Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility

Robert Hesse, D. Russell Luke and Patrick Neumann*

Abstract

The problem of finding a vector with the fewest nonzero elements that satisfies an underdetermined system of linear equations is an NP-complete problem that is typically solved numerically via convex heuristics or nicely-behaved nonconvex relaxations. In this work we consider elementary methods based on projections for solving a *sparse feasibility* problem without employing convex heuristics. It has been shown recently that, locally, the fundamental method of alternating projections *must* converge linearly to a solution to the sparse feasibility problem with an affine constraint. In this paper we apply different analytical tools that allow us to show *global* linear convergence of alternating projections under familiar constraint qualifications. These analytical tools can also be applied to other algorithms. This is demonstrated with the prominent Douglas-Rachford algorithm where we establish local linear convergence of this method applied to the sparse affine feasibility problem.

Index Terms

Compressed sensing, iterative methods, linear systems, optimization, projection algorithms, sparse Matrices

I. INTRODUCTION

NUMERICAL algorithms for nonconvex optimization models are often eschewed because the usual optimality criteria around which numerical algorithms are designed do not distinguish solutions from critical points. This issue comes into sharp relief with what has become known as the *sparsity optimization problem* [16, Equation (1.3)]:

$$\text{minimize } \|x\|_0 \text{ subject to } Mx = p, \tag{1}$$

where $m, n \in I\!N$, the nonnegative integers, with $m < n$, $M \in I\!R^{m \times n}$ is a real $m-$by$-n$ matrix of full rank and $\|x\|_0 := \sum_{j=1}^{n} |\text{sign}(x_j)|$ with $\text{sign}(0) = 0$ is the number of nonzero entries of a real vector $x \in I\!R^n$ of dimension n. The first-order necessary optimality condition for this problem is (formally)

$$0 \in \partial \left(\|x\|_0 + \iota_B(x) \right), \tag{2}$$

where ∂ is the *basic, or limiting, subdifferential* (see, for example, [33, Definition 1.77]),

$$B := \{x \in I\!R^n | \ Mx = p\} \tag{3}$$

and $\iota_B(x) = 0$ if $x \in B$ and $+\infty$ otherwise. The function $\| \cdot \|_0$ is subdifferentially regular so all of the varieties of the subdifferential – basic or limiting, Fréchet, Clarke – in (2) are equivalent. This fact, together with an explicit formulation of the subdifferential of $\| \cdot \|_0$ were given in [28]. It can be shown, however, that *every* point in B satisfies (2) and so this is uninformative as a basis for numerical algorithms [22].

In this note we explore the following question: when do elementary numerical algorithms for solving some related nonconvex problem converge locally and/or globally?

The current trend for solving this problem, sparked by the now famous paper of Candès and Tao [16], is to use convex relaxations. Convex relaxations have the advantage that every point satisfying the necessary optimality

criteria is also a solution to the relaxed optimization problem. This certainty comes at the cost of imposing difficult-to-verify restrictions on the affine constraints [39] in order to guarantee the correspondence of solutions to the relaxed problem to solutions to the original problem. Moreover, convex relaxations can lead to a tremendous increase in the dimensionality of the problem (see for example [15]).

In this work we present a different *nonconvex* approach; one with the advantage that the available algorithms are simple to apply, (locally) linearly convergent, and the problem formulation stays close in spirit if not in fact to the original problem, thus avoiding the curse of dimensionality. We also provide conditions under which fundamental algorithms applied to the nonconvex model are globally convergent. These conditions are in the same family of restrictions used to guarantee the success of convex relaxations, but come without the tax of dimensionality.

Many strategies for relaxing (1) have been studied in the last decade. In addition to convex, and in particular ℓ_1, relaxations, authors have studied dynamically reweighted ℓ_1 (see [14], [17]) as well as relaxations to ℓ_p semi-metric $(0 < p < 1)$ (see, for instance, [27]). The key to all relaxations, whether they be convex or not, is the correspondence between the relaxed problem and (1). Candès and Tao [16] introduced the restricted isometry property of the matrix M as a sufficient condition for the correspondence of solutions to (1) with solutions to the *convex* problem of finding the point x in the set B with smallest ℓ_1-norm. This condition was generalized in [12], [13], [11] in order to show *global* convergence of the simple projected gradient method for solving the problem

$$\text{minimize } \tfrac{1}{2}\|Mx - p\|_2^2 \text{ subject to } x \in A_s, \tag{4}$$

where

$$A_s := \left\{ x \in I\!\!R^n | \ \|x\|_0 \le s \right\}, \tag{5}$$

the set of s-sparse vectors for a fixed $s \le n$. Notable in this model is that the sparsity "objective" is in the constraint, and one must specify a priori the sparsity of the solution. Also notable is that the problem (4) is still nonconvex, although one can still obtain global convergence results.

Inspired by (4), and the desire to stay as close to (1) as possible, we model the optimization problem as a *feasibility* problem

$$\text{Find } \bar{x} \in A_s \cap B, \tag{6}$$

where A_s and B are given by (5) and (3), respectively. For a well-chosen sparsity parameter s, solutions to (6) exactly correspond to solutions to (1). Such an approach was also proposed in [18] where the authors proved local convergence of a simple alternating projections algorithm for feasibility with a sparsity set. Alternating projections is but one of a huge variety of projection algorithms for solving feasibility problems. The goal of this paper is to show *when* and *how fast* fundamental projection algorithms applied to this nonconvex problem converge. Much of this depends on the abstract geometric structure of the sets A_s and B; for affine sparse feasibility this is well-defined and surprisingly simple.

The set B is an affine subspace and A_s is a nonconvex set. However, the set A_s is the union of finitely many subspaces, each spanned by s vectors from the standard basis for $I\!\!R^n$ [9]. We show in (20) that one can easily calculate a *projection* onto A_s.

For $\Omega \subset I\!\!R^n$ closed and nonempty, we call the mapping $P_\Omega : I\!\!R^n \rightrightarrows \Omega$ the *projector* onto Ω defined by

$$P_\Omega x := \text{argmin}_{y \in \Omega}\|x - y\|. \tag{7}$$

This is in general a set-valued mapping, indicated by the notation "$\rightrightarrows$" [38, Chapter 5]. We call a point $\bar{x} \in P_\Omega x$ a *projection*. It is well known that if the set Ω is closed, nonempty and *convex* then the projector is single-valued. In a reasonable abuse of terminology and notation, we will write $P_\Omega x$ for the (there is only one) projection onto a convex set Ω. An operator closely related to the projector is the reflector. We call the (possibly set-valued) mapping $R_\Omega : I\!\!R^n \rightrightarrows I\!\!R^n$ the *reflector* across Ω defined by $R_\Omega x := 2P_\Omega x - x$. We call a point in $R_\Omega x$ a *reflection*. As with projections, when Ω is convex, we will write $R_\Omega x$ for the (there is only one) reflection. The projection/reflection methods discussed in this work are easy to implement, computationally efficient and lie at the foundation of many first-order methods for optimization.

Definition I.1 (alternating projections) *For two closed sets $\Omega_1, \Omega_2 \subset I\!\!R^n$ the mapping*

$$T_{AP}x := P_{\Omega_1}P_{\Omega_2}x \tag{8}$$

is called the alternating projections operator. The corresponding alternating projections algorithm is given by the iteration $x^{k+1} \in T_{AP}x^k$, $k \in I\!N$ with x^0 given.

Other well known algorithms, such as steepest descents for minimizing the average of squared distances between sets, can be formulated as instances of the alternating projections algorithm [36], [37]. We show below (Corollary III.13) that alternating projections corresponds to projected gradients for problems with special linear structure.

Definition I.2 (Douglas-Rachford) *For two closed sets $\Omega_1, \Omega_2 \subset I\!R^n$ the mapping*

$$T_{DR}x := \frac{1}{2}\left(R_{\Omega_1}R_{\Omega_2}x + x\right) \tag{9}$$

is called the Douglas-Rachford operator. The corresponding Douglas-Rachford algorithm is the fixed point iteration $x^{k+1} \in T_{DR}x^k$, $k \in I\!N$ with x^0 given.

The Douglas-Rachford algorithm [30] owes its prominence in large part to its relation via duality to the alternating directions method of multipliers for solving constrained optimization problems [20].

We present four main results, three of which are new. The first of these results, Theorem III.8, concerns local linear convergence of alternating projections to a solution of (6). This has been shown, with optimal rates, in [9]. Our proof uses fundamentally different tools developed in [21]. It is exactly these newer tools that enable us to prove the second of our main results, Theorem IV.7, namely local linear convergence of the Douglas-Rachford algorithm. Convergence of Douglas-Rachford, with rates, for sparse affine feasibility is a new result. In the remaining two main new results, Corollary III.13 and Theorem III.15, we specify classes of affine subspaces B for which alternating projections is *globally linearly convergent*. This shows that nonconvex models, in this case, can be a reasonable alternative to convex relaxations.

The outline of this paper is as follows. First we recall some definitions and results from variational analysis regarding alternating projections and Douglas-Rachford in Section 2. In Section 3 we show local linear convergence of alternating projections and we provide conditions on matrices M that guarantee global linear convergence of alternating projections. In the same section we formulate different conditions on the matrices M that guarantee global linear convergence of the same algorithm. In Section 4 we show that for most problems of interest in sparse optimization there exist fixed points of Douglas-Rachford that are *not* in the intersection $A_s \cap B$. On the other hand, we show that locally the iterates of Douglas-Rachford converge with linear rate to a fixed point whose *shadow* is a solution to (6). Finally in Section 5 we present numerical and analytical examples to illustrate the theoretical results.

II. PRELIMINARY DEFINITIONS AND RESULTS

We use the following notation, most of which is standard. We denote the *closed* ball of radius δ centered on $\bar{x}$ by $I\!B_\delta(\bar{x})$. We assume throughout that the matrix M is full rank in the definition of the affine subspace B (3). The nullspace of M is denoted $\ker M$ and $M^\dagger$ indicates the *Moore-Penrose inverse*, which, because of the full rank assumption, becomes

$$M^\dagger = M^\top \left(MM^\top\right)^{-1}. \tag{10}$$

The inner product of two points $x, y \in I\!R^n$ is denoted $\langle x, \ y \rangle$. The orthogonal complement to a nonempty affine set Ω is given by

$$\Omega^\perp := \{p \in I\!R^n \ | \ \langle p, \ v - w \rangle = 0 \ \forall \ v, w \in \Omega\}.$$

For two arbitrary sets $\Omega_1, \Omega_2 \subset I\!R^n$ we denote the Minkowski sum by $\Omega_1 + \Omega_2 := \{x_1 + x_2 \ | \ x_1 \in \Omega_1 \text{ and } x_2 \in \Omega_2\}$. The set of fixed points of a self-mapping T is given by Fix T. The identity mapping is denoted by Id. For a set $\Omega \subset I\!R^n$ we define the *distance* of a point $x \in I\!R^n$ to Ω by $d_\Omega(x) := \inf_{y \in \Omega} \|x - y\|$. When Ω is closed the distance is attained at a projection onto Ω, that is, $d_\Omega(x) = \|\bar{x} - x\|$ for $\bar{x} \in P_\Omega x$.

A. Tools and notions of regularity

Our proofs make use of some standard tools and notation from variational analysis which we briefly define here. We remind the reader of the definition of the projection onto a closed set (7). The following definition follows [8, Definition 2.1] and is based on [33, Definition 1.1 and Theorem 1.6].

Definition II.1 (normal cones) *The* proximal normal cone $N_\Omega^P(\bar{x})$ *to a closed nonemtpy set* $\Omega \subset \mathbb{R}^n$ *at a point* $\bar{x} \in \Omega$ *is defined by*

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

The limiting normal cone*, or simply the* normal cone $N_\Omega(\bar{x})$ *is defined as the set of all vectors 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 normal cone describes the local geometry of a set. What is meant by *regularity* of sets is made precise below.

Definition II.2 ((ε, δ)-subregularity) *A nonempty set* $\Omega \subset \mathbb{R}^n$ *is* (ε, δ)-subregular *at* $\bar{x}$ *with respect to* $U \subset \mathbb{R}^n$*, if there exist* $\varepsilon \geq 0$ *and* $\delta > 0$ *such that*

$$\langle v, z - y \rangle \leq \varepsilon \|v\| \|z - y\|$$

holds for all $y \in \Omega \cap \mathbb{B}_\delta(\bar{x})$, $z \in U \cap \mathbb{B}_\delta(\bar{x})$, $v \in N_\Omega(y)$. *We simply say* Ω *is* (ε, δ)-subregular *at* $\bar{x}$ *if* $U = \{\bar{x}\}$.

The definition of (ε, δ)-subregularity was introduced in [21] and is a generalization of the notion of (ε, δ)-regularity introduced in [8, Definition 8.1]. During the preparation of this article it was brought to our attention that a similar condition appears in the context of regularized inverse problems [24, Corollary 3.6].

We define next some notions of regularity of collections of sets that, together with (ε, δ)-subregularity, provide sufficient conditions for linear convergence of both alternating projections and Douglas-Rachford. In the case of Douglas-Rachford, as we shall see, these conditions are also necessary. Linear regularity, defined next, can be found in [2, Definition 3.13]. Local versions of this have appeared under various names in [23, Proposition 4], [34, Section 3], and [25, Equation (15)].

Definition II.3 (linear regularity)
A collection of closed, nonempty sets $(\Omega_1, \Omega_2, \ldots, \Omega_m) \subset \mathbb{R}^n$ *is called* locally linearly regular *at* $\bar{x} \in \cap_{j=1}^m \Omega_j$ *on* $\mathbb{B}_\delta(\bar{x})$ *($\delta > 0$) if there exists a* $\kappa > 0$ *such that*

$$d_{\cap_{j=1}^m \Omega_j}(x) \leq \kappa \max_{i=1,\ldots,m} d_{\Omega_i}(x), \quad \forall x \in \mathbb{B}_\delta(\bar{x}). \tag{11}$$

The infimum over all such κ *is called* modulus of regularity on $\mathbb{B}_\delta(\bar{x})$. *If there exists* $\kappa > 0$ *such that (11) holds at* $\bar{x}$ *for every* $\delta > 0$ *(that is, for all* $x \in \mathbb{R}^n$*), the collection of sets is said to be* linearly regular *there. The* modulus of regularity *of a collection of linearly regular sets is the infimum over all* κ *satisfying (11) for all* $\delta > 0$.

There is yet a stronger notion of regularity of collections of sets that we make use of called the *basic qualification condition for sets* in [33, Definition 3.2]. For the purposes of this paper we refer to this as *strong regularity*.

Definition II.4 (strong regularity) *The collection* (Ω_1, Ω_2) *is strongly regular at* $\bar{x}$ *if*

$$N_{\Omega_1}(\bar{x}) \cap -N_{\Omega_2}(\bar{x}) = \{0\}. \tag{12}$$

It can be shown that strong regularity implies local linear regularity (see, for instance [21]). Any collection of finite dimensional affine subspaces with nonempty intersection is linearly regular (see for instance [3, Proposition 5.9 and Remark 5.10]). Moreover, it follows from [4, Example 6.42], that if Ω_1 and Ω_2 are affine subspaces, we have $N_{\Omega_1} = \Omega_1^\perp, N_{\Omega_2} = -N_{\Omega_2} = \Omega_2^\perp$ and hence,

$$(\Omega_1, \Omega_2) \text{ is strongly regular at any } \bar{x} \in \Omega_1 \cap \Omega_2 \iff \Omega_1^\perp \cap \Omega_2^\perp = \{0\} \quad \text{and} \quad \Omega_1 \cap \Omega_2 \neq \emptyset. \tag{13}$$

In the case where Ω_1 and Ω_2 are affine subspaces we say that the collection is strongly regular without mention of any particular point in the intersection - as long as this is nonempty - since the collection is strongly regular at all points in the intersection.

B. General local linear convergence results

The algorithms that we consider here are fixed-point algorithms built upon projections onto sets. Using tools developed in [7] and [8], alternating projections applied to (6) was shown in [9] to be locally linearly convergent

with optimal rates in terms of the *Friedrichs angle* between A_s and B, and an estimate of the radius of convergence. Our approach, based on [21], is in line with [32] but does not rely on local firm nonexpansiveness of the fixed point mapping. It has the advantage of being general enough to be applied to *any* fixed point mapping, but the price one pays for this generality is in the rate estimates, which may not be optimal or easy to compute. We do not present the results of [21] in their full generality, but focus instead on the essential elements for affine feasibility with sparsity constraints.

Lemma II.5 (local linear convergence of alternating projections) *(See [21, Corollary 3.13].) Let the collection (Ω_1, Ω_2) be locally linearly regular at $\bar{x} \in \Omega := \Omega_1 \cap \Omega_2$ with modulus of regularity κ on $\mathbb{B}_\delta(\bar{x})$ and let Ω_1 and Ω_2 be $(\varepsilon, \delta)-$subregular at $\bar{x}$. For any $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$, generate the sequence $\left(x^k\right)_{k \in \mathbb{N}} \subset \mathbb{R}^n$ by alternating projections, that is, $x^{k+1} \in T_{AP} x^k$. Then*

$$d_\Omega(x^{k+1}) \leq \left(1 - \frac{1}{\kappa^2} + \varepsilon\right) d_\Omega\left(x^k\right).$$

In the analogous statement for the Douglas-Rachford algorithm, we defer, for the sake of simplicity, characterization of the constant in the asserted linear convergence rate. A more refined analysis of such rate constants and their geometric interpretation is the subject of future research.

Lemma II.6 (local linear convergence of Douglas-Rachford) *(See [21, Corollary 3.20].) For $T := \frac{1}{2}(R_{\Omega_1} R_{\Omega_2} + Id)$ and Ω_1, Ω_2 affine subspaces with $\Omega_1 \cap \Omega_2 \neq \emptyset$, Fix $T = \Omega_1 \cap \Omega_2$ if and only if (Ω_1, Ω_2) is strongly regular, in which case the corresponding Picard iteration converges linearly to the intersection.*

III. SPARSE FEASIBILITY WITH AN AFFINE CONSTRAINT: LOCAL AND GLOBAL CONVERGENCE OF ALTERNATING PROJECTIONS

We are now ready to apply the above general results to affine sparse feasibility. We begin with characterization of the regularity of the sets involved.

A. Regularity of sparse sets

We specialize to the case where B is an affine subspace defined by (3) and A_s defined by (5) is the set of vectors with at most s nonzero elements. Following [9] we decompose the set A_s into a union of subspaces. For $a \in \mathbb{R}^n$ define the *sparsity subspace* associated with a by

$$\text{supp}(a) := \{x \in \mathbb{R}^n | \; x_j = 0 \text{ if } a_j = 0\}, \tag{14}$$

and the mapping

$$I : \mathbb{R}^n \to \{1, \ldots, n\}, \; x \mapsto \{i \in \{1, \ldots, n\} | \; x_i \neq 0\}. \tag{15}$$

Define $\mathcal{J} := 2^{\{1,2,\ldots,n\}}$ and $\mathcal{J}_s := \{J \in \mathcal{J} | \; J \text{ has } s \text{ elements}\}$. The set A_s can be written as the union of all subspaces indexed by $J \in \mathcal{J}_s$ [9, Equation (27d)],

$$A_s = \bigcup_{J \in \mathcal{J}_s} A_J, \tag{16}$$

where $A_J := \text{span}\{e_i | \; i \in J\}$ and e_i is the $i-$th standard unit vector in $\mathbb{R}^n$. For $x \in \mathbb{R}^n$ we define the set of s largest coordinates in absolute value

$$C_s(x) := \left\{J \in \mathcal{J}_s \middle| \; \min_{i \in J} |x_i| \geq \max_{i \notin J} |x_i|\right\}. \tag{17}$$

The next elementary result will be useful later.

Lemma III.1 *(See [9, Lemma 3.4])*
Let $a \in A_s$ and assume $s \leq n - 1$. Then

$$\min\{d_{A_J}(a) \mid a \notin A_J, \; J \in \mathcal{J}_s\} = \min\{|a_j| \mid j \in I(a)\}. \tag{18}$$

Using the above notation, the normal cone to the sparsity set A_s at $a \in A_s$ has the following closed-form representation (see [9, Theorem 3.9] and [31, Proposition 3.6] for the general matrix representation).

$$
\begin{aligned}
N_{A_s}(a) &= \{\nu \in {I\!\!R}^n \mid \|\nu\|_0 \le n - s\} \cap (\operatorname{supp}(a))^\perp \\
&= \bigcup_{J \in \mathcal{J}_s, I(a) \subseteq J} A_J^\perp.
\end{aligned}
\tag{19}
$$

The normal cone to the affine set B also has a simple closed form, namely $N_B(x) = B^\perp$ (see for example [33, Proposition 1.5]). Let $y \in {I\!\!R}^n$ be a point such that $My = p$. Note that $\ker M$ is the subspace parallel to B, i.e. $\ker M = B + \{-y\}$.

This notation yields the following explicit representations for the projectors onto A_s [9, Proposition 3.6] and B:

$$
P_B x := x - M^\dagger(Mx - p) \quad \text{and} \quad P_{A_s} x := \bigcup_{J \in C_s(x)} P_{A_J} x,
\tag{20}
$$

where $M^\dagger$ is given by (10) and

$$
(P_{A_J} x)_i = \begin{cases} x_i, & i \in J, \\ 0, & i \notin J \end{cases}.
\tag{21}
$$

We collect next some facts about the projectors and reflectors of A_s and B. We remind the reader that, in a slight abuse of notation, since the set B is convex, we make no distinction between the projector $P_B x$ and the projection $\bar{x} \in P_B x$.

Lemma III.2 *Let A_s and B be defined by (5) and (3). Let $a \in A_s$ and $b \in B$. For any $\delta_a \in (0, \min\{|a_j| \mid j \in I(a)\})$ and $\delta_b \in (0, \infty)$ the following hold:*
   (i) $P_B x \in {I\!\!B}_{\delta_b}(b)$ for all $x \in {I\!\!B}_{\delta_b}(b)$;
   (ii) $P_{A_s} x \subset {I\!\!B}_{\delta_a/2}(a)$ for all $x \in {I\!\!B}_{\delta_a/2}(a)$;
   (iii) $R_B x \in {I\!\!B}_{\delta_b}(b)$ for all $x \in {I\!\!B}_{\delta_b}(b)$;
   (iv) $R_{A_s} x \subset {I\!\!B}_{\delta_a/2}(a)$ for all $x \in {I\!\!B}_{\delta_a/2}(a)$.

Proof: (i). This follows from the fact that the projector is nonexpansive, since B is convex and $\|P_B x - b\| = \|P_B x - P_B b\| \le \|x - b\|$.

(ii). Let $x \in {I\!\!B}_{\delta_a/2}(a)$. For any $i \in I^\circ(a) := \{i : a_i = 0\}$, we have $|x_i - a_i| = |x_i| \le \delta_a/2$. Moreover, for all $j \in I(a) := \{j : a_j \ne 0\}$, we have $|x_j - a_j| \le \delta_a/2$ and so $|x_j| > \delta_a/2$ for all $j \in I(a)$. Altogether this means that $|x_j| > |x_i|$ for all $i \in I^\circ(a), j \in I(a)$. Therefore the indices of the nonzero elements of a correspond exactly to the indices of the $|I(a)|$-largest elements of x, where $|I(a)|$ denotes the cardinality of the set $I(a)$. Since $|I(a)| \le s$, the projector of x need not be single-valued. (Consider the case $a = (1, 0, \ldots, 0)^\top$ and $x = (1, \delta_a/4, \delta_a/4, 0, \ldots, 0)^\top$ and $s = 2$.) Nevertheless, for all $x^+ \in P_{A_s} x$ we have $a \in \operatorname{supp}(x^+)$ where $\operatorname{supp}(x^+)$ is defined by (14). (Suppose not. Then $a \notin \operatorname{supp}(x^+)$ for some $x^+ \in P_{A_s} x$, that is, for some $i \in I(a)$ we have $a_i \ne 0$ while $x_i^+ = 0$. But then $\|a - x^+\| \ge |a_i| > \delta_a$, which is a contradiction to $x \in {I\!\!B}_{\delta_a/2}(a)$.) Since $\operatorname{supp}(x^+)$ is a subspace, x^+ is the orthogonal projection of x onto a subspace, hence by Pythagoras' Theorem

$$
\begin{aligned}
\|x - x^+\|_2^2 + \|x^+ - a\|_2^2 &= \|x - a\|_2^2 \\
\text{and} \quad \|x^+ - a\|_2 \le \|x - a\|_2 &\le \tfrac{\delta}{2}.
\end{aligned}
\tag{22}
$$

Thus $P_{A_s} x \subset {I\!\!B}_{\delta_a/2}(a)$.

(iii). Since the reflector R_B is with respect to an affine subspace containing b a simple geometric argument shows that for all x we have $\|R_B x - b\| = \|x - b\|$. The result follows immediately.

(iv). As in the proof of (ii), for all $x \in {I\!\!B}_{\delta_a/2}$ we have $a \in \operatorname{supp}(x^+)$ for each $x^+ \in P_{A_s} x$. In other words, the projector, and hence the corresponding reflector, is with respect to a subspace containing a. Thus, as in (iii), $\|R_{A_s} x - a\| = \|x - a\|$, though in this case only for $x \in {I\!\!B}_{\delta_a/2}$. ∎

The next lemma shows that around any point $\bar{x} \in A_s$ the set A_s is the union of subspaces in A_s containing $\bar{x}$. Hence around any point $\bar{x} \in A_s \cap B$ the intersection $A_s \cap B$ can be described locally as the intersection of subspaces and the affine set B, each containing $\bar{x}$.

Lemma III.3 *Let $\bar{x} \in A_s \cap B$ with $0 < \|\bar{x}\|_0 \leq s$. Then for all $\delta < \min\{|\bar{x}_i| : \bar{x}_i \neq 0\}$ we have*

$$A_s \cap I\!\!B_\delta(\bar{x}) = \bigcup_{J \in \mathcal{J}_s, \ I(\bar{x}) \subseteq J} A_J \cap I\!\!B_\delta(\bar{x}) \tag{23}$$

and hence

$$A_s \cap B \cap I\!\!B_\delta(\bar{x}) = \bigcup_{J \in \mathcal{J}_s, \ I(\bar{x}) \subseteq J} A_J \cap B \cap I\!\!B_\delta(\bar{x}). \tag{24}$$

If in fact $\|\bar{x}\|_0 = s$, then there is a unique $J \in \mathcal{J}_s$ such that for all $\delta < \min\{|\bar{x}_i| : \bar{x}_i \neq 0\}$ we have $A_s \cap I\!\!B_\delta(\bar{x}) = A_J \cap I\!\!B_\delta(\bar{x})$ and hence $A_s \cap B \cap I\!\!B_\delta(\bar{x}) = A_J \cap B \cap I\!\!B_\delta(\bar{x})$.

Proof: If $s = n$, then the set A_s is all of $I\!\!R^n$ and both statements are trivial. For the case $s \leq n - 1$, choose any $x \in I\!\!B_\delta(\bar{x}) \cap A_s$. From the definition of δ and Lemma III.1 we have that, for any $J \in \mathcal{J}_s$, if $\bar{x} \notin A_J$, then $d_{A_J}(\bar{x}) > \delta$, and, since $\|x - \bar{x}\| < \delta$, we have $x \notin A_J$. By contraposition, therefore, $x \in A_J$ implies that $\bar{x} \in A_J$, hence, for each $x \in I\!\!B_\delta(\bar{x}) \cap A_s$, we have $x \in I\!\!B_\delta(\bar{x}) \cap A_{I(x)}$ where $I(\bar{x}) \subseteq I(x) \in \mathcal{J}_s$. The intersection $I\!\!B_\delta(\bar{x}) \cap A_s$ is then the union over all such intersections as given by (23). Equation (24) is an immediate consequence of (23).

If, in addition $\|\bar{x}\|_0 = s$, then the cardinality of $I(\bar{x})$ is s and by [9, Lemma 3.5] $C_s(\bar{x}) = \{I(\bar{x})\}$, where $C_s(\bar{x})$ is given by (17). This means that if $\bar{x}$ has sparsity s, then there is exactly one subspace A_J with index set $J := I(\bar{x})$ in $\mathcal{J}_s$ containing $\bar{x}$. By (16) we have $A_s = \bigcup_{J \in \mathcal{J}_s} A_J$ and by Lemma III.1 we conclude $d_{A_K}(\bar{x}) \geq \min\{|\bar{x}_j| \mid j \in J\} > \delta$ for all $K \in \mathcal{J}_s \setminus J$. This gives us the equality $A_s \cap I\!\!B_\delta(\bar{x}) = A_J \cap I\!\!B_\delta(\bar{x})$ and hence $A_s \cap B \cap I\!\!B_\delta(\bar{x}) = A_J \cap B \cap I\!\!B_\delta(\bar{x})$, as claimed. ∎

We conclude this introductory section with a characterization of the sparsity set A_s.

Theorem III.4 (regularity of A_s) *At any point $\bar{x} \in A_s \backslash \{0\}$ the set A_s is $(0, \delta)$-subregular at $\bar{x}$ for $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. On the other hand, the set A_s is not $(0, \delta)$-subregular at $\bar{x} \in A_s \backslash \{0\}$ for any $\delta \geq \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$. In contrast, at 0 the set A_s is $(0, \infty)$-subregular.*

Proof: Choose any $x \in I\!\!B_\delta(\bar{x}) \cap A_s$ and any $v \in N_{A_s}(x)$. By the characterization of the normal cone in (19) there is some $J \in \mathcal{J}_s$ with $I(x) \subseteq J$ and $v \in A_J^\perp \subset N_{A_s}(x)$. As in the proof of Lemma III.3, for any $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$ we have $I(\bar{x}) \subseteq I(x)$, hence $\bar{x} - x \in A_J$ and thus $\langle v, \bar{x} - x \rangle = 0$. By the definition of (ε, δ)-regularity (Definition II.2) A_s is $(0, \delta)$-subregular as claimed.

That A_s is not $(0, \delta)$-subregular at $\bar{x} \in A_s \setminus \{0\}$ for any $\delta \geq \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$ follows from the failure of Lemma III.3 on balls larger than $\min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. Indeed, suppose $\delta \geq \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$, then by Lemma III.1 there is a point $x \in I\!\!B_\delta(\bar{x}) \cap A_s$ and there exists $J \in \mathcal{J}_s$ with $x \in A_J \subset A_s$ but $\bar{x} \notin A_J$. In particular, $A_J^\perp \cap A_{I(\bar{x})} \neq \{0\}$, so there exists a nonzero $v \in A_J^\perp \cap A_{I(\bar{x})}$. Then $v \in N_{A_s}(x)$ and $|\langle v, \bar{x} - x \rangle| = |\langle v, \bar{x} \rangle| > 0$. Since $N_{A_s}(x)$ is a union of subspaces, the sign of v can be chosen so that $\langle v, \bar{x} - x \rangle > 0$, in violation of $(0, \delta)$-subregularity.

For the case $\bar{x} = 0$, by (19) for any $x \in A_s$ and $v \in N_{A_s}(x)$ we have $\langle v, x \rangle = 0$, since $\text{supp}(x)^\perp \perp \text{supp}(x)$, which completes the proof. ∎

B. Regularity of the collection (A_s, B)

We show in this section that the collection (A_s, B) is locally linearly regular as long as the intersection is nonempty. We begin with a technical lemma.

Lemma III.5 (linear regularity under unions) *Let $(\Omega_1, \Omega_2, \ldots, \Omega_m, \Omega_{m+1})$ be a collection of nonempty subsets of $I\!\!R^n$ with nonempty intersection. Let $\bar{x} \in \left(\cap_{j=1}^m \Omega_j\right) \cap \Omega_{m+1}$. Suppose that, for some $\delta > 0$, the pair (Ω_j, Ω_{m+1}) is locally linearly regular with modulus κ_j on $I\!\!B_\delta(\bar{x})$ for each $j \in \{1, 2, \ldots, m\}$. Then the collection $\left(\bigcup_{j=1}^m \Omega_j, \Omega_{m+1}\right)$ is locally linearly regular at $\bar{x}$ on $I\!\!B_\delta(\bar{x})$ with modulus $\overline{\kappa} = \max_j\{\kappa_j\}$.*

Proof: Denote $\Gamma := \bigcup_{j=1}^m \Omega_j$. First note that for all $x \in I\!\!B_\delta(\bar{x})$ we have

$$d_{\Gamma \cap \Omega_{m+1}}(x) = \min_j \left\{ d_{\Omega_j \cap \Omega_{m+1}}(x) \right\} \leq \min_j \left\{ \kappa_j \max\{d_{\Omega_j}(x), \ d_{\Omega_{m+1}}(x)\} \right\}, \tag{25}$$

where the inequality on the right follows from the assumption that (Ω_j, Ω_{m+1}) is locally linearly regular with modulus κ_j on $\mathbb{B}_\delta(x)$. Let $\overline{\kappa} \geq \max_j\{\kappa_j\}$. Then

$$d_{\Gamma \cap \Omega_{m+1}}(x) \leq \overline{\kappa} \min_j \left\{ \max\{d_{\Omega_j}(x), \ d_{\Omega_{m+1}}(x)\} \right\} = \overline{\kappa} \max \left\{ \min_j\{d_{\Omega_j}(x)\}, \ d_{\Omega_{m+1}}(x) \right\}. \tag{26}$$

This completes the proof. ■

Theorem III.6 (regularity of (A_s, B)) *Let A_s and B be defined by (5) and (3) with $A_s \cap B \neq \emptyset$. At any $\bar{x} \in A_s \cap B$ and for any $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$ the collection (A_s, B) is locally linearly regular on $\mathbb{B}_{\delta/2}(\bar{x})$ with modulus of regularity $\overline{\kappa} = \max\limits_{J \in \mathcal{J}_s, I(\bar{x}) \subseteq J} \{\kappa_J\}$ where κ_J is the modulus of regularity of the collection (A_J, B).*

Proof: For any $\bar{x} \in A_s \cap B$ we have $\bar{x} \in A_J \cap B$ for all $J \in \mathcal{J}_s$ with $I(\bar{x}) \subseteq J$ and thus (A_J, B) is linearly regular with modulus of regularity κ_J [3, Proposition 5.9 and Remark 5.10]. Define

$$A_s^{\bar{x}} := \bigcup_{J \in \mathcal{J}_s, \ I(\bar{x}) \subseteq J} A_J.$$

Then by Lemma III.5 the collection $(A_s^{\bar{x}}, B)$ is linearly regular at $\bar{x}$ with modulus of regularity $\overline{\kappa} := \max\limits_{J \in \mathcal{J}_s, \ I(\bar{x}) \subseteq J} \{\kappa_J\}$. By Lemma III.3 $A_s \cap \mathbb{B}_{\delta/2}(\bar{x}) = A_s^{\bar{x}} \cap \mathbb{B}_{\delta/2}(\bar{x})$ for any $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$. Moreover, by Lemma III.2(ii), for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$, we have $P_{A_s} x \subset \mathbb{B}_{\delta/2}(\bar{x})$, and thus $P_{A_s} x = P_{A_s^{\bar{x}}} x$. In other words, $d_{A_s}(x) = d_{A_s^{\bar{x}}}(x)$ for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$, hence the collection (A_s, B) is locally linearly regular on $\mathbb{B}_\delta(\bar{x})$ with modulus $\overline{\kappa}$. This completes the proof. ■

Remark III.7 A simple example shows that the collection (A_s, B) need not be linearly regular. Consider the sparsity set A_1, the affine set $B = \{(1, \tau, 0)^\top \mid \tau \in \mathbb{R}\}$ and the sequence of points $(x^k)_{k \in \mathbb{N}}$ defined by $x^k = (0, k, 0)^\top$. Then $A_1 \cap B = \{(1, 0, 0)^\top\}$ and $\max\{d_{A_1}(x^k), d_B(x^k)\} = 1$ for all k while $d_{A_1 \cap B}(x^k) \to \infty$ as $k \to \infty$.

C. Local linear convergence of alternating projections

The next result shows the local linear convergence of alternating projections to a solution of (6). This was also shown in [9, Theorem 3.19] using very different techniques. The approach taken here based on the modulus of regularity κ on $\mathbb{B}_\delta(x)$ is more general, that is, it can be applied to other nonconvex problems, but the relationship between the modulus of regularity and the *angle of intersection* which is used to characterize the optimal rate of convergence [9, Theorem 2.11] is not fully understood.

Theorem III.8 *Let A_s and B be defined by (5) and (3) with nonempty intersection and let $\bar{x} \in A_s \cap B$. Choose $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. For $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$ the alternating projections iterates converge linearly to the intersection $A_s \cap B$ with rate $\left(1 - \frac{1}{\kappa^2}\right)$ where κ is the modulus of regularity of (A_s, B) on $\mathbb{B}_\delta(\bar{x})$ (Definition II.3).*

Proof: We show that we may apply Lemma II.5. By Lemma III.2(i) and (ii) the projections P_B and P_{A_s} each map $\mathbb{B}_{\delta/2}(\bar{x})$ to itself, hence their composition maps $\mathbb{B}_{\delta/2}(\bar{x})$ to itself. Moreover, the set B is $(0, +\infty)$-subregular at every point in B (i.e., convex) and, by Theorem III.4, the sparsity set A_s is $(0, \delta)-$ subregular at $\bar{x}$. Finally, by Theorem III.6 the pair (A_s, B) is locally linearly regular at $\bar{x}$ on $\mathbb{B}_\delta(\bar{x})$. The assertion then follows from Lemma II.5 with $\epsilon = 0$. ■

Remark III.9 The above result does *not* need an exact a priori assumption on the sparsity s. If there is a solution $\bar{x} \in A_s \cap B$, then $\|\bar{x}\|_0$ can be smaller than s and, geometrically speaking, $\bar{x}$ is on a crossing of linear subspaces contained in A_s. It is also worth noting that the assumptions are also *not* tantamount to local convexity. In the case that B is a subspace, the point 0 is trivially a solution to (6) (and, for that matter (1)). The set A_s is not convex on any neighborhood of 0, however the assumptions of Theorem III.8 hold, and alternating projections indeed converges locally linearly to 0, regardless of the size of the parameter s.

D. Global convergence of alternating projections

Following [11] where the authors consider problem (4), we present a sufficient condition for global linear convergence of the alternating projections algorithm for affine sparse feasibility. Though our presentation is modeled after [11] this work is predated by the nearly identical approach developed in [12], [13]. We also note that the arguments presented here do not use any structure that is particular to $\mathbb{R}^n$, hence the results can be extended, as they were in [11], to the problem of finding the intersection of the set of matrices with rank at most s and an affine subspace in the Euclidean space of matrices. Since this generalization complicates the local analysis, we have chosen to limit our scope to $\mathbb{R}^n$.

Key to the analysis of [12], [13], [11] are the following well-known restrictions on the matrix M.

Definition III.10 *The mapping $M : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the* restricted isometry property *of order s, if there exists $0 \leq \delta \leq 1$ such that*

$$(1-\delta)\|x\|_2^2 \leq \|Mx\|_2^2 \leq (1+\delta)\|x\|_2^2 \quad \forall x \in A_s. \tag{27}$$

The infimum δ_s of all such δ is the restricted isometry constant.
The mapping $M : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the scaled/asymmetric restricted isometry property *of order (s, α) for $\alpha > 1$, if there exist $\nu_s, \mu_s > 0$ with $1 \leq \frac{\mu_s}{\nu_s} < \alpha$ such that*

$$\nu_s\|x\|_2^2 \leq \|Mx\|_2^2 \leq \mu_s\|x\|_2^2 \quad \forall x \in A_s. \tag{28}$$

The restricted isometry property (27) was introduced in [16], while the asymmetric version (28) first appeared in [12, Theorem 4]. Clearly (27) implies (28), since if a matrix M satisfies (27) of order s with restricted isometry constant δ_s, then it also satisfies (28) of order (s, β) for $\beta > \frac{1+\delta_s}{1-\delta_s}$.

To motivate the projected gradient algorithm given below, note that any solution to (6) is also a solution to

$$\text{Find } \bar{x} \in S := \text{argmin}_{x \in A_s} \; \frac{1}{2}\|Mx - p\|_2^2. \tag{29}$$

Conversely, if $A_s \cap B \neq \emptyset$ and $\bar{x}$ is in S, then $\bar{x}$ solves (6).

Definition III.11 (projected gradients) *Given a closed set $A \subset \mathbb{R}^n$, a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a positive real number τ, the mapping*

$$T_{PG}(x;\tau) = P_A\left(x - \frac{1}{\tau}\nabla f(x)\right) \tag{30}$$

is called the projected gradient operator. The projected gradients algorithm is the fixed point iteration

$$x^{k+1} \in T_{PG}(x^k; \tau_k) = P_A\left(x^k - \frac{1}{\tau_k}\nabla f(x^k)\right), \; k \in \mathbb{N}$$

for x^0 given arbitrarily and a sequence of positive real numbers $(\tau_k)_{k \in \mathbb{N}}$.

In the context of linear least squares with a sparsity constraint, the projected gradient algorithm is equivalent to what is also known as the iterative hard thresholding algorithm (see for instance [12], [13], [26]) where the constraint $A = A_s$ and the projector given by (20) amounts to a thresholding operation on the largest elements of the iterate.

With these definitions we cite a result on convergence of the projected gradient algorithm applied to (29) (see [13, Theorem 4] and [11, Theorem 3 and Corollary 1]).

Theorem III.12 (global convergence of projected gradients/iterative hard thresholding) *Let M satisfy (28) of order $(2s, 2)$ and, for any given initial point x^0, let the sequence $(x^k)_{k \in \mathbb{N}}$ be generated by the projected gradient algorithm with $A = A_s$, $f(x) = \frac{1}{2}\|Mx - p\|_2^2$ and the constant step size $\tau \in [\mu_{2s}, 2\nu_{2s})$. Then there exists a unique global solution $\bar{x}$ to (29) and the iterates converge to $\bar{x}$ for any $x^0 \in \mathbb{R}^n$. Further, $f(x^k) \to 0$ linearly as $k \to \infty$ with rate $\rho = \left(\frac{\tau}{\nu_{2s}} - 1\right) < 1$, that is,*

$$f(x^{k+1}) \leq \rho f(x^k) \qquad (\forall k \in \mathbb{N}).$$

We specialize this theorem to alternating projections next.

Corollary III.13 (global convergence of alternating projections I) *Let the matrix M satisfy (28) of order $(2s, 2)$ with $\mu_{2s} = 1$ and $MM^\top = Id$. Then $A_s \cap B$ is a singleton $\bar{x}$ and alternating projections applied to (6) converges to $\bar{x}$ for every initial point x^0. Further, the values of $f(x^k) = \frac{1}{2} \left\| Mx^k - p \right\|_2^2$ converge linearly to 0 with rate $\rho = \left(\frac{1}{\nu_{2s}} - 1 \right) < 1$.*

Proof: For $f(x) = \frac{1}{2} \|Mx - p\|_2^2$ we have $\nabla f(x) = M^\top (Mx - p)$. The projected gradients iteration with constant step length $\tau = 1$ then takes the form

$$x^{k+1} \in P_{A_s} \left(x^k - \nabla f(x^k) \right) = P_{A_s} \left(x^k - M^\top (Mx^k - p) \right).$$

The projection onto the subspace B is given by (see (20))

$$P_B x = \left(\text{Id} - M^\top (MM^\top)^{-1} M \right) x + M^\top (MM^\top)^{-1} p.$$

Since $MM^\top = \text{Id}$ this simplifies to $x^k - M^\top (Mx^k - p) = P_B x^k$, hence

$$x^{k+1} \in P_{A_s} \left(x^k - \nabla f(x^k) \right) = P_{A_s} P_B x^k.$$

This shows that projected gradients III.11 with unit step length applied to (29) with $A = A_s$ and $f(x) = \frac{1}{2} \|Mx - p\|_2^2$ is equivalent to the method of alternating projections I.1 applied to (6).

To show convergence to a unique solution, we apply Theorem III.12, for which we must show that the step length $\tau = 1$ lies in the nonempty interval $[\mu_{2s}, 2\nu_{2s})$. By assumption M satisfies (28) of order $(2s, 2)$ with $\mu_{2s} = 1$. Hence $\frac{1}{2} < \nu_{2s} \leq 1$ and $\tau = 1$ lies in the nonempty interval $[1, 2\nu_{2s})$. The assumptions of Theorem III.12 are thus satisfied with $\tau = 1$, whence global linear convergence to the unique solution of (29), and hence (6), immediately follows. ∎

The restriction to matrices satisfying $MM^\top = \text{Id}$ is very strong indeed. We consider next a different condition that, in principle, can be more broadly applied to the alternating projections algorithm. The difference lies in our ansatz: while in [11] the goal is to minimize $f(x) := \frac{1}{2} \|Mx - p\|_2^2$ over $x \in A_s$, we solve instead

$$\underset{x \in A_s}{\text{minimize}} \ g(x) := \frac{1}{2} d_B (x)^2. \tag{31}$$

These are different objective functions, yet the idea is similar: both functions f and g take the value zero on A_s if and only if $x \in A_s \cap B$. The distance of the point x to B, however, is the space of signals, while f measures the distance of the image of x under M to the measurement. The former is more robust to bad conditioning of the matrix $M \in \mathbb{R}^{m \times n}$ with $m < n$, since a poorly-conditioned M could still yield a small residual $\frac{1}{2} \|Mx - p\|_2^2$.

Note also that the matrix $M^\dagger M$ is the orthogonal projection onto the subspace $\ker(M)^\perp$. This means that the operator norm of $M^\dagger M$ is 1 and so we have, for all $x \in \mathbb{R}^n$, that $\left\| M^\dagger Mx \right\|_2 \leq \|x\|_2$. Our second global result for alternating projections given below, involves a scaled/asymmetric restricted isometry condition analogous to (28) with M replaced by $M^\dagger M$. This only requires a *lower bound* on the operator norm of $M^\dagger M$ with respect to vectors of sparsity $2s$ since the upper bound analogous to (28) is automatic. Specifically, we assume that

$$M \text{ is full rank and } (1 - \delta_{2s}) \|x\|_2^2 \leq \left\| M^\dagger Mx \right\|_2^2 \quad \forall \, x \in A_{2s}. \tag{32}$$

The condition (32) can be reformulated in terms of the scaled/asymmetric restricted isometry property (28) and strong regularity of the range of $M^\top$ and the complement of each of the subspaces comprising A_{2s}. We remind the reader that $A_J := \text{span} \{ e_i | \ i \in J \}$ for $J \in \mathcal{J}_{2s} := \left\{ J \in 2^{\{1,2,\dots,n\}} | \ J \text{ has } 2s \text{ elements} \right\}$.

Proposition III.14 (scaled/asymmetric restricted isometry and strong regularity) *Let $M \in \mathbb{R}^{m \times n}$ with $m \leq n$ be full rank. Then M satisfies (32) with $\delta_{2s} \in [0, \frac{\alpha - 1}{\alpha})$ for some fixed $s > 0$ and $\alpha > 1$ if and only if $M^\dagger M$ satisfies the scaled/asymmetric restricted isometry property (28) of order $(2s, \alpha)$ with $\mu_{2s} = 1$ and $\nu_{2s} = (1 - \delta_{2s})$. Moreover, for M satisfying (32) with $\delta_{2s} \in [0, \frac{\alpha - 1}{\alpha})$ for some fixed $s > 0$ and $\alpha > 1$, for all $J \in \mathcal{J}_{2s}$ the collection $\left(A_J^\perp, \text{range}(M^\top) \right)$ is strongly regular (Definition II.4), that is,*

$$(\forall J \in \mathcal{J}_{2s}) \qquad A_J \cap \ker(M) = \{0\}. \tag{33}$$

Proof: The first statement follows directly from the definition of the scaled/asymmetric restricted isometry property.

For the second statement, note that, if M satisfies inequality (32) with $\delta_{2s} \in [0, \frac{\alpha-1}{\alpha})$ for some fixed $s > 0$ and $\alpha > 1$, then the only element in A_{2s} satisfying $M^\dagger M x = 0$ is $x = 0$. Recall that $M^\dagger M$ is the projector onto the space orthogonal to the nullspace of M, that is, the projector onto the range of $M^\top$. Thus

$$A_{2s} \cap [\text{range}(M^\top)]^\perp = \{0\}. \tag{34}$$

Here we have used the fact that the projection of a point x onto a subspace Ω is zero if and only if $x \in \Omega^\perp$. Now using the representation for A_{2s} given by (16) we have that (34) is equivalent to

$$A_J \cap \ker(M^\top) = \{0\} \quad \text{for all } J \in \mathcal{J}_{2s}. \tag{35}$$

But by (13) this is equivalent to the strong regularity of $\left(A_J^\perp, \text{range}(M^\top)\right)$ for all $J \in \mathcal{J}_{2s}$. ∎

We are now ready to prove one of our main new results.

Theorem III.15 (global convergence of alternating projections II) *For a fixed $s > 0$, let the matrix $M^\dagger M$ satisfy (32) with $\delta_{2s} \in [0, \frac{1}{2})$ for M in the definition of the affine set B given by (3) . Then $B \cap A_s$ is a singleton and, for any initial value $x^0 \in \mathbb{R}^n$, the sequence $(x^k)_{k \in \mathbb{N}}$ generated by alternating projections with $x^{k+1} \in P_{A_s} P_B x^k$ converges to $B \cap A_s$. Moreover, $d_B\left(x^k\right) \to 0$ as $k \to \infty$ at a linear rate with constant bounded by $\sqrt{\frac{\delta_{2s}}{1-\delta_{2s}}}$.*

Proof: From the correspondence between (32) and (28) in Proposition III.14, we can apply Theorem III.12 to the feasibility problem Find $x \in A_s \cap \hat{B}$, where $\hat{B} := \{x \mid M^\dagger M x = \hat{p}\}$ for $\hat{p} := M^\dagger p$. This establishes that the intersection is a singleton. But from (10) the set $\hat{B}$ is none other than B, hence (32) for $\alpha = 2$ implies existence and uniqueness of the intersection $A_s \cap B$. To establish convergence of alternating projections, for the iterate x^k define the mapping

$$q(x, x^k) := g(x^k) + \left\langle x - x^k, M^\dagger(M x^k - p) \right\rangle + \frac{1}{2} \left\| x - x^k \right\|_2^2,$$

where g is the objective function defined in (31). By definition of the projector, the iterate x^{k+1} is a solution to the problem $\min \left\{ q(x, x^k) \mid x \in A_s \right\}$. To see this, recall that, by the definition of the projection, $g(x^k) = \frac{1}{2}\|x^k - P_B x^k\|^2$. Together with (20), i.e. $x^k - P_B x^k = M^\dagger(M x^k - p)$, this yields

$$
\begin{aligned}
q(x, x^k) &= \frac{1}{2} \left\| x^k - P_B x^k \right\|_2^2 + \left\langle x - x^k, x^k - P_B x^k \right\rangle + \frac{1}{2} \left\| x - x^k \right\|_2^2 \\
&= \frac{1}{2} \left\| x - x^k + x^k - P_B x^k \right\|_2^2.
\end{aligned} \tag{36}
$$

Now, by definition of the alternating projections sequence,

$$x^{k+1} \in P_{A_s} P_B x^k = P_{A_s}\left(x^k - (\text{Id} - P_B)x^k\right),$$

which, together with (36), yields

$$x^{k+1} \in \underset{x \in A_s}{\text{argmin}} \left\{ \left\| x - \left(x^k - (\text{Id} - P_B)x^k\right) \right\|_2^2 \right\} = \underset{x \in A_s}{\text{argmin}}\{q(x, x^k)\}.$$

That is, x^{k+1} is a minimizer of $q(x, x^k)$ in A_s. On the other hand, by (20) and (31), we have

$$g(x^{k+1}) = \frac{1}{2} \left\| M^\dagger(M x^{k+1} - p) \right\|_2^2. \tag{37}$$

Expanding the above expression and using fact that $M^\dagger M$ is an orthogonal projection onto a subspace yields

$$
\begin{aligned}
g(x^{k+1}) &= \frac{1}{2} \left\| M^\dagger M(x^{k+1} - x^k) + M^\dagger(M x^k - p) \right\|_2^2 \\
&= g(x^k) + \left\langle M^\dagger M(x^{k+1} - x^k), M^\dagger(M x^k - p) \right\rangle + \frac{1}{2} \left\| M^\dagger M(x^{k+1} - x^k) \right\|_2^2
\end{aligned}
$$

$$\leq g(x^k) + \left\langle M^\dagger M(x^{k+1} - x^k), M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2. \tag{38}$$

Now, by (10), the right hand side of (38) is equivalent to

$$g(x^k) + \left\langle x^{k+1} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2 = q(x^{k+1}, x^k). \tag{39}$$

Altogether, (38)-(39) yield $g(x^{k+1}) \leq q(x^{k+1}, x^k)$. But since x^{k+1} minimizes $q(x, x^k)$ over A_s, we know that, for $\{\overline{x}\} = B \cap A_s$,

$$q(x^{k+1}, x^k) \leq q(\overline{x}, x^k). \tag{40}$$

Moreover, by assumption (32) we have

$$\begin{aligned}
q(\overline{x}, x^k) =& g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|\overline{x} - x^k\right\|_2^2 \\
\leq& g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2(1 - \delta_{2s})}\left\|M^\dagger M(\overline{x} - x^k)\right\|_2^2 \\
=& g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2(1 - \delta_{2s})}\left\|M^\dagger(p - Mx^k)\right\|_2^2.
\end{aligned} \tag{41}$$

By (20) and (31) we have

$$\begin{aligned}
& g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2(1 - \delta_{2s})}\left\|M^\dagger(p - Mx^k)\right\|_2^2 \\
& = \left(1 + \frac{1}{1 - \delta_{2s}}\right) g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle
\end{aligned} \tag{42}$$

Finally, by (10) and another application of (20) and (31) we have

$$\begin{aligned}
& \left(1 + \frac{1}{1 - \delta_{2s}}\right) g(x^k) + \left\langle \overline{x} - x^k, M^\dagger(Mx^k - p)\right\rangle \\
& = \left(1 + \frac{1}{1 - \delta_{2s}}\right) g(x^k) + \left\langle M^\dagger M(\overline{x} - x^k), M^\dagger(Mx^k - p)\right\rangle \\
& = \left(1 + \frac{1}{1 - \delta_{2s}}\right) g(x^k) - 2g(x^k) \\
& = \frac{\delta_{2s}}{1 - \delta_{2s}} g(x^k).
\end{aligned} \tag{43}$$

Together, (41)-(43) yield $q(\overline{x}, x^k) \leq \frac{\delta_{2s}}{1 - \delta_{2s}} g(x^k)$, which from (38)-(40) show that $g(x^{k+1}) \leq \frac{\delta_{2s}}{1 - \delta_{2s}} g(x^k)$. When $0 \leq \delta_{2s} < \frac{1}{2}$, as assumed, we have $0 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} < 1$. We then have that $d_B\left(x^k\right) \to 0$ as $k \to \infty$ at a linear rate for $0 \leq \delta_{2s} < \frac{1}{2}$, with constant bounded above by $\sqrt{\frac{\delta_{2s}}{1 - \delta_{2s}}} < 1$. Since the iterates x^k lie in A_s this proves convergence of the iterates to the intersection $A_s \cap B$, that is, to $\overline{x}$, as claimed. ∎

IV. SPARSE FEASIBILITY WITH AN AFFINE CONSTRAINT: LOCAL LINEAR CONVERGENCE OF DOUGLAS-RACHFORD

We turn our attention now to the Douglas-Rachford algorithm. First we present a result that could be discouraging since we show that the Douglas-Rachford operator has a set of fixed points that is too large in most interesting cases. However we show later that this set of fixed points has a nice structure guaranteeing local linear convergence of the iterates and thus convergence of the shadows to a solution of (6).

A. Fixed point sets of Douglas-Rachford

In contrast to the alternating projections algorithm, the iterates of the Douglas-Rachford algorithm are not actually the points of interest - it is rather the shadows of the iterates that are relevant. This results in an occasional incongruence between the fixed points of Douglas-Rachford and the intersection that we seek. Indeed, this mismatch occurs in the most interesting cases of the affine sparse feasibility problem as we show next.

Theorem IV.1 *Let A_s and B be defined by (5) and (3) and suppose there exists a point $\bar{x} \in A_s \cap B$ with $\|\bar{x}\|_0 = s$. If $s < \mathrm{rank}(M)$, then on all open neighborhoods $\mathcal{N}$ of $\bar{x} \in A_s \cap B$ there exist fixed points $z \in \mathrm{Fix}\, T_{DR}$ with $z \notin A_s \cap B$.*

Proof: Let $\bar{x} \in A_s \cap B$ with $\|\bar{x}\|_0 = s$ and set $\delta < \min\{|\bar{x}_j| \mid \bar{x}_j \neq 0\}$. By Lemma III.3 we have $A_s \cap B \cap I\!B_{\delta/2}(\bar{x}) = A_J \cap B \cap I\!B_{\delta/2}(\bar{x})$ for a unique $J := I(\bar{x}) \in \mathcal{J}_s$. Thus on the neighborhood $I\!B_{\delta/2}(\bar{x})$ the feasibility problems $\mathrm{Find}\, x \in A_J \cap B$, and $\mathrm{Find}\, x \in A_s \cap B$ have the same set of solutions. We consider the Douglas-Rachford operators applied to these two feasibility problems, for which we introduce the following notation: $T_J := \frac{1}{2}(R_{A_J} R_B + \mathrm{Id})$ and $T_s := \frac{1}{2}(R_{A_s} R_B + \mathrm{Id})$. Our proof strategy is to show first that the operators T_J and T_s restricted to $I\!B_{\delta/2}(\bar{x})$ are identical, hence their fixed point sets intersected with $I\!B_{\delta/2}(\bar{x})$ are identical. We then show that under the assumption $s < \mathrm{rank}\,(M)$ the set $\mathrm{Fix}\, T_J$ is strictly larger than the intersection $A_J \cap B$, hence completing the proof.

To show that the operators T_J and T_s applied to points $x \in I\!B_{\delta/2}(\bar{x})$ are identical, note that, by Lemma III.2(ii) and (iv), for all $x \in I\!B_{\delta/2}(\bar{x})$ we have $P_{A_s} x \subset I\!B_{\delta/2}(\bar{x})$ and $R_{A_s} x \subset I\!B_{\delta/2}(\bar{x})$. Moreover by Lemma III.3, since $\|\bar{x}\|_0 = s$ we have $A_s \cap I\!B_\delta(\bar{x}) = A_J \cap I\!B_\delta(\bar{x})$. Thus for all $x \in I\!B_{\delta/2}(\bar{x})$ we have $P_{A_s} x = P_{A_J} x \in I\!B_{\delta/2}(\bar{x})$ and $R_{A_s} x = R_{A_J} x \in I\!B_{\delta/2}(\bar{x})$. Also by Lemma III.2, $R_B x \in I\!B_{\delta/2}(\bar{x})$ for $x \in I\!B_{\delta/2}(\bar{x})$. Altogether, this yields

$$T_s x = \tfrac{1}{2}(R_{A_s} R_B + \mathrm{Id})\, x = \tfrac{1}{2}(R_{A_J} R_B + \mathrm{Id})\, x = T_J x \in I\!B_{\delta/2}(\bar{x}) \tag{44}$$

for all $x \in I\!B_{\delta/2}(\bar{x})$. Hence the operators T_s and T_J and their fixed point sets coincide on $I\!B_{\delta/2}(\bar{x})$.

We derive next an explicit characterization of $\mathrm{Fix}\, T_J$. By [5, Corollary 3.9] and (13) we have:

$$\begin{aligned}
\mathrm{Fix}\, T_J &= (A_J \cap B) + N_{A_J - B}(0) \\
&= (A_J \cap B) + (N_{A_J}(\bar{x}) \cap -N_B(\bar{x})) \\
&= (A_J \cap B) + \left(A_J^\perp \cap B^\perp\right).
\end{aligned} \tag{45}$$

The following equivalences show that $A_J^\perp \cap B^\perp$ is nontrivial if $s < \mathrm{rank}\,(M)$. Indeed,

$$\begin{aligned}
& & \mathrm{rank}(M) & > s \\
\Leftrightarrow & & \dim(\ker(M)^\perp) & > s \\
\Leftrightarrow & & n - s + \dim(\ker(M)^\perp) & > n \\
\Leftrightarrow & & \dim(A_J^\perp) + \dim(\ker(M)^\perp) & > n \\
\Leftrightarrow & & A_J^\perp \cap B^\perp & \neq \{0\}.
\end{aligned} \tag{46}$$

In other words, $\mathrm{Fix}\, T_J$ contains elements from the intersection $A_J \cap B$ and the nontrivial subspace $A_J^\perp \cap B^\perp$. This completes the proof. ∎

Remark IV.2 The inequality (46) shows that if $\mathrm{rank}(M) > s$ then the intersection $A_J \cap B$ is *not strongly regular*, or in other words, if $A_J \cap B$ is strongly regular then $\mathrm{rank}(M) \leq s$. This was also observed in [9, Remark 3.17] using tangent cones and *transversality*. The simple meaning of these results is that if the sparsity of a feasible point is less than the rank of the measurement matrix (the only interesting case in sparse signal recovery) then, since locally the affine feasibility problem is indistinguishable from simple linear feasibility at points $\bar{x} \in A_s$ with $\|\bar{x}\|_0 = s$, by Lemma II.6 the Douglas-Rachford algorithm may fail to converge *to the intersection* on all balls around a feasible point. As we noted in the beginning of this section, however, it is not the fixed points of Douglas-Rachford themselves but rather their shadows that are of interest. This leads to positive convergence results detailed in the next section.

B. Linear convergence of Douglas-Rachford

We begin with an auxiliary result that the Douglas-Rachford iteration applied to *linear subspaces* converges to its set of fixed points with linear rate. As the sparse feasibility problem reduces locally to finding the intersection of (affine) subspaces, by a translation to the origin, results for the case of subspaces will yield local linear convergence of Douglas-Rachford to fixed points associated with points $\bar{x} \in A_s \cap B$ such that $\|\bar{x}\|_0 = s$. Convergence of Douglas-Rachford for convex sets with nonempty intersection was proved first by Lions and Mercier [30], but without rate. (They do, however, achieve linear rates of convergence under strong assumptions that are not satisfied for convex

feasibility.) As surprising as it may seem, results on the *rate* of convergence of this algorithm even for the simple case of affine subspaces are very recent. Our proof, based on [21], is one of several independent results (with very different proofs) that we are aware of which have appeared in the last several months [1], [19]. Of these, [21] is the only work that includes nonconvex settings.

1) The linear case: The idea of our proof is to show that the set of fixed points of the Douglas-Rachford algorithm applied to the subspaces A and B can always be written as the intersection of different subspaces $\widetilde{A}$ and $\widetilde{B}$, the collection of which is *strongly regular*. We then show that the iterates of the Douglas-Rachford algorithm applied to the subspaces A and B are identical to those of the Douglas-Rachford algorithm applied to the subspaces $\widetilde{A}$ and $\widetilde{B}$. Linear convergence of Douglas-Rachford then follows directly from Lemma II.6.

We recall that the set of fixed points of Douglas-Rachford in the case of two linear subspaces $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ is by [5, Corollary 3.9] and (45) equal to

$$\operatorname{Fix} T_{DR} = (A \cap B) + \left(A^\perp \cap B^\perp \right)$$

for $T_{DR} := \frac{1}{2} \left(R_A R_B + \operatorname{Id} \right)$. For two linear subspaces $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ define the enlargements $\widetilde{A} := A + \left(A^\perp \cap B^\perp \right)$ and $\widetilde{B} := B + \left(A^\perp \cap B^\perp \right)$. By definition of the Minkowski sum these enlargements are given by

$$\widetilde{A} = \left\{ a + n \mid a \in A, n \in A^\perp \cap B^\perp \right\} \tag{47a}$$

and

$$\widetilde{B} = \left\{ b + n \mid b \in B, n \in A^\perp \cap B^\perp \right\}. \tag{47b}$$

The enlargements $\widetilde{A}$ and $\widetilde{B}$ are themselves subspaces of $\mathbb{R}^n$ as the Minkowski sum of subspaces.

Lemma IV.3 *The equation*

$$C := \left(A + \left(A^\perp \cap B^\perp \right) \right)^\perp \cap \left(B + \left(A^\perp \cap B^\perp \right) \right)^\perp = \{0\}$$

holds for any linear subspaces A and B of $\mathbb{R}^n$, and hence the collection $(\widetilde{A}, \widetilde{B})$ is strongly regular for any linear subspaces A and B.

Proof: Let v be an element of C. Because $C = \widetilde{A}^\perp \cap \widetilde{B}^\perp$, we know that

$$\langle v, \widetilde{a} \rangle = \left\langle v, \widetilde{b} \right\rangle = 0 \quad \text{for all } \widetilde{a} \in \widetilde{A}, \widetilde{b} \in \widetilde{B}. \tag{48}$$

Further, since $A \subset \widetilde{A}$ and $B \subset \widetilde{B}$ we have

$$\langle v, a \rangle = \langle v, b \rangle = 0 \quad \text{for all } a \in A, b \in B. \tag{49}$$

In other words, $v \in A^\perp$ and $v \in B^\perp$, so $v \in A^\perp \cap B^\perp$. On the other hand, $A^\perp \cap B^\perp \subset \widetilde{A}$ and $A^\perp \cap B^\perp \subset \widetilde{B}$, so we similarly have

$$\langle v, n \rangle = 0 \quad \text{for all } n \in A^\perp \cap B^\perp, \tag{50}$$

because A and B are subspaces and $v \in C$. Hence v is also an element of $\left(A^\perp \cap B^\perp \right)^\perp$. We conclude that v can only be zero. ∎

Lemma IV.4 *Let A and B be linear subspaces and let $\widetilde{A}$ and $\widetilde{B}$ be their corresponding enlargements defined by (47).*

(i) $R_A d = -d$ *for all* $d \in A^\perp$.
(ii) $R_A x = R_{\widetilde{A}} x$ *for all* $x \in A + B$.
(iii) $R_{\widetilde{B}} a \in A + B$ *for all* $a \in A$.
(iv) $R_{\widetilde{A}} R_{\widetilde{B}} x = R_A R_B x$ *for all* $x \in \mathbb{R}^n$.
(v) *For any $x \in \mathbb{R}^n$ the following equality holds:*

$$\frac{1}{2} \left(R_{\widetilde{A}} R_{\widetilde{B}} + \operatorname{Id} \right) x = \frac{1}{2} \left(R_A R_B + \operatorname{Id} \right) x.$$

Proof: To prove (i), let $d \in A^\perp$ be arbitrary. The projection $P_A d$ of d onto A is the orthogonal projection onto A. The orthogonal projection of $d \in A^\perp$ is the zero vector. This means that $R_A d = (2P_A - \text{Id})d = -d$.

To show (ii)[1] note that $(A^\perp \cap B^\perp) = (A + B)^\perp$ hence $\widetilde{A} = A + (A + B)^\perp$. Now by [6, Proposition 2.6], $P_{A+(A+B)^\perp} = P_A + P_{(A+B)^\perp}$. Hence for all $x \in A + B$, $P_{\widetilde{A}} x = P_A x$ and, consequently, $R_{\widetilde{A}} x = R_A x$, as claimed.

To prove (iii), let $a \in A$ and thus $a \in A + B$. We note that by (ii) with A replaced by B we have $R_B a = R_{\widetilde{B}} a$. Write a as a sum $b + v$ where $b = P_B a$ and $v = a - P_B a$. We note that $v \in A + B$ and so $-v \in A + B$. From (i) we conclude, since A in (i) can be replaced by B and $v \in B^\perp$, that $R_B v = -v$. Since $b \in B$, we have $R_B b = 2P_B b - b = b$ and so

$$R_{\widetilde{B}} a = R_B a = R_B b + R_B v = b - v \in A + B. \tag{51}$$

To see (iv) let $x \in \mathbb{R}^n$ be arbitrary. Define $D := A^\perp \cap B^\perp$. Then we can write as $x = a + b + d$ with $a \in A$, $b \in B$ and $d \in D$. This expression does not have to be unique since A and B may have a nontrivial intersection. In any case, we have the identity $\langle b, d \rangle = \langle a, d \rangle = 0$. Since A and B are linear subspaces, the Douglas-Rachford operator is a linear mapping which, together with parts (i)-(iii) of this lemma, yields

$$\begin{aligned}
R_A R_B x &= R_A (R_B a + R_B b + R_B d) \\
&= R_A (R_B a + b - d) \\
&= R_A R_B a + R_A b + R_A(-d) \\
&= R_A R_B a + R_A b + d \\
&= R_A R_{\widetilde{B}} a + R_{\widetilde{A}} b + d \\
&= R_{\widetilde{A}} R_{\widetilde{B}} a + R_{\widetilde{A}} b + d
\end{aligned} \tag{52}$$

But since $d \in A^\perp \cap B^\perp$, we have $d \in \widetilde{A} \cap \widetilde{B}$, and moreover, $b \in B \subset \widetilde{B}$, hence by (52) we have

$$\begin{aligned}
R_A R_B x &= R_{\widetilde{A}} (R_{\widetilde{B}} a + b + d) \\
&= R_{\widetilde{A}} (R_{\widetilde{B}} a + R_{\widetilde{B}} b + R_{\widetilde{B}} d) \\
&= R_{\widetilde{A}} R_{\widetilde{B}} x.
\end{aligned} \tag{53}$$

This proves (iv).

Statement (v) is an immediate consequence of (iv), which completes the proof. ∎

Proposition IV.5 *Let A and B be linear subspaces and let $\widetilde{A}$ and $\widetilde{B}$ be their corresponding enlargements defined by (47). The Douglas-Rachford iteration applied to the enlargements*

$$x^{k+1} = \widetilde{T}_{DR} x^k := \frac{1}{2} \left(R_{\widetilde{A}} R_{\widetilde{B}} + Id \right) x^k \tag{54}$$

converges with linear rate to Fix $\widetilde{T}_{DR}$ *for any starting point $x^0 \in \mathbb{R}^n$.*

Proof: By Lemma IV.3 we know that the only common element in $\left(A + \left(A^\perp \cap B^\perp\right)\right)^\perp$ and $\left(B + \left(A^\perp \cap B^\perp\right)\right)^\perp$ is the zero vector. By Lemma II.6 [21, Corollary 3.20] the sequence

$$\widetilde{x}_{k+1} := \frac{1}{2} \left(R_{\widetilde{A}} R_{\widetilde{B}} + \text{Id} \right) \widetilde{x}_k$$

converges linearly to the intersection $\widetilde{A} \cap \widetilde{B}$ for any starting point $\widetilde{x}_0 \in \mathbb{R}^n$. ∎

Combining these results we obtain the following theorem confirming *linear* convergence of the Douglas-Rachford algorithm for subspaces. Convergence of the Douglas-Rachford algorithm for *strongly regular* affine subspaces was proved in [21, Corollary 3.20] as a special case of a more general result [21, Theorem 3.18] about linear convergence of the Douglas-Rachford algorithm for a strongly regular collection of a *super-regular set* [29, Definition 4.3] and an affine subspace. Our result below shows that the iterates of the Douglas-Rachford algorithm for *linearly regular* affine subspaces (not necessarily strongly regular) converge linearly to the fixed point set.

---

[1] This proof is a simplification of our original proof suggested by an anonymous referee.

Theorem IV.6 *For any two affine subspaces* $A, B \subset I\!\!R^n$ *with nonempty intersection, the Douglas-Rachford iteration*

$$x^{k+1} = T_{DR}x^k := \frac{1}{2}\left(R_A R_B + \mathit{Id}\right)x^k \tag{55}$$

converges for any starting point x^0 *to a point in the fixed point set with linear rate. Moreover,* $P_B\bar{x} \in A \cap B$ *for* $\bar{x} = \lim_{k\to\infty} x^k$.

Proof: Without loss of generality, by translation of the sets A and B by $-\bar{x}$ for $\bar{x} \in A \cap B$, we consider the case of subspaces. By Proposition IV.5 Douglas-Rachford applied to the enlargements $\widetilde{A} = A + \left(A^\perp \cap B^\perp\right)$ and $\widetilde{B} = B + \left(A^\perp \cap B^\perp\right)$, namely (54), converges to the intersection $\widetilde{A} \cap \widetilde{B}$ with linear rate for any starting point $x^0 \in I\!\!R^n$. By [5, Corollary 3.9] and (13), the set of fixed points of the Douglas-Rachford algorithm (55) is

$$\text{Fix}_{T_{DR}} = (A \cap B) + \left(A^\perp \cap B^\perp\right) = \widetilde{A} \cap \widetilde{B}, \tag{56}$$

where the rightmost equality follows from repeated application of the identity $(\Omega_1 + \Omega_2)^\perp = (\Omega_1^\perp \cap \Omega_2^\perp)$, the definition of set addition and closedness of subspaces under addition. By Lemma IV.4(v) the iterates of (54) are the same as the iterates of (55). So the iterates of the Douglas-Rachford algorithm applied to A and B converge to a point in the set of its fixed points with linear rate. Finally, by [5, Corollary 3.9], $P_B\bar{x} \in A \cap B$ for any $\bar{x} \in$ Fix T_{DR}. ∎

2) Douglas-Rachford applied to sparse affine feasibility: We conclude with an application of the analysis for affine subspaces to the case of affine feasibility with a sparsity constraint.

Theorem IV.7 *Let* A_s *and* B *be defined by (5) and (3) with nonempty intersection and let* $\bar{x} \in A_s \cap B$ *with* $\|\bar{x}\|_0 = s$. *Choose* $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. *For* $x^0 \in I\!\!B_{\delta/2}(\bar{x})$ *the corresponding Douglas-Rachford iterates converge with linear rate to Fix* T_{DR}. *Moreover, for any* $\hat{x} \in$ *Fix* $T_{DR} \cap I\!\!B_{\delta/2}(\bar{x})$, *we have* $P_B\hat{x} \in A_s \cap B$.

Proof: By Lemma III.3 we have $A_s \cap B \cap I\!\!B_\delta(\bar{x}) = A_J \cap B \cap I\!\!B_\delta(\bar{x})$ for a unique $J \in \mathcal{J}_s$. Thus by (44) at all points in $I\!\!B_{\delta/2}(\bar{x})$ the Douglas-Rachford operator corresponding to A_s and B is equivalent to the Douglas-Rachford operator corresponding to A_J and B, whose intersection includes $\bar{x}$. Applying Theorem IV.6, shifting the subspaces appropriately, we see that the iterates converge to some point $\hat{x} \in$ Fix T_{DR} with linear rate for all initial points $x^0 \in I\!\!B_{\delta/2}(\bar{x})$. The last statement follows from (44) and Theorem IV.6. ∎

V. EXAMPLES

A. Numerical Demonstration

We demonstrate the above results on the following synthetic numerical example. We construct a sparse object with 328 uniform random positive and negative point-like sources in a 256-by-256 pixel field and randomly sample the Fourier transform of this object at a ratio of 1-to-8. This yields 8192 affine constraints. Local convergence results are illustrated in Figure 1 where the initial points x^0 are selected by uniform random $(-\delta/512, \delta/512)$ perturbations of the true solution in order to satisfy the assumptions of Theorems III.8 and IV.7. The alternating projections and Douglas-Rachford algorithms are shown respectively in panels (a)-(b) and (c)-(d) of Figure 1. We show both the step lengths per iteration as well as the *gap distance* at each iteration defined as

$$(\text{gap distance})^k := \|P_{A_s}x^k - P_Bx^k\|. \tag{57}$$

Monitoring the gap allows one to ensure that the algorithm is indeed converging to a point of intersection instead of just a best approximation point. In panels (a) and (c) we set the sparsity parameter $s = 328$, exactly the number of nonzero elements in the original image. Panels (b) and (d) demonstrate the effect of overestimating the sparsity parameter, $s = 350$, on algorithm performance. The convergence of Douglas-Rachford for the case $s = 350$ is not covered in our theory, however our numerical experiments indicate that one still achieves a linear-looking convergence over cycles, albeit with a very poor rate constant. This remains to be proven.
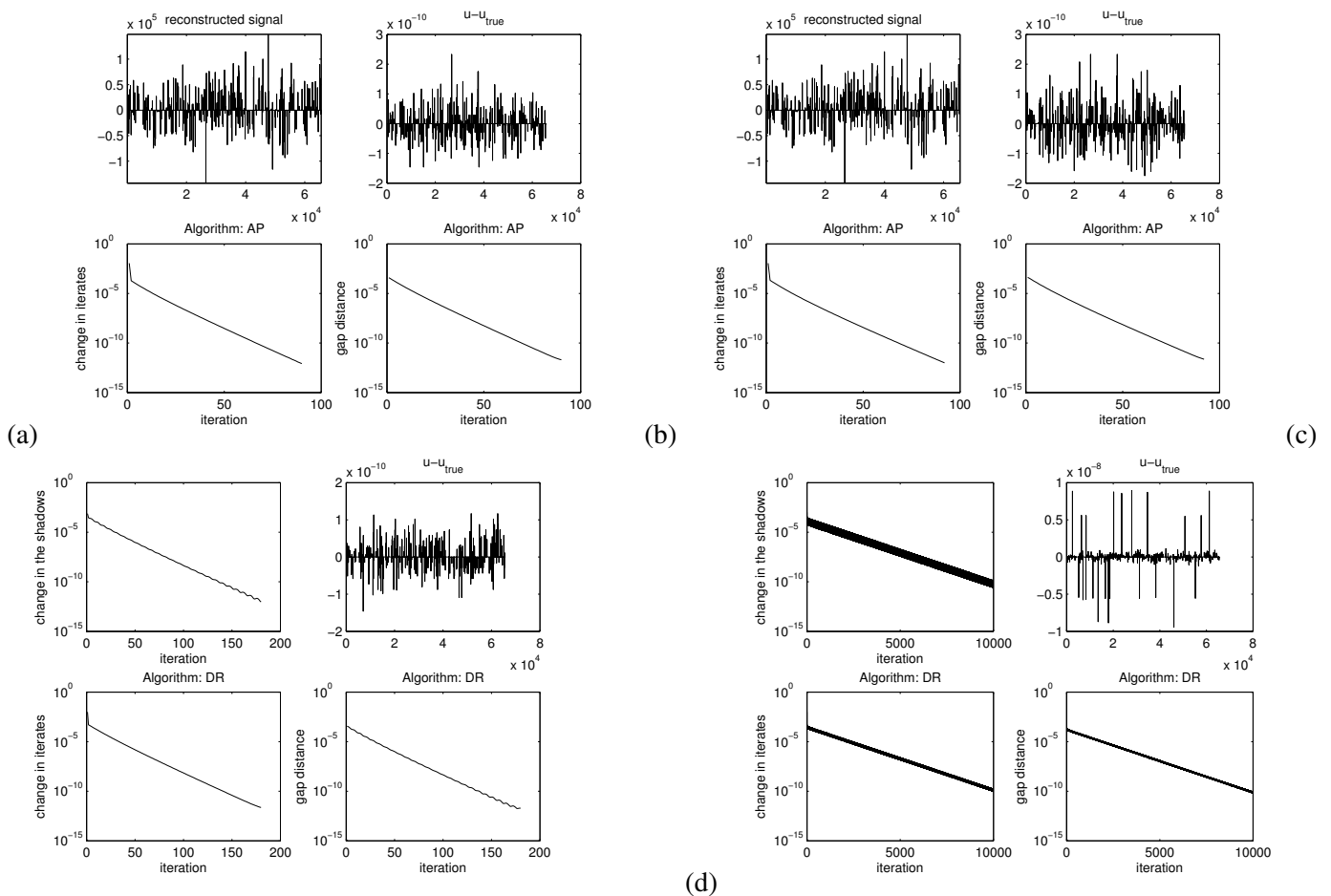
Fig. 1. Case (a) shows the convergence of alternating projections in the case where the sparsity is exact, $s = 328$. Case (b) shows the same with sparsity assumed too big, $s = 350$. In case (c) and (d) we have the corresponding plots for Douglas-Rachford. Case (d) is not covered by our theory.

The second synthetic example, shown in Figure 2, demonstrates global performance of the algorithms and illustrates the results in Theorem III.8, Theorem IV.7 and Corollary III.13. The solution is the vector $\bar{x} := (10, 0, 0, 0, 0, 0, 0, 0)^{\top}$ and the affine subspace is the one generated by the matrix in (58). This matrix fulfills the assumptions of Corollary III.13, as shown in Section V-B1. For the cases (a) and (c) the initial point x^0 can be written as $x^0 := \bar{x} + u$ where u is a vector with uniform random values from the interval $(-1, 1)$. The initial values hence fulfill the assumptions of Theorems III.8 and IV.7. For (b) and (d) again the initial point x^0 can be written as $x^0 := \bar{x} + u$ while u is now a vector with uniform random values from the interval $(-100, 100)$. As expected, the sequence of alternating projections converges to the true solution in (c). The case for Douglas-Rachford however, shown in (d), is not covered by our theory.
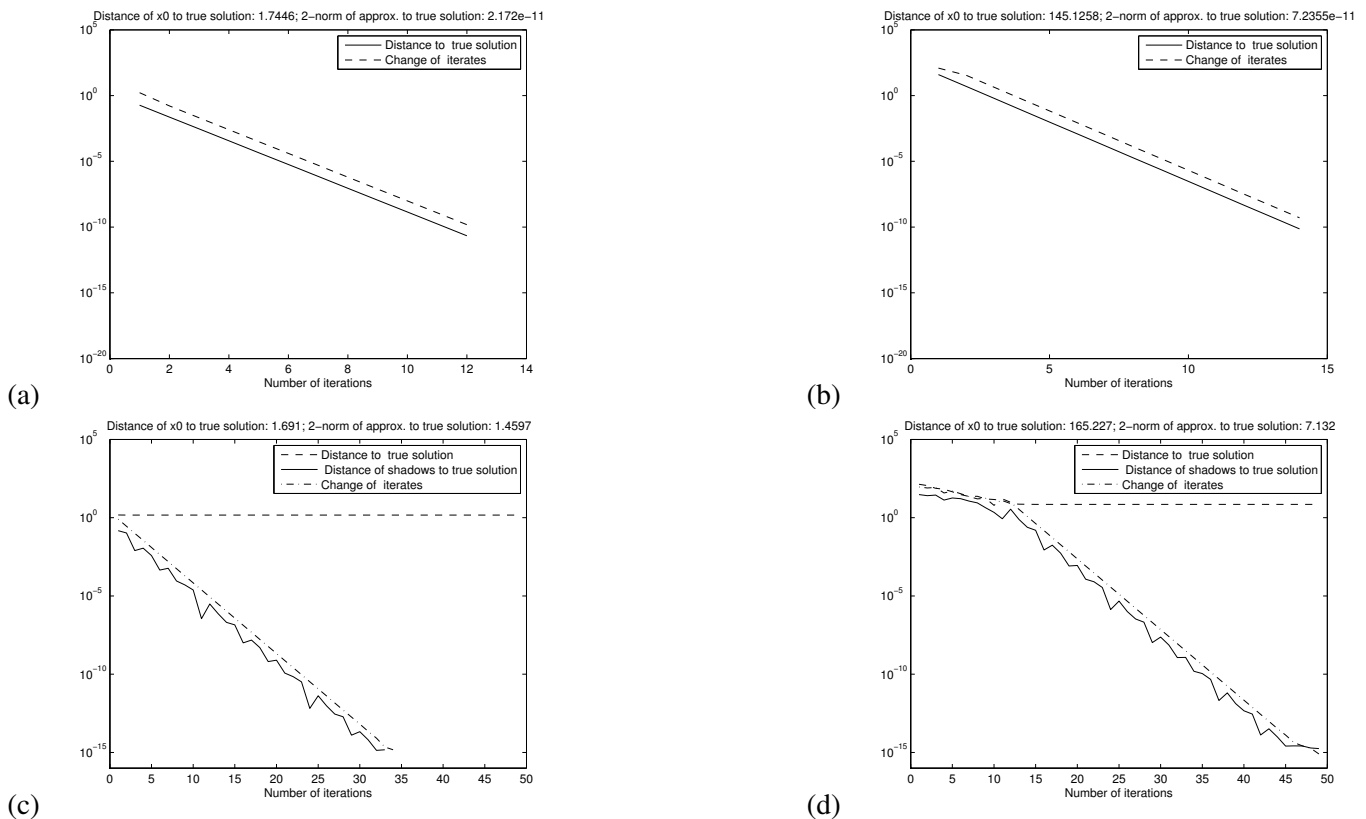
Fig. 2. Example with an affine subspace generated by the matrix from Section V-B1: (a) shows the local convergence as shown in Theorem III.8, (b) is an example of global convergence of alternating projections as stated in Corollary III.13. (c) is an example of local convergence of Douglas-Rachford to its fixed point set while the shadows converge to the intersection, as proven in Theorem IV.7. This example also shows that the iterates converge to a fixed point that is not in the intersection, as proven in Theorem IV.1. Plot (d) is an example where Douglas-Rachford appears to converge globally. This behavior is not covered by our theory.

B. Analytic examples

1) Example of a matrix satisfying assumptions of Corollary III.13: Finding nonsquare matrices satisfying (27) or deciding whether or not a matrix fulfills some similar condition is, in general, hard to do – but not impossible. In this section we provide a concrete example of a nonsquare matrix satisfying the assumptions of Corollary III.13.

We take the matrix

$$M = \frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \end{pmatrix}. \tag{58}$$

The rows of M are pairwise orthogonal and so $MM^\top = \mathrm{Id}_7$. We compute the constant δ in (27) for $s = 2$ to get a result for the recovery of 1-sparse vectors with alternating projections. Recall that s can be larger than the sparsest feasible solution (see Remark III.9). In general, a normalized 2-sparse vector in $\mathbb{R}^8$ has the form

$$x = (\cos(\alpha), \sin(\alpha), 0, 0, 0, 0, 0, 0)^\top,$$

where the position of the $\sin$ and of the $\cos$ are of course arbitrary. The squared norm of the product Mx is equal to

$$\|Mx\|_2^2 = \frac{1}{8} \sum_{i=1}^{7} |\cos(\alpha) + z_i \sin(\alpha)|^2,$$

where $z_i \in \{-1, 1\}$. We note that the inner products of distinct columns of M are $-1, 1$, so $\sum_{i=1}^{7} z_i = \pm 1$. Then

$$\frac{1}{8} \sum_{i=1}^{7} |\cos(\alpha) + z_i \sin(\alpha)|^2$$

$$= \frac{1}{8} \sum_{i=1}^{7} \cos(\alpha)^2 + 2z_i \sin(\alpha) \cos(\alpha) + \sin(\alpha)^2$$

$$= \frac{1}{8} (7 \pm \sin(2\alpha)) \in \left[\frac{3}{4}, 1 \right].$$

This means that $\frac{3}{4} \|x\|_2^2 \leq \|Mx\|_2^2 \leq \|x\|_2^2 \quad \forall x \in A_2$, where A_2 is the set of 2-sparse vectors in $\mathbb{R}^8$. In other words, we can recover any 1-sparse vector with the method of alternating projections.

2) An easy example where alternating projections and Douglas-Rachford iterates don't converge: The following example, discovered with help from Matlab's Symbolic Toolbox, shows some of the more interesting pathologies that one can see with these algorithms when not starting sufficiently close to a solution.

Let $n = 3, m = 2, s = 1$ and

$$M = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -1 \end{pmatrix}, \qquad p = \begin{pmatrix} -5 \\ 5 \end{pmatrix}.$$

The point $(0, 10, 0)^\top$ is the sparsest solution to the equation $Mx = p$ and the affine space B is

$$B = \begin{pmatrix} 0 \\ 10 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \qquad \text{with } \lambda \in \mathbb{R}.$$

From the initial point $(0, -2.5, -5)^\top$, the iterates of the Douglas-Rachford algorithm converge to the cycle $T_{DR}\bar{x} = \bar{x} + (-5, 0, 5)^\top$ and $T_{DR}^2 \bar{x} = \bar{x}$, where

$$\bar{x} = \begin{pmatrix} \frac{38894857328700073}{237684487542793012780631851008} \\[8pt] \frac{-29710560942850721475845458 0565}{1188422437713965506390315925504} \\[8pt] \frac{-1188422437713940163629828887893}{237684487542793012780631851008} \end{pmatrix} \approx \begin{pmatrix} 0 \\ -2.5 \\ -5 \end{pmatrix}. \tag{59}$$

Note that this example is different from the case in Theorem IV.1: in Theorem IV.1 we establish that, if $s < \mathrm{rank}(M)$, then the fix point set of T_{DR} is strictly larger than the solution set to problem (6). The concrete case detailed here also satisfies $s < \mathrm{rank}(M)$, however, with the given x^0 we are not near the set of fixed points, but in a *cycle* of T_{DR}.

If, on the other hand, we take the point $\hat{x}^0 = (-4, 0, 0)^\top$, then $P_B \hat{x}^0 = (-4, 2, -4)^\top$ and the set $P_{A_1} P_B \hat{x}^0$ is equal to $\left\{ (-4, 0, 0)^\top, (0, 0, -4)^\top \right\}$. The projection $P_B (0, 0, -4)^\top$ is again the point $(-4, 2, -4)^\top$. This shows that the alternating projection (8) iteration is stuck at the points $(-4, 0, 0)^\top$ and $(0, 0, -4)^\top$ which are clearly not in the intersection $A_1 \cap B = \left\{ (0, 10, 0)^\top \right\}$. This also highlights a manifestation of the multivaluedness of the projector P_{A_1}.

VI. CONCLUSION

The usual avoidance of nonconvex optimization over convex relaxations is not always warranted. In this work we have determined sufficient conditions under which simple algorithms applied to nonconvex sparse affine feasibility are guaranteed to converge globally at a linear rate. We have also shown local convergence of the prominent Douglas-Rachford algorithm applied to this problem. These results are intended to demonstrate the potential of recently developed analytical tools for understanding nonconvex fixed-point algorithms in addition to making the broader point that nonconvexity is not *categorically* bad. That said, the global results reported here rely heavily on the linear structure of the problem, and local results are of limited practical use. Of course, the decision about whether to favor a convex relaxation over a nonconvex formulation depends on the structure of the problem at hand and many open questions remain. First and foremost among these is: what are *necessary* conditions for global

convergence of simple algorithms to global solutions of nonconvex problems? The apparent robust global behavior of Douglas-Rachford has eluded explanation. What are conditions for global convergence of the Douglas-Rachford algorithm for this problem? What happens to these algorithms when the chosen sparsity parameter s is too small? At the heart of these questions lies a long-term research program into regularity of nonconvex variational problems, the potential impact of which is as broad as it is deep.

ACKNOWLEDGEMENTS

REFERENCES

[1] H. H. Bauschke, J. Y. Bello Cruz, T. T. A. Nghia, H. M. Phan, and X. Wang. The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle. *Journal of Approximation Theory*, to appear. DOI 10.1016/j.jat.2014.06.002.

[2] H. H. Bauschke and J. M. Borwein. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1(2):185–212, 1993.

[3] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[4] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[5] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Finding Best Approximation Pairs Relative to Two Closed Convex Sets in Hilbert Spaces. *Journal of Approximation Theory*, 127:178–314, 2004.

[6] H. H. Bauschke, P. L. Combettes, and D. R. Luke. A strongly convergent reflection method for finding the projection onto the intersection of two closed convex sets in a Hilbert space. *Journal of Approximation Theory*, 141(1):63–69, 2006.

[7] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted Normal Cones and the Method of Alternating Projections: Applications. *Set-Valued and Variational Analysis*, 21:475–501, 2013.

[8] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted Normal Cones and the Method of Alternating Projections: Theory. *Set-Valued and Variational Analysis*, 21:431–473, 2013.

[9] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and sparsity optimization with affine constraints. *Foundations of Computational Mathematics*, 14(1):63–83, 2014.

[10] H. H. Bauschke and D. Noll. On the local convergence of the Douglas-Rachford algorithm. http://arxiv.org/abs/1401.6188v1, 2014.

[11] A. Beck and M. Teboulle. A Linearly Convergent Algorithm for Solving a Class of Nonconvex/Affine Feasibility Problems. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 33–48. Springer New York, 2011.

[12] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[13] T. Blumensath and M. Davies. Normalised iterative hard thresholding; guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309, 2010.

[14] J. M. Borwein and D. R. Luke. Entropic regularization of the ℓ_0 function. In H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 65–92. Springer, 2011.

[15] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2011.

[16] E. J. Candès and T. Tao. Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[17] E. J. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14:877–905, 2007.

[18] P. L. 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*, 67:487–507, 1990.

[19] L. Demanet and X. Zhang. Eventual linear convergence of the Douglas Rachford iteration for basis pursuit. http://arxiv.org/abs/1301.0542v3.

[20] D. Gabay. Applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Solution of Boundary- Value Problems*, pages 299–340. North-Holland, 1983.

[21] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization*, 23-4:2397–2419, 2013.

[22] J.-B. Hiriart-Urruty. When only global optimization matters. *Journal of Global Optimization*, 56(3):761–763, 2013.

[23] A. D. Ioffe. Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501, 2000.

[24] B. Jin and D. A. Lorenz. Heuristic Parameter-Choice Rules for Convex Variational Regularization Based on Error Estimates. *SIAM Jounal on Numerical Analysis*, 48(3):1208–1229, 2010.

[25] A. Y. Kruger. About Regularity of Collections of Sets. *Set-Valued Analysis*, 14:187–206, 2006.

[26] A. Kyrillidis and V. Cevher. Matrix recipes for hard thresholding methods. *Journal of Mathematical Imaging and Vision*, 28(2):235–265, 2014.

[27] M.-J. Lai and J. Wang. An Unconstrained ℓ_q Minimization with $0 < q \leq 1$ for Sparse Solution of Underdetermined Linear Systems. *SIAM Journal on Optimization*, 21(1):82–101, 2011.

[28] H. Le. Generalized subdifferentials of the rank function. *Optimization Letters*, 7:1–13, 2012.

[29] A. S. Lewis, D. R. Luke, and J. Malick. Local Linear Convergence for Alternating and Averaged Nonconvex Projections. *Foundations of Computational Mathematics*, 9(4):485–513, 2009.

[30] P. L. Lions and B. Mercier. Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[31] D. Luke. Prox-Regularity of Rank Constraint Sets and Implications for Algorithms. *Journal of Mathematical Imaging and Vision*, 47(3):231–238, 2013.

[32] D. R. Luke. Finding best approximation pairs relative to a convex and a prox-regular set in a Hilbert space. *SIAM Journal on Optimization*, 19(2):714–739, 2008.

[33] B. Mordukhovich. *Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 2006.

[34] H. V. Ngai and M. Théra. Metric Inequality, Subdifferential Calculus and Applications. *Set-Valued Analysis*, 9:187–216, 2001.

[35] H. M. Phan. Linear convergence of the Douglas-Rachford method for two closed sets. http://arxiv.org/abs/1401.6509v1, 2014.

[36] G. Pierra. Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique. *Lecture Notes in Computer Science*, 41:200–218, 1976.

[37] G. Pierra. Decomposition through formalization in a product space. *Mathematical Programming*, 28:96–115, 1984.

[38] R. T. Rockafellar and R. J. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, 1998.

[39] A. Tillmann and M. E. Pfetsch. The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing. *IEEE Transactions on Information Theory*, 60(2):12481259, 2014.