NONCONVEX NOTIONS OF REGULARITY AND CONVERGENCE OF FUNDAMENTAL ALGORITHMS FOR FEASIBILITY PROBLEMS[*]

ROBERT HESSE[†] AND D. RUSSELL LUKE[†]

Abstract. We consider projection algorithms for solving (nonconvex) feasibility problems in Euclidean spaces. Of special interest are the method of alternating projections (AP) and the Douglas–Rachford algorithm (DR). In the case of convex feasibility, firm nonexpansiveness of projection mappings is a global property that yields global convergence of AP and for consistent problems DR. A notion of local subfirm nonexpansiveness with respect to the intersection is introduced for consistent feasibility problems. This, together with a coercivity condition that relates to the regularity of the collection of sets at points in the intersection, yields local linear convergence of AP for a wide class of nonconvex problems and even local linear convergence of nonconvex instances of the DR algorithm.

1. Introduction. In the last decade there has been significant progress in the understanding of convergence of algorithms for solving generalized equations for non-monotone operators, in particular those arising from variational problems such as minimization or maximization of functions, feasibility, variational inequalities, and minimax problems. Early efforts focused on the proximal point algorithm [32, 23, 12] and notions of *(co)hypomonotonicity*, which is closely related to *prox-regularity* of functions [33]. Other works have focused on *metric regularity* and its refinements [2, 3, 24]. Proximal-type algorithms have been studied for functions satisfying the Kurdyka–Łojasiewicz inequality in [4]. In a more limited context, the method of alternating projections (AP) has been investigated in [27, 8] with the aim of formulating *dual* characterizations of regularity requirements for linear convergence via the *normal cone* and its variants.

The framework we present here generalizes the tools for the analysis of fixed-point iterations of operators that violate the classical property of firm nonexpansiveness in some quantifiable fashion. As such, our approach is more closely related to the ideas of [32] and hypomonotonicity through the resolvent mapping; however, our application to AP bears a direct resemblance to the more classical *primal* characterizations of regularity described in [5] (in particular what they call "linear regularity"). There are also some parallels between what we call (S, ε)-firm nonexpansiveness and ε-enlargements of maximal monotone operators [11], though this is beyond the scope of this paper. Our goal is to introduce the essential tools we make use of with a cursory treatment of the connections to other concepts in the literature, and to apply these tools to the AP

[†]Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 37083 Göttingen, Germany (hesse@math.uni-goettingen.de, r.luke@math.uni-goettingen.de).

and Douglas–Rachford algorithms, comparing our results to the best known results at this time.

We review the basic definitions and classical definitions and results below. In section 2 we introduce our relaxations of the notions of firm nonexpansiveness and set regularity. Our main abstract result concerning fixed-point iterations of mappings that violate the classical firm nonexpansive assumptions is in section 3. We specialize in subsequent subsections to AP and the Douglas–Rachford algorithm. Our statement of linear convergence of AP is as general as the results reported in [8], with more elementary proofs, although our estimates for the radius of convergence are more conservative. The results on local linear convergence of nonconvex instances of Douglas–Rachford are new and provide some evidence supporting the conjecture that, asymptotically, Douglas–Rachford converges more slowly (albeit still linearly) than simple alternating projections. Our estimates of the rate of convergence for both AP and Douglas–Rachford are not optimal but to our knowledge are the most general to date.

We also show that strong regularity conditions on the collection of *affine* sets are in fact *necessary* for linear convergence of iterates of the Douglas–Rachford algorithm to the intersection, in contrast to AP, where the same conditions are sufficient but not necessary [8]. This may seem somewhat spurious to experts since, as is well known, the Douglas–Rachford iterates themselves are not of interest, but rather their shadows or projections onto one of the sets [28, 7] are. Indeed, in the convex setting where local is global, the shadows of the iterates of the Douglas–Rachford algorithm could converge even though the iterates themselves diverge. This happens in particular when the sets do not intersect but have instead *best approximation points* [7, Theorem 3.13]. The nonconvex setting is much less forgiving, however. Indeed, existence of local best approximation points does not guarantee convergence of the shadows to best approximation points in the nonconvex setting [29], and so convergence of the sequence itself is essential. As nonconvex settings are our principal interest, we focus on convergence of the iterates of the Douglas–Rachford algorithm instead of the shadows, in particular convergence of these iterates to the intersection of collections of sets. We leave a fuller examination of the shadow sequences to future work.

1.1. Basics/notation. $\mathbf{E}$ is a Euclidean space. We denote the closed unit ball centered at the origin by $\mathbb{B}$ and the ball of radius δ centered at $\bar{x} \in \mathbf{E}$ by $\mathbb{B}_\delta(\bar{x}) := \{x \in \mathbf{E} | \|x - \bar{x}\| \le \delta\}$. When the δ-ball is centered at the origin we write $\mathbb{B}_\delta$. The notation $\rightrightarrows$ indicates that this mapping in general is *multivalued*. The composition of two *multivalued* mappings T_1, T_2 is pointwise defined by $T_2\, T_1 x = \cup_{y \in T_1 x} T_2 y$. A nonempty subset K of $\mathbf{E}$ is a cone if for all $\lambda > 0$, $\lambda K := \{\lambda k \mid k \in K\} \subseteq K$. The smallest cone containing a set $\Omega \subset \mathbf{E}$ is denoted $\operatorname{cone}(S)$.

DEFINITION 1.1. *Let $\Omega \subset \mathbf{E}$ be nonempty, $x \in \mathbf{E}$. The distance of x to Ω is defined by*

$$(1.1) \qquad d(x, \Omega) := \inf_{y \in \Omega} \|x - y\|.$$

DEFINITION 1.2 (projectors/reflectors). *Let $\Omega \subset \mathbf{E}$ be nonempty and $x \in \mathbf{E}$. The (possibly empty) set of all best approximation points from x to Ω denoted $P_\Omega(x)$ (or $P_\Omega x$) is given by*

$$(1.2) \qquad P_\Omega(x) := \{y \in \Omega \mid \|x - y\| = d(x, \Omega)\}.$$

The mapping $P_\Omega \rightrightarrows \Omega$ is called the metric projector, *or* projector, *onto Ω. We call an element of $P_\Omega(x)$ a* projection. *The reflector $R_\Omega : \mathbf{E} \rightrightarrows \mathbf{E}$ to the set Ω is defined as*

$$(1.3) \qquad\qquad R_\Omega x := 2P_\Omega x - x$$

for all $x \in \mathbf{E}$.

Since we are on a Euclidean space $\mathbf{E}$ convexity and closedness of a subset $C \subset \mathbf{E}$ is sufficient for the projector (respectively, the reflector) to be single-valued. Closedness of a set Ω suffices for the set Ω being *proximinal,* i.e., $P_C x \neq \emptyset$ for all $x \in \mathbf{E}$. (For a modern treatment see [6, Corollary 3.13].)

DEFINITION 1.3 (method of alternating projections). *For two sets $A, B \subset \mathbf{E}$ we call the mapping*

$$(1.4) \qquad\qquad T_{AP}x = P_A P_B x$$

the AP operator. The AP algorithm is the corresponding Picard iteration

$$(1.5) \qquad\qquad x_{n+1} \in T_{AP}x_n, \quad n = 0, 1, 2, \dots,$$

for x_0 given.

DEFINITION 1.4 (averaged alternating reflections/Douglas–Rachford). *For two sets $A, B \subset \mathbf{E}$ we call the mapping*

$$(1.6) \qquad\qquad T_{DR}x = \frac{1}{2}\left(R_A R_B x + x\right),$$

the Douglas–Rachford operator. We call the Douglas–Rachford algorithm, *or simply* Douglas–Rachford, *the corresponding Picard iteration*

$$(1.7) \qquad\qquad x_{n+1} \in T_{DR}x_n, \quad n = 0, 1, 2, \dots,$$

for x_0 given.

Example 1.5. The following easy examples will appear throughout this work and serve to illustrate the regularity concepts we introduce and the convergence behavior of the algorithms under consideration.

(i) Two lines in $\mathbb{R}^2$:

$$A = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\right\} \subset \mathbb{R}^2,$$
$$B = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2\right\} \subset \mathbb{R}^2.$$

We will see that AP and Douglas–Rachford converge with a linear rate to the intersection.

(ii) Two lines in $\mathbb{R}^3$:

$$A = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 = 0, x_3 = 0\right\} \subset \mathbb{R}^3,$$
$$B = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 = x_2, x_3 = 0\right\} \subset \mathbb{R}^3.$$

After the first iteration step AP shows exactly the same convergence behavior as in the first example. Douglas–Rachford does not converge to $\{0\} = A \cap B$. All iterates from starting points on the line $\{t(0, 0, 1) \mid t \in \mathbb{R}\}$ are fixed points of the Douglas Rachford operator. On the other hand, iterates from starting points in $A + B$ stay in $A + B$, and the case then reduces to example (i).

(iii) A line and a ball intersecting in one point:

$$A = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0 \right\} \subset \mathbb{R}^2,$$
$$B = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + (x_2 - 1)^2 \leq 1 \right\}.$$

AP converges to the intersection, but not with a linear rate. Douglas–Rachford has fixed points that lie outside the intersection.

(iv) A *cross* and a subspace in $\mathbb{R}^2$:

$$A = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R},$$
$$B = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2 \right\}.$$

This example relates to the problem of *sparse-signal recovery*. Both AP and Douglas–Rachford converge globally to the intersection $\{0\} = A \cap B$, though A is nonconvex. The convergence of both methods is covered by the theory built up in this work.

(v) A circle and a line:

$$A = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_2 = \sqrt{2}/2 \right\} \subset \mathbb{R}^2,$$
$$B = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1 \right\}.$$

This example is of our particular interest, since it is a simple model case of the *phase retrieval problem*. So far the only *direct* nonconvex convergence results for Douglas–Rachford are related to this model case; see [1, 10]. Local convergence of AP is covered by [27, 8] as well as by the results in this work.

1.2. Classical results for (firmly) nonexpansive mappings. To begin, we recall (firmly) nonexpansive mappings and their natural association with projectors and reflectors on convex sets. We later extend this notion to nonconvex settings where the algorithms involve set-valued mappings.

DEFINITION 1.6. *Let $\Omega \subset \mathbf{E}$ be nonempty. $T : \Omega \to \mathbf{E}$ is called nonexpansive if*

$$\text{(1.8)} \qquad \qquad \|Tx - Ty\| \leq \|x - y\|$$

holds for all $x, y \in \Omega$.

$T : \Omega \to \mathbf{E}$ *is called* firmly nonexpansive *if*

$$\text{(1.9)} \qquad \|Tx - Ty\|^2 + \|(Id - T)x - (Id - T)y\|^2 \leq \|x - y\|^2$$

holds for all $x, y \in \Omega$.

LEMMA 1.7 (Proposition 4.2 in [6]). *Let $\Omega \subset \mathbf{E}$ be nonempty and let $T : \Omega \to \mathbf{E}$. The following are equivalent:*

(i) *T is firmly nonexpansive on Ω.*

(ii) *$2T - Id$ is nonexpansive on Ω.*

(iii) *$\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle$ for all $x, y \in \Omega$.*

THEOREM 1.8 (best approximation property, convex case). *Let $C \subset \mathbf{E}$ be nonempty and convex, $x \in \mathbf{E}$, and $\bar{x} \in C$. The point $\bar{x}$ is the projection $\bar{x} = P_C(x)$ if and only if*

$$\text{(1.10)} \qquad \qquad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0 \quad \forall\, y \in C.$$

If C is an affine subspace, then (1.10) *holds with equality.*

Proof. For (1.10) see Theorem 3.14 of [6], while equality follows from Corollary 3.20 of [6].   ☐

THEOREM 1.9 ((firm) nonexpansiveness of projectors/reflectors). *Let C be a closed, nonempty, and convex set. The projector $P_C : \mathbf{E} \to \mathbf{E}$ is a firmly nonexpansive mapping and hence the reflector R_C is nonexpansive. If, in addition, C is an affine subspace, then the following conditions hold:*

(i) *P_C is firmly nonexpansive with equality, i.e.,*

$$(1.11) \quad \|P_C x - P_C y\|^2 + \|(Id - P_C)x - (Id - P_C)y\|^2 = \|x - y\|^2$$

for all $x \in \mathbf{E}$.
(ii) *For all $x \in \mathbf{E}$*

$$(1.12) \qquad\qquad \|R_C x - c\| = \|x - c\|$$

holds for all $c \in C$.

Proof. For the first part of the statement see [15, Theorems 4.1 and 5.5], [19, Chapter 12], [20, Propositions 3.5 and 11.2], and [35, Lemma 1.1]. The well-known refinement for affine subspaces follows by a routine application of the definitions and Theorem 1.8.   ☐

2. (S, ε)-firm nonexpansiveness. Up to this point, the results have concerned only convex sets, and hence the projector and related algorithms have all been single-valued. In what follows, we generalize to nonconvex sets and therefore allow multi-valuedness of the projectors:

LEMMA 2.1. *Let $A, B \subset \mathbf{E}$ be nonempty and closed. Let $x \in \mathbf{E}$. For any element $x_+ \in T_{DR}x$ there is a point $\tilde{x} \in R_A R_B x$ such that $x_+ = \frac{1}{2}(\tilde{x} + x)$. Moreover, T_{DR} satisfies the following properties:*

(i)

$$(2.1) \quad \|x_+ - y_+\|^2 + \|(x - x_+) - (y - y_+)\|^2 = \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|\tilde{x} - \tilde{y}\|^2,$$

where x and y are elements of $\mathbf{E}$, x_+ and y_+ are elements of $T_{DR}x$ and $T_{DR}y$, respectively, and $\tilde{x} \in R_A R_B x$ and $\tilde{y} \in R_A R_B y$ are the corresponding points satisfying $x_+ = \frac{1}{2}(\tilde{x} + x)$ and $y_+ = \frac{1}{2}(\tilde{y} + y)$.
(ii) *For all $x \in \mathbf{E}$*

$$(2.2) \qquad\qquad T_{DR}x = \{P_A(2z - x) - z + x \mid z \in P_B x\}.$$

Proof. By Definition 1.4

$$
\begin{aligned}
x_+ &\in\ T_{DR}x \\
\iff\qquad x_+ &\in\ \tfrac{1}{2}(R_A R_B x + x) \\
\iff\quad 2x_+ - x &\in\ R_A R_B x.
\end{aligned}
$$

Defining $\tilde{x} = 2x_+ - x$ yields $x_+ = \frac{1}{2}(\tilde{x} + x)$, where $\tilde{x} \in R_A R_B x$.

(i) For $x_+ \in T_{DR}x$ (respectively, $y_+ \in T_{DR}y$) choose $\tilde{x} \in R_A R_B x$ (respectively, $\tilde{y}$) such that $x_+ = (\tilde{x} + x)/2$ (respectively, y_+). Then

$$\|x_+ - y_+\|^2 + \|(x - x_+) - (y - y_+)\|^2$$
$$= \left\|\frac{1}{2}\tilde{x} + \frac{1}{2}x - \frac{1}{2}\tilde{y} - \frac{1}{2}y\right\|^2 + \left\|\frac{1}{2}x - \frac{1}{2}\tilde{x} - \frac{1}{2}y + \frac{1}{2}\tilde{y}\right\|^2$$
$$= \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|\tilde{x} - \tilde{y}\|^2 + \frac{1}{2}\langle\tilde{x} - \tilde{y}, x - y\rangle - \frac{1}{2}\langle\tilde{x} - \tilde{y}, x - y\rangle$$
$$= \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|\tilde{x} - \tilde{y}\|^2.$$

(ii) This follows easily from the definitions. Indeed, we represent $v \in R_B x$ as $v = 2z - x$ for $z \in P_B x$ so that

$$T_{DR}x = \left\{\frac{1}{2}(R_A v + x) \mid v \in R_B x\right\}$$
$$= \left\{\frac{1}{2}(R_A(2z - x) + x) \mid z \in P_B x\right\}$$
$$= \left\{\frac{1}{2}(2P_A(2z - x) - (2z - x) + x) \mid z \in P_B x\right\}$$
$$= \{P_A(2z - x) - z + x \mid z \in P_B x\}. \qquad \square$$

Remark 2.2. In the case where A and B are convex, then as a consequence of Lemma 2.1(i) and the fact that the reflector R_Ω onto a convex set Ω is nonexpansive, we recover the well-known fact that T_{DR} is firmly nonexpansive and (2.1) reduces to

(2.3) $$\|T_{DR}x - T_{DR}y\|^2 + \|(Id - T_{DR})x - (Id - T_{DR})y\|^2$$
$$= \frac{1}{2}\|x - y\|^2 + \frac{1}{2}\|R_A R_B x - R_A R_B y\|^2,$$

while (2.2) reduces to

(2.4) $$T_{DR}x = x + P_A R_B x - P_B x.$$

We define next an analogue to firm nonexpansiveness in the nonconvex case with respect to a set S.

DEFINITION 2.3 ((S, ε)-(firmly) nonexpansive mappings). *Let D and S be nonempty subsets of $\mathbf{E}$ and let T be a (multivalued) mapping from D to $\mathbf{E}$.*
(i) *T is called (S, ε)-nonexpansive on D if*

(2.5) $$\|x_+ - \bar{x}_+\| \leq \sqrt{1 + \varepsilon}\|x - \bar{x}\|$$
$$\forall x \in D, \ \forall \bar{x} \in S, \ \forall x_+ \in Tx, \ \forall \bar{x}_+ \in T\bar{x}.$$

If (2.5) holds with $\varepsilon = 0$, then we say that T is S-nonexpansive on D.
(ii) *T is called (S, ε)-firmly nonexpansive on D if*

(2.6) $$\|x_+ - \bar{x}_+\|^2 + \|(x - x_+) - (\bar{x} - \bar{x}_+)\|^2 \leq (1 + \varepsilon)\|x - \bar{x}\|^2$$
$$\forall x \in D, \ \forall \bar{x} \in S, \ \forall x_+ \in Tx, \ \forall \bar{x}_+ \in T\bar{x}.$$

If (2.6) holds with $\varepsilon = 0$, then we say that T is S-firmly nonexpansive on D.

Note that, as with (firmly) nonexpansive mappings, the mapping T need not be a self-mapping from D to itself. In the special case where $S = FixT$, mappings satisfying (2.5) are also called *quasi (firmly) nonexpansive* [6]. Quasi nonexpansiveness is a restriction of another well-known concept, Fejér monotonicity, to $FixT$. Equation (2.6) is a relaxed version of firm nonexpansiveness (1.9). The aim of this work is to expand the theory for projection methods (and in particular AP and Douglas–Rachford) to the setting where one (or more) of the sets is nonconvex. The classical (firmly) nonexpansive operator on D is $(D, 0)$-(firmly) nonexpansive on D.

Analogous to the relation between firmly nonexpansive mappings and averaged mappings (see [6, Chapter 4] and references therein) we have the following relationship between (S, ε)-firmly nonexpansive mappings and their 1/2-averaged companion mapping.

LEMMA 2.4 (1/2-averaged mappings). *Let $D, S \subset \mathbf{E}$ be nonempty and $T : D \rightrightarrows \mathbf{E}$. The following are equivalent:*

(i) *T is (S, ε)-firmly nonexpansive on D.*

(ii) *The mapping $\widetilde{T} : D \rightrightarrows \mathbf{E}$ given by*

$$\widetilde{T}x := (2Tx - x) \quad \forall x \in D$$

is $(S, 2\varepsilon)$-nonexpansive on D, i.e., T can be written as

$$(2.7) \qquad Tx = \frac{1}{2}\left(x + \widetilde{T}x\right) \quad \forall x \in D.$$

Proof. For $x \in D$ choose $x_+ \in Tx$. Observe that by the definition of $\widetilde{T}$, there is a corresponding $\tilde{x} \in \widetilde{T}x$ such that $x_+ = \frac{1}{2}(x + \tilde{x})$, which is just formula (2.7). Let z be any point in S and select any $z_+ \in Tz$. Then

$$
\begin{aligned}
\|x_+ &- z_+\|^2 + \|x - x_+ - (z - z_+)\|^2 \\
&= \left\|\tfrac{1}{2}(x + \tilde{x}) - \tfrac{1}{2}(z + \tilde{z})\right\|^2 + \left\|\tfrac{1}{2}(x - \tilde{x}) - \tfrac{1}{2}(z - \tilde{z})\right\|^2 \\
&= \tfrac{1}{4}\left[\|x - z\|^2 + 2\langle x - z, \tilde{x} - \tilde{z}\rangle + \|\tilde{x} - \tilde{z}\|^2\right] \\
&\quad + \tfrac{1}{4}\left[\|x - z\|^2 - 2\langle x - z, \tilde{x} - \tilde{z}\rangle + \|\tilde{x} - \tilde{z}\|^2\right] \\
&= \tfrac{1}{2}\|x - z\|^2 + \tfrac{1}{2}\|\tilde{x} - \tilde{z}\|^2 \\
&\overset{!}{\leq} \tfrac{1}{2}\|x - z\|^2 + \tfrac{1}{2}(1 + 2\varepsilon)\|x - z\|^2 \\
&= (1 + \varepsilon)\|x - z\|^2,
\end{aligned}
$$

where the inequality holds if and only if $\widetilde{T}$ is $(S, 2\varepsilon)$-nonexpansive. By definition, it then holds that T is (S, ε)-firmly nonexpansive if and only if $\widetilde{T}$ is $(S, 2\varepsilon)$-nonexpansive, as claimed. ▯

We state the following theorem to suggest that the framework presented in this work can be extended to a more general setting, for example, the adaptive framework discussed in [5]. It shows that the (S, ε)-firm nonexpansiveness is preserved under convex combination of operators.

THEOREM 2.5. *Let T_1 be (S, ε_1)-firmly nonexpansive and T_2 be (S, ε_2)-firmly nonexpansive on D. The convex combination $\lambda T_1 + (1 - \lambda)T_2$ is (S, ε)-firmly nonexpansive on D, where $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$.*

Proof. Let $x, y \in D$. Let

$$x_+ \in \lambda T_1 x + (1 - \lambda) T_2 x \quad \text{and}$$
$$y_+ \in \lambda T_1 y + (1 - \lambda) T_2 y$$
$$\Rightarrow \quad x_+ = \lambda x_+^{(1)} + (1 - \lambda) x_+^{(2)}, \quad \text{where } x_+^{(1)} \in T_1 x, \ x_+^{(2)} \in T_2 x,$$
$$y_+ = \lambda y_+^{(1)} + (1 - \lambda) y_+^{(2)}, \quad \text{where } y_+^{(1)} \in T_1 y, \ y_+^{(2)} \in T_2 y.$$

By Lemma 2.4(ii) one has nonexpansiveness of the mappings given by $2T_1 x - x$ and $2T_2 x - x$, $x \in D$, that is,

$$\left\| \left[2x_+^{(1)} - x \right] - \left[2y_+^{(1)} - y \right] \right\| \le \sqrt{1 + 2\varepsilon_1} \, \|x - y\| ,$$
$$\left\| \left[2x_+^{(2)} - x \right] - \left[2y_+^{(2)} - y \right] \right\| \le \sqrt{1 + 2\varepsilon_2} \, \|x - y\| .$$

This implies

$$
\begin{aligned}
\| (2x_+ &- x) - (2y_+ - y) \| \\
&= \left\| \left(2 \left[\lambda x_+^{(1)} + (1 - \lambda) x_+^{(2)} \right] - x \right) - \left(2 \left[\lambda y_+^{(1)} + (1 - \lambda) y_+^{(2)} \right] - y \right) \right\| \\
&= \left\| \lambda \left(\left[2x_+^{(1)} - x \right] - \left[2y_+^{(1)} - y \right] \right) - (1 - \lambda) \left(\left[2x_+^{(2)} - x \right] - \left[2y_+^{(2)} - y \right] \right) \right\| \\
&\le \lambda \left\| \left[2x_+^{(1)} - x \right] - \left[2y_+^{(1)} - y \right] \right\| + (1 - \lambda) \left\| \left[2x_+^{(2)} - x \right] - \left[2y_+^{(2)} - y \right] \right\| \\
&\le \sqrt{1 + 2\varepsilon} \, \|x - y\| .
\end{aligned}
$$

Now using Lemma 2.4(i) the proof is complete. ⧠

2.1. Regularity of sets. To ensure property (2.6) for the projector and the Douglas–Rachford operator, we determine the inheritance of the regularity of the projector and reflectors from the regularity of the sets A and B upon which we project. We begin with some established notions of set regularity and introduce a new, weaker form that will be central to our analysis.

DEFINITION 2.6 (prox-regularity). *A nonempty (locally) closed set $\Omega \subset \mathbf{E}$ is prox-regular at a point $\bar{x} \in \Omega$ if the projector P_Ω is single-valued around $\bar{x}$.*

What we take as the definition of prox-regularity actually follows from the equivalence of prox-regularity of sets as defined in [33, Definition 1.1] and the single-valuedness of the projection operator on neighborhoods of the set [33, Theorem 1.3].

DEFINITION 2.7 (normal cones). *The proximal normal cone $N_\Omega^P(\bar{x})$ to a set $\Omega \subset \mathbf{E}$ at a point $\bar{x} \in \Omega$ is defined by*

$$(2.8) \qquad\qquad N_\Omega^P(\bar{x}) := \operatorname{cone}(P_\Omega^{-1}(\bar{x}) - \bar{x}).$$

The limiting normal cone $N_\Omega(\bar{x})$ is defined as any vector that can be written as the limit of proximal normals; that is, $\overline{v} \in N_\Omega(\bar{x})$ if and only if there exist sequences $(x_k)_{k \in \mathbb{N}}$ in Ω and $(v_k)_{k \in \mathbb{N}}$ in $N_\Omega^P(x_k)$ such that $x_k \to \bar{x}$ and $v_k \to \overline{v}$.

The construction of the *limiting normal cone* goes back to Mordukhovich (see [34, Chapter 6, commentary]).

PROPOSITION 2.8 (Mordukhovich). *The limiting normal cone or Mordukhovich normal cone is the smallest cone satisfying the two properties*
1. $P_\Omega^{-1}(\bar{x}) \subseteq (I + N_\Omega)(\bar{x})$, *where* $P_\Omega^{-1}(\bar{x})$ *is the preimage set of* $\bar{x}$ *under* P_Ω;
2. *for any sequence* $x_i \to \bar{x}$ *in* Ω *any limit of a sequence of normals* $v_i \in N_\Omega(x_i)$ *must lie in* $N_\Omega(\bar{x})$.

DEFINITION 2.9 ((ε, δ)-(sub)regularity).
(i) *A nonempty set* $\Omega \subset \mathbf{E}$ *is* (ε, δ)-*subregular at* $\hat{x}$ *with respect to* $S \subset \mathbf{E}$ *if*

$$(2.9) \qquad \langle v_x, \bar{x} - x \rangle \leq \varepsilon \, \|v_x\| \, \|\bar{x} - x\|$$

holds for all $x \in \mathbb{B}_\delta(\hat{x}) \cap \Omega$, $\bar{x} \in S \cap \mathbb{B}_\delta(\hat{x})$, $v_x \in N_\Omega^P(x)$. *We simply say* Ω *is* (ε, δ)-*subregular at* $\hat{x}$ *if* $S = \{\hat{x}\}$.
(ii) *If* $S = \Omega$ *in* (i), *then we say that the set* Ω *is* (ε, δ)-*regular at* $\hat{x}$.
(iii) *If for all* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that* (2.9) *holds for all* $x, \bar{x} \in \mathbb{B}_\delta(\bar{x}) \cap \Omega$ *and* $v_x \in N_\Omega(x)$, *then* Ω *is said to be* superregular.

The definition of (ε, δ)-regularity was introduced in [9, Definition 9.1] and is a generalization of the notion of superregularity introduced in [27, Definition 4.3]. More details about (ε, δ)-regularity can be seen in [9]. Of particular interest is the following proposition. Preparatory to this, we remind readers of another well-known type of regularity, *Clarke regularity*. To avoid introducing the Féchet normal which is conventionally used to define Clarke regularity, we follow [27], which uses proximal normals.

DEFINITION 2.10 (Clarke regularity). *A nonempty (locally) closed set* $\Omega \subset \mathbf{E}$ *is Clarke regular at a point* $\bar{x} \in \Omega$ *if for all* $\varepsilon > 0$, *any two points* u, z *close enough to* $\bar{x}$ *with* $z \in \Omega$, *and any point* $y \in P_\Omega(u)$, *satisfy* $\langle z - \bar{x}, u - y \rangle \leq \varepsilon \|z - \bar{x}\| \|u - y\|$.

PROPOSITION 2.11 (prox-regular implies superregular, Proposition 4.9 of [27]). *If a closed set* $\Omega \subset \mathbf{E}$ *is prox-regular at a point in* Ω, *then it is superregular at that point. If a closed set* $\Omega \subset \mathbf{E}$ *is superregular at a point in* Ω, *then it is Clarke regular at that point.*

Superregularity is something between Clarke regularity and amenability or prox-regularity. (ε, δ)-regularity is weaker still than Clarke regularity (and hence superregularity), as the next example shows.

Remark 2.12. (ε, δ)-regularity does not imply Clarke regularity
Proof.

$$(2.10) \qquad \Omega := \left\{ (x_1, x_2) \in \mathbb{R}^2 \,\middle|\, \begin{array}{ll} x_2 \leq -x_1 & \text{if } x_1 \leq 0 \\ x_2 \leq 0 & \text{if } x_1 > 0 \end{array} \right\}.$$

For any

$$x_- \in \partial_- \Omega := \{(x_1, x_2) \mid x_2 = -x_1, \ x_1 < 0\} \quad \text{and}$$
$$x_+ \in \partial_+ \Omega := \{(x_1, 0) \mid x_1 > 0\}$$

one has

$$N_\Omega(x_-) = \left\{ (\lambda, \lambda) \mid \lambda \in \mathbb{R}^+ \right\},$$
$$N_\Omega(x_+) = \left\{ (0, \lambda) \mid \lambda \in \mathbb{R}^+ \right\},$$

which implies $N_\Omega(0) = N_\Omega(x_-) \cup N_\Omega(x_+)$. Since $N_\Omega^P(0) = \{0\}$ the set Ω is not Clarke regular at 0. Define $\nu_- = (\sqrt{2}/2, \sqrt{2}/2) \in N_\Omega(x_-)$, $\nu_+ = (0, 1) \in N_\Omega(x_+)$ and note

$$\langle \nu_-, 0 - x_- \rangle = 0 \quad \text{and} \quad \langle \nu_+, 0 - x_+ \rangle = 0$$

to show that Ω is $(0, \infty)$-subregular at 0. By the use of these two inequalities one now has

$$\langle \nu_-, x_+ - x_- \rangle = \langle \nu_-, x_+ \rangle \leq \sqrt{2}/2 \, \|x_+\| \,,$$
$$\langle \nu_+, x_- - x_+ \rangle = \langle \nu_+, x_- \rangle \leq \sqrt{2}/2 \, \|x_-\|$$

and hence Ω is $\left(\frac{\sqrt{2}}{2}, \infty\right)$-regular. $\quad \square$

Remark 2.13 (Example 1.5(iv) revisited). The set

(2.11) $$A := \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$$

is a particularly easy pathological set that illustrates the distinction between our new notion of subregularity and previous notions found in the literature. Note that for $x_1 \in \mathbb{R} \times \{0\}$, $N_A(x_1) = N_A^P(x_1) = \{0\} \times \mathbb{R}$ and for $x_2 \in \{0\} \times \mathbb{R}$, $N_A(x_2) = N_A^P(x_2) = \mathbb{R} \times \{0\}$ and that $N_A(0) = A$ and $N_A^P(0) = 0$, which implies that at the origin A is not Clarke regular and therefore is neither superregular nor prox-regular there. In fact, it is not even (ε, δ)-regular at the origin for any $\varepsilon < 1$ and any $\delta > 0$. The set A is, however, $(0, \infty)$-*sub*regular at $\{0\}$. Indeed, for any $x_1 \in \mathbb{R} \times \{0\}$ one has $\nu_1 \in N_A(x_1) = \{0\} \times \mathbb{R}$ and therefore $\langle \nu_1, x_1 - 0 \rangle = 0$. Analogously for $x_2 \in \{0\} \times \mathbb{R}$, $\nu_2 \in N_A(x_2) = \mathbb{R} \times \{0\}$, and it follows that $\langle \nu_2, x_2 - 0 \rangle = 0$, which shows that A is $(0, \infty)$-subregular at 0.

2.2. Projectors and reflectors. We show in this section how (S, ε)(firm) non-expansiveness of projectors and reflectors is a consequence of (sub)regularity of the underlying sets.

THEOREM 2.14 (projectors and reflectors onto (ε, δ)-subregular sets). *Let $\Omega \subset \mathbf{E}$ be nonempty closed and (ε, δ)-subregular at $\hat{x}$ with respect to $S \subseteq \Omega \cap \mathbb{B}_\delta(\hat{x})$ and define*

(2.12) $$U := \{x \in \mathbf{E} \mid P_\Omega x \subset \mathbb{B}_\delta(\hat{x})\} \,.$$

  (i) *The projector P_Ω is $(S, \tilde{\varepsilon}_1)$-nonexpansive on U, that is,*
     $(\forall x \in U) \, (\forall x_+ \in P_\Omega x) \, (\forall \bar{x} \in S)$

(2.13) $$\|x_+ - \bar{x}\| \leq \sqrt{1 + \tilde{\varepsilon}_1} \, \|x - \bar{x}\| \,,$$

     where $\tilde{\varepsilon}_1 := 2\varepsilon + \varepsilon^2$.
  (ii) *The projector P_Ω is $(S, \tilde{\varepsilon}_2)$-firmly nonexpansive on $\mathbb{B}_\delta(\hat{x})$, that is, $(\forall x \in U)$
     $(\forall x_+ \in P_\Omega x) \, (\forall \bar{x} \in S)$*

(2.14) $$\|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \leq (1 + \tilde{\varepsilon}_2) \, \|x - \bar{x}\|^2 \,,$$

     where $\tilde{\varepsilon}_2 := 2\varepsilon + 2\varepsilon^2$.
  (iii) *The reflector R_Ω is $(S, \tilde{\varepsilon}_3)$-nonexpansive on $\mathbb{B}_\delta(\hat{x})$, that is,*
     $(\forall x \in U) \, (\forall x_+ \in R_\Omega x) \, (\forall \bar{x} \in S)$

(2.15) $$\|x_+ - \bar{x}\| \leq \sqrt{1 + \tilde{\varepsilon}_3} \, \|x - \bar{x}\| \,,$$

     where $\tilde{\varepsilon}_3 := 4\varepsilon + 4\varepsilon^2$.

Proof.
  (i) The projector is nonempty since Ω is closed. Then by the Cauchy–Schwarz inequality

(2.16) $$\|x_+ - \bar{x}\|^2 = \langle x - \bar{x}, x_+ - \bar{x} \rangle + \langle x_+ - x, x_+ - \bar{x} \rangle$$
$$\leq \|x - \bar{x}\| \, \|x_+ - \bar{x}\| + \langle x_+ - x, x_+ - \bar{x} \rangle.$$

Now for $x \in U$ we have also that $x_+ \in \mathbb{B}_\delta(\hat{x})$ and thus, by the definition of (ε, δ)-subregularity with respect to S, $(\forall x \in U)$ $(\forall x_+ \in P_\Omega x)$ $(\forall \bar{x} \in S)$

$$\begin{aligned}
(2.17) \qquad \langle x_+ - x, x_+ - \bar{x} \rangle &\leq \varepsilon \left\| x - x_+ \right\| \left\| x_+ - \bar{x} \right\| \\
&\leq \varepsilon \left\| x - \bar{x} \right\| \left\| x_+ - \bar{x} \right\|.
\end{aligned}$$

Combining this with (2.16) yields $(\forall x \in U)$ $(\forall x_+ \in P_\Omega x)$ $(\forall \bar{x} \in S)$

$$\begin{aligned}
\left\| x_+ - \bar{x} \right\| &\leq (1 + \varepsilon) \left\| x - \bar{x} \right\| \\
&= \sqrt{1 + (2\varepsilon + \varepsilon^2)} \left\| x - \bar{x} \right\|
\end{aligned}$$

as claimed.

(ii) Expanding and rearranging the norm yields $(\forall x \in U)$ $(\forall x_+ \in P_\Omega x)$ $(\forall \bar{x} \in S)$

(2.18)
$$\begin{aligned}
&\left\| x_+ - \bar{x} \right\|^2 + \left\| x - x_+ \right\|^2 \\
&= \left\| x_+ - \bar{x} \right\|^2 + \left\| x - \bar{x} + \bar{x} - x_+ \right\|^2 \\
&= \left\| x_+ - \bar{x} \right\|^2 + \left\| x - \bar{x} \right\|^2 + 2\langle x - \bar{x}, \bar{x} - x_+ \rangle + \left\| x_+ - \bar{x} \right\|^2 \\
&= 2 \left\| x_+ - \bar{x} \right\|^2 + \left\| x - \bar{x} \right\|^2 + 2 \underbrace{\langle x_+ - \bar{x}, \bar{x} - x_+ \rangle}_{= -\left\| x_+ - \bar{x} \right\|^2} + 2 \underbrace{\langle x - x_+, \bar{x} - x_+ \rangle}_{\leq \varepsilon \left\| x - x_+ \right\| \left\| x_+ - \bar{x} \right\|} \\
&\leq \left\| x - \bar{x} \right\|^2 + 2\varepsilon \left\| x_+ - \bar{x} \right\| \left\| x - x_+ \right\|,
\end{aligned}$$

where the last inequality follows from the definition of (ε, δ)-subregularity with respect to S. By definition, $\left\| x - x_+ \right\| = d(x, \Omega) \leq \left\| x - \bar{x} \right\|$. Combining (2.18) and (2.13) yields $(\forall x \in U)$ $(\forall x_+ \in P_\Omega x)$ $(\forall \bar{x} \in S)$

$$\left\| x_+ - \bar{x} \right\|^2 + \left\| x - x_+ \right\|^2 \leq (1 + 2\varepsilon(1 + \varepsilon)) \left\| x - \bar{x} \right\|^2.$$

(iii) By (ii) the projector is $(S, 2\varepsilon + 2\varepsilon^2)$-firmly nonexpansive on U, and so by Lemma 2.4(ii) $R_\Omega = 2P_\Omega - \mathrm{Id}$ is $(S, 4\varepsilon + 4\varepsilon^2)$-nonexpansive on U.
This completes the proof. $\qquad \square$

Note that $\tilde{\varepsilon}_1 < \tilde{\varepsilon}_2$ $(\varepsilon > 0)$ in the above theorem; in other words, the *degree* to which classical firm nonexpansiveness is violated is greater than the degree to which classical nonexpansiveness is violated. This is as one would expect since firm nonexpansiveness is a stronger property than nonexpansiveness.

We can now characterize the degree to which the Douglas–Rachford operator violates firm nonexpansiveness on neighborhoods of (ε, δ)-subregular sets.

THEOREM 2.15 $((S, \tilde{\varepsilon})$-*firm nonexpansiveness of* $T_{DR})$. *Let* $A, B \subset \mathbf{E}$ *be closed and nonempty. Let* A *and* B *be* (ε_A, δ)- *and* (ε_B, δ)-*subregular, respectively, at* $\hat{x}$ *with respect to* $S \subset \mathbb{B}_\delta(\hat{x}) \cap (A \cap B)$. *Let* $T_{DR} : \mathbf{E} \rightrightarrows \mathbf{E}$ *be the Douglas–Rachford operator defined by* (1.6) *and define*

$$(2.19) \qquad U := \{ z \in \mathbf{E} \mid P_B z \subset \mathbb{B}_\delta(\hat{x}) \text{ and } P_A R_B z \subset \mathbb{B}_\delta(\hat{x}) \}.$$

Then T_{DR} *is* $(S, \tilde{\varepsilon})$-*firmly nonexpansive on* U, *where*

$$(2.20) \qquad \tilde{\varepsilon} = 2\varepsilon_A(1 + \varepsilon_A) + 2\varepsilon_B(1 + \varepsilon_B) + 8\varepsilon_A(1 + \varepsilon_A)\varepsilon_B(1 + \varepsilon_B).$$

That is, $(\forall x \in U) \ (\forall x_+ \in T_{DR}x) \ (\forall \bar{x} \in S)$

$$\|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \leq (1 + \tilde{\varepsilon}) \|x - \bar{x}\|^2 \,. \tag{2.21}$$

Proof. Define $U_A := \{z \mid P_A z \subset \mathbb{B}_\delta(\hat{x})\}$. By Theorem 2.14(iii) $(\forall y \in U_A) \ (\forall \tilde{x} \in R_A y) \ (\forall \bar{x} \in S)$

$$\|\tilde{x} - \bar{x}\| \leq \sqrt{1 + 4\varepsilon_A(1 + \varepsilon_A)} \, \|y - \bar{x}\| \,. \tag{2.22}$$

Similarly, define $U_B := \{z \mid P_B z \subset \mathbb{B}_\delta(\hat{x})\}$ and again apply Theorem 2.14(iii) to get $(\forall x \in U_B) \ (\forall y \in R_B x) \ (\forall \bar{x} \in S)$

$$\|y - \bar{x}\| \leq \sqrt{1 + 4\varepsilon_B(1 + \varepsilon_B)} \, \|x - \bar{x}\| \,. \tag{2.23}$$

Now, we choose any $x \in U_B$ such that $R_B x \in U_A$, that is, $x \in U$, so that we can combine (2.22)–(2.23) to get $(\forall x \in U) \ (\forall \tilde{x} \in R_A R_B x) \ (\forall \bar{x} \in S)$

$$
\begin{aligned}
\|\tilde{x} - \bar{x}\| &\leq \sqrt{1 + 4\varepsilon_A(1 + \varepsilon_A)}\sqrt{1 + 4\varepsilon_B(1 + \varepsilon_B)} \, \|x - \bar{x}\| \\
&= \sqrt{1 + 2\tilde{\varepsilon}} \, \|x - \bar{x}\| \,.
\end{aligned}
\tag{2.24}
$$

Note that $R_A R_B \bar{x} = R_B \bar{x} = \bar{x}$ since $\bar{x} \in A \cap B$, so (2.24) says that the operator $\widetilde{T} := R_A R_B$ is $(S, \tilde{\varepsilon})$-nonexpansive on U. Hence by Lemma 2.4 $T_{DR} = \frac{1}{2}(\widetilde{T} + I)$ is $(S, 2\tilde{\varepsilon})$-firmly nonexpansive on U, as claimed. $\quad\square$

If one of the sets above is convex, say, B for instance, the constant $\tilde{\varepsilon}$ simplifies to $\tilde{\varepsilon} = 2\varepsilon_A(1 + \varepsilon_A)$ since B is $(0, \infty)$-subregular at $\bar{x}$.

3. Linear convergence of iterated (S, ε)-firmly nonexpansive operators. Our goal in this section is to establish the weakest conditions we can (at the moment) under which the AP and Douglas–Rachford algorithms converge locally linearly. The notions of regularity developed in the previous section are necessary but not sufficient. In addition to regularity of the operators, we need regularity of the fixed point sets of the operators. This is developed next.

Despite its simplicity, the following lemma is one of our fundamental tools.

LEMMA 3.1. *Let* $D \subset \mathbf{E}$, $S \subset Fix\,T$, $T : D \rightrightarrows \mathbf{E}$, *and* $U \subset D$. *If*
(a) T *is* (S, ε)-*firmly nonexpansive on* U *and*
(b) *for some* $\lambda > 0$, T *satisfies the* coercivity condition

$$\|x - x_+\| \geq \lambda \, d\,(x, S) \quad \forall \, x_+ \in Tx, \ \forall x \in U, \tag{3.1}$$

then

$$d\,(x_+, S) \leq \sqrt{(1 + \varepsilon - \lambda^2)} \, d\,(x, S) \quad \forall \, x_+ \in Tx, \ \forall x \in U. \tag{3.2}$$

Proof. For $x \in U$ choose any $x_+ \in Tx$, and define $\bar{x} := P_S x$. Combining equations (3.1) and (2.6) yields

$$
\begin{aligned}
\|x_+ - \bar{x}\|^2 + (\lambda \, \|x - \bar{x}\|)^2 &\overset{(b)}{\leq} \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \\
&\overset{(a)}{\leq} (1 + \varepsilon) \|x - \bar{x}\|^2 \,,
\end{aligned}
$$

which immediately yields

$$(3.3) \qquad \|x_+ - \bar{x}\|^2 \leq (1 + \varepsilon - \lambda^2) \|x - \bar{x}\|^2.$$

Since $\bar{x} \in S$ by definition one has $d(x_+, S) \leq \|x_+ - \bar{x}\|$. Inserting this into (3.3) and using the fact $\|x - \bar{x}\| = d(x, S)$ then proves (3.2). □

3.1. Regularity of intersections of collections of sets. To this point, we have shown how the regularity of sets translates to the degree of violation of (firm) nonexpansiveness of projection-based fixed-point mappings. What remains is to develop sufficient conditions for guaranteeing (3.1). For this we define a new notion of regularity of collections of sets which generalizes through localization two well-known concepts. The first concept, which we call *strong regularity* of the collection, has many different names in the literature, among them *linear regularity* [27]. We will use the term *linear regularity* of the collection to denote the second key concept upon which we build. Our generalization is called *local linear regularity*. Both terms "strong" and "linear" are overused in the literature but we have attempted, at the risk of some confusion, to conform to the usage that best indicates the heritage of the ideas.

DEFINITION 3.2 (strong regularity, Kruger [25]). *A collection of m closed, nonempty sets $\Omega_1, \Omega_2, \ldots, \Omega_m$ is strongly regular at $\bar{x}$ if there exists an $\alpha > 0$ and a $\delta > 0$ such that*

$$(3.4) \qquad (\cap_{i=1}^m (\Omega_i - \omega_i - a_i)) \cap \mathbb{B}_\rho \neq \emptyset$$

for all $\rho \in (0, \delta]$, $\omega_i \in \Omega_i \cap \mathbb{B}_\delta(\bar{x})$, $a_i \in B_{\alpha\rho}$, $i = 1, 2, \ldots, m$.

THEOREM 3.3 (Theorem 1 in [26]). *A collection of closed, nonempty sets $\Omega_1, \Omega_2, \ldots, \Omega_m$ is strongly regular at $\bar{x}$ if and only if there exists a $\kappa > 0$ and a $\delta > 0$ such that*

$$(3.5) \qquad d\left(x, \cap_{j=1}^m (\Omega_j - x_j)\right) \leq \kappa \max_{i=1,\ldots,m} d(x + x_i, \Omega_i) \quad \forall x \in \mathbb{B}_\delta(\bar{x})$$

for all $x \in \mathbb{B}_\delta(\bar{x})$, $x_i \in \mathbb{B}_\delta$, $i = 1, \ldots, m$.

THEOREM 3.4 (Theorem 1 in [25]). *A collection of closed sets $\Omega_1, \Omega_2, \ldots, \Omega_m \subset \mathbf{E}$ is strongly regular (3.4) at a point $\bar{x} \in \cap_i \Omega_i$ if the only solution to the system*

$$(3.6) \qquad \sum_{i=1}^m v_i = 0 \qquad \text{with } v_i \in N_{\Omega_i}(\bar{x}) \quad \text{for } i = 1, 2, \ldots, m$$

is $v_i = 0$ for $i = 1, 2, \ldots, m$. For two sets $\Omega_1, \Omega_2 \subset \mathbf{E}$ this can be written as

$$(3.7) \qquad N_{\Omega_1}(\bar{x}) \cap -N_{\Omega_2}(\bar{x}) = \{0\}$$

and is equivalent to the previous definition (3.4).

DEFINITION 3.5 (linear regularity). *A collection of closed, nonempty sets $\Omega_1, \Omega_2, \ldots, \Omega_m$ is locally linearly regular at $\hat{x} \in \cap_{j=1}^m \Omega_j$ on $\mathbb{B}_\delta(\hat{x})$ ($\delta > 0$) if there exists a $\kappa > 0$ such that for all $x \in \mathbb{B}_\delta(\hat{x})$,*

$$(3.8) \qquad d\left(x, \cap_{j=1}^m \Omega_j\right) \leq \kappa \max_{i=1,\ldots,m} d(x, \Omega_i).$$

The infimum over all κ such that (3.8) holds is called regularity modulus. *If there is a $\kappa > 0$ such that (3.8) holds for all $\delta > 0$ (that is, for all $x \in \mathbf{E}$) the collection of sets is called* linearly regular at $\hat{x}$.

2410 ROBERT HESSE AND D. RUSSELL LUKE

Remark 3.6. Since (3.8) is (3.5) with $x_j = 0$ for all $j = 1, 2, \ldots, m$, it is clear that strong regularity implies local linear regularity (for some $\delta > 0$) and is indeed a much more restrictive notion than local linear regularity. What we are calling local linear regularity at $\hat{x}$ has appeared in various forms elsewhere. See, for instance, [22, Proposition 4], [31, section 3], and [26, equation (15)]. Compare this to (bounded) linear regularity defined in [5, Definition 5.6]. Also compare this to the *basic constraint qualification for sets* in [30, Definition 3.2] and *strong regularity* of the collection in [26, Proposition 2], also called *linear regularity* in [27].

Remark 3.7. Based on strong regularity (more specifically, characterization (3.7)) Lewis, Luke, and Malick proved local linear convergence of AP in the nonconvex setting, where both sets A, B are closed and one of the sets is *superregular* [27]. This was refined later in [8]. The proof of convergence that will be given in this work is different from the one used in [27, 8] and more related to the one in [5]. Convergence is achieved using (local) linear regularity (3.8), which is described in [16, Theorem 4.5] as *the precise property equivalent to uniform linear convergence of the* cyclic projection algorithms. However, the rate of convergence achieved by the use of linear regularity is not optimal, while the one in [27, 8] is in some instances. An adequate description of the relation between the *direct/primal* techniques used here and the *dual* approach used in [27, 8] is a topic of future research.

THEOREM 3.8 (linear regularity of collections of convex cones). *Let $\Omega_1, \Omega_2, \ldots,$ Ω_m be a collection of closed, nonempty, convex cones. The following statements are equivalent:*

(i) *There is a $\delta > 0$ such that the collection is locally linearly regular at $\hat{x} \in \cap_{j=1}^m \Omega_j$ on $\mathbb{B}_\delta(\hat{x})$.*

(ii) *The collection is linearly regular at $\hat{x} \in \cap_{j=1}^m \Omega_j$.*

Proof. See [5, Proposition 5.9]. $\quad\square$

Example 3.9 (Example 1.5 revisited). The collection of sets in Example 1.5(i) is strongly regular at 0 ($c = \sqrt{2}/2$) and linearly regular at ($\kappa = \sqrt{2}/4$). The same collection of sets embedded in a higher-dimensional space is still linearly regular but loses its strong regularity. This can be seen by shifting one of the sets in Example 1.5(ii) in the x_3-direction, as this renders the intersection empty. This shows that *linear regularity does not imply strong regularity.* The collection of sets in Example 1.5(iii) is neither strongly regular nor linearly regular. The collection of sets in Example 1.5(iv) is strongly regular at the intersection. One has $N_B(0) = \{(\lambda, -\lambda) \mid \lambda \in \mathbb{R}\}$ and by Remark 2.13 $N_A(0) = A$ and this directly shows $N_A(0) \cap -N_B(0) = \{0\}$. In Example 1.5(v) one of the sets is nonconvex, but the collection of sets is still well behaved in the sense that it is both strongly and linearly regular. It is worth emphasizing, however, that the set A in Example 1.5(iv) is not Clarke regular at the origin. This illustrates the fact that collections of classically "irregular" sets can still be quite regular at points of intersection.

3.2. Linear convergence of AP. In the case of the AP operator, the connection between local linear regularity of the collection of sets and the coercivity of the operator with respect to the intersection is natural, as the next result shows.

PROPOSITION 3.10 (coercivity of the projector). *Let A, B be nonempty and closed subsets of $\mathbf{E}$, $\hat{x} \in S := A \cap B$ and let the collection $\{A, B\}$ be locally linearly regular at $\hat{x}$ on $\mathbb{B}_\delta(\hat{x})$ with constant κ for some $\delta > 0$. One has*

$$\|x - x_+\| \geq \gamma \, d(x, S) \quad \forall x_+ \in P_B x, \ \forall x \in A \cap \mathbb{B}_\delta(\hat{x}),$$

where $\gamma = 1/\kappa$.

Proof. By the definition of the distance and the projector one has, for $x \in A$ and any $x_+ \in P_B x$,

$$\|x - x_+\| = d(x, B)$$
$$= \max \left\{ d(x, B), \underbrace{d(x, A)}_{=0} \right\}$$
$$\geq \gamma \, d(x, S).$$

The inequality follows by Definition 3.5 (local linear regularity at $\hat{x}$ on $\mathbb{B}_\delta(\hat{x})$ with constant κ), since $x \in \mathbb{B}_\delta(\hat{x})$.   □

THEOREM 3.11 (projections onto a (ε, δ)-subregular set). *Let A, B be nonempty and closed subsets of $\mathbf{E}$ and let $\hat{x} \in S := A \cap B$. If*
  (a) *B is (ε, δ)-subregular at $\hat{x}$ with respect to S and*
  (b) *the collection $\{A, B\}$ is locally linearly regular at $\hat{x}$ on $\mathbb{B}_\delta(\hat{x})$,*
then

(3.9) $$d(x_+, S) \leq \sqrt{1 + \tilde{\varepsilon} - \gamma^2} \, d(x, S) \quad \forall x_+ \in P_B x, \ \forall x \in U,$$

where $\gamma = 1/\kappa$ with κ the regularity modulus on $\mathbb{B}_\delta(\hat{x})$, $\tilde{\varepsilon} = 2\varepsilon + 2\varepsilon^2$, and

(3.10) $$U \subset \{x \in A \cap \mathbb{B}_\delta(\hat{x}) \mid P_B x \subset \mathbb{B}_\delta(\hat{x})\}.$$

Proof. Since B is (ε, δ)-subregular at $\hat{x}$ with respect to S one can apply Theorem 2.14 to show that the projector P_B is $(S, 2\varepsilon + 2\varepsilon^2)$-firmly nonexpansive on U. Moreover, condition (b) and Proposition 3.10 yield

$$\|x_+ - x\| \geq \gamma d(x, S) \quad \forall x_+ \in P_B x, \ \forall x \in U.$$

Combining (a) and (b) and applying Lemma 3.1 then gives

$$d(x_+, S) \leq \sqrt{1 + 2\tilde{\varepsilon} - \gamma^2} d(x, S) \quad \forall x_+ \in P_B x, \ \forall x \in U.   □$$

COROLLARY 3.12 (projections onto a convex set [21]). *Let A and B be nonempty, closed subsets of $\mathbf{E}$. If*
  (a) *the collection $\{A, B\}$ is locally linearly regular at $\hat{x} \in A \cap B$ on $\mathbb{B}_\delta(\hat{x})$ with regularity modulus $\kappa > 0$ and*
  (b) *B is convex,*
then

(3.11) $$d(x_+, S) \leq \sqrt{1 - \gamma^2} \, d(x, S) \quad \forall x_+ \in P_B x, \ \forall x \in A \cap \mathbb{B}_\delta(\hat{x}),$$

where $\gamma = 1/\kappa$.

Proof. By convexity of B the projector P_B is nonexpansive and it follows that $P_B x \in \mathbb{B}_\delta(\hat{x})$ for all $x \in \mathbb{B}_\delta(\hat{x})$. Saying that B is convex is equivalent to saying that B is $(0, +\infty)$-regular and hence $\tilde{\varepsilon} = 0$ in Theorem 3.11.   □

COROLLARY 3.13 (linear convergence of AP). *Let A, B be closed nonempty subsets of $\mathbf{E}$ and let the collection $\{A, B\}$ be locally linearly regular at $\hat{x} \in S := A \cap B$ on $\mathbb{B}_\delta(\hat{x})$ with regularity modulus $\kappa > 0$. Define $\gamma := 1/\kappa$ and let $x_0 \in A$. Generate the sequence $\{x_n\}_{n \in \mathbb{N}}$ by*

(3.12) $$x_{2n+1} \in P_B x_{2n} \ \text{and} \ x_{2n+2} \in P_A x_{2n+1} \quad \forall n = 0, 1, 2, \ldots.$$

(a) *If A and B are (ε, δ)-subregular at $\hat{x}$ with respect to S and $\tilde{\varepsilon} := 2\varepsilon + 2\varepsilon^2 \leq \gamma^2$, then*

$$d\left(x_{2n+2}, S\right) \leq (1 - \gamma^2 + \tilde{\varepsilon})d\left(x_{2n}, S\right) \quad \forall n = 0, 1, 2, \ldots$$

for all $x_0 \in \mathbb{B}_{\delta/2}(\hat{x}) \cap A$.

(b) *If A is (ε, δ)-subregular with respect to S, B is convex, and $\tilde{\varepsilon} := 2\varepsilon + 2\varepsilon^2 \leq (2\gamma - \gamma^2)/(1 - \gamma^2)$, then*

$$d\left(x_{2n+2}, S\right) \leq \sqrt{1 - \gamma^2 + \tilde{\varepsilon}}\sqrt{1 - \gamma^2}d\left(x_{2n}, S\right) \quad \forall n = 0, 1, 2, \ldots$$

for all $x_0 \in \mathbb{B}_{\delta/2}(\hat{x}) \cap A$.

(c) *If A and B are convex, then*

$$d\left(x_{2n+2}, S\right) \leq (1 - \gamma^2)d\left(x_{2n}, S\right) \quad \forall n = 0, 1, 2, \ldots$$

for all $x_0 \in \mathbb{B}_{\delta}(\hat{x}) \cap A$.

Proof. (a) First one has to show that all iterates remain close to $\hat{x}$ for x_0 close to $\hat{x}$, that is, we have to show that all iterates remain in the set U defined by (3.10). Note that for any $x_0 \in \mathbb{B}_{\delta/2}(\hat{x})$ and $x_1 \in P_B x_0$ one has

$$\|x_0 - x_1\| = d\left(x_0, B\right) \leq \|x_0 - \hat{x}\|$$

since $\hat{x} \in B$. Thus

$$\|x_1 - \hat{x}\| \leq \|x_0 - x_1\| + \|x_0 - \hat{x}\| \leq \|x_0 - \hat{x}\| + \|x_0 - \hat{x}\| \leq \delta,$$

which shows that $P_B x_0 \subset \mathbb{B}_{\delta}(\hat{x})$ *for all* $x_0 \in \mathbb{B}_{\delta/2}(\hat{x})$. One can now apply Theorem 3.11 to conclude that

$$d\left(x_1, S\right) \leq \sqrt{1 - \gamma^2 + \tilde{\varepsilon}} \, d\left(x_0, S\right).$$

The last equation then implies that $x_1 \in \mathbb{B}_{\delta/2}(\hat{x})$ as long as $\gamma^2 \geq \tilde{\varepsilon}$ and therefore the same argument can be applied to x_1 to conclude that

(3.13) $$d\left(x_2, S\right) \leq \sqrt{1 - \gamma^2 + \tilde{\varepsilon}} \, d\left(x_1, S\right).$$

Combining the last two equations, (a) then follows by induction.

(b) Applying Corollary 3.12 yields

$$d\left(x_1, S\right) \leq \sqrt{1 - \gamma^2} \, d\left(x_0, S\right)$$

and analogous to (a) note that (3.13) is still valid for $\tilde{\varepsilon} \leq (2\gamma - \gamma^2)/(1 - \gamma^2)$. By $\tilde{\varepsilon} \leq (2\gamma - \gamma^2)/(1 - \gamma^2)$ it follows that

$$\begin{aligned}
\sqrt{1 - \gamma^2 + \tilde{\varepsilon}}\sqrt{1 - \gamma^2} &\leq \sqrt{1 - \gamma^2 + (2\gamma - \gamma^2)/(1 - \gamma^2)}\sqrt{1 - \gamma^2} \\
&\leq \sqrt{1 - 2\gamma + \gamma^2 + (2\gamma - \gamma^2)} \\
&\leq 1
\end{aligned}$$

and therefore by induction (b) follows.

(c) The proof is an immediate consequence of Corollary 3.12. $\square$

3.3. Linear convergence of Douglas–Rachford. We now turn to the Douglas–Rachford algorithm. This algorithm is notoriously difficult to analyze and our results reflect this in considerably more circumscribed conditions than are required for the AP algorithm. Nevertheless, to the best of our knowledge the following convergence results are the most general to date. The first result gives sufficient conditions for the coercivity condition (3.1) to hold.

LEMMA 3.14. *Let the collection of closed subsets A, B of $\mathbf{E}$ be locally linearly regular at $\hat{x} \in S := A \cap B$ on $\mathbb{B}_\delta(\hat{x})$ with constant $\kappa > 0$ for some $\delta > 0$. Suppose further that B is a subspace and that for some constant $c \in (0,1)$ the following condition holds:*

$$(3.14) \qquad \left. \begin{array}{ll} x \in \mathbb{B}_\delta(\hat{x}), \ y = P_B x, & \quad u \in N_A(z) \cap \mathbb{B} \\ z \in P_A(2y - x) & and \quad v \in N_B(y) \cap \mathbb{B} \end{array} \right\} \quad \Rightarrow \quad \langle u, v \rangle \geq -c.$$

Then T_{DR} satisfies

$$(3.15) \qquad \|x - x_+\| \geq \frac{\sqrt{1-c}}{\kappa} d(x, S) \quad \forall x_+ \in T_{DR} x, \ \forall \, x \in U,$$

where

$$(3.16) \qquad U \subset \{ x \in \mathbb{B}_\delta(\hat{x}) \mid P_A R_B x \subset \mathbb{B}_\delta(\hat{x}) \}.$$

Proof. In what follows we will use the notation $R_B x$ for $2y - x$ with $y = P_B x$, which is unambiguous, if a slight abuse of notation, since B is convex. We will use (3.14) to show that for all $x \in U$ with $y = P_B x$ and $z \in P_A R_B x$,

$$(3.17) \qquad \begin{aligned} \|x - x_+\|^2 &= \|z - y\|^2 \\ &\geq (1-c) \big[\|z - R_B x\|^2 + \|R_B x - y\|^2 \big]. \end{aligned}$$

We will then show that for all $x \in U$ with $y = P_B x$ and $z \in P_A R_B x$

$$(3.18) \qquad \|z - R_B x\|^2 + \|R_B x - y\|^2 \geq \frac{1}{\kappa^2} d(R_B x, S)^2.$$

Combining inequalities (3.17) and (3.18) yields

$$\|x - x_+\|^2 \geq \frac{1-c}{\kappa^2} d(R_B x, S)^2 \quad \forall \, x \in U.$$

Let $\tilde{x} \in P_S(R_B x)$ and note that $d(R_B x, S) = \|R_B x - \tilde{x}\|$. Since B is a subspace, by (1.12) one has $d(R_B x, S) = \|R_B x - \tilde{x}\| = \|x - \tilde{x}\|$. Moreover, $\|x - \tilde{x}\| \geq \min_{y \in S} \|x - y\| = d(x, S)$, hence

$$\|x - x_+\| \geq \frac{\sqrt{1-c}}{\kappa} d(x, S) \quad \forall \, x \in U$$

as claimed.

What remains is to prove (3.17) and (3.18) for all $x \in U$ with $y = P_B x$ and $z \in P_A R_B x$.

Proof of (3.17). Using Lemma 2.1, (2.2), one has for $x \in \mathbb{B}_\delta(\hat{x})$ with $y = P_B x$ and $z \in P_A R_B$

$$
(3.19) \quad \|x - x_+\|^2 = \|z - y\|^2
$$
$$
= \|z - R_B x + R_B x - y\|^2
$$
$$
= \|z - R_B x\|^2 + \|R_B x - y\|^2 + 2 \left\langle \underbrace{z - R_B x}_{\in -N_A(z)}, \ \underbrace{R_B x - y}_{=y - x \in -N_B(y)} \right\rangle
$$
$$
\overset{(3.14)}{\geq} \|z - R_B x\|^2 + \|R_B x - y\|^2 - 2c \|z - R_B x\| \|R_B x - y\|
$$
$$
= (1 - c) \big[\|z - R_B x\|^2 + \|R_B x - y\|^2 \big]
$$
$$
+ c \big[\|z - R_B x\|^2 + \|R_B x - y\|^2 \big] - 2c \|z - R_B x\| \|R_B x - y\|
$$
$$
= (1 - c) \big[\|z - R_B x\|^2 + \|R_B x - y\|^2 \big]
$$
$$
+ c \big[\|z - R_B x\| - \|R_B x - y\| \big]^2
$$
$$
\geq (1 - c) \big[\|z - R_B x\|^2 + \|R_B x - y\|^2 \big].
$$

Proof of (3.18). First note that if $x \in \mathbb{B}_\delta(\hat{x})$, since B is a subspace, by (1.12) $R_B x \subset \mathbb{B}_\delta(\hat{x})$ and by convexity of $\mathbb{B}_\delta(\hat{x})$ it follows that $y = P_B x \subset \mathbb{B}_\delta(\hat{x})$ and hence (3.14) is localized to $\mathbb{B}_\delta(\hat{x})$ (to which the dividends of linear regularity of the intersection extend) as long as $P_A R_B x \subset \mathbb{B}_\delta(\hat{x})$, that is, as long as $x \in U$. By definition of the projector $\|R_B x - y\| \geq \|R_B x - P_B(R_B x)\|$. Local linear regularity at $\hat{x}$ with radius δ and constant κ yields for $x \in U$ with $y = P_B x$ and $z \in P_A R_B x$

$$
(3.20) \qquad \|z - R_B x\|^2 + \|R_B x - y\|^2 \geq \|z - R_B x\|^2 + \|R_B x - P_B R_B x\|^2
$$
$$
= d(R_B x, A)^2 + d(R_B x, B)^2
$$
$$
\overset{(3.8)}{\geq} \frac{1}{\kappa^2} d(R_B x, S)^2 .
$$

This completes the proof of (3.18) and the theorem. □

Remark 3.15. The coercivity constant in (3.15) is not tight, even if κ were chosen to the regularity modulus of the intersection. Remember, by linearity, that $y = P_B x = P_B R_B x$. We use the inequality $\big[\|z - R_B x\| - \|R_B x - y\|\big]^2 \geq 0$ in line (3.19), while we use $\|z - R_B x\|^2 + \|R_B x - y\|^2 \geq \max\{\|z - R_B x\|^2, \|R_B x - y\|^2\}$ in line (3.20). If the first inequality is tight, then this is the worst possible result in the second inequality, since $\|z - R_B x\| = \|R_B x - y\|$, that is, this second inequality is satisfied not only strictly, but the inequality is as large as it can possibly be. On the other hand if the second inequality is tight this means $\|z - R_B x\|^2 = 0$ *or* $\|R_B x - y\|^2 = 0$ and this means that the first inequality is strict. In any event, it is impossible to achieve equality in the argumentation of the proof. This is a technical limitation of the logic of the proof and does not preclude improvements.

Lemma 3.14 with the added assumption of (ε, δ)-regularity of the nonconvex set yields local linear convergence of the Douglas–Rachford algorithm in this special case.

THEOREM 3.16. *Let the collection of closed subsets A, B of $\mathbf{E}$ be locally linearly regular at $\hat{x} \in S := A \cap B$ on $\mathbb{B}_\delta(\hat{x})$ with constant $\kappa > 0$ for some $\delta > 0$. Suppose*

further that B is a subspace and that A is (ε, δ)-regular at $\hat{x}$ with respect to S. Assume that for some constant $c \in (0, 1)$ the following condition holds:

$$(3.21) \qquad \left. \begin{array}{ll} z \in A \cap \mathbb{B}_\delta(\hat{x}), & u \in N_A(z) \cap \mathbb{B} \\ y \in B \cap \mathbb{B}_\delta(\hat{x}), & v \in N_B(y) \cap \mathbb{B} \end{array} \right\} \quad \Rightarrow \quad \langle u, v \rangle \geq -c.$$

If $x \in \mathbb{B}_{\delta/2}(\hat{x})$, then

$$d(x_+, S) \leq \sqrt{1 + \tilde{\varepsilon} - \eta} \, d(x, S) \quad \forall \, x_+ \in T_{DR}x$$

with $\eta := \frac{(1-c)}{\kappa^2}$ and $\tilde{\varepsilon} = 2\varepsilon + 2\varepsilon^2$.

Proof. First one has to show requirement (3.16). Since $\hat{x} \in A \cap B$ note that for any $x \in \mathbb{B}_{\delta/2}(\hat{x})$ for all $z \in P_A R_B x$ by definition $\|z - R_B x\| = d(R_B x, A) \leq \|R_B x - \hat{x}\|$ and by (1.12) $\|R_B x - \hat{x}\| = \|x - \hat{x}\|$ holds. This now implies

$$\|z - \hat{x}\| \leq \|z - R_B x\| + \|R_B x - \hat{x}\| \leq 2\|x - \hat{x}\| \leq \delta,$$

and therefore $z \in \mathbb{B}_\delta(\hat{x})$.

Now for B a subspace (3.21) and (3.14) are equivalent, and so by Lemma 3.14 the coercivity condition (3.1)

$$(3.22) \qquad \|x - x_+\| \geq \frac{\sqrt{1 - c}}{\kappa} d(x, S)$$

is satisfied on $\mathbb{B}_{\delta/2}(\hat{x})$. Moreover, since A is (ε, δ)-regular and B is $(0, \infty)$-regular, by Theorem 2.15 T_{DR} is $(S, \tilde{\varepsilon})$-firmly nonexpansive with $\tilde{\varepsilon} = 2\varepsilon(1 + \varepsilon)$, that is, $(\forall x \in \mathbb{B}_{\delta/2}(\hat{x}))$ $(\forall x_+ \in T_{DR}x)$ $(\forall \bar{x} \in S)$

$$\|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \leq (1 + \tilde{\varepsilon}) \|x - \bar{x}\|^2.$$

Lemma 3.1 then applies to yield $(\forall x \in \mathbb{B}_{\delta/2}(\hat{x}))$ $(\forall x_+ \in T_{DR}x)$

$$d(x_+, S) \leq \sqrt{1 + \tilde{\varepsilon} - \eta} \, d(x, S),$$

where $\eta := \frac{1-c}{\kappa^2}$. $\quad\square$

The next lemma establishes sufficient conditions under which (3.21) holds.

LEMMA 3.17 (See [27, Theorem 5.16]). *Assume $B \subset \mathbf{E}$ is a subspace and that $A \subset \mathbf{E}$ is closed and super-regular at $\hat{x} \in A \cap B$. If the collection $\{A, B\}$ is strongly regular at $\hat{x}$, then there is a $\delta > 0$ and a constant $c \in (0, 1)$ such that (3.21) holds on $\mathbb{B}_\delta(\hat{x})$.*

Proof. Condition (3.14) can be shown using (3.7). For more details see [27]. $\quad\square$

We summarize this discussion with the following convergence result for the Douglas–Rachford algorithm in the case of an affine subspace and a superregular set.

THEOREM 3.18. *Assume $B \subset \mathbf{E}$ is a subspace and that $A \subset \mathbf{E}$ is closed and superregular at $\hat{x} \in S := A \cap B$. If the collection $\{A, B\}$ is strongly linearly regular at S, then there is a $\delta > 0$ and an $\varepsilon > 0$ such that (3.21) holds for some $c \in (0, 1)$, A is (ε, δ)-subregular at $\hat{x}$ and*

$$(3.23) \qquad \frac{(1 - c)}{\kappa^2} > 2\varepsilon + 2\varepsilon^2.$$

Hence

$$d\left(x_+, S\right) \leq \tilde{c}\, d\left(x, S\right) \quad \forall\, x_+ \in T_{DR}x,$$

with $\tilde{c} = \sqrt{1 + 2\varepsilon + 2\varepsilon^2 - \frac{(1-c)}{\kappa^2}} < 1$ *for all* $x \in \mathbb{B}_{\delta/2}(\hat{x})$.

   Proof. Strong regularity of the collection implies linear regularity with constant κ on $\mathbb{B}_{\delta_1}(\hat{x})$ (see Remark 3.6). Lemma 3.17 guarantees the existence of constants $\delta_2 > 0$ and $c \in (0, 1)$ such that (3.21) holds on $\mathbb{B}_{\delta_2}(\hat{x})$. Now, by superregularity at $\hat{x}$, for any ε there exists a δ_3 such that A is (ε, δ_3)-subregular at $\hat{x}$. In other words, for c and κ determined by the regularity of the collection $\{A, B\}$ at $\hat{x}$, we can always choose ε (generating a corresponding δ_3 radius) so that (3.23) is satisfied on $\mathbb{B}_{\delta_3}(\hat{x})$. Then for $\delta := \min\{\delta_1, \delta_2, \delta_3\}$, the requirements of Theorem 3.16 are satisfied on $\mathbb{B}_\delta(\hat{x})$, which completes the proof of linear convergence on $\mathbb{B}_{\delta/2}(\hat{x})$. $\qquad \square$

   Remark 3.19. Example 1.5(v) has been studied by Borwein and coauthors [1, 10], who achieve global characterizations of convergence with rates. Our work does not directly overlap with [1, 10] since our results are local, and the order of the reflectors is reversed: we must reflect first across the subspace, then reflect across the nonconvex set; Borwein and coauthors reflect first across the circle.

   3.4. Douglas–Rachford on subspaces. We finish this section with the fact that strong regularity of the intersection is *necessary*, not just sufficient for convergence of the iterates of the Douglas–Rachford algorithm to the intersection in the affine case.

   COROLLARY 3.20. *Let N, M be two affine subspaces with $N \cap M \neq \emptyset$. Let A be the subspace parallel to N and B be the subspace parallel to M. Douglas–Rachford converges for any starting point $x_0 \in \mathbf{E}$ with linear rate to the intersection $N \cap M$ if and only if $A^\perp \cap B^\perp = \{0\}$.*

   Proof. Without loss of generality for $\hat{x} \in N \cap M$ by shifting the subspaces by $\hat{x}$ we consider the case of linear subspaces A, B. By (3.7), on subspaces the condition $A^\perp \cap B^\perp = \{0\}$ is equivalent to strong regularity of the collection $\{A, B\}$.

   If the intersection is strongly regular and A and B are subspaces, then the requirements of Theorem 3.18 are globally satisfied, so Douglas–Rachford converges with linear rate

$$\tilde{c} = \sqrt{1 - \frac{(1-c)}{\kappa^2}} < 1,$$

where $c \in [0, 1)$ (compare (3.14)) now becomes

$$(3.24) \qquad c = \max\langle u, v \rangle, \qquad u \in A^\perp, \ \|u\| = 1, \ v \in B^\perp, \ \|v\| = 1,$$

and κ is an associated *global* constant of linear regularity, the existence of which follows by Theorem 3.8.

   On the other hand, for $\hat{x} \in A \cap B$ by [7, Theorem 3.5] we get the characterization

$$\begin{aligned}
\operatorname{Fix} T_{DR} &= (A \cap B) + N_{A-B}(0) \\
&= (A \cap B) + (N_A(\hat{x}) \cap -N_B(\hat{x})) \\
&= (A \cap B) + A^\perp \cap B^\perp,
\end{aligned}$$

and so the fixed-point set of T_{DR} does not coincide with the intersection unless the collection $\{A, B\}$ is strongly regular. In other words, if the intersection is not strongly

regular, then convergence to the intersection cannot be linear, thus proving the reverse implication by the contrapositive.   □

Remark 3.21 (Friedrichs angle [18]). We would like to make a final remark about connection between the notion of the angle of the sets at the intersection and the regularity of the collection of sets at points in the intersection. The operative notion of angle is the *Friedrichs angle*. For two subspaces A and B the *Friedrichs angle* is the angle $\alpha(A, B)$ in $[0, \pi/2]$ whose cosine is defined by

$$(3.25) \qquad c_F(A, B) := \sup \left\{ |\langle x, y \rangle| \; \middle| \; \begin{array}{l} x \in A \cap (A \cap B)^{\perp}, \;\; \|x\| \leq 1 \\ y \in B \cap (A \cap B)^{\perp}, \;\; \|y\| \leq 1 \end{array} \right\}.$$

The Friedrichs angle being less than 1 is not sufficient for convergence of Douglas–Rachford. This can be seen by Example 1.5(ii). The Friedrichs angle in this example is the same as in Example 1.5(i), but for $x_0 \notin \mathbb{R}^2 \times \{0\}$ the Douglas–Rachford algorithm does not converge to $\{0\} = A \cap B$. Another interesting observation is that if, on the other hand, $A^{\perp} \cap B^{\perp} = \{0\}$, then $(A^{\perp} \cap B^{\perp})^{\perp} = \mathbf{E}$, which implies that $c_F(A^{\perp}, B^{\perp})$ coincides with (3.24) and by [14, Theorem 2.16] $c_F(A^{\perp}, B^{\perp})$ then coincides with $c_F(A, B)$. So if the Douglas–Rachford algorithm on subspaces converges linearly, then the rate of convergence is dependent on the Friedrichs angle. A detailed analysis regarding the relation between the Friedrichs angle and linear convergence of AP can be found in [17, 16].

4. Concluding remarks. In the time that has passed since first submitting our manuscript for publication we learned about an optimal linear convergence result for Douglas–Rachford applied to ℓ_1 optimization with an affine constraint using different techniques [13]. The modulus of linear regularity does not recover optimal convergence results (see Remark 3.7), but we suspect this is an artifact of our proof technique. The question remains whether there is a quantitative primal definition of an angle between two sets that recovers the same results for AP as [8]. This could also be useful to achieve optimal linear convergence results for Douglas–Rachford in general. Also, as we noted in the introduction, it is well known that the fixed-point set of the Douglas–Rachford operator is in general bigger than the intersection of the sets, and Corollary 3.20 stating that the iterates converge to the intersection if and only if the collection of sets is strongly regular is a consequence of this. In the convex case, the *shadows* of the iterates still converge. We leave a fuller investigation of the shadows of the iterates and the angles between the sets at the intersection in the nonconvex setting to future work.

Another direction of future work will be to extend this analysis more generally to fixed-point mappings built upon functions and more general set-valued mappings, but also in particular *proximal* operators and reflectors. The generality of our approach makes such extensions quite natural. Indeed, local linear regularity of collections of sets can be shown to be related to *strong metric subregularity* of set-valued mappings which guarantees that condition (3.1) of Lemma 3.1 is satisfied. Of course, the difficulty remains to show that the mappings are indeed metrically subregular.

REFERENCES

[1] F. A. ARTACHO AND J. BORWEIN, *Global Convergence of a Non-convex Douglas–Rachford Iteration*, arXiv:1203. 2392, 2012.
[2] F. A. ARTACHO, A. DONTCHEV, AND M. GEOFFROY, *Convergence of the proximal point method for metrically regular mappings*, ESAIM Proc., 17 (2007), pp. 1–8.

[3] F. A. ARTACHO AND M. GEOFFROY, *Uniformity and inexact version of a proximal method for metrically regular mappings*, J. Math. Anal. Appl., 335 (2007), pp. 168–183.

[4] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.

[5] H. H. BAUSCHKE AND J. M. BORWEIN, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[6] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[7] H. H. BAUSCHKE, P. L. COMBETTES, AND D. R. LUKE, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*, J. Approx. Theory, 127 (2004), pp. 178–314.

[8] H. H. BAUSCHKE, D. R. LUKE, H. M. PHAN, AND X. WANG, *Restricted normal cones and the method of alternating projections: Applications*, Set-Valued Variational Anal., 21(2013), pp. 475–501.

[9] H. H. BAUSCHKE, D. R. LUKE, H. M. PHAN, AND X. WANG, *Restricted normal cones and the method of alternating projections: Theory*, Set-Valued Variational Anal., 21(2013), pp. 431–473.

[10] J. M. BORWEIN AND B. SIMS, *The Douglas–Rachford algorithm in the absence of convexity*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Optim. Appl., 49 Springer, New York, 2011, pp. 93–109.

[11] R. S. BURACHIK AND A. N. IUSEM, *Set-Valued Mappings and Enlargements of Monotone Operators*, Optim. Appl. 8, Springer, New York, 2008.

[12] P. L. COMBETTES AND T. PENNANEN, *Proximal methods for cohypomonotone operators*, SIAM J. Control Optim., 43 (2004), pp. 731–742.

[13] L. DEMANET AND X. ZHANG, *Eventual Linear Convergence of the Douglas-Rachford Iteration for Basis Pursuit*, arXiv:1301.0542[math.NA], 2013.

[14] F. DEUTSCH, *The Angle Between Subspaces of a Hilbert Space*, in Approximation Theory, Wavelets and Applications, S. Singh, ed., NATO Sci. Ser. 454, Springer, 1995, pp. 107–130.

[15] F. DEUTSCH, *Best Approximation in Inner Product Spaces*, Springer, New York, 2001.

[16] F. DEUTSCH AND H. HUNDAL, *The rate of convergence for the cyclic projections algorithm* III*: Regularity of convex sets*, J. Approx. Theory, 155 (2008), pp. 155–184.

[17] F. DEUTSCH AND H. HUNDAL, *The rate of convergence for the cyclic projections algorithm* III*: Regularity of convex sets*, J. Approx. Theory, 155 (2008), pp. 155–184.

[18] K. FRIEDRICHS, *On certain inequalities and characteristic value problems for analytic functions and for functions of two variables*, Trans. Amer. Math. Soc., 41 (1937), pp. 321–364.

[19] K. GOEBEL AND W. A. KIRK, *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, UK, 1990.

[20] K. GOEBEL AND S. REICH, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Pure Appl. Math., 83, CRC Press, Boca Raton, FL, 1984.

[21] L. G. GUBIN, B. T. POLYAK, AND E. RAIK, *The method of projections for finding the common point of convex sets*, Comput. Math. Math. Phys., 7 (1967), pp. 1–24.

[22] A. D. IOFFE, *Metric regularity and subdifferential calculus*, Russian Math. Surveys, 55 (2000), pp. 501–558.

[23] A. IUSEM, T. PENNANEN, AND B. SVAITER, *Inexact versions of the proximal point algorithm without monotonicity*, SIAM J. Optim., 13 (2003), pp. 1080–1097.

[24] D. KLATTE AND B. KUMMER, *Optimization methods and stability of inclusions in Banach spaces*, Math. Program. B, 117 (2009), pp. 305–350.

[25] A. Y. KRUGER., *Weak stationarity: Eliminating the gap between necessary and sufficient conditions*, Optimization, 53 (2004), pp. 147–164.

[26] A. Y. KRUGER, *About regularity of collections of sets*, Set-Valued Anal., 14 (2006), pp. 187–206.

[27] A. S. LEWIS, D. R. LUKE, AND J. MALICK, *Local linear convergence for alternating and averaged nonconvex projections*, Found. Comput. Math., 9 (2009), pp. 485–513.

[28] P. L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.

[29] D. R. LUKE, *Finding best approximation pairs relative to a convex and a prox-regular set in a Hilbert space*, SIAM J. Optim., 19 (2008), pp. 714–739.

[30] B. MORDUKHOVICH, *Variational Analysis and Generalized Differentiation,* I*: Basic Theory*, Grundlehren Math. Wiss., Springer, New York, 2006.

[31] H. V. Ngai and M. Théra, *Metric inequality, subdifferential calculus and applications*, Set-Valued Anal., 9 (2001), pp. 187–216.

[32] T. Pennanen, *Local convergence of the proximal point algorithm and multiplier methods without monotonicity*, Math. Oper. Res., 27 (2002), pp. 170–191.

[33] R. A. Poliquin, R. T. Rockafellar, and L. Thibault, *Local differentiability of distance functions*, Trans. Amer. Math. Soc., 352 (2000), pp. 5231–5249.

[34] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss., Springer, New York, 1998.

[35] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 237–424.