Restricted normal cones and sparsity optimization with affine constraints

Heinz H. Bauschke,[*] D. Russell Luke,[†] Hung M. Phan,[‡] and Xianfu Wang[§]

May 2, 2012

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 paper we consider the elementary method of alternating projections (MAP) for solving the sparsity optimization problem without employing convex heuristics. In a parallel paper we recently introduced the restricted normal cone which generalizes the classical Mordukhovich normal cone and reconciles some fundamental gaps in the theory of sufficient conditions for local linear convergence of the MAP algorithm. We use the restricted normal cone together with the notion of superregularity, which is naturally satisfied for the affine sparse optimization problem, to obtain local linear convergence results with estimates for the radius of convergence of the MAP algorithm applied to sparsity optimization with an affine constraint.

2010 Mathematics Subject Classification: Primary 49J52, 49M20, 90C26; Secondary 15A29, 47H09, 65K05, 65K10, 94A08.

Keywords: Compressed sensing, constraint qualification, Friedrichs angle, linear convergence, method of alternating projections, normal cone, projection operator, restricted normal cone, sparsity optimization, superregularity, underdetermined.

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: `heinz.bauschke@ubc.ca`.

[†]Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestr. 16–18, 37083 Göttingen, Germany. E-mail: `r.luke@math.uni-goettingen.de`.

[‡]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: `hung.phan@ubc.ca`.

[§]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: `shawn.wang@ubc.ca`.

1 Introduction

We consider the problem of sparsity optimization with affine constraints:

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

where m and n are integers such that $1 \le m < n$, M is a real m-by-n matrix, denoted $M \in \mathbb{R}^{m\times n}$, and $\|x\|_0 := \sum_{j=1}^{n} |\operatorname{sgn}(x_j)|$ counts[1] the number of nonzero entries of real vectors x of length n, denoted by $x \in \mathbb{R}^n$.

If there is some a priori bound on the desired sparsity of the solution, represented by an integer s, where $1 \le s \le n$, then one can relax (1) to the feasibility problem

$$\text{find } c \in A \cap B, \tag{2}$$

where

$$A := \left\{x \in \mathbb{R}^n \;\middle|\; \|x\|_0 \le s\right\} \quad \text{and} \quad B := \left\{x \in \mathbb{R}^n \;\middle|\; Mx = p\right\}. \tag{3}$$

The sparsity subspace associated with $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is

$$\operatorname{supp}(a) := \left\{x \in \mathbb{R}^n \;\middle|\; x_j = 0 \text{ whenever } a_j = 0\right\}. \tag{4}$$

Also, we define

$$I\colon \mathbb{R}^n \to \{1, \ldots, n\}\colon x \mapsto \left\{i \in \{1, \ldots, n\} \;\middle|\; x_i \neq 0\right\}, \tag{5}$$

and we denote the i^{th} standard unit vector by e_i for every $i \in \{1, \ldots, n\}$.

Problem (1) is in general NP-complete [21] and so convex and nonconvex relaxations are typically employed for its solution. For a primal-dual convex strategy see [6]; for relaxations to ℓ_p $(0 < p < 1)$ see [15]; see [8] for a comprehensive review and applications. In this paper we apply recent tools developed by the authors in [3] to prove local linear convergence of an elementary algorithm applied to the feasibility formulation of the problem (2), that is, we do not use convex heuristics or conventional smooth relaxations. The key to our results is a new normal cone called the *restricted normal cone*. A central feature of our approach is the decomposition of the original nonconvex set into collections of simpler (indeed, linear) sets which can be treated separately. Ours is not the first result on local linear convergence for sparsity optimization with affine constraints. Indeed the problem was considered more than twenty years ago by Combettes and Trussell who show local convergence of alternating projections [11]. The problem was recently used to illustrate the application of analytical tools developed in [17] and [18]. Other approaches that also yield convergence results for different algorithms can be found in [1] and [5], with the latter of these being notable in that they obtain *global* convergence results with additional assumptions (restricted isometry) that we do not consider here. The novelty of the results we report here, based principally on the works [17], [16] and [3], is that we obtain not only convergence rates but

[1] We set $\operatorname{sgn}(0) := 0$.

also radii of convergence when all conventional sufficient conditions for local linear convergence, notably those of [17] and [16], fail. In this sense, our criteria for convergence are more robust and yield richer information than other available notions.

The remainder of the paper is organized as follows. In Section 2, we define the restricted normal cones and corresponding constraint qualifications for sets and collections of sets first introduced in [3] as well as the notion of superregularity introduced in [16] adapted to the restricted normal cones. A few of the many properties of these objects developed in [3] are restated in preparation for Section 3 where we apply these tools to a convergence analysis of the method of alternating projections (MAP) for the problem of finding a vector $c \in \mathbb{R}^n$ satisfying an affine constraint and having sparsity no greater than some a priori bound, that is, we solve (2) for A and B defined by (3). Given a starting point $b_{-1} \in X$, MAP sequences $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ are generated as follows:

$$(6) \qquad (\forall k \in \mathbb{N}) \qquad a_k := P_A b_{k-1}, \quad b_k := P_B a_k.$$

We do not attempt to review the history of the MAP, its many extensions, and its rich and convergence theory; the interested reader is referred to, e.g., [2], [10], [12], and the references therein. We consider the MAP iteration to be a prototype for more sophisticated approaches, both of projection type or more generally subgradient algorithms, hence our focus on this simple algorithm.

Notation

Our notation is standard and follows largely [2],[7], [20], [22], and [23] to which the reader is referred for more background on variational analysis. Throughout this paper, we assume that $X = \mathbb{R}^n$ with inner product $\langle\cdot,\cdot\rangle$, induced norm $\|\cdot\|$, and induced metric d. The real numbers are $\mathbb{R}$, the integers are $\mathbb{Z}$, and $\mathbb{N} := \{z \in \mathbb{Z} \mid z \geq 0\}$. Further, $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x \geq 0\}$, $\mathbb{R}_{++} := \{x \in \mathbb{R} \mid x > 0\}$. Let R and S be subsets of X. Then the closure of S is $\overline{S}$, the interior of S is $\operatorname{int}(S)$, the boundary of S is $\operatorname{bdry}(S)$, and the smallest affine and linear subspaces containing S are $\operatorname{aff} S$ and $\operatorname{span} S$, respectively. If Y is an affine subspace of X, then $\operatorname{par} Y$ is the unique linear subspace parallel to Y. The negative polar cone of S is $S^{\ominus} = \{u \in X \mid \sup\langle u, S\rangle \leq 0\}$. We also set $S^{\oplus} := -S^{\ominus}$ and $S^{\perp} := S^{\oplus} \cap S^{\ominus}$. We also write $R \oplus S$ for $R + S := \{r + s \mid (r,s) \in R \times S\}$ provided that $R \perp S$, i.e., $(\forall (r,s) \in R \times S)\ \langle r, s\rangle = 0$. We write $F\colon X \rightrightarrows X$, if F is a mapping from X to its power set, i.e., $\operatorname{gr} F$, the graph of F, lies in $X \times X$. Abusing notation slightly, we will write $F(x) = y$ if $F(x) = \{y\}$. A nonempty subset K of X is a cone if $(\forall \lambda \in \mathbb{R}_+)\ \lambda K := \{\lambda k \mid k \in K\} \subseteq K$. The smallest cone containing S is denoted $\operatorname{cone}(S)$; thus, $\operatorname{cone}(S) := \mathbb{R}_+ \cdot S := \{\rho s \mid \rho \in \mathbb{R}_+, s \in S\}$ if $S \neq \varnothing$ and $\operatorname{cone}(\varnothing) := \{0\}$. If $z \in X$ and $\rho \in \mathbb{R}_{++}$, then $\operatorname{ball}(z;\rho) := \{x \in X \mid d(z,x) \leq \rho\}$ is the closed ball centered at z with radius ρ while $\operatorname{sphere}(z;\rho) := \{x \in X \mid d(z,x) = \rho\}$ is the (closed) sphere centered at z with radius ρ. If u and v are in X, then $[u,v] := \{(1-\lambda)u + \lambda v \mid \lambda \in [0,1]\}$ is the line segment connecting u and v.

2 Foundations

We review in this section some of the fundamental tools used in the analysis of projection algorithms, and in particular MAP, for the solution of feasibility problems like (2). The tools below are intended for more general situations where the sets A and B might admit decompositions into unions of sets, in which case we consider the feasibility problem

$$\text{find} \quad c \in \Big(\bigcup_{i\in I} A_i\Big) \cap \Big(\bigcup_{j\in J} B_j\Big) \tag{7}$$

Central to the convergence analysis of the MAP algorithm for solving (7) is the notion of regularity of the intersection and the regularity of neighborhoods of the intersection. These ideas are developed in detail in [3]. We review the main points relevant to our application here.

Normal cones are used to provide information about the orientation and local geometry of subsets of X. There are many species of normal cones, the key ones for our purposes are defined here. In addition to the classical notions (*proximal*, *Fréchet*, *Mordukhovich*) we define the *restricted* normal cone introduced and developed in [3].

Definition 2.1 (normal cones) *Let A and B be nonempty subsets of X, and let a and u be in X. If $a \in A$, then various normal cones of A at a are defined as follows:*

(i) *The B-*restricted proximal normal cone *of A at a is*

$$\widehat{N}^B_A(a) := \operatorname{cone}\Big(\big(B \cap P_A^{-1}a\big) - a\Big) = \operatorname{cone}\Big(\big(B - a\big) \cap \big(P_A^{-1}a - a\big)\Big). \tag{8}$$

(ii) *The (classical)* proximal normal cone *of A at a is*

$$N^{\text{prox}}_A(a) := \widehat{N}^X_A(a) = \operatorname{cone}\big(P_A^{-1}a - a\big). \tag{9}$$

(iii) *The B-*restricted normal cone *$N^B_A(a)$ is implicitly defined by $u \in N^B_A(a)$ if and only if there exist sequences $(a_k)_{k\in\mathbb{N}}$ in A and $(u_k)_{k\in\mathbb{N}}$ in $\widehat{N}^B_A(a_k)$ such that $a_k \to a$ and $u_k \to u$.*

(iv) *The* Fréchet normal cone *$N^{\text{Fré}}_A(a)$ is implicitly defined by $u \in N^{\text{Fré}}_A(a)$ if and only if $(\forall \varepsilon > 0)$ $(\exists\, \delta > 0)$ $(\forall x \in A \cap \operatorname{ball}(a;\delta))$ $\langle u, x - a\rangle \le \varepsilon\|x - a\|$.*

(v) *The* convex normal from convex analysis *$N^{\text{conv}}_A(a)$ is implicitly defined by $u \in N^{\text{conv}}_A(a)$ if and only if $\sup\langle u, A - a\rangle \le 0$.*

(vi) *The* Mordukhovich normal cone *$N_A(a)$ of A at a is implicitly defined by $u \in N_A(a)$ if and only if there exist sequences $(a_k)_{k\in\mathbb{N}}$ in A and $(u_k)_{k\in\mathbb{N}}$ in $N^{\text{prox}}_A(a_k)$ such that $a_k \to a$ and $u_k \to u$.*

If $a \notin A$, then all normal cones are defined to be empty.

The following elementary calculus rules are a restatement of [3, Proposition 3.7].

Proposition 2.2 *Let A, A_1, A_2, B, B_1, and B_2 be nonempty subsets of X, let $c \in X$, and suppose that $a \in A \cap A_1 \cap A_2$. Then the following hold:*

(i) *If A and B are convex, then $\widehat{N}_A^B(a)$ is convex.*

(ii) $\widehat{N}_A^{B_1 \cup B_2}(a) = \widehat{N}_A^{B_1}(a) \cup \widehat{N}_A^{B_2}(a)$ *and* $N_A^{B_1 \cup B_2}(a) = N_A^{B_1}(a) \cup N_A^{B_2}(a)$.

(iii) *If $B \subseteq A$, then $\widehat{N}_A^B(a) = N_A^B(a) = \{0\}$.*

(iv) *If $A_1 \subseteq A_2$, then $\widehat{N}_{A_2}^B(a) \subseteq \widehat{N}_{A_1}^B(a)$.*

(v) $-\widehat{N}_A^B(a) = \widehat{N}_{-A}^{-B}(-a)$, $-N_A^B(a) = N_{-A}^{-B}(-a)$, *and* $-N_A(a) = N_{-A}(-a)$.

(vi) $\widehat{N}_A^B(a) = \widehat{N}_{A-c}^{B-c}(a-c)$ *and* $N_A^B(a) = N_{A-c}^{B-c}(a-c)$.

The *constraint qualification-*, or *CQ-number* defined next is built upon the normal cone and quantifies classical notions of constraint qualifications for set intersections that indicate sufficient regularity of the intersection.

Definition 2.3 ((joint) CQ-number) *Let A, $\widetilde{A}$, B, $\widetilde{B}$, be nonempty subsets of X, let $c \in X$, and let $\delta \in \mathbb{R}_{++}$. The* CQ-number *at c associated with $(A, \widetilde{A}, B, \widetilde{B})$ and δ is*

$$\theta_\delta := \theta_\delta\big(A, \widetilde{A}, B, \widetilde{B}\big) := \sup\left\{ \langle u, v\rangle \;\middle|\; \begin{array}{l} u \in \widehat{N}_A^{\widetilde{B}}(a), v \in -\widehat{N}_B^{\widetilde{A}}(b), \|u\| \le 1, \|v\| \le 1, \\ \|a-c\| \le \delta, \|b-c\| \le \delta. \end{array}\right\}. \tag{10}$$

The limiting CQ-number *at c associated with $(A, \widetilde{A}, B, \widetilde{B})$ is*

$$\overline{\theta} := \overline{\theta}\big(A, \widetilde{A}, B, \widetilde{B}\big) := \lim_{\delta \downarrow 0} \theta_\delta\big(A, \widetilde{A}, B, \widetilde{B}\big). \tag{11}$$

For nontrivial collections[2] $\mathcal{A} := (A_i)_{i\in I}$, $\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i\in I}$, $\mathcal{B} := (B_j)_{j\in J}$, $\widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j\in J}$ *of nonempty subsets of X, the* joint-CQ-number *at $c \in X$ associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ and $\delta > 0$ is*

$$\theta_\delta = \theta_\delta\big(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}\big) := \sup_{(i,j)\in I\times J} \theta_\delta\big(A_i, \widetilde{A}_i, B_j, \widetilde{B}_j\big), \tag{12}$$

and the limiting joint-CQ-number at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ is

$$\overline{\theta} = \overline{\theta}\big(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}\big) := \lim_{\delta \downarrow 0} \theta_\delta\big(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}\big). \tag{13}$$

The CQ-number is obviously an instance of the joint-CQ-number when I and J are singletons. When the arguments are clear from the context we will simply write θ_δ and $\overline{\theta}$.

[2]The collection $(A_i)_{i\in I}$ is said to be *nontrivial* if $I \neq \varnothing$.

Using Proposition 2.2(vi), we see that, for every $x \in X$,

$$\theta_\delta\big(A, \widetilde{A}, B, \widetilde{B}\big) \text{ at } c \quad = \quad \theta_\delta\big(A - x, \widetilde{A} - x, B - x, \widetilde{B} - x\big) \text{ at } c - x. \tag{14}$$

The CQ-number is based on the behavior of the restricted proximal normal cone in a neighborhood of a given point. A related notion is that of the exact CQ-number, defined next, which is based on the restricted normal cone at the point instead of nearby restricted proximal normal cones. In both instances, the important case to consider is when $c \in A \cap B$ (or when $c \in A_i \cap B_j$ in the joint-CQ case).

Definition 2.4 (exact CQ-number and exact joint-CQ-number) *Let $c \in X$.*

(i) *Let A, $\widetilde{A}$, B and $\widetilde{B}$ be nonempty subsets of X. The* exact CQ-number *at c associated with $(A, \widetilde{A}, B, \widetilde{B})$ is*

$$\overline{\alpha} := \overline{\alpha}\big(A, \widetilde{A}, B, \widetilde{B}\big) := \sup\left\{ \langle u, v\rangle \;\middle|\; u \in N_A^{\widetilde{B}}(c), v \in -N_B^{\widetilde{A}}(c), \|u\| \leq 1, \|v\| \leq 1 \right\}. \tag{15}$$

where we define $\overline{\alpha} = -\infty$ in the case that $c \notin A \cap B$ which is consistent with the convention $\sup \varnothing = -\infty$.

(ii) *Let $\mathcal{A} := (A_i)_{i\in I}$, $\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i\in I}$, $\mathcal{B} := (B_j)_{j\in J}$ and $\widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j\in J}$ be nontrivial collections of nonempty subsets of X. The* exact joint-CQ-number *at c associated with $(\mathcal{A}, \mathcal{B}, \widetilde{\mathcal{A}}, \widetilde{\mathcal{B}})$ is*

$$\overline{\alpha} := \overline{\alpha}(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}) := \sup_{(i,j)\in I\times J} \overline{\alpha}(A_i, \widetilde{A}_i, B_j, \widetilde{B}_j). \tag{16}$$

The next result, which we quote from [3, Theorem 7.8], establishes relationships between the condition numbers defined above.

Theorem 2.5 *Let $\mathcal{A} := (A_i)_{i\in I}$, $\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i\in I}$, $\mathcal{B} := (B_j)_{j\in J}$ and $\widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j\in J}$ be nontrivial collections of nonempty subsets of X. Set $A := \bigcup_{i\in I} A_i$ and $B := \bigcup_{j\in J} B_j$, and suppose that $c \in A \cap B$. Denote the exact joint-CQ-number at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ by $\overline{\alpha}$, the joint-CQ-number at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ and $\delta > 0$ by θ_δ, and the limiting joint-CQ-number at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ by $\overline{\theta}$. Then the following hold:*

(i) *If $\overline{\alpha} < 1$, then the $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$-CQ condition holds at c.*

(ii) $\overline{\alpha} \leq \theta_\delta$.

(iii) $\overline{\alpha} \leq \overline{\theta}$.

If in addition I and J are finite, then the following hold:

(iv) $\overline{\alpha} = \overline{\theta}$.

(v) *The* $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$*-joint-CQ condition holds at* c *if and only if* $\overline{\alpha} = \overline{\theta} < 1$.

The CQ-number is related to the angle of intersection of the sets. The case of linear subspaces underscores the subtleties of this idea and illustrates the connection between the CQ-number and the *correct* notion of an angle of intersection. The *Friedrichs angle* [14] (or simply the *angle*) between subspaces A and B is the number in $[0, \frac{\pi}{2}]$ whose cosine is given by

$$c(A,B) := \sup\left\{ |\langle a,b\rangle| \;\middle|\; a \in A \cap (A\cap B)^{\perp}, b \in B \cap (A \cap B)^{\perp}, \|a\| \le 1, \|b\| \le 1 \right\}, \tag{17}$$

and we set $c(A,B) := c(\operatorname{par} A, \operatorname{par} B)$ if A and B are two intersecting affine subspaces of X. The following result is a consolidation of [3, Theorem 8.12 and Corollary 8.13].

Theorem 2.6 (CQ-number of two (affine) subspaces and Friedrichs angle) *Let* A *and* B *be affine subspaces of* X*, and let* $\delta > 0$*. Then*

$$\theta_\delta(A,A,B,B) = \theta_\delta(A,X,B,B) = \theta_\delta(A,A,B,X) = c(A,B) < 1, \tag{18}$$

where the CQ-number at 0 *is defined as in* (10).

Moreover, if A *and* B *are* affine *subspaces of* X *with* $c \in A \cap B$*, and* $\delta > 0$*, then* (18) *holds at* c.

An easy consequence of Theorem 2.6 is the case of two distinct lines through the origin for which the CQ-number is simply the cosine of the angle between them ([3, Proposition 7.3]).

Corollary 2.7 (two distinct lines through the origin) *Suppose that* w_a *and* w_b *are two vectors in* X *such that* $\|w_a\| = \|w_b\| = 1$*. Let* $A := \mathbb{R}w_a$*,* $B := \mathbb{R}w_b$*, and* $\delta > 0$*. Assume that* $A \cap B = \{0\}$*. Then the CQ-number at* 0 *is*

$$\theta_\delta(A,A,B,B) = \theta_\delta(A,X,B,B) = \theta_\delta(A,A,B,X) = c(A,B) = |\langle w_a, w_b\rangle| < 1. \tag{19}$$

Convergence of MAP requires also a certain regularity on neighborhoods of the corresponding fixed points. For this we used a notion of regularity of the sets that is an adaptation to restricted normal cones of type of regularity introduced in [16].

Definition 2.8 ((joint-) regularity and (joint-) superregularity) *Let* A *and* B *be nonempty subsets of* X*, let* $\mathcal{B} := (B_j)_{j\in J}$ *be a nontrivial collection of nonempty subsets of* X*, and let* $c \in X$.

(i) *We say that* B *is* (A,ε,δ)-regular *at* $c \in X$ *if* $\varepsilon \ge 0$*,* $\delta > 0$*, and*

$$\left.\begin{array}{c} (y,b) \in B \times B, \\ \|y - c\| \le \delta, \|b - c\| \le \delta, \\ u \in \widehat{N}^A_B(b) \end{array}\right\} \quad\Rightarrow\quad \langle u, y-b\rangle \le \varepsilon \|u\| \cdot \|y - b\|. \tag{20}$$

If B *is* (X,ε,δ)*-regular at* c*, then we also simply speak of* (ε,δ)*-regularity.*

(ii) *The set B is called A-*superregular *at $c \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that B is (A, ε, δ)-regular at c. Again, if B is X-superregular at c, then we also say that B is superregular at c.*

(iii) *We say that $\mathcal{B}$ is (A, ε, δ)-joint-regular at c if $\varepsilon \geq 0$, $\delta > 0$, and for every $j \in J$, B_j is (A, ε, δ)-regular at c.*

(iv) *The collection $\mathcal{B}$ is A-joint-superregular at c if for every $j \in J$, B_j is A-superregular at c. We omit the prefix A if $A = X$.*

Joint-(super)regularity can be easily checked by any of the following conditions.

Proposition 2.9 *Let $\mathcal{A} := (A_j)_{j\in J}$ and $\mathcal{B} := (B_j)_{j\in J}$ be nontrivial collections of nonempty subsets of X, let $c \in X$, let $(\varepsilon_j)_{j\in J}$ be a collection in $\mathbb{R}_+$, and let $(\delta_j)_{j\in J}$ be a collection in $]0, +\infty]$. Set $A := \bigcap_{j\in J} A_j$, $\varepsilon := \sup_{j\in J} \varepsilon_j$, and $\delta := \inf_{j\in J} \delta_j$. Then the following hold:*

(i) *If $\delta > 0$ and $(\forall j \in J)$ B_j is $(A_j, \varepsilon_j, \delta_j)$-regular at c, then $\mathcal{B}$ is (A, ε, δ)-joint-regular at c.*

(ii) *If J is finite and $(\forall j \in J)$ B_j is $(A_j, \varepsilon_j, \delta_j)$-regular at c, then $\mathcal{B}$ is (A, ε, δ)-joint-regular at c.*

(iii) *If J is finite and $(\forall j \in J)$ B_j is A_j-superregular at c, then $\mathcal{B}$ is A-joint-superregular at c.*

If in addition $\mathcal{B} := (B_j)_{j\in J}$ is a nontrivial collection of nonempty convex *subsets of X then, for $A \subseteq X$, $\mathcal{B}$ is $(0, +\infty)$-joint-regular, $(A, 0, +\infty)$-joint-regular, joint-superregular, and A-joint-superregular at $c \in X$.*

The framework of restricted normal cones allows for a great deal of flexibility in how one decomposes problems. Whatever the chosen decomposition, the following properties will be required.

$$
\left\{
\begin{array}{l}
\mathcal{A} := (A_i)_{i\in I} \text{ and } \mathcal{B} := (B_j)_{j\in J} \text{ are nontrivial collections} \\
\quad \text{of nonempty closed subsets of } X; \\
A := \bigcup_{i\in I} A_i \text{ and } B := \bigcup_{j\in J} B_j \text{ are closed;} \\
c \in A \cap B; \\
\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i\in I} \text{ and } \widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j\in J} \text{ are collections} \\
\quad \text{of nonempty subsets of } X \text{ such that} \\
\qquad (\forall i \in I) \;\; P_{A_i}\big((\operatorname{bdry} B) \smallsetminus A\big) \subseteq \widetilde{A}_i, \\
\qquad (\forall j \in J) \;\; P_{B_j}\big((\operatorname{bdry} A) \smallsetminus B\big) \subseteq \widetilde{B}_j; \\
\widetilde{A} := \bigcup_{i\in I} \widetilde{A}_i \text{ and } \widetilde{B} := \bigcup_{j\in J} \widetilde{B}_j.
\end{array}
\right.
\tag{21}
$$

With the above assumptions one can establish rates of convergence for the MAP algorithms.

Theorem 2.10 (convergence rate, Corollary 10.8 of [3]) *Assume that* (21) *holds and that there exists* $\delta > 0$ *such that*

(i) $\mathcal{A}$ *is* $(\widetilde{B}, 0, 3\delta)$*-joint-regular at* c*;*

(ii) $\mathcal{B}$ *is* $(\widetilde{A}, 0, 3\delta)$*-joint-regular at* c*; and*

(iii) $\theta < 1$*, where* $\theta := \theta_{3\delta}$ *is the joint-CQ-number at* c *associated with* $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ *(see Definition 2.3).*

Suppose also that the starting point of the MAP b_{-1} *satisfies* $\|b_{-1} - c\| \leq \frac{(1-\theta)\delta}{6(2-\theta)}$*. Then* $(a_k)_{k\in\mathbb{N}}$ *and* $(b_k)_{k\in\mathbb{N}}$ *converge linearly to some point in* $\bar{c} \in A \cap B$ *with* $\|\bar{c} - c\| \leq \delta$ *and rate* θ^2*; in fact,*

$$(\forall k \geq 1) \quad \max\left\{\|a_k - \bar{c}\|, \|b_k - \bar{c}\|\right\} \leq \frac{\delta}{2-\theta}\left(\theta^2\right)^{k-1}. \tag{22}$$

3 Sparse feasibility with an affine constraint

We now move to the application of feasibility with a sparsity set and an affine subspace, problem (2). Our main result on the convergence of MAP is given in Theorem (3.19). Along the way we develop explicit representations of the projections, normal cones, and tangent cones to the sparsity set (3) and motivate our decomposition of the problem.

Properties of sparsity sets

Lemma 3.1 *Let* x *and* y *be in* $\mathbb{R}^n$*, and let* $\lambda \in \mathbb{R}$*. Then the following hold:*

(i) $\operatorname{supp}(x) = \operatorname{span}\left\{e_i \mid i \in I(x)\right\}$ *and* $\|x\|_0 = \operatorname{card}(I(x)) = \dim \operatorname{supp}(x)$*.*

(ii) $x \in \operatorname{supp}(y) \Leftrightarrow I(x) \subseteq I(y) \Leftrightarrow \operatorname{supp}(x) \subseteq \operatorname{supp}(y) \Rightarrow \|x\|_0 \leq \|y\|_0$*.*

(iii) $I(x+y) \subseteq I(x) \cup I(y)$ *and* $I(\lambda x) = \begin{cases} I(x), & \text{if } \lambda \neq 0; \\ \varnothing, & \text{otherwise.} \end{cases}$

(iv) $I((1-\lambda)x + \lambda y) \subseteq I(x) \cup I(y)$*.*

(v) $\operatorname{supp}(\lambda x) = \lambda \operatorname{supp}(x)$ *and* $\|\lambda x\|_0 = |\operatorname{sgn}(\lambda)| \cdot \|x\|_0$*.*

(vi) $\operatorname{supp}(x+y) \subseteq \operatorname{supp}(x) + \operatorname{supp}(y)$ *and* $\|x+y\|_0 \leq \|x\|_0 + \|y\|_0$*.*

(vii) *If* $\operatorname{supp}(x) \subseteq \operatorname{supp}(y)$ *and* $z \in \operatorname{supp}(y)$*, then there exist* u *and* v *in* $\mathbb{R}^n$ *such that* $z = u + v$*,* $u \in \operatorname{supp}(x)$ *and* $\|v\|_0 \leq \|y\|_0 - \|x\|_0$*.*

(viii) *Let* $\delta \in \left]0, \min\left\{|x_i| \mid i \in I(x)\right\}\right[$ *and* $y \in x + [-\delta, +\delta]^n$*, then* $\operatorname{supp}(x) \subseteq \operatorname{supp}(y)$*.*

(ix) *If* $I(x) \not\subseteq I(y)$ *and* $I(y) \not\subseteq I(x)$, *then*

$$\|x+y\|^2 \geq \min_{i\in I(x)\smallsetminus I(y)} |x_i|^2 + \min_{j\in I(y)\smallsetminus I(x)} |y_j|^2 \geq \min_{i\in I(x)} |x_i|^2 + \min_{j\in I(y)} |y_j|^2. \tag{23}$$

(x) $\|\cdot\|_0$ *is lower semicontinuous.*

Proof. (i)–(v): These follow readily from the definitions.

(vi): By (ii), $I(x+y) \subseteq I(x) \cup I(y)$. Hence $\operatorname{supp}(x+y) \subseteq \operatorname{supp}(x) + \operatorname{supp}(y)$; on the other hand, taking cardinality and using (i) yields $\|x+y\|_0 \leq \|x\|_0 + \|y\|_0$.

(vii): By (ii), we have $I(x) \subseteq I(y)$. Write $I(y) = I(x) \dot\cup J$ as disjoint union, where $J = I(y) \smallsetminus I(x)$, and note that that $\operatorname{card}(J) = \operatorname{card}(I(y)) - \operatorname{card}(I(x)) = \|y\|_0 - \|x\|_0$. Then $\operatorname{supp}(y) = \operatorname{supp}(x) \oplus \operatorname{span}\{e_i \mid i \in J\}$. Now since $z \in \operatorname{supp}(y)$, we can write $z = u + v$, where $u \in \operatorname{supp}(x)$ and $v \in \operatorname{span}\{e_i \mid i \in J\}$ and $\|v\|_0 \leq \operatorname{card}(J) = \|y\|_0 - \|x\|_0$.

(viii): If $i \in I(x)$, then $|y_i| \geq |x_i| - |x_i - y_i| > \delta - |x_i - y_i| \geq 0$ and hence $y_i \neq 0$. It follows that $I(x) \subseteq I(y)$. Now apply (ii).

(ix): Let $i_0 \in I(x) \smallsetminus I(y)$ and $j_0 \in I(y) \smallsetminus I(x)$. Then $y_{i_0} = 0$ and $x_{j_0} = 0$, and hence

$$\|x+y\|^2 \geq |x_{i_0} + y_{i_0}|^2 + |x_{j_0} + y_{j_0}|^2 \tag{24a}$$

$$\geq \min_{i\in I(x)\smallsetminus I(y)} |x_i|^2 + \min_{j\in I(y)\smallsetminus I(x)} |y_j|^2 \tag{24b}$$

$$\geq \min_{i\in I(x)} |x_i|^2 + \min_{j\in I(y)} |y_j|^2, \tag{24c}$$

as claimed.

(x): Indeed, borrowing the notation below, we see that $\{z \in X \mid \|z\|_0 \leq \rho\} = \bigcup_{J\in\mathcal{J}_r} A_J$, where $r = \lfloor \rho \rfloor$, is closed as a union of finitely many (closed) linear subspaces. ■

In order to apply Theorem 2.10 to MAP for solving (2) we must choose a suitable decomposition, $\mathcal{A}$ and $\mathcal{B}$, and restrictions, $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$, and verify the assumptions of the theorem. We now abbreviate

$$\mathcal{J} := 2^{\{1,2,\ldots,n\}} \quad \text{and} \quad \mathcal{J}_s := \mathcal{J}(s) := \{J \in \mathcal{J} \mid \operatorname{card}(J) = s\} \tag{25a}$$

and set

$$(\forall J \in \mathcal{J}) \quad A_J := \operatorname{span}\{e_j \mid j \in J\}. \tag{25b}$$

Define the collections

$$\mathcal{A} := \widetilde{\mathcal{A}} := (A_J)_{J\in\mathcal{J}_s} \quad \text{and} \quad \mathcal{B} := \widetilde{\mathcal{B}} := (B). \tag{25c}$$

Clearly,

$$A := \widetilde{A} := \bigcup_{J\in\mathcal{J}_s} A_J = \{x \in \mathbb{R}^n \mid \|x\|_0 \leq s\} \quad \text{and} \quad B = \widetilde{B} := \{x \in X \mid Mx = p\}. \tag{25d}$$

The proofs of the following two results are elementary and thus omitted.

Proposition 3.2 (properties of A_J) *Let J, J_1, and J_2 be in $\mathcal{J}$, and let $x \in X$. Then the following hold:*

(i) $A_{J_1} \cup A_{J_2} \subseteq A_{J_1 \cup J_2} = \operatorname{span}(A_{J_1} \cup A_{J_2})$.

(ii) $J_1 \subseteq J_2 \Leftrightarrow A_{J_1} \subseteq A_{J_2}$.

(iii) $x \in A_{I(x)} = \operatorname{supp}(x)$.

(iv) $I(x) \subseteq J \Leftrightarrow x \in A_J$.

(v) $I(x) \cap J = \varnothing \Leftrightarrow x \in A_J^{\perp}$.

(vi) $s \leq n-1 \Leftrightarrow \operatorname{int} A = \varnothing$.

Proposition 3.3 *Let $J \in \mathcal{J}$, let $x = (x_1, \ldots, x_n) \in X$, and set $y := P_{A_J}x$. Then*

$$(\forall i \in \{1, \ldots, n\}) \quad y_i = \begin{cases} x_i, & \text{if } i \in J; \\ 0, & \text{if } i \notin J, \end{cases} \tag{26}$$

and

$$d_{A_J}^2(x) = \sum_{j \in \{1,\ldots,n\} \smallsetminus J} |x_j|^2 = \sum_{j \in I(x) \smallsetminus J} |x_j|^2. \tag{27}$$

The following technical result will be useful later.

Lemma 3.4 *Let $c \in A$, and assume that $s \leq n-1$. Then*

$$\min\big\{d_{A_J}(c) \;\big|\; c \notin A_J, J \in \mathcal{J}_s\big\} = \min\big\{|c_j| \;\big|\; j \in I(c)\big\}. \tag{28}$$

Proof. First, let $J \in \mathcal{J}_s$ such that $c \notin A_J \Leftrightarrow I(c) \not\subseteq J$ by Proposition 3.2(iv). So $I(c) \smallsetminus J \neq \varnothing$. By (27), $d_{A_J}^2(c) = \sum_{j \in I(c) \smallsetminus J} |c_j|^2 \geq \min\big\{|c_j|^2 \;\big|\; j \in I(c)\big\}$. Hence

$$\min\big\{d_{A_J}(c) \;\big|\; c \notin A_J, J \in \mathcal{J}_s\big\} \geq \min\big\{|c_j| \;\big|\; j \in I(c)\big\}. \tag{29}$$

Since $1 \leq 1 + s - \|c\|_0 \leq n - \|c\|_0 = \operatorname{card}(\{1, \ldots, n\} \smallsetminus I(c))$, there exists a nonempty subset K of $\{1, \ldots, n\} \smallsetminus I(c)$ with $\operatorname{card}(K) = s - \|c\|_0 + 1$. Let $j \in I(c)$ such that $|c_j| = \min_{i \in I(c)} |c_i|$ and set

$$J := (I(c) \smallsetminus \{j\}) \cup K. \tag{30}$$

Then $c \notin A_J$ and $\operatorname{card}(J) = \operatorname{card}(I(c)) - 1 + \operatorname{card}(K) = \|c\|_0 - 1 + s - \|c\|_0 + 1 = s$. Hence $J \in \mathcal{J}_s$. Because $I(c) \smallsetminus J = \{j\}$, it follows again from (27) that $d_{A_J}^2(c) = \sum_{i \in I(c) \smallsetminus J} |c_i|^2 = |c_j|^2$. Therefore $d_{A_J}(c) = |c_j| = \min_{i \in I(c)} |c_i|$, which yields the inequality complementary to (30). ■

Now let $x = (x_1, ..., x_n) \in X$, and set

$$\mathcal{C}_s(x) := \big\{J \in \mathcal{J}_s \bigm| \min_{j\in J} |x_j| \geq \max_{k\notin J} |x_k|\big\}; \tag{31}$$

in other words, $J \in \mathcal{C}_s(y)$ if and only if J contains s indices to the s largest coordinates of x in absolute value.

The proof of the next result is straightforward.

Lemma 3.5 *Let $x = (x_1, \ldots, x_n) \in X$ such that $\|x\|_0 = \operatorname{card}(I(x)) \geq s$, and let $J \in \mathcal{C}_s(x)$. Then $J \subseteq I(x)$ and $\min_{j\in J} |x_j| \geq \min_{j\in I(x)} |x_j| > 0$. If $\|x\|_0 = \operatorname{card}(I(x)) = s$, then $\mathcal{C}_s(x) = \{I(x)\}$.*

Projections

The decomposition of the sparsity set defined by (25) yields a natural expression for the projection onto this set.

Proposition 3.6 (Projection onto A and its inverse) *Let $x = (x_1, \ldots, x_n) \in X$, and define $A := \big\{x \in X \bigm| \|x\|_0 \leq s\big\}$. Then the following hold:*

(i) *The distance from x to A is solely determined by $\mathcal{C}_s(x)$; more precisely,*

$$(\forall J \in \mathcal{J}_s) \quad d_{A_J}(x) \begin{cases} = d_A(x), & \text{if } J \in \mathcal{C}_s(x); \\ > d_A(x), & \text{if } J \notin \mathcal{C}_s(x). \end{cases} \tag{32}$$

(ii) *The projection of x on A is solely determined by $\mathcal{C}_s(x)$; more precisely,*

$$P_A(x) = \bigcup_{J\in\mathcal{C}_s(x)} P_{A_J}(x) = \bigcup_{J\in\mathcal{C}_s(x)} \left\{ y = (y_1, \ldots, y_n) \in X \;\middle|\; (\forall j \in \{1, \ldots, n\})\ y_j = \begin{cases} x_j, & \text{if } j \in J; \\ 0, & \text{if } j \notin J. \end{cases} \right\} \tag{33}$$

(iii) $(\forall y \in P_A(x))\ \|y\|_0 = \min\{\|x\|_0, s\}$.

(iv) *If $x \notin A$, then $(\forall y \in P_A(x))\ I(y) \in \mathcal{C}_s(x)$ and $\|y\|_0 = s$.*

(v) *If $a \in A$ and $\|a\|_0 = s$, then*

$$P_A^{-1}(a) = \left\{ y = (y_1, \ldots, y_n) \in X \;\middle|\; \begin{array}{l} (\forall j \in I(a))\ y_j = a_j \\ \max_{k\notin I(a)} |y_k| \leq \min_{j\in I(a)} |a_j|. \end{array} \right\} \tag{34}$$

(vi) *If $a \in A$ and $\|a\|_0 < s$, then $P_A^{-1}(a) = a$.*

Proof. The following observation will be useful. If $J \in \mathcal{J}_s$, $j \in J$, and $k \notin J$, then $K := (J \smallsetminus \{j\}) \cup \{k\} \in \mathcal{J}_s$ and (27) implies

$$d^2_{A_K}(x) = \sum_{l \notin K} |x_l|^2 = \|x\|^2 - \sum_{l \in K} |x_l|^2 = \|x\|^2 - \sum_{l \in J \cap K} |x_l|^2 - |x_k|^2 \tag{35a}$$

$$= \|x\|^2 - \sum_{l \in J \cap K} |x_j|^2 - |x_j|^2 + \big(|x_j|^2 - |x_k|^2\big) \tag{35b}$$

$$= \|x\|^2 - \sum_{l \in J} |x_l|^2 + \big(|x_j|^2 - |x_k|^2\big) = \sum_{l \notin J} |x_l|^2 + \big(|x_j|^2 - |x_k|^2\big) \tag{35c}$$

$$= d^2_{A_J}(x) + |x_j|^2 - |x_k|^2. \tag{35d}$$

(i): It is clear that

$$d_A(x) = \min\big\{ d_{A_J}(x) \;\big|\; J \in \mathcal{J}_s \big\}. \tag{36}$$

Let $K \in \mathcal{J}_s$ and assume that $K \notin \mathcal{C}_s(x)$. Then there exists j and k in $\{1, \dots, n\}$ such that $k \in K$, $j \notin K$, and $|x_k| < |x_j|$. Now define $J = (K \smallsetminus \{k\}) \cup \{j\}$. Then $J \in \mathcal{J}_s$ and

$$d^2_{A_K}(x) = d^2_{A_J}(x) + |x_j|^2 - |x_k|^2 > d^2_{A_J}(x) \tag{37}$$

by (35). It follows that index sets in $\mathcal{J}_s \smallsetminus \mathcal{C}_s(x)$ do not contribute to the computation of $d_A(x)$.

Now assume that J and K both belong to $\mathcal{C}_s(x)$ and that $J \neq K$. Then $\operatorname{card}(J \smallsetminus K) = \operatorname{card}(K \smallsetminus J)$. Take $j \in J \smallsetminus K$ and $k \in K \smallsetminus J$. Since $j \in J \in \mathcal{C}_s(x)$ and $k \notin J$, we have $|x_j| \geq |x_k|$. On the other hand, since $k \in K \in \mathcal{C}_s(x)$ and $j \notin K$, we also have $|x_k| \geq |x_j|$. Altogether, $|x_j| = |x_k|$. Thus

$$d^2_{A_J}(x) = \|x\|^2 - \sum_{l \in J} |x_l|^2 = \|x\|^2 - \sum_{l \in J \cap K} |x_l|^2 - \sum_{l \in J \smallsetminus K} |x_l|^2 \tag{38a}$$

$$= \|x\|^2 - \sum_{l \in K \cap J} |x_l|^2 - \sum_{l \in K \smallsetminus J} |x_l|^2 = \|x\|^2 - \sum_{l \in K} |x_l|^2 = d^2_{A_K}(x). \tag{38b}$$

This completes the proof of (32).

(ii): This follows from (32) and (26).

(iii): *Case 1:* $\|x\|_0 = \operatorname{card}(I(x)) \leq s$. Then, by definition, $x \in A$. Thus $P_A(x) = x$ and hence $\|P_A(x)\|_0 = \|x\|_0 = \min\{\|x\|_0, s\}$.

Case 2: $\|x\|_0 = \operatorname{card}(I(x)) > s$. Let $J \in \mathcal{C}_s(x)$. Lemma 3.5 implies $\min_{j \in J} |x_j| > 0$. It follows from (33) that there exists $y = (y_1, \dots, y_n) \in P_A(x)$ such that

$$(\forall j \in J)\ |y_j| = |x_j| > 0 \quad \text{and} \quad (\forall j \notin J)\ y_j = 0. \tag{39}$$

So

$$I(y) = J, \tag{40}$$

and hence $\|y\|_0 = \operatorname{card}(J) = s = \min\{\operatorname{card}(I(x)), s\}$.

(iv): Let $y \in P_A(x)$. Since $x \notin A$, we have $\|x\|_0 > s$ and hence (iii) implies that $\|y\|_0 = s$. By (33), there exists $J \in \mathcal{C}_s(x)$ such that $I(y) \subseteq J$. But card $I(y) = s =$ card J, and hence $I(y) = J$.

(v): Denote the right-hand side of (34) by R. "$\supseteq$": for every $y \in R$, we have $I(a) \in \mathcal{C}_s(y)$. By (33), $a \in P_A y$. Hence $y \in P_A^{-1}(a)$. This establishes $R \subseteq P_A^{-1}(a)$. "$\subseteq$": Suppose that $y \in P_A^{-1}(a)$, i.e., $a \in P_A(y)$. Again by (33), there exists $J \in \mathcal{C}_s(y)$ such that

$$(\forall j \in J)\ a_j = y_j \quad \text{and} \quad (\forall j \notin J)\ a_j = 0. \tag{41}$$

Since $\|a\|_0 = s$, Lemma 3.5 implies that $J = I(a)$. Hence, by (41), $(\forall j \in I(a))\ y_j = a_j$. On the other hand, by definition of $\mathcal{C}_s(y)$, we have $\min_{j \in J} |y_j| \geq \max_{k \notin J} |y_k|$. Altogether, $y \in R$.

(vi): Let $y \in P_A^{-1}a$, i.e., $a \in P_A y$. The hypothesis and (iii) imply $s > \|a\|_0 = \min\{\|y\|_0, s\}$, Hence $\|y\|_0 < s$; therefore, $y \in A$ and so $a = P_A y = y$. ■

Proposition 3.7 (projection onto B) (See [4, Lemma 4.1].) *Recall that $B = \{x \in X \mid Mx = p\}$. Then the projection onto B is given by*

$$P_B \colon X \to X \colon x \mapsto x - M^\dagger(Mx - p), \tag{42}$$

where $M^\dagger$ denotes the Moore-Penrose inverse of M.

Normal and tangent cones

Proposition 3.8 (proximal normal cone to A)

$$(\forall a \in A) \quad N_A^{\text{prox}}(a) = \begin{cases} (\text{supp}(a))^\perp, & \text{if } \|a\|_0 = s; \\ \{0\}, & \text{if } \|a\|_0 < s. \end{cases} \tag{43}$$

Proof. Combine the definition of $N_A^{\text{prox}}(a)$ with Proposition 3.6(v)&(vi). ■

The following is a special case of a more general normal cone formulation for the set of matrices with rank bounded above by s given in [19].

Theorem 3.9 (Mordukhovich normal cone to A)

$$(\forall a \in A) \quad N_A(a) = \{u \in \mathbb{R}^n \mid \|u\|_0 \leq n - s\} \cap (\text{supp}(a))^\perp = \bigcup_{I(a) \subseteq J \in \mathcal{J}_s} A_J^\perp. \tag{44}$$

Consequently, if $\|a\|_0 = s$, then $N_A(a) = (\text{supp}(a))^\perp = A_{I(a)}^\perp$.

Proof. Let $a \in A$, and let $\varepsilon \in \left]0, \min\{a_j \mid j \in I(a)\}\right[$. Let $x = (x_1, \ldots, x_n) \in A \cap (a + [-\varepsilon, +\varepsilon]^n)$. Then $\|x\|_0 \leq s$ and, by Lemma 3.1(viii), $\text{supp}(a) \subseteq \text{supp}(x)$. Hence, using Proposition 3.8, we deduce that

$$N_A^{\text{prox}}(x) = \begin{cases} (\text{supp}(x))^\perp, & \text{if } \|x\|_0 = s; \\ \{0\}, & \text{if } \|x\|_0 < s \end{cases} \subseteq (\text{supp}(a))^\perp. \tag{45}$$

Note that if $\|x\|_0 = s$, then (45) yields $\dim(\operatorname{supp}(x))^\perp = n - s$; in either case,

$$(46) \qquad \big(\forall u \in N_A^{\mathrm{prox}}(x)\big) \quad \|u\|_0 \leq n - s.$$

Let $u \in X$. We assume first that $u \in N_A(a)$. Then there exist sequences $(x_k)_{k\in\mathbb{N}}$ in $A \cap \big(a + [-\varepsilon, +\varepsilon]^n\big)$ and $(u_k)_{k\in\mathbb{N}}$ in X such that $x_k \to a$, $u_k \to u$, and $(\forall k \in \mathbb{N})$ $u_k \in N_A^{\mathrm{prox}}(x_k)$. It follows from (45), (46), and Lemma 3.1(x) that $u \in (\operatorname{supp}(a))^\perp$ and $\|u\|_0 \leq n - s$. Thus

$$(47) \qquad N_A(a) \subseteq \big\{u \in \mathbb{R}^n \bigm| \|u\|_0 \leq n - s\big\} \cap \big(\operatorname{supp}(a)\big)^\perp.$$

We now assume that $u \in (\operatorname{supp}(a))^\perp$ and $\|u\|_0 \leq n - s$. Since $u \in (\operatorname{supp}(a))^\perp$, we have $I(a) \cap I(u) = \varnothing$ and hence $I(a) \subset \{1, 2, \ldots, n\} \smallsetminus I(u)$. Since $a \in A$ and $\operatorname{card} I(u) = \|u\|_0 \leq n - s$, we have $\operatorname{card} I(a) \leq s \leq \operatorname{card}(\{1, 2, \ldots, n\} \smallsetminus I(u))$. Let $J \in \mathcal{J}_s$ such that $I(a) \subseteq J \subseteq \{1, 2, \ldots, n\} \smallsetminus I(u)$. By Proposition 3.2(v), $u \in A_J^\perp$. We have established that

$$(48) \qquad \big\{u \in \mathbb{R}^n \bigm| \|u\|_0 \leq n - s\big\} \cap \big(\operatorname{supp}(a)\big)^\perp \subseteq \bigcup_{I(a)\subseteq J \in \mathcal{J}_s} A_J^\perp.$$

Finally, assume that $u \in A_J^\perp$, where $\operatorname{card} J = s$ and $I(a) \subseteq J$. Set

$$(49) \qquad (\forall \varepsilon \in \mathbb{R}_{++})(\forall j \in \{1, 2, \ldots, n\}) \quad x_{\varepsilon,j} := \begin{cases} a_j, & \text{if } j \in I(a); \\ \varepsilon, & \text{if } j \in J \smallsetminus I(a); \\ 0 & \text{otherwise.} \end{cases}$$

This defines a bounded net $(x_\varepsilon)_{\varepsilon\in]0,1[}$ in X with $x_\varepsilon \to a$ as $\varepsilon \to 0$. Note that $(\forall \varepsilon \in]0, 1[)$ $I(x_\varepsilon) = J$; hence, $x_\varepsilon \in \operatorname{supp}(x_\varepsilon) = A_J \subseteq A$ and, by Proposition 3.8, $u \in A_J^\perp = (\operatorname{supp}(x_\varepsilon))^\perp = N_A^{\mathrm{prox}}(x_\varepsilon)$. Thus $u \in N_A(a)$. We have established the inclusion

$$(50) \qquad \bigcup_{I(a)\subseteq J \in \mathcal{J}_s} A_J^\perp \subseteq N_A(a).$$

This completes the proof of (44).

Finally, if $\|a\|_0 = s$, then $\operatorname{card} I(a) = s$ and the only choice for J in (44) is $I(a)$. ■

We now turn to the classical tangent cone of A.

Definition 3.10 (tangent cone) *Let C be a nonempty subset of X, and let $c \in C$. Then a vector $v \in X$ belongs to the* tangent cone *to C at c, denoted $T_C(c)$, if there exist sequences $(x_k)_{k\in\mathbb{N}}$ in C and $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{R}_{++}$ such that $x_k \to c$, $t_k \to 0$, and $(x_k - c)/t_k \to v$.*

The proof of the following result is elementary and hence omitted.

Lemma 3.11 *Let C be a nonempty subset of X, let $c \in C$, and assume that $(Y_k)_{k\in K}$ a finite collection of affine subspaces such that $y \in \bigcap_{k\in K} Y_k \subseteq Y := \bigcup_{k\in K} Y_k$. Then the following hold:*

(i) $(\forall \rho \in \mathbb{R}_{++})$ $T_C(c) = T_{(C\cap \text{ball}(c;\rho))}(c)$.

(ii) $T_Y(y) = \bigcup_{k\in K} \text{par}(Y_k)$.

(iii) *If each* Y_k *is a linear subspace, then* $T_Y(y) = Y$.

Lemma 3.12 *Let* $a = (a_1, \ldots, a_n) \in A$ *and suppose that* $0 < \rho \leq \min_{j\in I(a)} |a_j|$. *Then*

$$\text{ball}(a;\rho) \cap A = \text{ball}(a;\rho) \cap \bigcup_{I(a)\subseteq J\in \mathcal{J}_s} A_J. \tag{51}$$

Proof. The inclusion "$\supseteq$" is clear. To prove "$\subseteq$", let $x \in A \cap \text{ball}(a;\rho)$. If $I(x) \not\subseteq I(a)$ and $I(a) \not\subseteq I(x)$, then Lemma 3.1(ix) implies $\rho^2 \geq \|x-a\|^2 \geq \min_{i\in I(x)} |x_i|^2 + \min_{j\in I(a)} |a_j|^2 > \rho^2$, which is absurd. Therefore, $I(x) \subseteq I(a)$ or $I(a) \subseteq I(x)$. Furthermore, there exists $J \in \mathcal{J}_s$ such that $I(a) \subseteq I(a) \cup I(x) \subseteq J$. By Proposition 3.2(iv), $x \in A_J$. This completes the proof. ■

Corollary 3.13 *Let* $a \in A$. *If* $s = n$, *then* A *is superregular at* a; *otherwise,* A *is superregular at* $a \Leftrightarrow \|a\|_0 = s$.

Proof. Since $A = X$ if $s = n$, the first statement is clear. We now consider two cases. *Case 1:* $\|a\|_0 \leq s-1$. By (44),

$$N_A(a) = \bigcup_{I(a)\subseteq J\in \mathcal{J}_s} A_J^{\perp}. \tag{52}$$

Since card $I(a) < s$, $N_A(a)$ is therefore the finite union of two or more different linear subspaces of X all of the same dimension $n-s$. Hence $N_A(a)$ cannot be convex. On the other hand, $N_A^{\text{Fré}}(a)$ is always convex. Altogether, $N_A^{\text{Fré}}(a) \neq N_A(a)$. Thus, by [23, Definition 6.4], A is not Clarke regular at a. Hence [16, Corollary 4.5] implies that A is not superregular at a.

Case 2: $\|a\|_0 = s$. Let ρ be as in Lemma 3.12. Then Lemma 3.12 implies that

$$\text{ball}(a;\rho) \cap A = \text{ball}(c;\rho) \cap A_{I(a)} \tag{53}$$

is convex because it is the intersection of a ball and a linear subspace. By [3, Remark 9.2(vii)], A is superregular at c. ■

Lemma 3.14 *Let* $a \in A$. *Then*

$$\bigcup_{I(a)\subseteq J\in \mathcal{J}_s} A_J = \text{supp}(a) + \big\{x \in X \;\big|\; \|x\|_0 \leq s - \|a\|_0\big\}. \tag{54}$$

Proof. "$\subseteq$": Let $z \in A_J$, where $I(a) \subseteq J \in \mathcal{J}_s$. Write $J = I(a) \dot{\cup} K$, where $K := J \smallsetminus I(a)$ and the union is disjoint. Then $z = y + x$, where $y \in A_{I(a)} = \text{supp}(a)$, $x \in A_K$, and $\|x\|_0 \leq \text{card}(K) = \text{card}(J) - \text{card}(I(a)) = s - \|a\|_0$.

"$\supseteq$": Let $x \in X$ be such that $\|x\|_0 \leq s - \|a\|_0$, and let $y \in \operatorname{supp}(a)$. By Lemma 3.1, $I(y) \subseteq I(a)$, $I(x+y) \subseteq I(x) \cup I(y) \subseteq I(x) \cup I(a)$ and $\|x+y\|_0 \leq \|x\|_0 + \|y\|_0 \leq (s - \|a\|_0) + \|a\|_0 = s$. Hence, there exists $J \in \mathcal{J}_s$ such that $I(x) \cup I(a) \subset J$, and therefore $x + y \in A_{I(x)\cup I(a)} \subseteq A_J$. ■

Theorem 3.15 (tangent cone to A) *Let $a = (a_1, \ldots, a_n) \in A$. Then*

$$T_A(a) = \bigcup_{I(a)\subseteq J\in\mathcal{J}_s} A_J = \operatorname{supp}(a) + \big\{x \in X \;\big|\; \|x\|_0 \leq s - \|a\|_0\big\}; \tag{55}$$

consequently,

$$\|a\|_0 = s \quad \Leftrightarrow \quad T_A(a) = A_{I(a)} = \operatorname{supp}(a). \tag{56}$$

Proof. Set

$$\rho := \min_{j\in I(a)} |a_j| > 0 \quad \text{and} \quad A(a) := \bigcup_{I(a)\subseteq J\in\mathcal{J}_s} A_J = \bigcup_{a\in A_J, J\in\mathcal{J}_s} A_J. \tag{57}$$

Lemma 3.11(i) and Lemma 3.12 imply

$$T_A(a) = T_{A\cap \operatorname{ball}(a;\rho)}(a) = T_{A(a)\cap \operatorname{ball}(a;\rho)}(a) = T_{A(a)}(a). \tag{58}$$

On the other hand, by Lemma 3.11(iii), $T_{A(a)}(a) = A(a)$. Altogether, $T_A(a) = A(a)$ and we have established the first equality in (55). The second equality is precisely Lemma 3.14. Finally, the "consequently" part is clear from (55). ■

Remark 3.16 For the affine set B, the normal and tangent cones are much simpler to derive: indeed, because $\operatorname{par}(B) = \ker M$, it follows that $T_B(x) = \ker M$ and $N_B(x) = (\ker M)^\perp = \operatorname{ran} M^T$, for every $x \in B$.

Remark 3.17 (transversality) Recall (2) and assume that $c \in A \cap B$. By (55), Remark 3.16, and e.g. [2, Lemma 1.43(i)], we have the implications

(59a)

$$T_A(c) + T_B(c) = \mathbb{R}^n \Leftrightarrow \left(\bigcup_{I(c)\subseteq J\in\mathcal{J}_s} A_J\right) + \ker(M) = \mathbb{R}^n$$

$$\Leftrightarrow \bigcup_{I(c)\subseteq J\in\mathcal{J}_s} \big(A_J + \ker(M)\big) = \mathbb{R}^n \tag{59b}$$

$$\Leftrightarrow \operatorname{int}\left(\bigcup_{I(c)\subseteq J\in\mathcal{J}_s} \big(A_J + \ker(M)\big)\right) = \mathbb{R}^n \tag{59c}$$

$$\Rightarrow \overline{\operatorname{int}\left(\bigcup_{I(c)\subseteq J\in\mathcal{J}_s} \big(A_J + \ker(M)\big)\right)} = \overline{\bigcup_{I(c)\subseteq J\in\mathcal{J}_s} \operatorname{int}\big(A_J + \ker(M)\big)} = \mathbb{R}^n. \tag{59d}$$

Let us assume momentarily that $T_A(c) + T_B(c) = \mathbb{R}^n$. By (59), there exists $J \in \mathcal{J}_s$ such that $I(c) \subseteq J$ and $A_J + \ker(M) = \mathbb{R}^n$. Hence $s + \dim\ker(M) = \dim A_J + \dim\ker(M) \geq \dim(A_J + \ker(M)) = \dim\mathbb{R}^n = n = \dim\ker(M) + \operatorname{rank}(M)$. We have established the implication

$$T_A(c) + T_B(c) = \mathbb{R}^n \quad \Rightarrow \quad s \geq \operatorname{rank}(M); \tag{60}$$

that is, *transversality* imposes a lower bound on s and is thus at odds with the objective of finding the *sparsest* points in $A \cap B$.

The MAP for the sparse feasibility problem

We begin with an example illustrating shortcomings of previous approaches.

Example 3.18 Suppose that

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and} \quad s = 1; \tag{61}$$

thus, $m = 2$ and $n = 3$. Then $B = (1, 0, 0) + \mathbb{R}(-1, 1, 0)$ and hence the set of all solutions to (2) consists of[3] $x^* := (1, 0, 0)$ and $y^* := (0, 1, 0)$. Since $\|x^*\| = \|y^*\| = s$, Theorem 3.9 yields

$$N_A(x^*) = \{0\} \times \mathbb{R} \times \mathbb{R} \quad \text{and} \quad N_A(y^*) = \mathbb{R} \times \{0\} \times \mathbb{R}. \tag{62}$$

On the other hand, $(\forall x \in B)\ N_B(x) = \operatorname{ran} M^T = \operatorname{span}\{(1, 1, 1), (1, 1, 0)\}$ by Remark 3.16. Altogether,

$$N_A(x^*) \cap \big(-N_B(x^*)\big) = N_A(y^*) \cap \big(-N_B(y^*)\big) = \{0\} \times \{0\} \times \mathbb{R} \neq \{(0, 0, 0)\}. \tag{63}$$

Consequently, neither the Lewis-Luke-Malick framework [16] nor the framework proposed in [18] is able to deal with this case. Furthermore, in view of (60), the transversality condition

$$T_A(c) + T_B(c) = \mathbb{R}^n \tag{64}$$

proposed by Lewis and Malick [17] also always fails because $s = 1 \not\geq 2 = \operatorname{rank}(M)$.

Finally, readers familiar with sparse optimization will also note that the usual sufficient conditions for the correspondence of solutions to the nonconvex problem to those of convex relaxations—namely the restricted isometry property [9] or the mutual coherence condition [13]—are not satisfied either. Constraint qualifications as developed in the present work have no apparent relation to conditions like restricted isometry or mutual coherence conditions used to guarantee the correspondence between solutions to convex surrogate problems and solutions to the problem with the original $\|\cdot\|_0$ objective. Indeed, if the matrix M is changed for instance to

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \end{pmatrix} \tag{65}$$

the mutual coherence condition is satisfied and a unique sparsest solution exists, but still the constraint qualifications (63) and (64) are not satisfied.

[3] When there is no cause for confusion, we shall write column vectors as row vectors for space reasons.

We are now ready for our main result, which is very general and which in particular is applicable to the setting of Example 3.18.

Theorem 3.19 (main result for sparse affine feasibility and linear local convergence of MAP) *Let A, $\mathcal{A}$, $\widetilde{\mathcal{A}}$ B, $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be defined by* (25)*. Suppose that $s \leq n-1$, that $c \in A \cap B$, and fix $\delta \in \left]0, \overline{\delta}\right[$ for $\overline{\delta} := \frac{1}{3} \min\left\{d_{A_J}(c) \mid c \notin A_J, J \in \mathcal{J}_s\right\}$. Then*

$$\overline{\delta} = \tfrac{1}{3} \min\left\{|c_j| \mid j \in I(c)\right\} \tag{66}$$

and

$$\overline{\alpha} = \overline{\theta} = \theta_{3\delta}(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}) = \max\left\{c(A_J, B) \mid c \in A_J, J \in \mathcal{J}_s\right\} < 1, \tag{67}$$

where $\theta_{3\delta}$, $\overline{\theta}$, $\overline{\alpha}$ denote the joint-CQ-number, the limiting joint-CQ-number and the exact joint-CQ-number ((12), (13) *and* (16) *respectively) at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$. Suppose the starting point of the MAP b_{-1} satisfies $\|b_{-1} - c\| \leq \frac{(1-\overline{\theta})\delta}{6(2-\overline{\theta})}$. Then $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ converge linearly to some point in $\bar{c} \in A \cap B \cap \operatorname{ball}(c;\delta)$ with rate $\overline{\theta}^2$.*

Proof. Observe that (66) follows from Lemma 3.4. Let $J \in \mathcal{J}_s$. If $c \notin A_J$, then $\operatorname{ball}(c; 3\delta) \cap A_J = \varnothing$ and hence $\theta_{3\delta}(A_J, A_J, B, B) = -\infty$. On the other hand, if $c \in A_J$, then $c \in A_J \cap B$ and hence $\theta_{3\delta}(A_J, A_J, B, B) = c(A_J, B) < 1$ by Theorem 2.6. Combining this with Theorem 2.5(iv), we obtain (67). Because A_J is a linear subspace and hence convex, Proposition 2.9 yields the $(0, +\infty)$-joint-regularity of $\mathcal{A}$; in particular, $\mathcal{A}$ is $(\widetilde{B}, 0, 3\delta)$-joint-regular. Analogously, $\widetilde{B} = (B)$ is $(\widetilde{A}, 0, 3\delta)$-joint-regular. Now apply Theorem 2.10 to complete the proof. ∎

Remark 3.20 Some comments regarding Theorem 3.19 are in order.

(i) Note that regularity of the intersection is not an assumption of the theorem, but is rather *automatically* satisfied. This is in contrast to the results of [17] and [16] where the required regularity is assumed to hold. In view of Example 3.18, which illustrated that the notions of regularity developed in [17] and [16] are *not* satisfied, it is clear that Theorem 3.19 is new and has a genuinely wider range of applicability.

(ii) In contrast to [16] and [17], our analysis yields a quantification of the neighborhood on which local linear convergence is guaranteed.

(iii) Finding the local neighborhood on which linear convergence is guaranteed is not an easy task, and may well be tantamount of finding the sparsest solution; however, it does open the door to justify combining the MAP with more aggressive algorithms such as Douglas-Rachford in order to find such neighborhoods.

(iv) Consider again Example 3.18 and its notation. Since $s = 1$, $\mathcal{A} = (A_1, A_2, A_3)$, where $A_i = \mathbb{R}e_i$, while $B = e_1 + \mathbb{R}(e_2 - e_1)$. Hence $c(A_1, B) = c(\mathbb{R}e_1, \mathbb{R}(e_2 - e_1)) = |\langle e_1, (e_2 - e_1)/\sqrt{2}\rangle| = 1/\sqrt{2}$ by Theorem 2.6 and Corollary 2.7. Similarly, $c(A_2, B) = 1/\sqrt{2}$ while $A_3 \cap B = \varnothing$. Let $c \in \{x^*, y^*\}$. Then $\overline{\theta} = 1/\sqrt{2}$ and (66) implies that $\overline{\delta} = 1/3$. The predicted rate of linear convergence is $\overline{\theta}^2 = 1/2$.

(v) The projectors P_A and P_B given by (33) and (42) are easy to implement numerically, which we have done. Indeed, for random initial guesses b_{-1} in the neighborhood $\operatorname{ball}(c;(\sqrt{2}-1)/(18(2\sqrt{2}-1)))$ the observed ratios $\|a_{k+1}-c\|/\|a_k-c\|$ and $\|b_{k+1}-c\|/\|b_k-c\|$ for $a_k = P_A b_k$ ($k \in \mathbb{N}$, $b_0 = P_B b_{-1}$) and $b_k = P_B a_{k-1} \in B$ ($k \in \mathbb{N} \setminus \{0\}$) are $1/2 + |O(10^{-13})|$). The observed rate corresponds nicely to the theory under the assumption of exact evaluation of the projections. However, exact projections are not in fact computed in practice (in particular the projection onto the affine set B), so the numerical illustration is not precisely applicable. Inexact alternating projections is beyond the scope of this work.

Conclusion

We have applied new tools in variational analysis to the problem of finding sparse vectors in an affine subspace. The key tool is the restricted normal cone which generalizes classical normal cones. The restricted normal cones are used to define constraint qualifications, and notions of regularity that provide sufficient conditions for local convergence of iterates of the elementary method of alternating projections applied to the lower level sets of the function $\|\cdot\|_0$ and an affine set. Key ingredients were suitable restricting sets ($\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$). The coarsest choice, $(\widetilde{A}, \widetilde{B}) = (X, X)$, recovers the framework by Lewis, Luke, and Malick [16]. We show, however, that the corresponding regularity conditions are not satisfied in general for the sparse feasibility problem (2). The tighter (and hence more powerful) choice of $(\widetilde{\mathcal{A}}, \widetilde{\mathcal{B}}) = (\mathcal{A}, \mathcal{B})$ recovers local linear convergence and yields an estimate of the radius of convergence.

Acknowledgments

HHB was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. This research was initiated when HHB visited the Institut für Numerische und Angewandte Mathematik, Universität Göttingen because of his study leave in Summer 2011. HHB thanks DRL and the Institut for their hospitality. DRL was supported in part by the German Research Foundation grant SFB755-A4. HMP was partially supported by the Pacific Institute for the Mathematical Sciences and and by a University of British Columbia research grant. XW was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kuryka-Lojasiewicz inequality, *Mathematics of Operations Research* 35 (2010), 438–457.

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

[3] H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang, Restricted normal cones and the method of alternating projections, arXiv preprint, May 2012, `http://arxiv.org`

[4] H.H. Bauschke and S.G. Kruk, Reflection-projection method for convex feasibility problems with an obtuse cone, *Journal of Optimization Theory and Applications* 120 (2004), 503–531.

[5] A. Beck and M. Teboulle, A Linearly Convergent Algorithm for Solving a Class of Nonconvex/Affine Feasibility Problems, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke, R.S. Burachick, P.L. Combettes, V. Elser, D.R. Luke and H. Wolkowicz (editors), Springer, 33–48, 2011.

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

[7] J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, 2005.

[8] A.M. Bruckstein, D.L. Donoho and M. Elad, From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images, *SIAM Review* 51 (2009), 34–81.

[9] E. Candes and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Transactions on Information Theory* 52 (2006), 5406–5425.

[10] Y. Censor and S.A. Zenios, *Parallel Optimization*, Oxford University Press, 1997.

[11] 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 (1990), 487–507.

[12] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer, 2001.

[13] D.L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization, *Proceedings of the National Academy of Sciences of the USA* 100 (2003), 2197–2202.

[14] K. Friedrichs, On certain inequalities and characteristic value problems for analytic functions and for functions of two variables, *Transactions of the AMS* 41 (1937), 321–364.

[15] M.J. Lai and J. Wang, An Unconstrained ℓ_q Minimization for Sparse Solution of Underdetermined Linear Systems, *SIAM Journal of Optimization*, 21 (2010), 82–101.

[16] A.S. Lewis, D.R. Luke, and J. Malick, Local linear convergence for alternating and averaged nonconvex projections, *Foundations of Computational Mathematics* 9 (2009), 485–513.

[17] A.S. Lewis and J. Malick, Alternating projection on manifolds, *Mathematics of Operations Research* 33 (2008), 216–234.

[18] D.R. Luke, Local linear convergence and approximate projections onto regularized sets, *Nonlinear Analysis* 75 (2012), 1531–1546.

[19] D.R. Luke, Prox-regularity of Rank-constraint Sets and Implications for Algorithms, preprint arXiv:1112.0526v1, December 2011, `http://arxiv.org/abs/1112.0526`

[20] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I*, Springer-Verlag, 2006.

[21] B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM Journal on Computation*, 24(1995), 227–234.

[22] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[23] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer, corrected 3rd printing, 2009.