\tilde{T}(x)-x\\&\overset{a)}{=}2\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}(x)-x\right)+2x-2\mathbf{prox}_{\gamma f}(x)-x\\&=2\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}(x)-x\right)+x-2\mathbf{prox}_{\gamma f}(x),\end{aligned} \tag{10}$$
where $a)$ uses (9). Hence, for every $x\in\mathbf{R}^n$
$$\begin{aligned}\tilde{S}(x)&=\left(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n\right)\left(2\mathbf{prox}_{\gamma f}-I_n\right)(x)\\&=\left(2\tilde{\mathbf{\Pi}}_{\mathcal{C}}-I_n\right)\underbrace{\left(2\mathbf{prox}_{\gamma f}(x)-x\right)}_{=y\text{ (let)}}\\&=2\tilde{\mathbf{\Pi}}_{\mathcal{C}}(y)-y\\&=2\tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f}(x)-x\right)-2\mathbf{prox}_{\gamma f}(x)+x\\&\overset{a)}{=}\tilde{R}x,\end{aligned}$$
where $a)$ uses (10).

(ii): In (i), $\tilde{R}=\tilde{S}$ implying $\mathbf{fix}\,\tilde{R}=\mathbf{fix}\,\tilde{S}$. Now $x\in\mathbf{fix}\,\tilde{T}\Leftrightarrow\tilde{T}(x)=x\Leftrightarrow2\tilde{T}(x)=2x\Leftrightarrow2\tilde{T}(x)-x=x\Leftrightarrow\left(2\tilde{T}-I_n\right)(x)=x\Leftrightarrow x=\mathbf{fix}\,\tilde{R}$. So $\mathbf{fix}\,\tilde{T}=\mathbf{fix}\,\tilde{R}=\mathbf{fix}\,\tilde{S}$. ∎

Remark 3 (**nonconvex Cayley operator $\tilde{R}$ is not nonexpansive**). From Remark 2 and Proposition 1, it follows that $\tilde{R}$ is not a nonexpansive operator in general. This plays an important role in our convergence analysis; in particular, the sufficient conditions for convergence of (NC-DRS) are dictated by the squared expansiveness of $\tilde{R}$ over the iterates of (NC-DRS).

C. Characterization of global minimizers

Minimizers of (OPT) are characterized via the nonconvex Douglas-Rachford operator as follows.

Theorem 2 (**global minimizers of** (OPT))**.** *For* (OPT)*, let* $\tilde{T}$ *be the nonconvex Douglas-Rachford operator with parameter* $\gamma > 0$ *defined in* (4)*. Then,*

(i) *sum of the functions* $f + \delta_{\mathcal{C}}$ *is proper,* $\partial f(x) + \partial \delta_{\mathcal{C}}(x) \subseteq \partial(f + \delta_{\mathcal{C}})(x)$ *for every* $x \in \mathbf{R}^n$*, and both* $\partial f + \partial \delta_{\mathcal{C}}$ *and* $\partial(f + \delta_{\mathcal{C}})$ *are monotone operators,*

(ii) $\mathbf{zer}\big(\partial f + \partial \delta_{\mathcal{C}}\big) = \mathbf{prox}_{\gamma f}\big(\mathbf{fix}(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f})\big)$*, and*

(iii) *if* $\mathbf{fix}\,\tilde{T} \neq \emptyset$*, then* $\mathbf{prox}_{\gamma f}\big(\mathbf{fix}\,\tilde{T}\big) \subseteq \operatorname{argmin}(f + \delta_{\mathcal{C}})$*.*

Proof: (i): The indicator function of a closed set is closed [1, Example 1.25], and the indicator function of a nonempty set is proper [12, pages 6-7]. Hence, $\delta_{\mathcal{C}}$ is closed and proper. Also, we have $\mathbf{dom}\, f \cap \mathbf{dom}\,\delta_{\mathcal{C}} \neq \emptyset$, otherwise (OPT) is infeasible. So, using Lemma 1, the function $f + \delta_{\mathcal{C}}$ is proper, $\partial f(x) + \partial\delta_{\mathcal{C}}(x) \subseteq \partial(f + \delta_{\mathcal{C}})(x)$ for every $x \in \mathbf{R}^n$, and both $\partial f + \partial \delta_{\mathcal{C}}$ and $\partial(f + \delta_{\mathcal{C}})$ are monotone operators.

(ii): This proof is based on [1, Proposition 25.1 (ii)]. For every $\gamma > 0$, we have

$$
\begin{aligned}
& x \in \mathbf{zer}\big(\partial f + \partial \delta_{\mathcal{C}}\big) \\
\Leftrightarrow & (\exists y \in \mathbf{R}^n)\ x - y \in \gamma\partial\delta_{\mathcal{C}}(x) \text{ and } y - x \in \gamma \partial f(x) \\
\Leftrightarrow & (\exists y \in \mathbf{R}^n)\ \ 2x - y \in (I_n + \gamma\partial\delta_{\mathcal{C}})(x) \text{ and} \\
& \qquad\qquad\quad y \in (I_n + \gamma\partial f)(x) \\
\Leftrightarrow & (\exists y \in \mathbf{R}^n)\ \ \underbrace{(I_n + \gamma\partial\delta_{\mathcal{C}})^{-1}}_{=J_{\gamma\partial\delta_{\mathcal{C}}}}(2x - y) \ni x \text{ and} \\
& \qquad\qquad\quad \underbrace{(I_n + \gamma\partial f)^{-1}}_{=J_{\gamma\partial f}}(y) \ni x \\
\overset{a)}{\Leftrightarrow} & (\exists y \in \mathbf{R}^n)\ \ x \in J_{\gamma\partial\delta_{\mathcal{C}}}(2x - y) \text{ and } x = J_{\gamma\partial f}(y), \\
\overset{b)}{\Leftrightarrow} & (\exists y \in \mathbf{R}^n)\ \ x \in J_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(y) \text{ and } x = J_{\gamma\partial f}(y) \qquad (11)
\end{aligned}
$$

where $a)$ uses the facts that $J_{\gamma\partial f}$ is a single-valued operator (from Lemma 2), and $J_{\gamma\partial\delta_{\mathcal{C}}}$ is a set-valued operator (from Corollary 1), and $b)$ uses the observation that $x = J_{\gamma\partial f}(y)$ can be expressed as

$$x = J_{\gamma\partial f}(y) \Leftrightarrow 2x - y = \big(2J_{\gamma\partial f} - I_n\big)y = R_{\gamma\partial f}(y).$$

Also, using the last expression, we can write the first term of (11) as

$$
\begin{aligned}
& J_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(y) \ni x \\
\Leftrightarrow & 2J_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(y) - y \ni 2x - y = R_{\gamma\partial f}(y) \\
\Leftrightarrow & y \in 2J_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(y) - R_{\gamma\partial f}(y) \\
& \quad = \big(2J_{\gamma\partial\delta_{\mathcal{C}}} - I_n\big)\big(R_{\gamma\partial f}(y)\big) \\
& \quad = R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(y) \\
\Leftrightarrow & y \in \mathbf{fix}\,\big(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big). \qquad (12)
\end{aligned}
$$

Using (11), (12), and $J_{\gamma\partial f} = \mathbf{prox}_{\gamma f}$ (from Lemma 2) we have

$$
\begin{aligned}
& x \in \mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) \\
\Leftrightarrow & (\exists y \in \mathbf{R}^n)\ \ y \in \mathbf{fix}\,\big(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big) \text{ and } x = \mathbf{prox}_{\gamma f}(y) \\
\Leftrightarrow & x \in \mathbf{prox}_{\gamma f}\big(\mathbf{fix}(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f})\big).
\end{aligned}
$$

Thus, $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) = \mathbf{prox}_{\gamma f}\big(\mathbf{fix}(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f})\big)$.

(iii): We have

$$
\begin{aligned}
& x \in \mathbf{zer}\,(\partial f + \partial\delta_{\mathcal{C}}) \\
\Leftrightarrow & 0 \in \partial f(x) + \partial\delta_{\mathcal{C}}(x) \overset{a)}{\subseteq} \partial(f + \delta_{\mathcal{C}})(x) \\
\Rightarrow & x \in \mathbf{zer}\,\partial(f + \delta_{\mathcal{C}}),
\end{aligned}
$$

where $a)$ uses $\partial f(x) + \partial\delta_{\mathcal{C}}(x) \subseteq \partial(f + \delta_{\mathcal{C}})(x)$ proven in (i). So, $\mathbf{zer}\,(\partial f + \partial\delta_{\mathcal{C}}) \subseteq \mathbf{zer}\,(\partial(f + \delta_{\mathcal{C}}))$. Combining the last statement with $\mathbf{zer}\,(\partial(f + \delta_{\mathcal{C}})) = \operatorname{argmin}(f + \delta_{\mathcal{C}})$ (from Theorem 1) and (ii), we have

$$
\begin{aligned}
& \mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) \\
= & \mathbf{prox}_{\gamma f}\Big(\mathbf{fix}\,\big(R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big)\Big) \qquad (13) \\
\subseteq & \operatorname{argmin}\big(f + \delta_{\mathcal{C}}\big). \qquad (14)
\end{aligned}
$$

Recall from Lemma 2 that $\mathbf{prox}_{\gamma f}$ is a single-valued operator. Thus,

$$
\begin{aligned}
\mathbf{prox}_{\gamma f}\big(\mathbf{fix}\,\tilde{S}\big) &= \bigcup_{x \in \mathbf{fix}\,\tilde{S}} \mathbf{prox}_{\gamma f}(x) \\
&= \bigcup_{x : x = \tilde{S}(x)} \mathbf{prox}_{\gamma f}(x) \\
&\overset{a)}{\subseteq} \bigcup_{x : x \in R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(x)} \mathbf{prox}_{\gamma f}(x) \\
&= \bigcup_{x : x \in \mathbf{fix}\, R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}} \mathbf{prox}_{\gamma f}(x) \\
&= \mathbf{prox}_{\gamma f}\big(\mathbf{fix}\, R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big),
\end{aligned}
$$

where $a)$ uses $\tilde{S}(x) \subseteq R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}(x)$ for every $x \in \mathbf{R}^n$ (from Lemma 7). But, $\mathbf{fix}\,\tilde{S} = \mathbf{fix}\,\tilde{T}$ from Proposition 1. So,

$$
\begin{aligned}
\mathbf{prox}_{\gamma f}\big(\mathbf{fix}\,\tilde{S}\big) &= \mathbf{prox}_{\gamma f}\big(\mathbf{fix}\,\tilde{T}\big) \\
&\subseteq \mathbf{prox}_{\gamma f}\big(\mathbf{fix}\, R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big).
\end{aligned}
$$

Combining the last equation with (14), we have

$$
\begin{aligned}
\mathbf{prox}_{\gamma f}\big(\mathbf{fix}\,\tilde{T}\big) &\subseteq \mathbf{prox}_{\gamma f}\big(\mathbf{fix}\, R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}\big) \\
&\subseteq \operatorname{argmin}\big(f + \delta_{\mathcal{C}}\big).
\end{aligned}
$$

■

Remark 4 (**nonemptiness of** $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big)$). A necessary condition for nonemptiness of $\mathbf{fix}\,\tilde{T}$ is nonemptiness of $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big)$. This necessary condition $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) \neq \emptyset$ is stronger than the existence of a minimizer, because, even in a convex setup, $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) \neq \mathbf{zer}\,(\partial(f + \delta_{\mathcal{C}}))$, in general [1, Remark 16.7]. Nevertheless, we will assume that $\mathbf{zer}\,\big(\partial f + \partial\delta_{\mathcal{C}}\big) \neq \emptyset$ for the rest of our development, as this seems to be a standard assumption even in convex optimization literature [17].

D. Main convergence result

We remind the reader that the nonconvex Cayley operator $\tilde{R}$ is not nonexpansive in general (Remark 3). To characterize the deviation of $\tilde{R}$ from being a nonexpansive operator, recalling §II-B, we use expansiveness and squared expansiveness of $\tilde{R}$ at each x, y in $\mathbf{R}^n$, denoted by ε_{xy} and σ_{xy}, respectively; here we have dropped the superscript $(\tilde{R})$ to reduce notational burden. So, from (2) and (3), for every x, y in $\mathbf{R}^n$,

$$\|\tilde{R}(x) - \tilde{R}(y)\| \leq \|x - y\| + \varepsilon_{xy}, \text{and} \tag{15}$$
$$\|\tilde{R}(x) - \tilde{R}(y)\|^2 \leq \|x - y\|^2 + \sigma_{xy}^2. \tag{16}$$

Also, the closed ball with center $x \in \mathbf{R}^n$ and finite radius $r > 0$, denoted by $B(x; r)$, is defined as $B(x; r) = \{y \mid \|x-y\| \leq r\}$; a closed ball in $\mathbf{R}^n$ with finite radius is compact [1, §2.4]. Now we present our main convergence result.

Theorem 3 (main convergence result). *For* (OPT)*, let* $(z_n)_{n\in\mathbf{N}}$ *be the sequence of vectors generated by* (NC-DRS)*. Suppose that, for the chosen initial point* z_0*, there exists a* $z \in \mathbf{fix}\,\tilde{T}$*, such that* $\sum_{n=0}^{\infty} \sigma_{z_n z}^2$ *is bounded above, and* $\|z_0 - z\|^2$ *is finite. Define* $r := \sqrt{\|z_0 - z\|^2 + \frac{1}{2}\sum_{n=0}^{\infty} \sigma_{z_n z}^2}$*. Then, one of the following holds:*

(i) *the sequence* $(z_n)_{n\in\mathbf{N}}$ *converges to a point* $z^\star \in B(z; r)$*. In this case, suppose also that* $\underline{\lim}_{n\to\infty} \sigma_{z_n z^\star}^2 = 0$*. Then,* $\mathbf{prox}_{\gamma f}(z^\star)$ *is an optimal solution of* (OPT)*, and the sequence* $(x_n)_{n\in\mathbf{N}}$ *generated by* (NC-DRS) *converges to* $\mathbf{prox}_{\gamma f}(z^\star)$*.*

(ii) *the set of cluster points of* $(z_n)_{n\in\mathbf{N}}$ *forms a nontrivial continuum in* $B(z; r)$*.*

Proof: **Step 1.** First, we show that the sequence $(z_n)_{n\in\mathbf{N}}$ stays in the compact set $B(z; r)$. For every $n \in \mathbf{N}$,

$$\begin{aligned}
\|z_{n+1} - z\|^2 &\overset{a)}{=} \|z_n + \frac{1}{2}\big(\tilde{R}z_n - z_n\big) - z\|^2 \\
&= \|\frac{1}{2}\big(z_n - z\big) + \frac{1}{2}\big(\tilde{R}z_n - z\big)\|^2 \\
&\overset{b)}{=} \frac{1}{2}\|z_n - z\|^2 + \frac{1}{2}\|\tilde{R}z_n - z\|^2 \\
&\quad - \frac{1}{4}\|(z_n - z) - (\tilde{R}z_n - z)\|^2 \\
&\overset{c)}{=} \frac{1}{2}\|z_n - z\|^2 + \frac{1}{2}\|\tilde{R}z_n - \tilde{R}z\|^2 \\
&\quad - \frac{1}{4}\|z_n - \tilde{R}z_n\|^2 \\
&\overset{d)}{\leq} \frac{1}{2}\|z_n - z\|^2 + \frac{1}{2}\|z_n - z\|^2 \\
&\quad + \frac{1}{2}\sigma_{z_n z}^2 - \frac{1}{4}\|z_n - \tilde{R}z_n\|^2 \\
&\leq \|z_n - z\|^2 - \frac{1}{4}\|z_n - \tilde{R}z_n\|^2 + \frac{1}{2}\sigma_{z_n z}^2 \qquad (17) \\
&\overset{e)}{\leq} \|z_n - z\|^2 + \frac{1}{2}\sigma_{z_n z}^2, \qquad (18)
\end{aligned}$$

where $a)$ uses (Compact-NC-DRS), $b)$ uses the identity $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x - y\|^2$ for every $x, y \in \mathbf{R}^n$ and every $\alpha \in \mathbf{R}$ [1, Corollary 2.14], $c)$ uses $z \in \mathbf{fix}\,\tilde{T}$, and $\mathbf{fix}\,\tilde{T} = \mathbf{fix}\,\tilde{R}$ (from Proposition 1(ii)), $d)$ uses (16), and $e)$ is obtained by removing the nonpositive term $-\frac{1}{4}\|z_n - \tilde{R}z_n\|^2$. From (18), we have

$$\begin{aligned}
\|z_n - z\|^2 &\leq \|z_{n-1} - z\|^2 + \frac{1}{2}\sigma_{z_{n-1}z}^2 \\
&\leq \|z_{n-2} - z\|^2 + \frac{1}{2}\sigma_{z_{n-2}z}^2 + \frac{1}{2}\sigma_{z_{n-1}z}^2 \\
&\leq \|z_0 - z\|^2 + \frac{1}{2}\sum_{i=0}^{n-1}\sigma_{z_i z}^2 \\
&\leq \|z_0 - z\|^2 + \frac{1}{2}\sum_{i=0}^{\infty}\sigma_{z_i z}^2, \qquad (19)
\end{aligned}$$

where the final term is bounded, because $\sum_{i=0}^{\infty} \sigma_{z_i z}^2$ is bounded above. Hence, the sequence $(z_n)_{n\in\mathbf{N}}$ stays in the compact set $B(z; r)$.

Step 2. Next, we show that $\lim_{n\to\infty} \|\tilde{R}z_n - z_n\| = 0$. From (17),

$$\begin{aligned}
&\frac{1}{4}\|\tilde{R}z_n - z_n\|^2 \leq \big(\|z_n - z\|^2 - \|z_{n+1} - z\|^2\big) + \frac{1}{2}\sigma_{z_n z}^2 \\
\Rightarrow &\frac{1}{4}\sum_{n=0}^{m}\|\tilde{R}z_n - z_n\|^2 \leq \sum_{n=0}^{m}\big(\|z_n - z\|^2 - \|z_{n+1} - z\|^2\big) \\
&\qquad + \frac{1}{2}\sum_{n=0}^{m}\sigma_{z_n z}^2 \\
&\qquad \overset{a)}{=} \big(\|z_0 - z\|^2 - \|z_{m+1} - z\|^2\big) \\
&\qquad + \frac{1}{2}\sum_{n=0}^{m}\sigma_{z_n z}^2 \\
&\qquad \overset{b)}{\leq} \|z_0 - z\|^2 + \frac{1}{2}\sum_{n=0}^{m}\sigma_{z_n z}^2,
\end{aligned}$$

where $a)$ uses the telescopic sum, and $b)$ is obtained by removing the negative term $\|z_{m+1} - z\|^2$. If $m \to \infty$, then the right hand side of the last inequality is bounded above, because $\sum_{n=0}^{\infty} \sigma_{z_n z}^2$ is bounded above. Thus, $\sum_{n=0}^{\infty} \|\tilde{R}z_n - z_n\|^2$ is bounded above, and using Lemma 3, we have $\lim_{n\to\infty} \|\tilde{R}z_n - z_n\|^2 = 0$, *i.e.,* $\lim_{n\to\infty} \|\tilde{R}z_n - z_n\| = 0$.

Step 3. We show that sequence $(z_n)_{n\in\mathbf{N}}$ either converges to a point or its set of cluster points forms a nontrivial continuum. In step 2, we have shown that, $\lim_{n\to\infty} \|\tilde{R}z_n - z_n\| = 0$. On the other hand, $\|\tilde{R}z_n - z_n\| = 2\|z_{n+1} - z_n\|$ from (Compact-NC-DRS), so $\lim_{n\to\infty} \|z_{n+1} - z_n\| = 0$. Thus, the sequence $(z_n)_{n\in\mathbf{N}}$ stays in a compact set $B(z; r)$ and satisfies $\lim_{n\to\infty} \|z_{n+1} - z_n\| = 0$. So, due to Lemma 6, the sequence $(z_n)_{n\in\mathbf{N}}$ either converges to a point $z^\star \in B(z; r)$ or the set of cluster points of $(z_n)_{n\in\mathbf{N}}$ forms a nontrivial continuum in $B(z; r)$. *This proves the first part of claim* (i) *and claim* (ii).

Step 4. Now we prove the second part of claim (i). Under the additional condition $\underline{\lim}_{n\to\infty} \sigma_{z_n z^\star}^2 = 0$, we show that $z_n - \tilde{R}z_n \to 0, z_n \to z^\star$ implies $z^\star \in \mathbf{fix}\,\tilde{T}$, $\mathbf{prox}_{\gamma f}(z^\star)$ is an optimal solution of (OPT), and $x_n \to \mathbf{prox}_{\gamma f}(z^\star)$. For every $n \in \mathbf{N}$,

$$\|z^\star - \tilde{R}z^\star\|^2 \overset{a)}{=} \|z_n - \tilde{R}z^\star\|^2 - \|z_n - z^\star\|^2$$

$$
\begin{aligned}
&- 2\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \rangle \\
&= \|(z_n - \tilde{R}z_n) + (\tilde{R}z_n - \tilde{R}z^\star)\|^2 \\
&\quad - \|z_n - z^\star\|^2 - 2\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \rangle \\
&= \|z_n - \tilde{R}z_n\|^2 + \|\tilde{R}z_n - \tilde{R}z^\star\|^2 \\
&\quad + 2\langle z_n - \tilde{R}z_n \mid \tilde{R}z_n - \tilde{R}z^\star \rangle \\
&\quad - \|z_n - z^\star\|^2 - 2\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \rangle \\
&\overset{b)}{\leq} \|z_n - \tilde{R}z_n\|^2 + \cancel{\|z_n - z^\star\|^2} + \sigma^2_{z_n z^\star} \\
&\quad + 2\langle z_n - \tilde{R}z_n \mid \tilde{R}z_n - \tilde{R}z^\star \rangle - \cancel{\|z_n - z^\star\|^2} \\
&\quad - 2\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \rangle \\
&= \|z_n - \tilde{R}z_n\|^2 + 2\langle z_n - \tilde{R}z_n \mid \tilde{R}z_n - \tilde{R}z^\star \rangle \\
&\quad - 2\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \rangle + \sigma^2_{z_n z^\star},
\end{aligned} \tag{20}
$$

where, $a)$ uses the identity

$$
\begin{aligned}
\|z_n - \tilde{R}z^\star\|^2 &= \|(z_n - z^\star) + (z^\star - \tilde{R}z^\star)\|^2 \\
&= \|z_n - z^\star\|^2 + \|z^\star - \tilde{R}z^\star\|^2 \\
&\quad + 2\left\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star \right\rangle,
\end{aligned}
$$

and $b)$ uses (16). We now compute the limit (or the limit inferior) for each of the terms on the right-hand side of (20). As $z_n - \tilde{R}z_n \to 0$ and $z_n \to z^\star$, subtracting them we have $\tilde{R}z_n \to z^\star$, hence $\tilde{R}z_n - \tilde{R}z^\star \to z^\star - \tilde{R}z^\star$. Combining the last statement with $z_n - \tilde{R}z_n \to 0$, we have $\langle z_n - \tilde{R}z_n \mid \tilde{R}z_n - \tilde{R}z^\star\rangle \to 0$. Also, $z_n - z^\star \to 0$ implies $\langle z_n - z^\star \mid z^\star - \tilde{R}z^\star\rangle \to 0$. Additionally, $\underline{\lim}_{n\to\infty} \sigma^2_{z_n z^\star} = 0$. So, using Lemma 5, limit inferior of the right hand side (20) goes to zero. Hence, we conclude that $z^\star - \tilde{R}z^\star = 0$, *i.e.,* $z^\star \in \mathbf{fix}\,\tilde{R}$. But, $\mathbf{fix}\,\tilde{R} = \mathbf{fix}\,\tilde{T}$ from Proposition 1(ii). So, $z^\star \in \mathbf{fix}\,\tilde{T}$. We now recall from Lemma 2 that $\mathbf{prox}_{\gamma f}$ is continuous everywhere on $\mathbf{R}^n$. So, using the definition of continuity, $z_n \to z^\star \in \mathbf{fix}\,\tilde{T}$ implies $x_{n+1} = \mathbf{prox}_{\gamma f}(z_n) \to \mathbf{prox}_{\gamma f}(z^\star) \in \mathbf{prox}_{\gamma f}(\mathbf{fix}\,\tilde{T})$. But, $\mathbf{prox}_{\gamma f}(\mathbf{fix}\,\tilde{T}) \subseteq \operatorname{argmin}(f + \delta_{\mathcal{C}})$ from Theorem 2(iii). *Thus we have arrived at the second part of claim* (i). ■

1) Notes on Theorem 3. We make the following notes on Theorem 3.

• **Nonemptiness of $\mathbf{fix}\,\tilde{T}$.** Note that Theorem 3 assumes that $\mathbf{fix}\,\tilde{T}$ is nonempty. This is a standard assumption in monotone operator theory even in a convex setup [1, §5.2].

• **Relation to a convex setup.** In our convergence analysis, the constraint set is nonempty and compact, but not necessarily convex. However, our convergence analysis is also applicable to a convex setup. Let $\mathcal{C}$ be convex. Then, both $(2\tilde{\mathbf{\Pi}}_{\mathcal{C}} - I_n)$ and $(2\mathbf{prox}_{\gamma f} - I_n)$ are nonexpansive operators, hence, their composition $\tilde{S} = (2\tilde{\mathbf{\Pi}}_{\mathcal{C}} - I_n)(2\mathbf{prox}_{\gamma f} - I_n)$ is a nonexpansive operator. In such a convex setup, $\tilde{R}$ is a nonexpansive operator, because $\tilde{S} = \tilde{R}$ from Proposition 1 (where the relationship is established irrespective of convexity). Then, recalling Remark 1, expansiveness of $\tilde{R}$ is zero everywhere, *i.e.,* $\sigma_{xy} = \varepsilon_{xy} = 0$ at every x, y in $\mathbf{R}^n$. As a result, the iteration scheme (Compact-NC-DRS) corresponds to an averaged iteration of a nonexpansive operator $\tilde{R}$, which guarantees convergence of the sequence to a fixed point of $\tilde{R}$ for any initial point [18]. Also, the additional condition in the second part of claim (i) are automatically satisfied. This guarantees the convergence of (NC-DRS) to an optimal solution for any initial point if we assume that $\mathcal{C}$ is convex.

• **Comments on the conditions.** Once we move from a convex setup to a nonconvex setup, $\tilde{R}$ is not nonexpansive anymore (recall Remark 3). Roughly speaking, convergence in such a case requires that the total squared expansiveness of $\tilde{R}$ stays bounded for the iterates with respect to at least one fixed point of the nonconvex Douglas-Rachford operator. More precisely, $\sum_{n=0}^{\infty} \sigma^2_{z_n z}$ needs to be bounded, where the sum represents the total deviation of $\tilde{R}$ from being a nonexpansive operator over the sequence $\{(z_n, z)\}_{n\in\mathbf{N}}$. If the stated condition is satisfied, then $(z_n)_{n\in\mathbf{N}}$ is bounded in $B(z; r)$ and $\|z_{n+1} - z_n\| \to 0$, but it does not necessarily guarantee convergence to a point due to the lack of nonexpansiveness of $\tilde{R}$, and this is why the cluster points of $(z_n)_{n\in\mathbf{N}}$ may form a nontrivial continuum in $B(z; r)$.

Suppose now that $(z_n)_{n\in\mathbf{N}}$ converges to a point $z^\star$. Whether $z^\star$ is related to an optimal solution of (OPT) would depend on $\underline{\lim}_{n\to\infty} \sigma^2_{z_n z^\star}$. If it is zero, then $\mathbf{prox}_{\gamma f}(z^\star)$ is an optimal solution, and the iterate x_n in (NC-DRS) converges to this optimal solution. Roughly speaking, $\underline{\lim}_{n\to\infty} \sigma^2_{z_n z^\star} = 0$ means that over $\{(z_n, z^\star)\}_{n\in\mathbf{N}}$, $\tilde{R}$ acts as a nonexpansive operator in the lower limit.

IV. Construction and convergence of (NC-ADMM)

In this section, we discuss how (NC-ADMM) can be constructed from the Douglas-Rachford splitting algorithm and comment on how the construction influences the convergence properties of the former. First, in §IV-A we present some preliminary results to be used later. Then, in §IV-B we describe how (NC-ADMM) is constructed from the Douglas-Rachford splitting algorithm. Finally, in §IV-C we comment on convergence of (NC-ADMM), and we compare it with (NC-DRS).

A. Preliminaries

First, we describe the Douglas-Rachford splitting algorithm for minimizing sum of two CPC functions; we will use it in the first step of constructing (NC-ADMM). Then, we review the necessary background on conjugate and biconjugate functions, and we present two lemmas to be referenced in the second step of constructing (NC-ADMM).

1) Douglas-Rachford splitting algorithm for minimizing sum of two CPC functions. Consider the convex optimization problem

$$
\text{minimize} \quad g(x) + h(x), \tag{21}
$$

where both $g : \mathbf{R}^n \to \overline{\mathbf{R}}$ and $h : \mathbf{R}^n \to \overline{\mathbf{R}}$ are CPC functions, and $x \in \mathbf{R}^n$ is the optimization variable. The Douglas-Rachford splitting algorithm for this problem is

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma h}(z_n) \\
y_{n+1} &= \mathbf{prox}_{\gamma g}(2x_{n+1} - z_n)
\end{aligned} \tag{Convex-DRS}
$$

$$z_{n+1} = z_n + y_{n+1} - x_{n+1},$$

where n is the iteration counter, and γ is a positive parameter. In this convex setup, both x_n and y_n converge to an optimal solution of (21) for any initial point [1, Corollary 27.4].

2) Conjugate and biconjugate of a function. Let $g : \mathbf{R}^n \to \{-\infty\} \cup \overline{\mathbf{R}}$. The **conjugate** of g, denoted by $g^\star$, is defined as $g^\star(y) = \sup_{x\in\mathbf{R}^n} (\langle x \mid y\rangle - g(x))$, which is closed and convex irrespective of the convexity of g [1, Proposition 13.11]. Also, the conjugate of a CPC function is CPC [19, Theorem 4.3, Theorem 4.5]. Similarly, the **biconjugate** of g, denoted by $g^{\star\star}$, is defined as $g^{\star\star}(y) = \sup_{x\in\mathbf{R}^n} (\langle x \mid y\rangle - g^\star(x))$. Additionally, if the function is CPC, then its biconjugate is equal to the function itself [19, Lemma 4.8]. Finally, the relationship between the proximal operator of a CPC function f with the proximal operator of its conjugate is given by **Moreau's decomposition**: $\mathbf{prox}_f(x) + \mathbf{prox}_{f^\star}(x) = x$ for every $x \in \mathbf{R}^n$. Moreau's decomposition does not hold for a nonconvex function.

Next, we present the following lemmas about conjugate functions in the context of (OPT). Here we use the notation $g^\vee$, which denotes the reversal of a function g, and it is defined as $g^\vee(x) = g(-x)$ for every $x \in \mathbf{R}^n$.

Lemma 8 (**proximal operator of** $f^{\star\vee}$)**.** *Let $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ be the cost function in* (OPT)*. Then, for every $\gamma > 0$ and for every $x \in \mathbf{R}^n$,*

$$\mathbf{prox}_{\gamma f^{\star\vee}}(x) = x + \gamma\, \mathbf{prox}_{\gamma^{-1} f}\left(-\gamma^{-1}x\right).$$

Proof: Recall that f is CPC. For every $\gamma > 0$ and for every $x \in \mathbf{R}^n$,

$$\begin{aligned}\mathbf{prox}_{\gamma f^{\star\vee}}(x) &\overset{a)}{=} \mathbf{prox}_{\gamma (f^\vee)^\star}(x)\\ &\overset{b)}{=} x - \gamma\, \mathbf{prox}_{\gamma^{-1} f^\vee}\left(\gamma^{-1}x\right)\\ &\overset{c)}{=} x + \gamma\, \mathbf{prox}_{\gamma^{-1} f}\left(-\gamma^{-1}x\right),\end{aligned}$$

where $a)$ follows from $f^{\star\vee} = f^{\vee\star}$ [1, Proposition 13.20(v)], $b)$ follows from [1, Proposition 23.29(viii)] and the fact that $f^\vee$ is CPC, and $c)$ directly follows from [1, Proposition 23.29(v)]. ■

In the following Lemma **convex hull** of a nonempty set $\mathcal{C}$, which is the smallest convex set containing $\mathcal{C}$, is denoted by $\mathbf{conv}\,\mathcal{C}$. Closure of $\mathbf{conv}\,\mathcal{C}$ is denoted by $\overline{\mathbf{conv}}\mathcal{C}$.

Lemma 9 (**conjugate and biconjugate of indicator function of** $\mathcal{C}$)**.** *Let $\mathcal{C}$ be the constraint set in* (OPT)*. Then,*

(i) $\delta_{\mathcal{C}}^{\star\star} = \delta_{\mathbf{conv}\,\mathcal{C}}$, *and*

(ii) $\mathbf{prox}_{\gamma\delta_{\mathcal{C}}^\star}(x) = x - \gamma \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(\gamma^{-1}x\right)$.

Proof: (i): From [19, Example 4.2, Example 4.9], we have $\delta_{\mathcal{C}}^{\star\star} = \delta_{\overline{\mathbf{conv}}\mathcal{C}}$. The constraint set $\mathcal{C}$ is compact, hence its convex hull $\mathbf{conv}\,\mathcal{C}$ is also compact, hence closed [12, Corollary 2.30]. So, $\overline{\mathbf{conv}}\mathcal{C} = \mathbf{conv}\,\mathcal{C}$, and we conclude that $\delta_{\mathcal{C}}^{\star\star} = \delta_{\mathbf{conv}\,\mathcal{C}}$.

(ii): As the constraint set $\mathcal{C}$ is nonempty and compact, its indicator function $\delta_{\mathcal{C}}$ is closed [1, Example 1.25] and proper [12, page 7]. Hence, its conjugate $\delta_{\mathcal{C}}^\star$, which is called the **support function** of the set $\mathcal{C}$, is CPC (closed and convex due to [1, Proposition 13.11], proper because $\mathcal{C}$ is bounded). As the conjugate of a CPC function is CPC [19, Theorem 4.3, Theorem 4.5], the function $\delta_{\mathcal{C}}^{\star\star}$ is CPC. Using Moreau's decomposition for every $x \in \mathbf{R}^n$,

$$\begin{aligned}\mathbf{prox}_{\gamma\delta_{\mathcal{C}}^\star}(x) &= x - \mathbf{prox}_{\left(\gamma\delta_{\mathcal{C}}^\star\right)^\star}(x)\\ &\overset{a)}{=} x - \mathbf{prox}_{\gamma\delta_{\mathcal{C}}^{\star\star}\left(\gamma^{-1}(\cdot)\right)}(x)\\ &\overset{b)}{=} x - \gamma\, \mathbf{prox}_{\gamma^{-1}\delta_{\mathcal{C}}^{\star\star}}\left(\gamma^{-1}x\right)\\ &\overset{c)}{=} x - \gamma \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(\gamma^{-1}x\right),\end{aligned}$$

where $a)$ follows from [1, Proposition 13.20(i)], $b)$ follows from [1, Proposition 23.29(iii)], and $c)$ follows from combining $\delta_{\mathcal{C}}^{\star\star} = \delta_{\mathbf{conv}\,\mathcal{C}}$ in (i) and Corollary 1. ■

B. Constructing (NC-ADMM) *from Douglas-Rachford splitting*

This subsection is organized as follows. First, by applying (Convex-DRS) to the convex dual of (OPT) we construct a relaxed version of (NC-ADMM), where the projection is onto $\mathbf{conv}\,\mathcal{C}$ rather than $\mathcal{C}$. Then, we show that the relaxed version (NC-ADMM) minimizes f over $\mathbf{conv}\,\mathcal{C}$. Next, we discuss construction of (NC-ADMM) from the relaxed variant by restricting the latter's projection step onto $\mathcal{C}$. Finally, we comment on the convergence properties of (NC-ADMM) and relate it to (NC-DRS).

1) Constructing dual of (OPT)*.* Using indicator function, we write (OPT) as

$$\begin{array}{ll}\text{minimize} & f(x) + \delta_{\mathcal{C}}(y)\\ \text{subject to} & x - y = 0,\end{array}$$

where $x, y \in \mathbf{R}^n$ are the optimization variables. Denote the optimal value of the problem above by $p^\star$. The dual of the reformulated problem, which is a convex optimization problem [20, §5.1.6], is

$$\text{maximize} \quad -f^{\star\vee}(\nu) - \delta_{\mathcal{C}}^\star(\nu),$$

where $\nu \in \mathbf{R}^n$ is the optimization variable. Denote the optimal value of the dual problem by $d^\star$. Due to weak duality, we have, $d^\star \leq p^\star$, and, as the primal problem is nonconvex, the duality gap $p^\star - d^\star$ is strict in general. For convenience, we write the dual problem in minimization form:

$$\text{minimize} \quad f^{\star\vee}(\nu) + \delta_{\mathcal{C}}^\star(\nu), \qquad \text{(Dual-OPT)}$$

with optimal value $-d^\star$ and same set of optimal solutions as the dual problem. As f is CPC, $f^{\star\vee}$ is also CPC (follows from §IV-A2 and [1, Proposition 8.20]). Furthermore, from Lemma 9(ii), $\delta_{\mathcal{C}}^\star$ is also CPC. Thus we can apply (Convex-DRS) to (Dual-OPT).

2) Applying Douglas-Rachford splitting to (Dual-OPT). By setting $g := f^{\star\vee}$ and $h := \delta_{\mathcal{C}}^{\star}$ in (21), we have the following Douglas-Rachford splitting algorithm for the dual problem:

$$\begin{aligned} \zeta_{n+1} &= \mathbf{prox}_{\gamma\delta_{\mathcal{C}}^{\star}}(\psi_n) \\ \xi_{n+1} &= \mathbf{prox}_{\gamma f^{\star\vee}}(2\zeta_{n+1} - \psi_n) \\ \psi_{n+1} &= \psi_n + \xi_{n+1} - \zeta_{n+1}. \end{aligned} \quad \text{(Dual-DRS)}$$

Using Lemma 9 and Lemma 8, we simplify the first two iterates of (Dual-DRS) as

$$\begin{aligned} &\zeta_{n+1} = \psi_n - \gamma\,\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\big(\gamma^{-1}\psi_n\big), \text{ and} \\ &\xi_{n+1} = 2\zeta_{n+1} - \psi_n + \gamma\,\mathbf{prox}_{\gamma^{-1}f}\big(-\gamma^{-1}(2\zeta_{n+1} - \psi_n)\big). \end{aligned}$$

Using these simplified iterates and introducing intermediate iterates $\tilde{y}_{n+1} = \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(\gamma^{-1}\psi_n\right)$ and $\tilde{x}_{n+1} = \mathbf{prox}_{\gamma^{-1}f}\big(-\gamma^{-1}(2\zeta_{n+1} - \psi_n)\big)$, we can write (Dual-DRS) as

$$\begin{aligned} \tilde{y}_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\big(\gamma^{-1}\psi_n\big) \\ \zeta_{n+1} &= \psi_n - \gamma\tilde{y}_{n+1} \\ \tilde{x}_{n+1} &= \mathbf{prox}_{\gamma^{-1}f}\big(-\gamma^{-1}(2\zeta_{n+1} - \psi_n)\big) \\ &= \mathbf{prox}_{\gamma^{-1}f}\big(-\gamma^{-1}(\psi_n - 2\gamma\tilde{y}_{n+1})\big) \\ \xi_{n+1} &= 2\zeta_{n+1} - \psi_n + \gamma\tilde{x}_{n+1} \\ &= \psi_n - 2\gamma\tilde{y}_{n+1} + \gamma\tilde{x}_{n+1} \\ \psi_{n+1} &= \psi_n + \xi_{n+1} - \zeta_{n+1} \\ &= \psi_n - \gamma\tilde{y}_{n+1} + \gamma\tilde{x}_{n+1}. \end{aligned}$$

Note that the iterates ζ_n and ξ_n do not have any explicit dependence, hence they can be removed. Furthermore, introduce a new iterate, $z_n = \frac{1}{\gamma}\psi_n - \tilde{x}_n$. Substituting $\psi_n := \gamma\left(z_n + \tilde{x}_n\right)$ in the iteration scheme above, we get

$$\begin{aligned} \tilde{y}_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(z_n + \tilde{x}_n\right) \\ \tilde{x}_{n+1} &= \mathbf{prox}_{\gamma^{-1}f}\left(-\left(z_n + \tilde{x}_n - 2\tilde{y}_{n+1}\right)\right) \\ &\overset{a)}{=} \mathbf{prox}_{\gamma^{-1}f}\left(-\left(z_{n+1} - \tilde{y}_{n+1}\right)\right) \\ z_{n+1} &= z_n + \tilde{x}_n - \tilde{y}_{n+1}, \end{aligned} \quad (22)$$

where $a)$ follows from (22).

Finally, we swap the order of of $\tilde{x}_{n+1}$ and z_{n+1} to obtain the correct dependency:

$$\begin{aligned} \tilde{y}_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(z_n + \tilde{x}_n\right) \\ z_{n+1} &= z_n + \tilde{x}_n - \tilde{y}_{n+1}, \\ \tilde{x}_{n+1} &= \mathbf{prox}_{\gamma^{-1}f}\left(\tilde{y}_{n+1} - z_{n+1}\right). \end{aligned}$$

We now substitute $\tilde{x}_n := x_{n+1}$, $\tilde{y}_n := y_n$, and $\frac{1}{\gamma} := \tilde{\gamma}$ in the iterates above to obtain a relaxed version (NC-ADMM):

$$\begin{aligned} x_{n+1} &= \mathbf{prox}_{\tilde{\gamma}f}\left(y_n - z_n\right) \\ y_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(z_n + x_{n+1}\right) \\ z_{n+1} &= z_n - y_{n+1} + x_{n+1}, \end{aligned} \quad \text{(Relaxed-NC-ADMM)}$$

which is similar to (NC-ADMM), except the projection is onto $\mathbf{conv}\,\mathcal{C}$ rather than onto $\mathcal{C}$.

3) Constructing (NC-ADMM) *from* (Relaxed-NC-ADMM). Now we discuss how we can arrive at (NC-ADMM) from (Relaxed-NC-ADMM). The first step requires the observation that (Relaxed-NC-ADMM) finds a minimizer of f over $\mathbf{conv}\,\mathcal{C}$. To see that, construct the dual of (Dual-OPT), which is

$$\text{maximize} \quad -\left(f^{\star\vee}\right)^{\star\vee}(x) - \delta_{\mathcal{C}}^{\star\star}(x)\,, \quad \text{(Double-Dual)}$$

where $x \in \mathbf{R}^n$ is the optimization variable. As both (Dual-OPT) and (Double-Dual) are convex optimization problems, strong duality usually holds (under constraint qualifications), where both problems have the same optimal value $-d^{\star}$. Now, $(f^{\star\vee})^{\star\vee} \overset{a)}{=} (f^{\vee\star})^{\star\vee} = \big((f^{\vee})^{\star\star}\big)^{\vee} \overset{b)}{=} f^{\vee\vee} \overset{c)}{=} f$,where $a)$ follows from $f^{\star\vee} = f^{\vee\star}$ for CPC function f [1, Proposition 13.20(v)], $b)$ follows from the fact that the biconjugate of a CPC function is equal to the function itself [19, Lemma 4.8], and $c)$ follows from the fact that applying reversal operation twice on a function returns the original function. Furthermore, $\delta_{\mathcal{C}}^{\star\star} = \delta_{\mathbf{conv}\,\mathcal{C}}$ from Lemma 9(i). Hence, the dual of (Double-Dual), written as a minimization problem, is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{conv}\,\mathcal{C}, \end{array}$$

where $x \in \mathbf{R}^n$ is the optimization variable with optimal value $d^{\star}$. So, under strong duality between (Dual-OPT) and (Double-Dual), (Relaxed-NC-ADMM) finds a minimizer of f over the set $\mathbf{conv}\,\mathcal{C}$, which appears in the projection step of (Relaxed-NC-ADMM). So, to solve the original problem (OPT) where we seek a minimizer of f over $\mathcal{C}$, an intuitive modification to (Relaxed-NC-ADMM) is replacing $\mathbf{conv}\,\mathcal{C}$ with $\mathcal{C}$ (hence $\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}$ with $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$), which results in (NC-ADMM). Roughly speaking, (NC-ADMM) is constructed by first relaxing the constraint set of the original problem to its convex hull, then applying the Douglas-Rachford splitting algorithm for the relaxed problem and finally restricting the resultant algorithm on the original constraint set.

The construction procedure also provides an alternative explanation behind why, when compared with exact solvers, (NC-ADMM) often achieves lower objective values in many numerical experiments performed in [2], [4], [5], [6]. In these works, these lower objective values are attributed to the superior performance of (NC-ADMM) in solving nonconvex problems based on empirical evidence. An alternative explanation could be that the heuristic is solving a modified dual problem, which, in the absence of strong duality, is guaranteed to yield an objective value that is smaller than or equal to that of the original problem.

C. *Convergence of* (NC-ADMM)

Now we comment on convergence properties of (NC-ADMM) in comparison with (NC-DRS).

• **Convergence to an optimal solution.** For (NC-DRS), the fixed point set of $R_{\gamma\partial\delta_{\mathcal{C}}}R_{\gamma\partial f}$ acts as a bridge between global minimizers of (OPT) and the fixed point set of the nonconvex Douglas-Rachford operator (Theorem 2). Though

(Convex-DRS) is equivalent to (Relaxed-NC-ADMM) under strong duality, no such equivalence seems to exist between (NC-DRS) and (NC-ADMM), because there is a strict duality gap between (OPT) and (Dual-OPT), and $\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}} \neq \tilde{\mathbf{\Pi}}_{\mathcal{C}}$. This lack of equivalence prevents connecting the fixed point set of the underlying (NC-ADMM) operator to global minimizers of (OPT) through the fixed point set of $R_{\gamma\partial\delta_{\mathcal{C}}} R_{\gamma\partial f}$.

• **Convergence to a point.** Furthermore, the lack of equivalence between (NC-DRS) and (NC-ADMM) makes it harder to comment analogously on convergence of (NC-ADMM) to a general point (not necessarily an optimal solution) as well. As shown in the proof of Theorem 3, establishing convergence to a general point for (NC-DRS) depends on the interrelationship between the nonconvex Douglas-Rachford operator and the nonconvex Peaceman-Rachford operator. Unfortunately, such relationship may break down for (NC-ADMM), because constructing such a relationship would require Moreau's decomposition to hold for nonconvex functions.

V. Future work

Future research directions include conducting numerical experiments to compare the performance of (NC-DRS) with (NC-ADMM).

References

[1] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[2] S. Diamond, R. Takapoui, and S. Boyd, "A general system for heuristic minimization of convex functions over non-convex sets," *Optimization Methods and Software*, vol. 33, no. 1, pp. 165–193, 2018.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[4] R. Takapoui, "The alternating direction method of multipliers for mixed-integer optimization applications," Ph.D. dissertation, Stanford University, 2017.

[5] R. Takapoui, N. Moehle, S. Boyd, and A. Bemporad, "A simple effective heuristic for embedded mixed-integer quadratic programming," *International Journal of Control*, pp. 1–11, 2017.

[6] ——, "A simple effective heuristic for embedded mixed-integer quadratic programming," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 5619–5625.

[7] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2909–2913, 2016.

[8] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, pp. 1–35, 2015.

[9] G. Li and T. K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.

[10] D. Gabay, "Applications of the method of multipliers to variational inequalities," *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems,*, 1983.

[11] S. Boyd, J. Duchi, and L. Vandenberghe, "Subgradients: Notes for EE364b, Stanford University, Spring 2014-15," http://web.stanford.edu/class/ee364b/lectures/subgradients_notes.pdf.

[12] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.

[13] K. R. Davidson and A. P. Donsig, *Real analysis and applications: theory in practice.* Springer Science & Business Media, 2009.

[14] B. T. Polyak, "Introduction to optimization. Translations series in mathematics and engineering," *Optimization Software*, 1987.

[15] H. Cornean, "Basic properties of limsup and liminf," http://people.math.aau.dk/~cornean/analyse2_F14/limsup-liminf.pdf, 2014.

[16] P. Combettes and H. J. Trussell, "Method of successive projections for finding a common point of sets in metric spaces," *Journal of optimization theory and applications*, vol. 67, no. 3, pp. 487–507, 1990.

[17] D. Davis and W. Yin, "Convergence rate analysis of several splitting schemes," in *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, 2016, pp. 115–163.

[18] E. K. Ryu and S. Boyd, "Primer on monotone operator methods," *Appl. Comput. Math*, vol. 15, no. 1, pp. 3–43, 2016.

[19] A. Beck, *First-Order Methods in Optimization*. SIAM, 2017, vol. 25.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.