Noname manuscript No.
(will be inserted by the editor)

A note on alternating projections for ill-posed semidefinite feasibility problems

Dmitriy Drusvyatskiy · Guoyin Li · Henry Wolkowicz

Received: date / Accepted: date

Abstract We observe that Sturm's error bounds readily imply that for semidefinite feasibility problems, the method of alternating projections converges at a rate of $\mathcal{O}\left(k^{-\frac{1}{2^{d+1}-2}}\right)$, where d is the singularity degree of the problem — the minimal number of facial reduction iterations needed to induce Slater's condition. Consequently, for almost all such problems (in the sense of Lebesgue measure), alternating projections converge at a worst-case rate of $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$.

Keywords Error bounds · regularity · alternating projections · sublinear convergence · linear matrix inequality (LMI) · semi-definite program (SDP)

Mathematics Subject Classification (2010) 90C22 · 65K10 · 49M20MSC

1 Introduction

In this short note, we revisit a basic result of semi-definite programming due to Sturm [23]: denoting by $\mathcal{V}$ an affine subspace of symmetric matrices having a nonempty intersection with the positive semi-definite cone $\mathcal{S}^n_+$, the semi-definite feasibility problem

$$X \in \mathcal{V} \cap \mathcal{S}^n_+$$

always admits a *Hölder error bound*, meaning that on any compact subset U of $\mathcal{S}^n$, the distance of any putative solution $X \in U$ to the true solution set $\mathcal{V} \cap \mathcal{S}^n_+$ is bounded by a multiple of a certain power of the distance of X to the affine space $\mathcal{V}$ and to $\mathcal{S}^n_+$, separately. Most interestingly, Sturm showed that the power (*Hölder exponent*) can be set to 2^{-d}, where d is the *singularity degree* of the problem — the minimal number of *facial reduction* iterations needed to induce Slater's condition. For a discussion on facial reduction see the original work [10] or the more

Research of G. Li was supported by an ARC Future Fellowship. Research of H. Wolkowicz was supported by The Natural Sciences and Engineering Research Council of Canada and AFOSR.

D. Drusvyatskiy
Department of Mathematics, University of Washington, Seattle, WA 98195
E-mail: ddrusv@uw.edu
URL: math.washington.edu/~ddrusv

G. Li
Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia
E-mail: g.li@unsw.edu.au
URL: web.maths.unsw.edu.au/~gyli

H. Wolkowicz
Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
E-mail: hwolkowi@uwaterloo.ca
URL: orion.uwaterloo.ca/~hwolkowi

recent manuscripts [20, 26]. What is striking here is that the exponent *only* depends on the singularity degree, and not say on the size or the rank of the matrices.

In this note, we combine Sturm's error bounds with the recent work [6] to conclude that the classical method of alternating projections (that of von Neumann [25]) converges at a rate of $\mathcal{O}\left(k^{-\frac{1}{2^{d+1}-2}}\right)$, where d is the singularity degree of the problem. This result is neither very surprising, since the sublinear rate at which alternating projections converge is intimately tied to the Hölder regularity of the intersection $\mathcal{V} \cap \mathcal{S}^n_+$, nor does it advocate the use of alternating projections for ill-conditioned problems. Nonetheless, the rate is notable: contrary to the common belief that alternating projections can converge "arbitrarily slowly" for ill-posed problems, the asymptotic rate of convergence for semi-definite feasibility problems is very precise. Many problems of interest that are degenerate only due to poor modeling choices have singularity degree at most one. For such problems, the method converges at the worst-case rate $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$. Moreover, we observe that this worst-case rate is, in a precise mathematical sense, typical for semi-definite problems (even ones that are infeasible).

2 Sublinear convergence of alternating projections

Consider a Euclidean space $\mathbf{E}$ (finite-dimensional real inner product space), along with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $|\cdot|$. Given a convex set $Q \subset \mathbf{E}$, we define the *distance function*

$$\operatorname{dist}(x, Q) := \min_{y \in Q} |x - y|$$

and the *projection mapping*

$$\operatorname{proj}(x, Q) := \{y \in Q : |x - y| = \operatorname{dist}(x, Q)\}.$$

We let $\operatorname{cl} Q$, $\operatorname{ri} Q$, and $\operatorname{rb} Q$ denote the closure, relative interior, and relative boundary of Q, respectively. It is standard that if Q is closed and convex, then the set $\operatorname{proj}(x, Q)$ is a singleton; see for example [7].

Fix now two closed, convex sets A and B and consider the feasibility problem:

$$\text{find some point} \quad x \in A \cap B. \tag{2.1}$$

When working with an infeasible problem, it is useful to define the *displacement vector*, denoted $\operatorname{disp}(A, B)$, to be the minimal norm element of $\operatorname{cl}(A - B)$. Observe that $A - B$ may not be closed and therefore the closure operation may be necessary; see for example [22, Section 9] for a discussion. We say that the pair (A, B) is *weakly infeasible* if the origin lies in the closure $\operatorname{cl}(A - B)$ but not in the difference $A - B$ itself. Weak infeasibility can be a point of concern in numerical optimization since this property is difficult to detect. When the vector $\operatorname{disp}(A, B)$ is attained, meaning that $\operatorname{disp}(A, B)$ actually lies in $A - B$, we have for any points $a \in A$ and $b \in B$ the equivalence

$$a - b = \operatorname{disp}(A, B) \qquad \Longleftrightarrow \qquad a - b \in N_B(b) \quad \text{and} \quad b - a \in N_A(a).$$

Here

$$N_A(a) = \{v : \langle v, x - a \rangle \le 0 \text{ for all } x \in A\}$$

is the usual normal cone of convex analysis. See [1, Facts 1.1] for details.

This note revolves around the *method of alternating projections*, originating with von Neumann [25], for solving the feasibility problem (2.1). Given a current point $a_k \in A$, the method simply iterates the following two steps

$$\begin{aligned} &\text{choose} \quad b_k \in \operatorname{proj}_B(a_k) \\ &\text{choose} \quad a_{k+1} \in \operatorname{proj}_A(b_k). \end{aligned}$$

In studying the convergence rate of the method, the following stability property of the intersection $A \cap B$ appears naturally.

Definition 2.1 (Hölder regularity) Consider two closed, convex subsets A and B of $\mathbf{E}$. We say that the pair (A, B) *is γ-Hölder regular* if for any compact set U, there is a constant $c > 0$ so that

$$\operatorname{dist}(x, A \cap B) \leq c \cdot \Big(\operatorname{dist}^{\gamma}(x, A) + \operatorname{dist}^{\gamma}(x, B)\Big) \qquad \text{for all } x \in U.$$

We say that (A, B) *is γ-Hölder regular, up to displacement,* if the pair $\big(A - \operatorname{disp}(A, B), B\big)$ is γ-Hölder regular.

Clearly when (A, B) is γ-Hölder regular, the intersection $A \cap B$ must be nonempty. Basic convergence guarantees (with no rate) of alternating projections appear in [11]. Linear convergence under linear regularity is discussed in [1–3]. A sublinear convergence rate of alternating projections under Hölder regularity was proved in [6], in part using techniques of [1, 3] and [18, Lemmas 3, 4]. Since the result and its proof are somewhat scattered throughout the text [6], we provide a proof sketch of the salient points for the reader.

Theorem 2.2 (Convergence rate of alternating projections) *Consider two closed convex sets A and B in $\mathbf{E}$, and let $\{a_k, b_k\}$ be a sequence of iterates generated by alternating projections. Then exactly one of the following two situations holds:*

(1) *The iterates $\{a_k\}$ and $\{b_k\}$ are unbounded in norm, in which case the infimum $\inf\{|a - b| : a \in A, b \in B\}$ is not attained.*
(2) *There exist points $\bar{a} \in A$ and $\bar{b} \in B$ satisfying $a_k \to \bar{a}$ and $b_k \to \bar{b}$ and $\bar{a} - \bar{b} = \operatorname{disp}(A, B)$.*

In the second case, if the pair (A, B) is γ-Hölder regular, up to displacement, then the sequence $\{a_k, b_k\}$ converges at the sublinear rate

$$\max\{|a_k - \bar{a}|, |b_k - \bar{b}|\} = \mathcal{O}\Big(k^{-\frac{1}{2\gamma^{-1}-2}}\Big). \tag{2.2}$$

Moreover, if the pair (A, B) is linearly regular ($\gamma = 1$), up to displacement, then the convergence is R-linear.

Proof. The fact that only the two claimed situations can hold is well-known; see for example [1, Facts 1.2]. Suppose now that the second alternative holds, and define $v := \operatorname{disp}(A, B) = \bar{a} - \bar{b}$. Suppose also that the pair $(A - v, B)$ admits a γ-Hölderian error bound. Define for convenience $A_v := A - v$. Then a short computation (see [6, Middle of the proof of Theorem 4.10]) shows

$$\operatorname{dist}^2(b_k, A_v) \leq \operatorname{dist}^2(b_k, A_v \cap B) - \operatorname{dist}^2(b_{k+1}, A_v \cap B). \tag{2.3}$$

Taking also into account that the pair (A_v, B) is γ-Hölder regular, we deduce that there exists a constant c so that

$$\begin{aligned} c^{-2\gamma^{-1}} \cdot \operatorname{dist}^{2\gamma^{-1}}(b_k, A_v \cap B) &\leq \operatorname{dist}^2(b_k, A_v) \\ &\leq \operatorname{dist}^2(b_k, A_v \cap B) - \operatorname{dist}^2(b_{k+1}, A_v \cap B). \end{aligned} \tag{2.4}$$

Thus the constants $\beta_k := \operatorname{dist}^2(b_k, A_v \cap B)$ satisfy the recursion

$$\beta_{k+1} \leq \beta_k\big(1 - \frac{1}{c^{2\gamma^{-1}}}\beta_k^{\gamma^{-1}-1}\big). \tag{2.5}$$

Then by [6, Lemma 4.1], the constants β_k satisfy

$$\beta_k = \mathcal{O}\Big((\delta + k)^{-\frac{1}{\gamma^{-1}-1}}\Big),$$

for some δ. In the case $\gamma < 1$, the additive term δ can clearly be set to zero. On the other hand, [1, Example 3.2] shows that $(b_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $A_v \cap B$, and therefore by the standard estimate [1, Theorem 3.3(iv)], we have

$$|b_k - \bar{b}| \leq 2 \operatorname{dist}(b_k, A_v \cap B) = \mathcal{O}\Big(k^{-\frac{1}{2\gamma^{-1}-2}}\Big).$$

In the case $\gamma = 1$, inequality (2.5) shows a geometric decay in β_k. Appealing to [1, Theorem 3.3(iv)] again, linear convergence follows. ☐

We next turn to the singularity degree of set intersections – a term coined by Sturm [23] and rooted in [10]. From now on, we will exclusively consider the problem

$$\text{find some point} \quad x \in \mathcal{V} \cap \mathcal{K}, \tag{2.6}$$

where $\mathcal{V}$ is an affine subspace of $\mathbf{E}$ and $\mathcal{K}$ is a closed convex cone. Assume that this problem is feasible and that $\mathcal{K}$ has a nonempty interior (for simplicity). We then say that the *Slater condition* holds if $\mathcal{V}$ meets the interior of $\mathcal{K}$. Whenever the Slater condition fails, one would like to detect this pathology and to regularize the problem somehow. With this in mind, Borwein and Wolkowicz [10] introduced the following procedure, called *facial reduction*, to successively embed the problem (2.6) in a smaller dimensional space, relative to which the Slater condition does hold. Here, we describe the conceptual mechanics of the procedure; a rigorous numerical study of facial reduction appears in [12]. To this end, consider some representation

$$\mathcal{V} = \{x : \mathcal{A}(x) = b\}$$

for some linear mapping $\mathcal{A} \colon \mathbf{E} \to \mathbf{R}^m$ and for some vector $b \in \mathbf{R}^m$. Then the first iteration of facial reduction consists of solving the *auxiliary problem*: find $y \in \mathbf{R}^m$ satisfying

$$0 \neq \mathcal{A}^* y \in \mathcal{K}^* \quad \text{and} \quad \langle y, b \rangle = 0, \tag{2.7}$$

where $\mathcal{A}^*$ denotes the adjoint and $\mathcal{K}^* = \{z : \langle z, x\rangle \geq 0 \text{ for all } x \in K\}$ is the dual cone. The auxiliary problem is feasible if and only if the Slater condition fails. Supposing the latter, let y solve the auxiliary problem (2.7). Then the entire feasible region $\mathcal{V} \cap \mathcal{K}$ is contained in the slice $\mathcal{K} \cap (\mathcal{A}^* y)^{\perp}$. We now replace $\mathcal{K}$ with $\mathcal{K} \cap (\mathcal{A}^* y)^{\perp}$ and $\mathbf{E}$ with the linear span of $\mathcal{K} \cap (\mathcal{A}^* y)^{\perp}$, and repeat the procedure. The minimal number of facial reduction iterations needed to obtain a problem satisfying the Slater condition is the *singularity degree of the pair* $(\mathcal{V}, \mathcal{K})$. In general, the singularity degree is no greater than $n-1$, and there are problems that require exactly $n-1$ facial reduction iterations [24, Section 2.6] — a property closely tied to possible failure of strict complementarity for SDP [13, Section 4.4.2].

It will be convenient to extend the definition of singularity degree to situations where $\mathcal{V}$ and $\mathcal{K}$ may not intersect. To this end, when the displacement vector $\operatorname{disp}(\mathcal{V}, \mathcal{K})$ is attained, the translated affine subspace $\mathcal{V} - \operatorname{disp}(\mathcal{V}, \mathcal{K})$ and the cone $\mathcal{K}$ do intersect and we define the *singularity degree of* $(\mathcal{V}, \mathcal{K})$*, with displacement*, to be the singularity degree of the pair $\big(\mathcal{V} - \operatorname{disp}(\mathcal{V}, \mathcal{K}), \mathcal{K}\big)$. When $\operatorname{disp}(\mathcal{V}, \mathcal{K})$ is unattained, we say that the singularity degree of $(\mathcal{V}, \mathcal{K})$, with displacement, is $+\infty$.

An important instance of (2.6), and one that we focus on, is the semi-definite feasibility problem:

$$\text{find some matrix} \quad X \in \mathcal{V} \cap \mathcal{S}^n_+,$$

where $\mathcal{V}$ is an affine subspace of the Euclidean space of $n \times n$-symmetric matrices $\mathcal{S}^n$ and $\mathcal{S}^n_+$ is the convex cone of $n \times n$ positive semi-definite matrices. We will always endow $\mathcal{S}^n$ with the trace inner product $\langle X, Y \rangle = \operatorname{tr} XY$ and the Frobenius norm $\|X\| = \sqrt{\langle X, X\rangle}$. In [23], Jos F. Sturm discovered a surprising connection between Hölder regularity and singularity degree in the semi-definite feasibility problem.

Theorem 2.3 (Sturm's error bounds for SDP) *Given an affine subspace $\mathcal{V}$ of $\mathcal{S}^n$, the pair $(\mathcal{V}, \mathcal{S}^n_+)$ is $\frac{1}{2^d}$−Hölder regular, with displacement, where d is the singularity degree of $(\mathcal{V}, \mathcal{S}^n_+)$, with displacement.*

Combining Sturm's result with Theorem 2.2, we immediately deduce the main contribution of this section.

Theorem 2.4 (Convergence rate of alternating projections for SDP) *Given an affine subspace $\mathcal{V}$ of $\mathcal{S}^n$, consider the semi-definite feasibility problem:*

$$\textit{find some matrix} \quad X \in \mathcal{V} \cap \mathcal{S}^n_+.$$

Letting $\{X_k, Y_k\}$ be the sequence of iterates generated by the method of alternating projections, exactly one of the following two situations holds:

(1) *The iterates* $\{X_k\}$ *and* $\{Y_k\}$ *are unbounded in norm, in which case the displacement vector* $\mathrm{disp}(\mathcal{V}, \mathcal{S}^n_+)$ *is not attained.*
(2) *There exist matrices* $\bar{X}$ *and* $\bar{Y}$ *satisfying* $X_k \to \bar{X}$ *and* $Y_k \to \bar{Y}$*, with* $\bar{X} - \bar{Y} = \mathrm{disp}(\mathcal{V}, \mathcal{S}^n_+)$.

In the second case, the iterates $\{X_k, Y_k\}$ *converge at a rate* $\mathcal{O}\Big(k^{-\frac{1}{2^{d+1}-2}}\Big)$*, where* d *is the singularity degree of the pair* $(\mathcal{V}, \mathcal{S}^n_+)$*, with displacement. Moreover, if the problem is feasible and the Slater condition holds, then the convergence is R-linear.*

3 Typical singularity degree and convergence of alternating projections for SDP

Consider the feasibility problem

$$\text{find some point} \quad x \in \mathcal{K} \cap \{x \in \mathbf{E} : \mathcal{A}(x) = b\},$$

where $\mathcal{K}$ is a closed convex cone and $\mathcal{A} \colon \mathbf{E} \to \mathbf{R}^m$ is a linear mapping. It is well known that the Slater condition holds for "typical" parameters $(\mathcal{A}, b)$ among all parameters $(\mathcal{A}, b)$ for which the problem is feasible. For a discussion of various generic properties of such problems, see for example [5, 16, 21]. In this brief section, in contrast, we consider the more realistic setting of when perturbations in parameters can yield an infeasible problem, with an eye towards the singularity degree.

We first note that the displacement vector of the problem is typically attained. Indeed, this is a direct consequence of [9]. From now on, all references to a measure on $\mathbf{E}$ will refer specifically to the Lebesgue measure on $\mathbf{E}$.

Proposition 3.1 (Displacement vector is typically attained) *Consider a closed, convex cone* $\mathcal{K} \subset \mathbf{E}$ *and a vector* $b \in \mathbf{R}^m$*. Then for an open, full-measure set of linear transformations* $\mathcal{A} \colon \mathbf{E} \to \mathbf{R}^m$*, the infimum*

$$\inf\{|x - y| : x \in \mathcal{K} \quad \textit{and} \quad \mathcal{A}(y) = b\}$$

is attained.

Proof. Define

$$\mathcal{L}_{\mathcal{A},b} = \{x \in \mathbf{E} : \mathcal{A}(x) = b\},$$

and $v := \mathrm{disp}(\mathcal{L}_{\mathcal{A},b}, \mathcal{K})$. Then the set $\mathcal{L}_{\mathcal{A},b} - v - \mathcal{K}$ is closed if and only if $(\ker \mathcal{A}) - \mathcal{K}$ is closed. A well-known theorem of Abrams (see for example [4, Lemma 3.1] or [19, Lemma 17H]) states that the latter holds if and only if the image $\mathcal{A}(\mathcal{K})$ is closed. On the other hand [8,9] show that the image $\mathcal{A}(\mathcal{K})$ is closed for some open, full-measure set of transformations $\mathcal{A}$. □

Next, we observe that though we cannot typically expect Slater's condition to hold (or feasibility to hold for that matter), the singularity degree of the problem, with displacement, is usually at most one. To this end, consider a closed, convex cone $\mathcal{K} \subset \mathbf{E}$ and define the affine space $\mathcal{V} := \{x \in \mathbf{E} : \mathcal{A}(x) = b\}$. Then the "strict complementarity" condition:

$$\text{there exist } x \in \mathcal{V} \cap \mathcal{K} \text{ and } y \in \mathbf{R}^m \text{ satisfying } 0 \neq \mathcal{A}^* y \in \mathrm{ri}\, N_{\mathcal{K}}(x).$$

is sufficient for the singularity degree of $(\mathcal{V}, \mathcal{K})$ to be one, provided that $\mathcal{K}$ is facially exposed (see [22, Section 18] for the definition). Indeed, if such x and y exist, then standard convex analysis shows that y solves the auxiliary problem (2.7), and moreover that $\mathcal{K} \cap (\mathcal{A}^* y)^\perp$ is the minimal exposed face of $\mathcal{K}$ containing x; see e.g. [17, Theorem A.2]. Hence, in particular, x lies in the relative interior of $\mathcal{K} \cap (\mathcal{A}^* y)^\perp$ and the Slater condition will hold for the second iteration. We are now ready to prove the main result.

Proposition 3.2 (Singularity degree is typically at most one)
Consider a closed, facially exposed, convex cone $\mathcal{K} \subset \mathbf{E}$ *and a vector* $b \in \mathbf{R}^m$*. Then for a dense set of linear transformations* $\mathcal{A} \colon \mathbf{E} \to \mathbf{R}^m$*, the feasibility system*

$$\mathcal{K} \cap \{x : \mathcal{A}(x) = b\}$$

has singularity degree, with displacement, of at most one.

Proof. Define
$$\mathcal{L}_{\mathcal{A},b} = \{x \in \mathbf{E} : \mathcal{A}(x) = b\},$$
and $v := \mathrm{disp}(\mathcal{L}_{\mathcal{A},b}, \mathcal{K})$. By Proposition 3.1, we may assume that the displacement vector v is attained, that is we may write $v = y - x$ for some $y \in \mathcal{L}_{\mathcal{A},b}$ and some $x \in \mathcal{K}$. Clearly we may suppose that the Slater condition fails, since otherwise $\mathcal{A}$ would certainly lie in the claimed dense set. Moreover, we can assume $v \neq 0$, since the complement of the set of matrices $\mathcal{A}$ for which the problem is feasible but Slater condition fails is a dense set.

Observe that v lies in $N_{\mathcal{K}}(x) \cap \mathrm{rge}\,\mathcal{A}^*$. Consequently if v actually lies in $\mathrm{ri}\, N_{\mathcal{K}}(x)$, then the pair $(\mathcal{L}_{\mathcal{A},b} - v, \mathcal{K})$ has singularity degree at most one. Suppose this is not the case, that is we have $v \in \mathrm{rb}\, N_{\mathcal{K}}(x)$. Choose then an arbitrary vector $w \in \mathrm{ri}\, N_{\mathcal{K}}(x)$ and define an orthogonal transformation $U : \mathbf{E} \to \mathbf{E}$, whose restriction to $\mathrm{span}\,\{v, w\}$ is a rotation sending v to w, and whose restriction to $\mathrm{span}\,\{v, w\}^{\perp}$ coincides with the identity mapping. Define the linear transformation $\widehat{\mathcal{A}} := \mathcal{A} \circ U^T$ and a point $\hat{y} := Uy$. Consider the perturbed system
$$\mathcal{K} \cap \{x : \widehat{\mathcal{A}}(x) = b\}. \tag{3.1}$$
The following properties are then easy to verify:
$$\hat{y} \in \mathcal{L}_{\widehat{\mathcal{A}},b}, \quad Ux = x, \quad \mathrm{rge}\,\widehat{\mathcal{A}}^* = U(\mathrm{rge}\,\mathcal{A}^*).$$
Observe
$$\hat{y} - x = U(y - x) = Uv = w.$$
Consequently, the inclusion $w \in \big(\mathrm{ri}\, N_{\mathcal{K}}(x)\big) \cap \mathrm{rge}\,\widehat{\mathcal{A}}^*$ holds. We deduce that the pair $(\mathcal{L}_{\widehat{\mathcal{A}},b}, \mathcal{K})$ has singularity degree, with displacement, of at most one. Letting w tend to v, the matrices $\widehat{\mathcal{A}}$ tend to $\mathcal{A}$, and the result follows. □

When $\mathcal{V}$ is an affine subspace and the cone $\mathcal{K}$ is semi-algebraic (e.g., the positive semi-definite cone) – meaning that it can be written as a union of finitely many sets, each defined by finitely many polynomial inequalities – basic quantifier elimination (see [14, Section 2.1.2]) shows that the set of transformations $\mathcal{A}$ for which the singularity degree, with displacement, of the pair $(\mathcal{V}, \mathcal{K})$ is at most one, is a semi-algebraic set. On the other hand, it is a direct consequence of existence of semi-algebraic stratifications [14, Section 2.3] that any dense semi-algebraic set necessarily contains an open full-measure set; see for example [15, Section 3] for a discussion. Note that this is in stark contrast to the non semi-algebraic setting. Thus when $\mathcal{K}$ is semi-algebraic (e.g., for $\mathcal{K} = \mathcal{S}^n_+$), the dense set of transformations $\mathcal{A}$ in Proposition 3.2 is actually an open, full-measure set. In particular, the following typical behavior of alternating projections is now immediate.

Corollary 3.3 (Generic convergence of alternating projections in SDP)
For a full-measure set of linear transformations $\mathcal{A} : \mathcal{S}^n \to \mathbf{R}^m$, the semi-definite program (for any $b \in \mathbf{R}^m$):
$$\mathcal{S}^n_+ \cap \{X : \mathcal{A}(X) = b\}$$
has singularity degree, with displacement, of at most one, and consequently the iterates generated by alternating projections converge at the rate of $\mathcal{O}\Big(\frac{1}{\sqrt{k}}\Big)$.

We end this section with a numerical experiment illustrating the above result and the interplay between alternating projections and error bounds, more broadly. Setting the groundwork, fix two intersecting closed convex sets A and B in $\mathbf{E}$. Consider now three successive iterates of alternating projections, namely points b, a^+, and b^+ satisfying $b \in B$, $a^+ = P_A(b)$, and $b^+ = P_B(a^+)$. We will use the following distance estimate; for completeness, we provide a quick proof.

Lemma 3.4 (Distance to the intersection) *Suppose $a^+ \neq b^+$. Then the inequality*
$$\mathrm{dist}(a^+; A \cap B) \geq \sin^{-1}(\theta) \cdot |b^+ - a^+|$$
holds, where θ is the angle between the two vectors $b - a^+$ and $b^+ - a^+$.

Proof. Clearly we can assume that the intersection $A \cap B$ is nonempty. Since by definition we have $a^+ = P_A(b)$ and $b^+ = P_B(a^+)$, the intersection $A \cap B$ lies in the polyhedron

$$L := \{x \in \mathbf{E} : \langle x - a^+, b - a^+ \rangle \le 0 \quad \text{and} \quad \langle x - b^+, a^+ - b^+ \rangle \le 0\}.$$

Let x be the projection of a^+ onto L, and note $x \neq a^+$, since otherwise from the definition of L we would deduce $a^+ = b^+$, a contradiction. Observe now the inequality $\mathrm{dist}(a^+; A \cap B) \ge |x - a^+|$. Writing the optimality conditions for the projection problem onto L, we deduce there are some real numbers $\lambda, \mu \ge 0$ satisfying the stationarity and complementarity conditions:

$$a^+ - x = \lambda(b - a^+) + \mu(a^+ - b^+),$$

$$0 = \lambda\langle x - a^+, b - a^+ \rangle \qquad \text{and} \qquad 0 = \mu\langle x - b^+, a^+ - b^+ \rangle.$$

Multiplying the first equation through by $\lambda(b - a^+)$, we obtain

$$0 = \lambda^2 |b - a^+|^2 + \lambda\mu\langle a^+ - b^+, b - a^+ \rangle. \tag{3.2}$$

Similarly multiplying through by $\mu(a^+ - b^+)$, we obtain

$$\begin{aligned} \mu|a^+ - b^+|^2 &= \langle (a^+ - b^+) + (b^+ - x), \mu(a^+ - b^+) \rangle \\ &= \lambda\mu\langle b - a^+, a^+ - b^+ \rangle + \mu^2 |a^+ - b^+|^2. \end{aligned} \tag{3.3}$$

We claim $\lambda \neq 0$. To see this, suppose otherwise $\lambda = 0$. The from (3.3), we deduce $\mu = 0$ or $\mu = 1$. The first case is impossible since we would deduce $a^+ = x$. Hence supposing $\mu = 1$, we deduce $x = b^+$. Since then both b and b^+ lie in L, the inequalities $\langle b^+ - a^+, b - a^+ \rangle \le 0$ and $\langle b - b^+, a^+ - b^+ \rangle \le 0$ must hold. Adding the two inequalities, we obtain the contradiction $a^+ = b^+$. Thus we conclude $\lambda \neq 0$.

Clearly μ is nonzero, since otherwise by (3.2) we would deduce $b = a^+$. Solving equations (3.2) and (3.3) for μ and λ, we obtain

$$\mu = \sin^{-2}(\theta) \quad \text{and} \quad \lambda = \sin^{-2}(\theta)\cos(\theta)\frac{|a^+ - b^+|}{|b - a^+|}.$$

Finally taking into account that $x - a^+$ and $b - a^+$ are orthogonal, we deduce

$$|x - a^+|^2 = \mu^2 |a^+ - b^+|^2 - \lambda^2 |b - a^+|^2 = \sin^{-2}(\theta)|a^+ - b^+|^2,$$

as claimed. □

Suppose that the pair (A, B) is γ-Hölder regular and that the alternating projection method is initiated at some $a_0 \in A$ and does not terminate finitely. Then Lemma 3.4 shows that there is some constant $c > 0$ so that all the iterates satisfy the inequality

$$\frac{|b^+ - a^+|}{\sin(\theta)} \le \mathrm{dist}(a^+; A \cap B) \le c \cdot \mathrm{dist}^\gamma(a^+; B) \le c \cdot |a^+ - b^+|^\gamma,$$

and hence

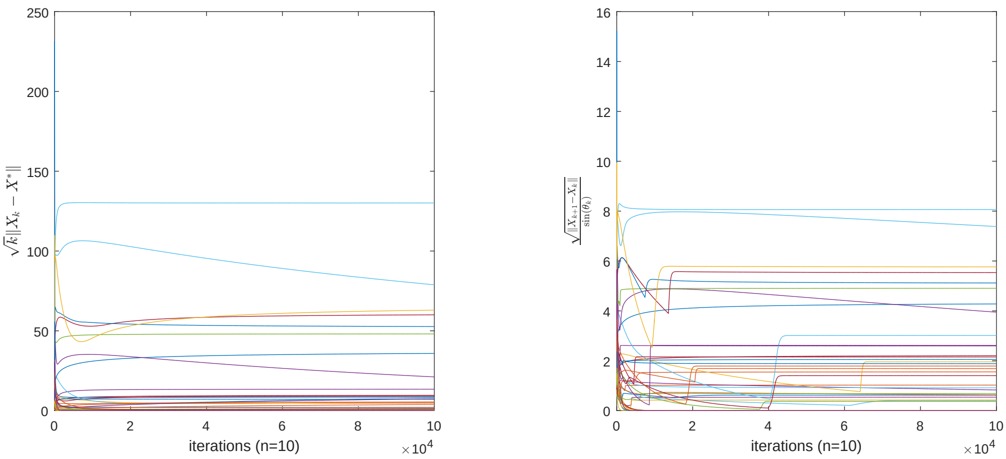
$$\frac{|b^+ - a^+|^{1-\gamma}}{\sin(\theta)} \le c, \tag{3.4}$$

Here θ is the angle between the two vectors $b - a^+$ and $b^+ - a^+$. Tracking these quotients provides a convenient way of testing Hölder error bounds.

We now construct a "generic" SDP feasible region

$$\{X \in \mathcal{S}^{10}_+ : \langle A_i, X \rangle = b \text{ for } i = 1, \dots, 40\}.$$

Namely we generate a vector $b \in \mathbf{R}^{40}$ and matrices $A_i \in \mathcal{S}^{10}$ for $i = 1, \dots, 40$ according to a standard normal distribution. We compute the displacement vector (after the projection algorithms terminate) and translate the affine constraints to achieve feasibility. We then run alternating projections on the resulting instance (starting with a random starting point), plotting the quantities $\sqrt{k}\|X_k - X^*\|$ and the quotients (3.4) with $\gamma = 1/2$. Repeating this experiment 40 times, we plot the results in the figure below.



The figure provides compelling evidence that the curves are bounded, verifying that the singularity degree with displacement is generically one. Moreover, many of the curves in the figure on the right are bounded away from zero, providing convincing evidence that the Hölder exponent for those instances is exactly 1/2.

Acknowledgments

We thank Levent Tunçel for insightful discussions, and the two anonymous referees for their comments and suggestions.

References

1. H.H. Bauschke and J.M. Borwein. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Anal.*, 1(2):185–212, 1993.
2. H.H. Bauschke and J.M. Borwein. Dykstra's alternating projection algorithm for two sets. *J. Approx. Theory*, 79(3):418–443, 1994.
3. H.H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.
4. A. Berman. *Cones, matrices and mathematical programming*. Springer-Verlag, Berlin-New York, 1973. Lecture Notes in Economics and Mathematical Systems, Vol. 79.
5. J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semialgebraic convex programs. *Math. Oper. Res.*, 36(1):55–70, 2011.
6. J. Borwein, G. Li, and L. Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, 24(1):498–527, 2014.
7. J.M. Borwein. Proximality and Chebyshev sets. *Optimization Letters*, 1(1):21–32, 2006.
8. J.M. Borwein and W.B. Moors. Stability of closedness of convex cones under linear mappings. *J. Convex Anal.*, 16(3-4):699–705, 2009.
9. J.M. Borwein and W.B. Moors. Stability of closedness of convex cones under linear mappings II. *J. Nonlinear Anal. Optim.*, 1(1):1–7, 2010.
10. J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81.
11. L.M. Bregman. The method of successive projection for finding a common point of convex sets. *Sov. Math. Dokl*, 6:688–692, 1965.
12. Y.-L. Cheung, S. Schurr, and H. Wolkowicz. Preprocessing and regularization for degenerate semidefinite programs. In *Computational and analytical mathematics*, volume 50 of *Springer Proc. Math. Stat.*, pages 251–303. Springer, New York, 2013.
13. Y.L. Cheung. *Preprocessing and Reduction for Semidefinite Programming Via Facial Reduction: Theory and Practice*. Ph.D. Thesis, University of Waterloo, 2013.

14. M. Coste. *An Introduction to Semialgebraic Geometry.* RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.
15. D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Generic Minimizing Behavior in Semialgebraic Optimization. *SIAM J. Optim.*, 26(1):513–534, 2016.
16. D. Drusvyatskiy and A.S. Lewis. Generic nondegeneracy in convex optimization. *Proc. Amer. Math. Soc.*, 139(7):2519–2527, 2011.
17. D. Drusvyatskiy, S.A. Vavasis, and H. Wolkowicz. Extreme point inequalities and geometry of the rank sparsity ball. *arXiv:1401.4774*, 2014.
18. L.G. Gubin, B.T. Polyak, and E.V. Raik. The method of projections for finding the common point of convex sets. *USSR Computational Mathematics and Mathematical Physics*, 7(6):1–24, 1967.
19. R.B. Holmes. *Geometric functional analysis and its applications.* Springer-Verlag, New York-Heidelberg, 1975. Graduate Texts in Mathematics, No. 24.
20. G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In *Computational and analytical mathematics*, volume 50 of *Springer Proc. Math. Stat.*, pages 613–634. Springer, New York, 2013.
21. G. Pataki and L. Tunçel. On the generic properties of convex optimization problems in conic form. *Math. Program.*, 89(3, Ser. A):449–457, 2001.
22. R.T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
23. J.F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248 (electronic), 2000.
24. L. Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs.* American Mathematical Society, Providence, RI; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2010.
25. J. von Neumann. *Functional Operators. II. The Geometry of Orthogonal Spaces.* Annals of Mathematics Studies, no. 22. Princeton University Press, Princeton, N. J., 1950.
26. H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013.