Alternating projections and Sturm's error bounds in semidefinite programming

Dmitriy Drusvyatskiy* Guoyin Li† Henry Wolkowicz‡

September 16, 2014

Abstract

We observe that Sturm's error bounds readily imply that for semidefinite programs, 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 (in the sense of Lebesgue measure) semi-definite programs, alternating projections converge at a worst-case sublinear rate $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$. We then use alternating projections to numerically test whether Sturm's error bounds are tight.

Key words: Error bounds, regularity, alternating projections, sublinear convergence, linear matrix inequality, semi-definite program

AMS 2000 Subject Classification: 90C22, 65K10, 49M20

Contents

*Department of Mathematics, University of Washington, Seattle, WA 98195; Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; `people.orie.cornell.edu/dd379`. Research supported by AFOSR.

†Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia; `web.maths.unsw.edu.au/~gyli`. Research supported by an ARC Future Fellowship.

‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; `orion.uwaterloo.ca/~hwolkowi`. Research supported by The Natural Sciences and Engineering Research Council of Canada and AFOSR.

List of Figures

1 Introduction

HW ♣ Should we add a comment early about ???? ... linear rate for subspaces? e.g. Aronszajn result that convergence rate is linear with rate $(\cos\theta)^2$; $\leq (\cos\theta)^{2k-1}\|P_{psd}\|$ and convergence is linear in convex case provided intersection of rel-int nonempty ... B-B '93????

In this work, we revisit a basic result of semi-definite programming due to Sturm [20]: 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 $X \in U$ from the true solution set $\mathcal{V} \cap \mathcal{S}^n_+$ is bounded by a multiple of a certain power of the distance of X from the affine space $\mathcal{V}$ and from $\mathcal{S}^n_+$, separately. Most interestingly, Sturm showed that the power (*Hölder exponent*) can be set to $\frac{1}{2^d}$, 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 [9] or the more recent manuscripts [17,22]. 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 short 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 [21]) 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 is not very surprising, since the sublinear rate at which alternating projections converge clearly has to do with the Hölder regularity of the intersection $\mathcal{V} \cap \mathcal{S}^n_+$, a fact made precise in [6]. Nonetheless, this is notable: contrary to the usual mantra that alternating projections converge "arbitrarily slowly", for semi-definite feasibility problems the rate of convergence is very specific. Many problems of interest, especially those that are degenerate only due to poor modeling choices, have singularity degree at most one. Consequently, for such problems, the method converges at the worst-case rate $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$. Moreover, we will show that this worst-case rate is, in a precise mathematical sense, typical for semi-definite problems (even ones that are infeasible). Coming back full circle, we complete the paper by using alternating projections to numerically test whether Sturm's error bounds are tight.

The outline of the paper is as follows. In Section 2, we discuss convergence guarantees of the method of alternating projections, while in Section 3, we observe that typically the singularity degree is no more than one. Finally in Section 4, we use alternating projections to numerically test whether Sturm's error bounds are tight.

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*

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

and the *projection mapping*

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

We let $\mathrm{cl}\, Q$, $\mathrm{ri}\, Q$, and $\mathrm{rb}\, Q$ denote the closure, relative interior, and relative boundary of Q, respectively. It is standard that if Q is closed, then $\mathrm{proj}(x, Q)$ is a singleton.

For the rest of the paper, we will consider two closed, convex sets A and B and focus on 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 $\mathrm{disp}(A, B)$, to be the minimal norm element of $\mathrm{cl}\,(A - B)$. Observe that $A - B$ may not be closed (e.g., the difference of the Lorentz cone $\{(x, y, r) : |(x, y)| \le r\}$ and a span of one of its bounding rays), and therefore the closure operation may be necessary. We say that the pair (A, B) is *weakly infeasible* if the origin lies in the closure $\mathrm{cl}\,(A - B)$ 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 $\mathrm{disp}(A, B)$ is attained, meaning that $\mathrm{disp}(A, B)$ actually lies in $A - B$, we have for any points $a \in A$ and $b \in B$ the equivalence

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

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* for solving the feasibility problem (2.1). Given a current point $a_k \in A$, the method simply iterates the following two steps

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

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

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

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

HW ♣ ???why clearly???attained????can we add - “by compactness assumption” ???

Clearly when (A, B) is γ-Hölder regular, the intersection $A \cap B$ must be nonempty. Moreover, if the pair (A, B) is γ-Hölder regular, up to displacement, the displacement vector $\mathrm{disp}(A, B)$ must be attained by some points $a \in A$ and $b \in B$.

Basic convergence guarantees (with no rate) of alternating projections appear in [10].

HW ♣ should a reference to classical Von Neumann result be added????

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 [3] and [15, Lemmas 3 and 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 $\bar{a} - \bar{b} = \mathrm{disp}(A, B)$.*

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 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 $v := \mathrm{disp}(A, B) = \bar{a} - \bar{b}$. 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

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

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

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

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

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

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

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

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 [20] and rooted in [9]. 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. We will 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 and to somehow regularize the problem. With this in mind, Borwein and Wolkowicz [9] 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 give an easy conceptual overview; a rigorous numerical study of facial reduction appears in [11]. 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 polar cone. The auxiliary problem is feasible if and only if Slater fails. Supposing the latter, let y solve the auxiliary problem. 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

> HW ♣ should we say something on how the definition of $\mathcal{A}$ changes? is it just the restriction to the new span? I forget the details but I think there was some subtlety to defining the new restriction.

repeat the procedure. The minimal number of facial reduction iterations needed to end up with a problem satisfying Slater's condition is the *singularity degree of the pair* $(\mathcal{V}, \mathcal{K})$.

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 $\mathrm{disp}(\mathcal{V},\mathcal{K})$ is attained, the translated affine subspace $\mathcal{V}-\mathrm{disp}(\mathcal{V},\mathcal{K})$ 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 $\big(\mathcal{V}-\mathrm{disp}(\mathcal{V},\mathcal{K}),\mathcal{K}\big)$. When $\mathrm{disp}(\mathcal{V},\mathcal{K})$ is unattained, we say that the singularity degree of $(\mathcal{V},\mathcal{K})$, with displacement, is equal to $+\infty$.

We next turn to 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 $\langle X,Y\rangle = \mathrm{tr}\,XY$ and the Frobenius norm $\|X\|=\sqrt{\langle X,X\rangle}$. In [20], 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 2.4, 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 Slater's condition holds, then the convergence is 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 Slater's 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, 13, 18]. In this brief section, in contrast, we consider the more realistic situation of when perturbations in parameters can yield an infeasible problem, with an eye towards the singularity degree.

We first note that displacement vector of the problem is typically attained. Indeed, this is a direct consequence of [8]. 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 usually 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 $v := \operatorname{disp}(\mathcal{L}_{\mathcal{A},b}, \mathcal{K})$. 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 [16, Lemma 17H]) states that the latter holds if and only if the image $\mathcal{A}(\mathcal{K})$ is closed. On the other hand, in [7, 8] the authors show that the image $\mathcal{A}(\mathcal{S}^n_+)$ 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 $\mathcal{K} \in \mathbf{E}$ and define the affine space $\mathcal{V} := \{x \in \mathbf{E} : \mathcal{A}(x) = b\}$. Then the "strict complementarity" condition:

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

is sufficient for the singularity degree of $(\mathcal{V}, \mathcal{K})$ to be one, provided that $\mathcal{K}$ is facially exposed (see [19, 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 [14, Theorem A.2]. Hence, in particular, x lies in the relative interior of $\mathcal{K}\cap(\mathcal{A}^*y)^\perp$ and 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 $v := \operatorname{disp}(\mathcal{L}_{\mathcal{A},b}, \mathcal{K})$. 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 fails is a dense set.

Observe that v lies in $N_{\mathcal{K}}(x) \cap \operatorname{rge} \mathcal{A}^*$. Consequently if v actually lies in $\operatorname{ri} N_{\mathcal{K}}(x)$, 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 $v \in \operatorname{rb} N_{\mathcal{K}}(x)$. Choose then an arbitrary vector $w \in \operatorname{ri} N_{\mathcal{K}}(x)$ and define an orthogonal transformation $U\colon \mathbf{E} \to \mathbf{E}$, whose restriction to $\operatorname{span}\{v, w\}$ is a rotation sending v to w, and whose restriction to $\operatorname{span}\{v, w\}^\perp$ 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

$$\mathcal{K} \cap \{x : \widehat{\mathcal{A}}(x) = b\}. \tag{3.8}$$

The following properties are then easy to verify:

$$\hat{y} \in \mathcal{L}_{\widehat{\mathcal{A}},b}, \quad Ux = x, \quad \operatorname{rge} \widehat{\mathcal{A}}^* = U(\operatorname{rge} \mathcal{A}^*).$$

Observe

$$\hat{y} - x = U(y - x) = Uv = w.$$

Consequently, the inclusion $w \in \big(\operatorname{ri} N_{\mathcal{K}}(x)\big) \cap \operatorname{rge} \widehat{\mathcal{A}}^*$ 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 finite union of finitely many set, each defined by finitely many polynomial inequalities – basic quantifier elimination (see [12]) shows that the set of transformation 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, dense semi-algebraic sets necessarily contain an open full-measure set. Thus in this case (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 $\mathcal{A}\colon \mathcal{S}^n \to \mathbf{R}^m$, 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 $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$.

4 Verifying Sturm's error bounds.

Section to be augmented............

Theorem 4.1 (Distance to the intersection)**.** *Consider two closed convex sets A and B in $\mathbf{E}$ and a building block $b \to a^+ \to b^+$. 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. Since by definition we have $a^+ = P_A(b)$ and $b^+ = P_B(a^+)$, the intersection $A \cap B$ lies in the polyhedron

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

Let x be the projection of a^+ onto L. Clearly we have the lower bound $\mathrm{dist}(a^+; A \cap B) \geq |x - a^+|$. We may write

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

for some real numbers $\lambda, \mu \geq 0$. It is easy to verify $\lambda, \mu > 0$ and therefore

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

Multiplying through by $b - a^+$, we obtain

$$0 = \lambda|b - a^+|^2 + \mu\langle a^+ - b^+, b - a^+\rangle$$

Similarly multiplying through by $a^+ - b^+$, we obtain

$$|a^+ - b^+|^2 = \lambda\langle b - a^+, a^+ - b^+\rangle + \mu|a^+ - b^+|^2$$

We deduce

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

Finally taking into account the equality $\langle a^+ - b^+, b - a^+\rangle = -\frac{\lambda}{\mu}|b - a^+|^2$, we obtain

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

The result follows. □

4.1 Figures

4.1.1 Random instances

For each figure we generate a random $b \in \mathbf{R}^m$. We then obtain 40 random instances by generating random $A_i \in \mathcal{S}^n, i = 1, \ldots, m$. This gives us our random linear manifold $\mathcal{L} := \{P \in \mathcal{S}^n : \mathcal{A}P = b\}$. Figures 1, 2, 3 use the same 40 random instances.

For each random instance, we start with a random starting point P_0 and *accurately* attempt to find a point in $\mathcal{L} \cap \mathcal{S}^n_+$. This results in limit points $P_{psd} \succeq 0$ and $P_\ell \in \mathcal{L}$ which yields the displacement vector $D := P_{psd} - P_\ell$.

In Figure 1, page 12, we look at the rates of convergence for feasible instances obtained after shifting $b \leftarrow \mathcal{A}P_{psd}$. The sequence for each of the 40 instances that we plot is

$$\sqrt{k}\|P_k - P_{psd}\|,$$

where $P_{psd} \succeq 0$ is the point for each instance found when solving for the displacement vector. These sequences are bounded which empirically verifies the worst case sublinear $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ convergence rate. Note that the shift implies that the Slater condition fails here which accounts for the slower sublinear convergence rate.

Figure 2, page 13, similarly shifts using $b \leftarrow \mathcal{A}P_{psd}$ to obtain feasibility. we look at the rate of convergence using Theorem 4.1 in combination with Theorem 2.4. Here P_k, L_k are the k-th iterates for the projections onto the semidefinite cone and linear manifold, respectively, and θ_k is the angle between the vectors $P_k - L_k, P_{k+1} - L_k$. Note that

$$\begin{array}{rcl} \frac{\|P_k - L_k\|}{\sin(\theta_k)} & \leq & \operatorname{dist}(P_k; \mathcal{S}^n_+ \cap \mathcal{L}) \\ & \leq & C\operatorname{dist}^\gamma(P_k; \mathcal{S}^n_+) + C\operatorname{dist}^\gamma(P_k; \mathcal{L}) \\ & = & C\|P_k - L_k\|^\gamma. \end{array} \tag{4.9}$$

Therefore, for each instance we look at the sequence

$$\frac{\|P_{k+1} - L_k\|^{1-\gamma}}{\sin(\theta_k)}, \gamma = \frac{1}{2}.$$

We see that this result is *tight* in the sense that the ratio stays bounded above and also bounded from below but above zero.

Figure 3, page 14, looks at the norm rate of convergence *before shifting for feasibility*. We see that the rate is consistently better than that obtained after the shift in Figure 1.

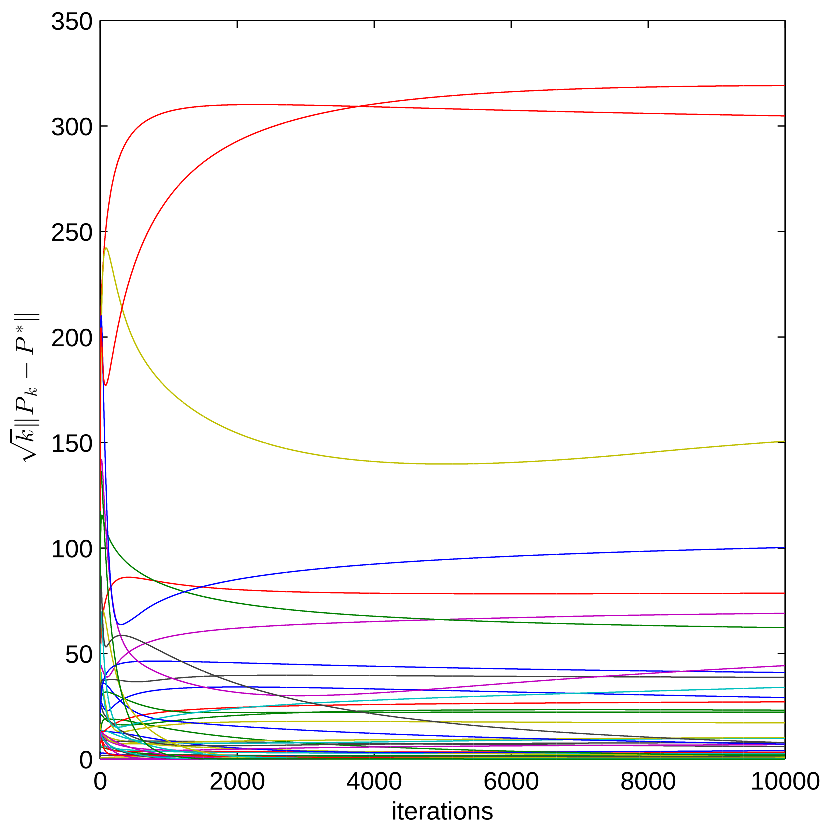

Figure 1: After feasibility shift $b \leftarrow \mathcal{A}P_{psd}$; 40 random instances

4.1.2 Instances with specific degree of singularity

We start with the *worst case* instance for the dual form $\mathcal{A}^*y \preceq C, C = 0$, presented in [**?**, Pg ?]. To form $\mathcal{A}$, let

$$A_1 = e_1e_1^T, A_2 = e_1e_2^T + e_2e_1^T, \quad A_i = e_1e_1^T + e_1e_i^T + e_ie_1^T, i = 3, \ldots, n_d \in \mathcal{S}^{n_d}.$$

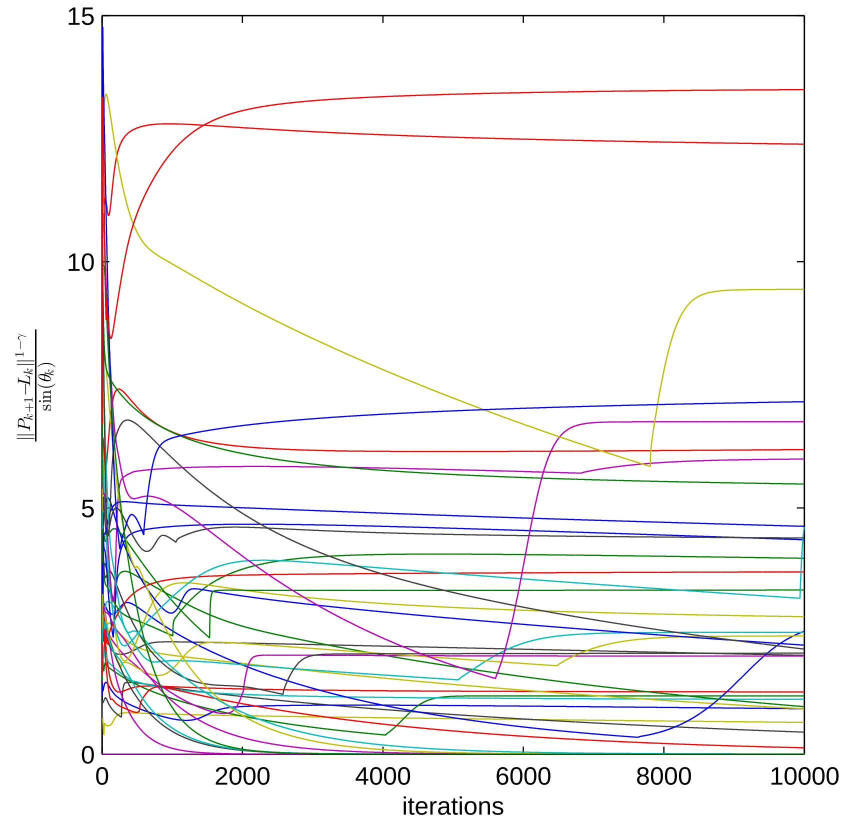

Figure 2: After feasibility shift $b \leftarrow \mathcal{A}P$; 40 random instances

The feasibility problem $\mathcal{A}^*y \preceq 0$ has degree of singularity $d = n_d - 1$. We form an equivalent primal form problem as follows. We form linear independent matrices $A_i, i = n+1, \ldots, n(n+1)/2$ that are orthogonal to $\{A_i, i = 1, \ldots, n\}$. We then choose random $B_i \in \mathcal{S}^{n_t}, i = n+1, \ldots, n(n+1)/2$ and set $A_i \leftarrow \text{blkdiag}(A_i, B_i)$, where blkdiag forms the block diagonal matrix from its arguments. we now set $n = n_d + t_n$ and, for each instance we let Q be a random $n \times n$ orthogonal matrix. The equivalent feasibility problem is

$$\mathcal{A}X = 0, X \succeq 0,$$

where $\mathcal{A}$ is formed using $QA_iQ^T, i = n+1, \ldots, n(n+1)/2$.

In Figures 8, page 19 and 10, page 21 we present plots for the degenerate cases where Slater's condition fails and the degree of singularity $d = n - 1$, i.e., this is the *worst case*.

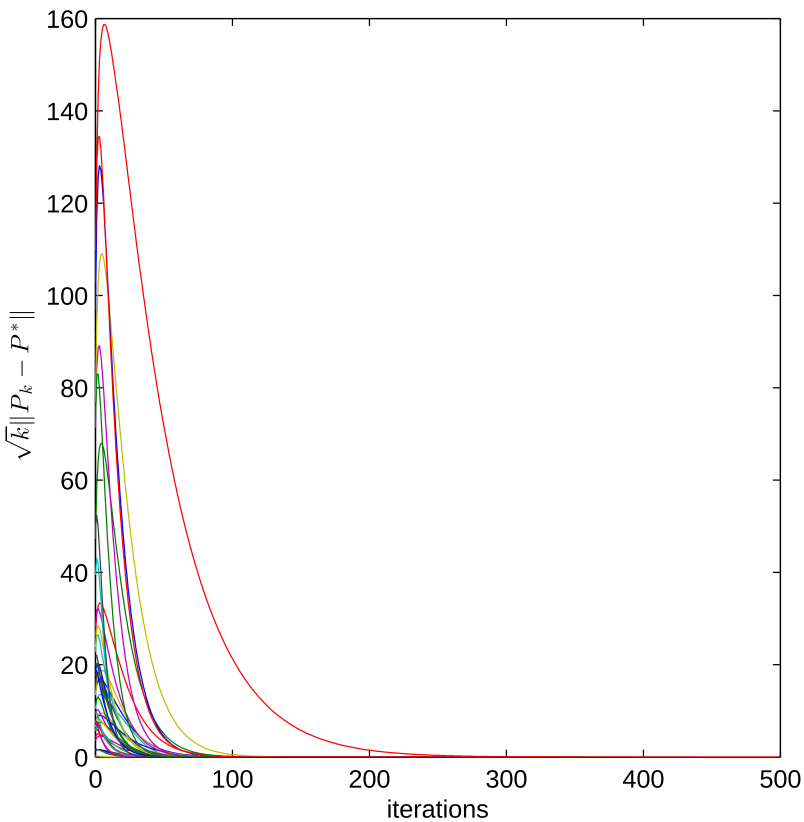

Figure 3: Before feasibility shift $b \leftarrow \mathcal{A}P$; 40 random instances

4.2 ????suggested numerics????

2 sets of pics with 2 pics each

1. first set: one picture with sine ratios and one pic with convergence rate norms for many instances side by side ... increase number of tests to see best picture
2. second set: 1 pic with the convergence rate using norms ... to see increased speed ... without shifting ...
3. number of facial reductions??? generate instance with predefined degeneracy degree $k^{\frac{1}{2^{d+1}-2}}$ e.g. from 1,2,3,4,...

Acknowledgements

We are grateful to Levent Tunçel for suggesting including the contents of Section 4.

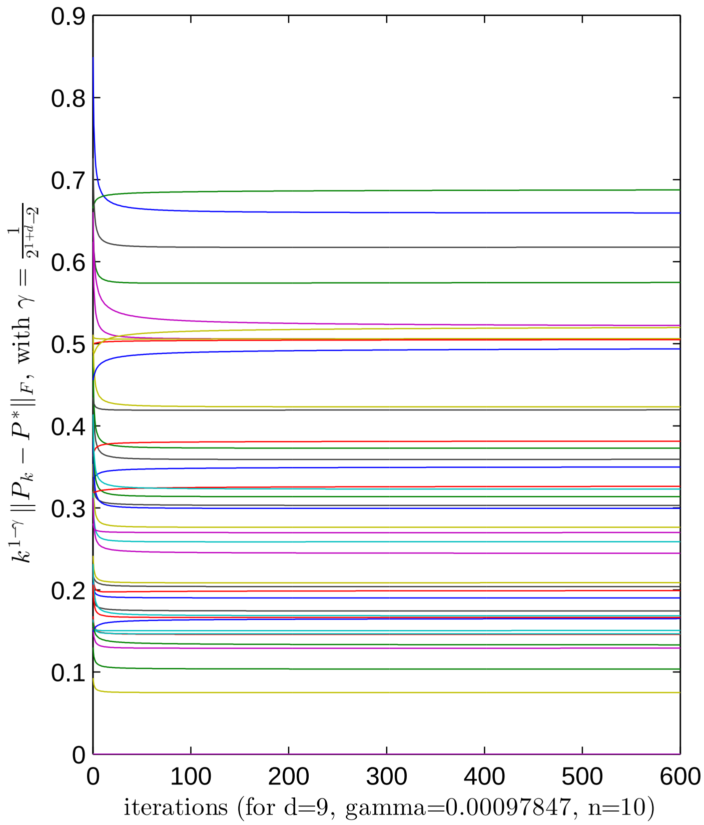

Figure 4: $\gamma = \frac{1}{2^{d+1}-2}, d = n - 1$ for degenerate case

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.

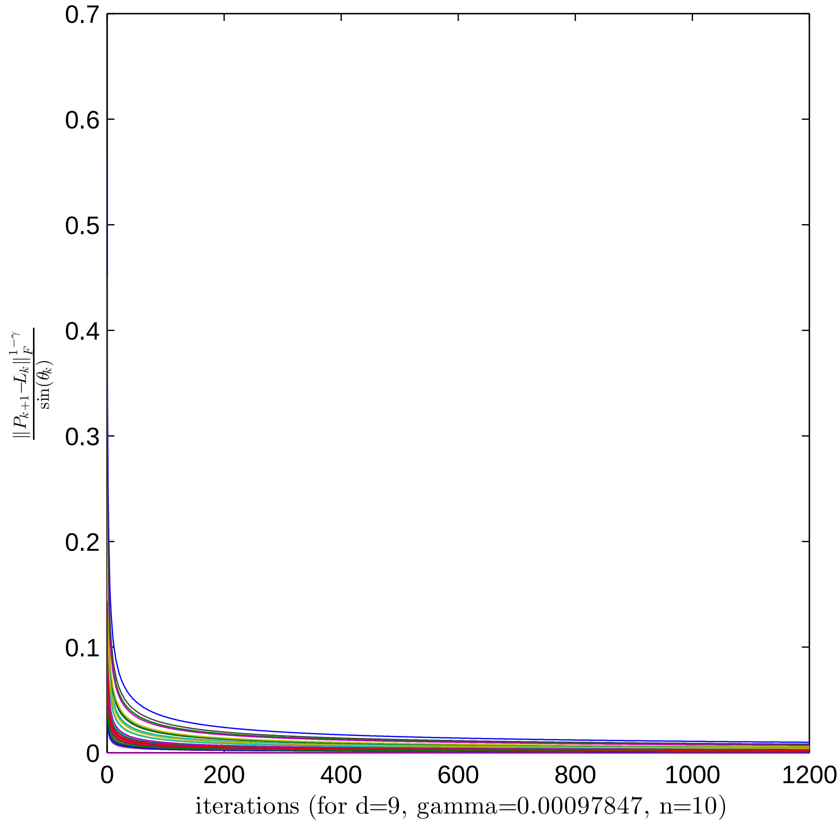

Figure 5: $\gamma = \frac{1}{2^{d+1}-2}, d = n - 1$ for degenerate case

[5] J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semi-algebraic 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 and W.B. Moors. Stability of closedness of convex cones under linear mappings. *J. Convex Anal.*, 16(3-4):699–705, 2009.

[8] 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.

[9] 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.

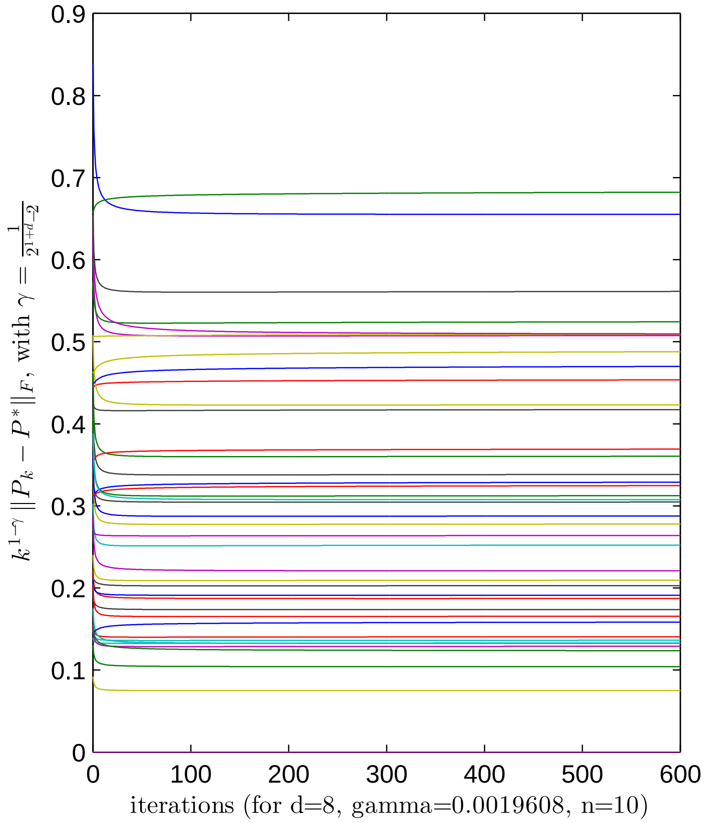

Figure 6: $\gamma = \frac{1}{2^{d+1}-2}$ for degenerate case

[10] L.M. Bregman. The method of successive projection for finding a common point of convex sets. *Sov. Math. Dokl*, 6:688–692, 1965.

[11] 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.

[12] M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

[13] D. Drusvyatskiy and A.S. Lewis. Generic nondegeneracy in convex optimization. *Proc. Amer. Math. Soc.*, 139(7):2519–2527, 2011.

[14] D. Drusvyatskiy, S.A. Vavasis, and H. Wolkowicz. Extreme point inequalities and geometry of the rank sparsity ball. *arXiv:1401.4774*, 2014.

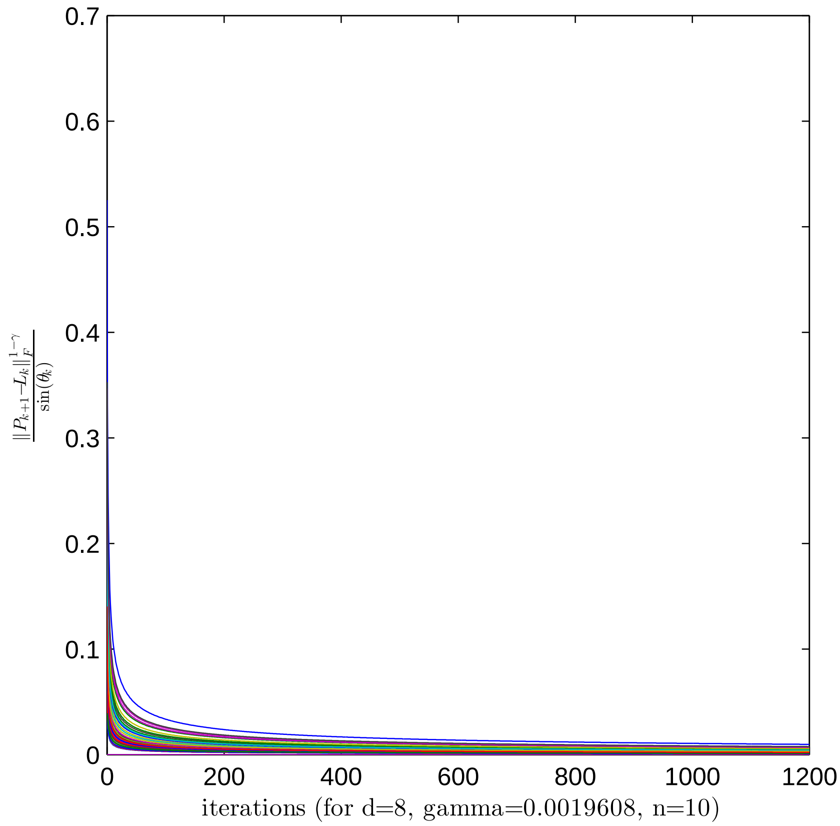

Figure 7: $\gamma = \frac{1}{2^{d+1}-2}$ for degenerate case

[15] 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.

[16] R.B. Holmes. *Geometric functional analysis and its applications*. Springer-Verlag, New York-Heidelberg, 1975. Graduate Texts in Mathematics, No. 24.

[17] 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.

[18] 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.

[19] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

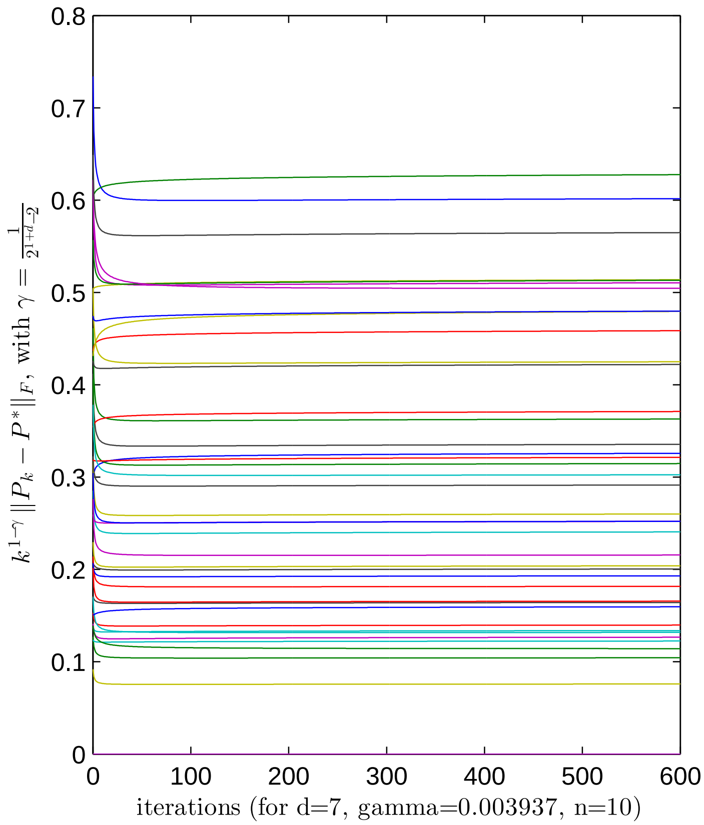

Figure 8: $\gamma = \frac{1}{2^{d+1}-2}$ for degenerate case

[20] J.F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248 (electronic), 2000.

[21] J. von Neumann. *Functional Operators. II. The Geometry of Orthogonal Spaces.* Annals of Mathematics Studies, no. 22. Princeton University Press, Princeton, N. J., 1950.

[22] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013.

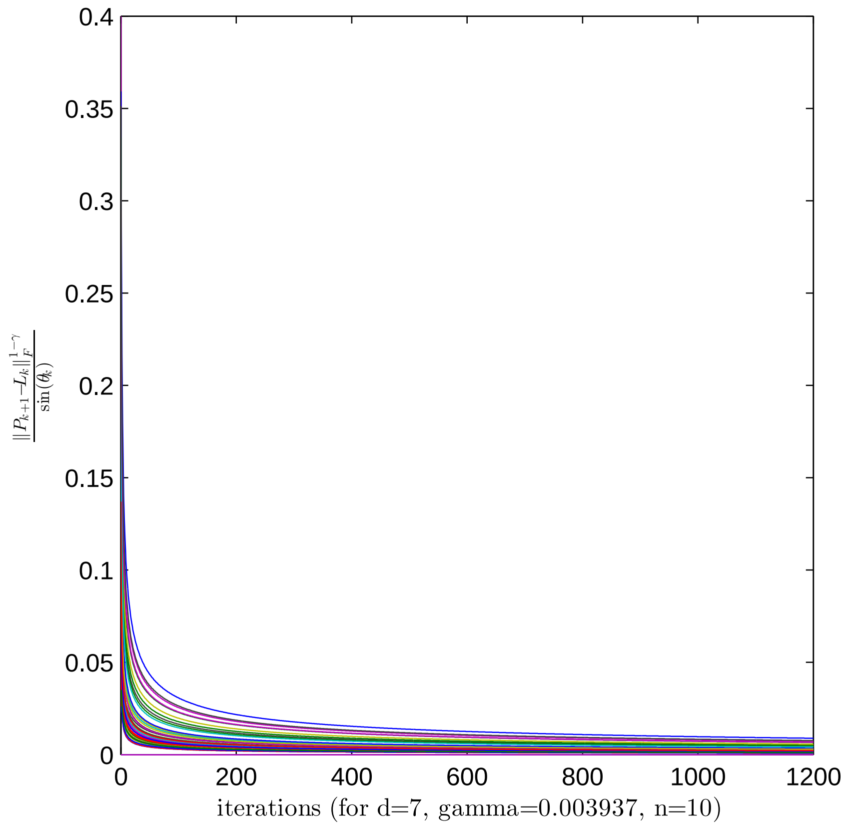

Figure 9: $\gamma = \frac{1}{2^{d+1}-2}$ for degenerate case

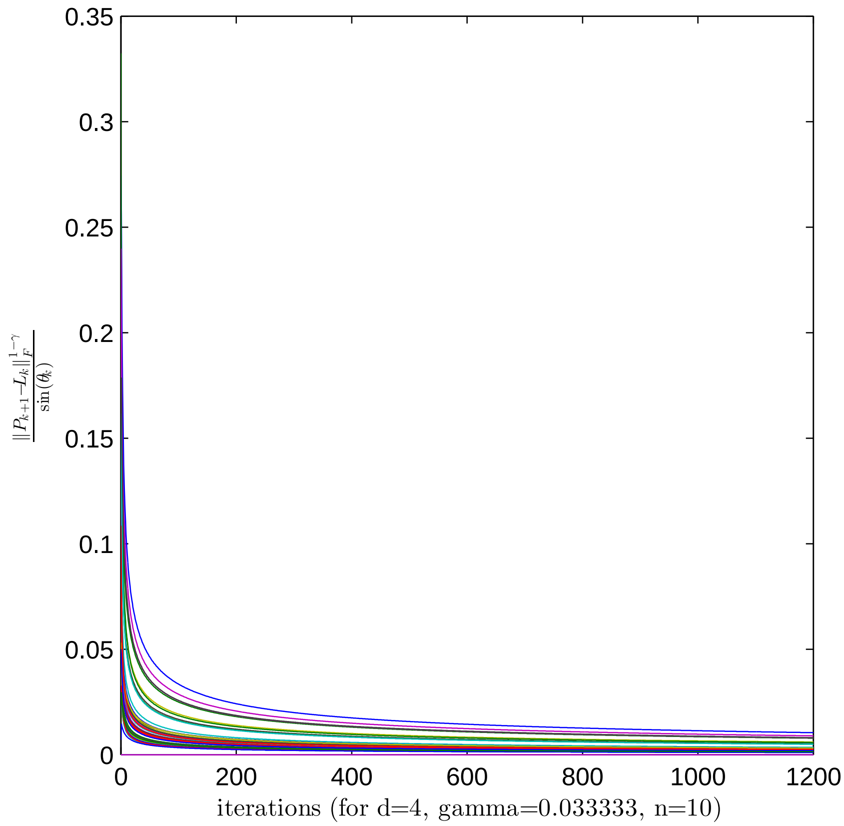

Figure 10: $\gamma = \frac{1}{2^{d+1}-2}$ for degenerate case