PROJECTION METHODS FOR SPARSE AFFINE FEASIBILITY: RESULTS AND COUNTEREXAMPLES

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 non-convex relaxations. In this work we consider elementary methods based on projections for solving a *sparse feasibility* problem without employing convex heuristics. We show that the fundamental method of alternating projections *must* converge locally linearly to a solution to the sparse feasibility problem with an affine constraint. Our analysis provides the radius of convergence and rate based on the angle of intersection of the sparsity set and the affine constraint. Under stronger assumptions on the intersection we also show local linear convergence, with radius of convergence, of the Douglas-Rachford algorithm. The stronger assumptions are in fact *necessary* for convergence of the Douglas-Rachford for affine feasibility. We show that these assumptions are *not* satisfied for most sparsity problems of interest, indicating that methods related to the Douglas-Rachford Algorithm, such as Alternating Directions Method of Multipliers, are not a promising algorithmic approach asymptotically for sparsity optimization.

Key words. non-convex feasibility, Projection Methods, Method of Alternating Projections, Douglas-Rachford Algorithm

AMS subject classifications. Primary 49J52, 49M20; Secondary 47H09, 65K05, 65K10, 90C26.

1. Introduction.

We consider the problem of sparsity optimization with affine constraints:

$$\text{minimize } \|x\|_0 \text{ subject to } Mx = p \tag{1.1}$$

where $m, n \in \mathbf{N}$, $m < n$, $M \in \mathbf{R}^{m\times n}$ is a real $m-$by$-n$ matrix and $\|x\|_0 := \sum_{j=1}^n |\operatorname{sgn} x_j|$ is the number of nonzero entries of a real vector $x \in \mathbf{R}^n$ of dimension n. Given an a priori bound $s \in \mathbf{N}$ on the desired sparsity of the solution one can relax problem (1.1) to the feasibility problem

$$\text{find } \bar{x} \in A \cap B \tag{1.2}$$

where

$$A := \{x \in \mathbf{R}^n |\ \|x\|_0 \le s\}, B := \{x \in \mathbf{R}^n |\ Mx = p\}. \tag{1.3}$$

The set B is an affine subspace, whilst A is a non-convex set. However the set A (locally) has a nice structure in the sense that one can explicitly calculate the *projection* onto the set, i.e. for a given point x one can find $\bar{x} \in A$ such that

$$\|x - \bar{x}\| = d_A(x) := \inf_{y\in A} \|x - y\|.$$

*Institut für Numerische und Angewandte Mathematik Universität Göttingen, Lotzestr. 16–18, 37083 Göttingen, Germany. E-mail: `hesse@math.uni-goettingen.de`. This author was supported by DFG grant SFB 755-C2.

†Institut für Numerische und Angewandte Mathematik Universität Göttingen, Lotzestr. 16–18, 37083 Göttingen, Germany. E-mail: `r.luke@math.uni-goettingen.de`. This author was supported by DFG grants SFB 755-C2 and GRK1023.

‡Institut für Numerische und Angewandte Mathematik Universität Göttingen, Lotzestr. 16–18, 37083 Göttingen, Germany. E-mail: `p.neumann@math.uni-goettingen.de`. This author was supported by DFG grant GRK1023.

We call the (set-valued) mapping $P_A : \mathbf{R}^n \to A$, $x \mapsto \text{argmin}_{y \in A} \textit{projector} \|x - \bar{x}\|$, while a point $\bar{x} \in P_A(x)$ is called a projection. Projection Methods as discussed in this work are easy to implement at a low cost computationally effort. We will discuss the applicability of two fundamental Projection Methods to the feasibility problem (1.2).

DEFINITION 1.1 (Method of Alternating Projections). *For two sets $A, B \subset \mathbf{E}$ we call the mapping*

$$T_{MAP}x = P_A P_B x \tag{1.4}$$

the Method of Alternating Projections operator. We call the MAP algorithm, or simply MAP, the corresponding Picard iteration,

$$x_{n+1} \in T_{MAP}x_n, \quad n = 0, 1, 2, \ldots \tag{1.5}$$

for x_0 given. □

An operator closely related to the projection is the reflection. We call the (possibly set-valued) mapping $R_A : \mathbf{R}^n \to \mathbf{R}^n$, $x \mapsto 2P_A x - x$ the reflection on A.

DEFINITION 1.2 (Averaged Alternating Reflections/Douglas Rachford). *For two sets $A, B \subset \mathbf{E}$ we call the mapping*

$$T_{AAR}x = \frac{1}{2}\left(R_A R_B x + x\right) \tag{1.6}$$

the Averaged Alternating Reflection (AAR) operator. We call the AAR algorithm, or simply AAR, the corresponding Picard iteration,

$$x_{n+1} \in T_{AAR}x_n, \quad n = 0, 1, 2, \ldots \tag{1.7}$$

for x_0 given.

Further, define the mapping

$$T_{RAAR}x = \frac{\beta}{2}\left(R_A R_B x + x\right) + (1 - \beta) P_B x \tag{1.8}$$

the Relaxed Averaged Alternating Reflection (RAAR) operator. The RAAR algorithm, or simply RAAR, is defined similarly. □

2. Algorithms for Nonconvex Feasibility - local convergence results.

2.1. Regularity of Sets.

DEFINITION 2.1 ((ε, δ)-subregularity). *A nonempty set $\Omega \subset \mathbf{E}$ is (ε, δ)-*subregular *at $\bar{x}$ with respect to $S \subset \mathbf{E}$, if there exists $\varepsilon > 0, \delta > 0$, and*

$$\langle v, z - y \rangle \leq \varepsilon \|v\| \|z - y\| \tag{2.1}$$

holds for all $y \in B_\delta(\bar{x}) \cap \Omega$, $z \in S \cap B_\delta(\bar{x})$, $v \in N_\Omega(y)$. We simply say Ω is (ε, δ)-subregular at $\bar{x}$ if $S = \{\bar{x}\}$. The definition of (ε, δ)-subregularity was introduced in [5] and is a generalization of the notion of (ε, δ)-regularity introduced in [3, Definition 9.1]. In contrast to [2, Proposition 3.13] the definition of subregularity applies to the set A for any $s \leq n$ and the convergence results of [5] can be applied directly.

2.2. Regularity of Intersections. We will state some Definitions of regularity of intersections, that provide sufficient and necessary conditions for linear convergence of both MAP and RAAR.

DEFINITION 2.2 (linear regularity).
A family of closed, nonempty sets $\Omega_1, \Omega_2, \dots, \Omega_m$ *has* locally linearly regular intersection *at* $\bar{x} \in \cap_{j=1}^m \Omega_j$ *if there exists a* $\kappa > 0$ *and a* $\delta > 0$ *such that*

$$d_{\cap_{j=1}^m \Omega_j}(x) \le \kappa \max_{i=1,\dots,m} d_{\Omega_i}(x), \quad \forall x \in B_\delta(\bar{x}). \tag{2.2}$$

If (2.2) holds for any $\delta > 0$ *the intersection is called* linearly regular*. The infimum over all* κ *such that (2.2) holds is called* regularity modulus*.* □

THEOREM 2.3 (strong regularity). *The intersection of* A, B *at* $\bar{x}$ *is* strongly regular *if*

$$N_A(\bar{x}) \cap -N_B(\bar{x}) = \{0\} \tag{2.3}$$

2.3. Linear Convergence.

PROPOSITION 2.4 ([5] Corollary 37). *Let* A, B *have* locally linearly regular intersection *at* $\bar{x} \in S := A \cap B$ *and let* A *and* B *be* $(\varepsilon, \delta)-$*subregular w.r.t.* $\bar{x}$*. For any* $x_0 \in B_\delta(\bar{x})$*, generate the sequence* $\{x_n\}_{n \in \mathbf{N}}$ *by*

$$x_{2n+1} \in P_A x_{2n} \text{ and } x_{2n+2} \in P_B x_{2n+1} \quad (\forall n = 0, 1, 2, \dots). \tag{2.4}$$

Then

$$d_S(x_{2n+2}) \le (1 - \frac{1}{\kappa^2} + \varepsilon) d_S((x_{2n}),$$

where κ *is the regularity modulus given in Definition* 2.2*.*

Proposition 2.4 is a generalization of the work [1] to the non-convex setting. Local linear regularity of the intersection $A \cap B$ is described in [4] as the *precise property equivalent to uniform linear convergence* of MAP.

[3] and [2] showed local linear convergence of MAP applied to (1.2) using a *relaxed version* of property (2.3). The framework used in that papers is a bit more restrictive, however the rate of convergence they achieve is optimal, whilst the one we discuss here is not. There still is a *gap* between this *more dual* result and the one used in [5] based on linear regularity.

PROPOSITION 2.5 ([5] Corollary 44). *Let* A, B *be two affine subspaces with* $A \cap B$*. AAR converges to a point in* $A \cap B$ *for all* $x_0 \in \mathbf{E}$ *with linear rate* $\tilde{c} < 1$ *if and only if the intersection* $A \cap B$ *is strongly regular.* This result is for our knowledge -so far- the *best* convergence result for AAR on subspaces. However the linear rate of convergence is not optimal as discussed in [5]

3. Sparse Feasibility with an Affine Constraint. In order to apply Proposition 2.4 to MAP and Proposition 2.5 to AAR for solving (1.2) we must discuss a decomposition of the set A.

The following tools can be reviewed in more details in in [2].

For $a \in \mathbf{R}^n$ define the *sparsity subspace* associated with a

$$\operatorname{supp}(a) := \{x \in \mathbf{R}^n \mid x_j = 0 \text{ if } a_j = 0\} \tag{3.1}$$

and the mapping

$$I : \mathbf{R}^n \to \{1, \dots, n\}, \quad x \mapsto \{i \in \{1, \dots, n\} |\ x_i \neq 0\} \tag{3.2}$$

Define

$$\mathcal{J} := 2^{\{1,2,\dots,n\}} \text{ and } \mathcal{J}_s := \{J \in \mathcal{J} |\ \#J = s\} \tag{3.3}$$

and note that the set A can be decomposed in

$$A = \bigcup_{J \in \mathcal{J}_s} A_J \tag{3.4}$$

where $A_J := \operatorname{span}\{e_i |\ i \in J\}$ and e_i is the $i-$th standart unit vector in $\mathbf{R}^n$.

and for $x \in \mathbf{R}^n$ the set of s largest coordinate in absolute value

$$C_s(x) := \left\{ I \in \mathcal{J}_s \,\middle|\, \min_{i \in I} x_i \geq \max_{i \notin I} x_i \right\}. \tag{3.5}$$

Proposition 3.1 ([2] Lemma 3.4).
Let $a \in A$ and assume $s \leq n-1$. Then

$$\min\{d_{A_J}(a) \mid a \notin A_J,\ J \in \mathcal{J}_s\} = \min\{|a_j| \mid j \in I(a)\}. \tag{3.6}$$

3.1. Variational Analysis.

Proposition 3.2 (Normal cones to A, [2] Theorem 3.9 and 3.15).
For $a \in A$ the (Mordukhovich) normal cone to A is

$$N_A(a) = \{\nu \in \mathbf{R}^n |\ \|\nu\|_0 \leq n-s\} \cap (\operatorname{supp}(a))^{\perp} = \bigcup_{I(a) \subseteq J \in \mathcal{J}_s} A_J^{\perp}. \tag{3.7}$$

Proposition 3.3 (normal cones to B).
For the affine set B one has for $x \in B$

$$\operatorname{par} B = \ker M, \tag{3.8}$$

where $\operatorname{par} B = B - b$ *for arbitrary b is the subspace parallel to B.*

Theorem 3.4. *At any point $\bar{x} \in A \backslash \{0\}$ the set A is $(0, \delta)$-subregular at $\bar{x}$ for* $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(c)\}$ *such. At 0 the set A is $(0, \infty)$-subregular.*

Proof. If $s = n$ the set A is all of $\mathbf{R}^n$ and the statement is trivial. For the case $s \leq n-1$, choose any $J \in \mathcal{J}_s$ and assume that $x \in B_\delta(\bar{x}) \cap A$. We show first $x \in A_J$ implies that $\bar{x} \in A_J$ by proving the contrapositive statement, namely that $\bar{x} \notin A_J$ implies that $x \notin A_J$. But this follows from the definition of δ and Proposition 3.1 as claimed. By the characterization of the normal cone in (3.7) we have $N_A(x) \subset A_J^{\perp}$ and hence, for all $x \in B_\delta(\bar{x}) \cap A$ and $\nu_x \in N_A(x)$,

$$\langle \underbrace{\nu_x}_{\in A_J^{\perp}}, \underbrace{\bar{x} - x}_{\in A_J} \rangle = 0.$$

Thus by the definition of (ε, δ)-regularity (Definition 2.1) A is $(0, \delta)$-subregular as claimed.

For $\bar{x} = 0$ by (3.7), for any $\nu_x \in N_A(x)$ $\langle \underbrace{\nu_x}_{\in \operatorname{supp}(x)^{\perp}}, \underbrace{x}_{\in \operatorname{supp}(x)} \rangle = 0$ holds. □

3.2. Projections onto A and B.

PROPOSITION 3.5 ([2] Theorem 3.5). *The projections onto the set A is given by*

$$P_A(x) = \bigcup_{J \in C_s(x)} P_{A_J}(x). \tag{3.9}$$

Note that $(P_{A_J}(x))_i = x_i, i \in J,\ (P_{A_J}(x))_i = 0, j \notin J$.

The pojection onto $B = \{x \in \mathbf{R}^n |\ Mx = p\}$ is given by

$$P_B x := x - M^\dagger(Mx - p), \tag{3.10}$$

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

3.3. Linear Convergence of MAP. We can now prove one of our main results.

THEOREM 3.6. *Let A and B be as above and let $\bar{x} \in A \cap B \neq \emptyset$. Choose $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(c)\}$. For $x \in B_\delta(\bar{x}) \cap U$, $U = \{x \mid P_A P_B x \subset B_\delta(\bar{x})\}$ the MAP iterates converge locally with linear rate.*

Proof. The intersection of two affine subspaces (finite dimension) is locally linearly regular (See for instance [4, Theorem 3.28]), so the intersection of a A (finite collection of affine subspaces) and B ist locally linearly regular. Theorem 3.4 showed that A is locally subregular, so we can apply Theorem 2.4. ◻ The rate of convergence given by the last theorem is not optimal, whilst the one achieved in [2] is and is explicitly given by the Friedrichs angle. However the result in this work, based on $(0, \delta)-$subregularity provides the convergence result more directly, omitting decompositions as in [3, 2]. Linear regularity is a more *quantitative* property and due to this fact the rate in this work is not optimal. As meantioned earlier it would be interesting to connect the results in [3, 2] and the results in a quantitative fashion [5].

3.4. Failure of DR. In contrast the the MAP algorithm, the Douglas Rachford algorithm fails in the most interesting cases, as we show next.

THEOREM 3.7. *Let A and B be as above and let $\bar{x} \in A \cap B \neq \emptyset$. Choose $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(c)\}$. For $x \in B_\delta(\bar{x}) \cap U$, $U = \{x \mid T_{AAR}x \subset B_\delta(\bar{x})\}$ the AAR iterates do not converge to $A \cap B$ if AAR does not converge if $s < \text{rank} M$.*

Proof. Since A, B are $(0, \delta)$-subregular T_{AAR} is $(0, \{\bar{x}\})$-firmly nonexpansive (see [5]) and therefore $x_+ \in B_\delta(\bar{x})$ for $x_+ \in T_{AAR}x$. Furthermore one has $x_+ = (R_{A_J} R_B x + x)/2$ for some $J \in \mathcal{C}_s$. By Proposition 2.5 AAR (applied to A_J and B) convergeces if and only if the intersection is strongly regular.

Assume that the intersection $A_J \cap B$ is strongly regular (2.3). $A_J \cap B = \{0\}$ implies

$$\begin{array}{rrl} & A_J^\perp + \dim(\ker(M)^\perp) & \leq n \\ \Leftrightarrow & n - s + \dim(\ker(M)^\perp) & \leq n \\ \Leftrightarrow & \dim(\ker(M)^\perp) & \leq s \\ \Leftrightarrow & \dim(\ker(M)) & \geq n - s \\ \Leftrightarrow & \text{rank}(M) & \leq s. \end{array}$$

This means if $s < \text{rank} M$ AAR cannot converge. Since this holds for any $x_+ \in T_{AAR}x$ the proof is complete. ◻

This gives us a lower bound for s in terms of the rank of the matrix M. Of course this condition is not very helpful when dealing with sparse solutions recovery.

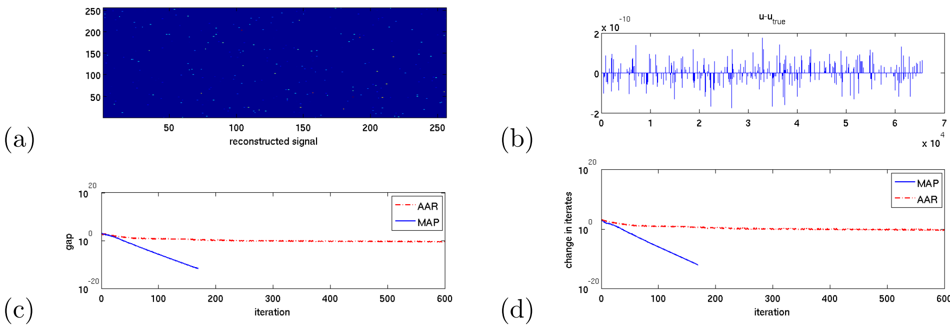

Fig. 4.1. *Comparision of MAP and AAR iteratations for sparse image reconstruction. (a) shows the reconstructed image with error shown in (b). The gap distance between the sparsity set and the affine constraint is shown in (c) and the change in the iterates is shown in (d).*

3.4.1. Counterexample for Douglas Rachford. The example given here is the same as in [2]. Define the matrix M and the measurement p to be

$$M := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \qquad p := \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The sparsest solutions to the problem "find $x \in \{v \in \mathbf{R}^3 \mid Mv = p\}$" are the points $\bar{x} = (1,0,0)^t$ and $y^* = (0,1,0)^t$. We choose a starting point $x_0 = \bar{x} + \lambda(0,0,1)^t$ with $\lambda < \frac{2-\sqrt{2}}{18(4-\sqrt{2})}$. The point x_0 is contained in the fixed point set of the AAR-algoritm. To see this, we apply AAR on x_0:

$$x_1 = T_{AAR}(x_0) = \frac{1}{2}(R_A R_B + Id)x_0 = \frac{1}{2}(x_0 + x_0) = x_0.$$

4. Numerical Demonstration. We show demonstrate the above counterexamples for more realistic applications as one might observe in practice. We construct a sparse object with 328 positive and negative point-like sources in a 256-by-256 pixel field and under-sample the Fourier transform of this object at a ratio of 1-to-8 for 8192 affine constraints. We set the sparsity set at exactly the number of nonzero elements in the original image, though in practice the sparsity of the original image is not known precisely. The problem is therefore consistent, and MAP converges locally linearly as shown in Theorem 3.6. Since the sparsity is less than the number of affine constraints, strong regularity of the intersection cannot hold. We showed in Theorem 3.7 that strong regularity is in fact *necessary* for convergence of the AAR method. Indeed, our numerical demonstration bears this out. What is quite unexpected in our numerical experiment, however, is that MAP appears to be much more robust than the theory predicts. Having a starting value within the radius of convergence as required in Theorem 3.6 is tantamount to solving the problem. Nevertheless, in our numerical experiments we have never been able to construct a starting point for which the MAP algorithm does not converge to the exact solution. This suggests that our estimates for the radius of convergence are too conservative and that further investigation of the non-convex structures will yield more information about the underlying regularity of the problem.

REFERENCES

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

[2] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and sparsity optimization with affine constraints. arXiv:1205.0320, 2012.

[3] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections. arXiv:1205.0318, 2012.

[4] F. Deutsch and H. Hundal. The rate of convergence for the cyclic projections algorithm III: Regularity of convex sets. *J. Approx. Theory*, 155(1):155–184, 2008.

[5] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. arXiv:1212.3349, 2012.