Smoothed Analysis of Condition Numbers and Complexity Implications for Linear Programming*

John Dunagan[†]
Microsoft Research

Daniel A. Spielman [‡]
Department of Computer Science
Program in Applied Mathematics
Yale University

Shang-Hua Teng [§]
Department of Computer Science
Boston University

March 30, 2009

Abstract

We perform a smoothed analysis of Renegar's condition number for linear programming by analyzing the distribution of the distance to ill-posedness of a linear program subject to a slight Gaussian perturbation. In particular, we show that for every n-by-d matrix $\bar{\mathbf{A}}$, n-vector $\bar{\boldsymbol{b}}$, and d-vector $\bar{\boldsymbol{c}}$ satisfying $\left\|\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}}\right\|_F \leq 1$ and every $\sigma \leq 1$,

$$\mathop{\mathbf{E}}_{\mathbf{A},\boldsymbol{b},\boldsymbol{c}} [\log C(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})] = O(\log(nd/\sigma)),$$

where $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are Gaussian perturbations of $\bar{\mathbf{A}}$, $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{c}}$ of variance σ^2 and $C(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ is the condition number of the linear program defined by $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$. From this bound, we obtain a smoothed analysis of interior point algorithms. By combining this with the smoothed analysis of finite termination of Spielman and Teng (Math. Prog. Ser. B, 2003), we show that the smoothed complexity of interior point algorithms for linear programming is $O(n^3 \log(nd/\sigma))$.

1 Introduction

In [ST04], Spielman and Teng introduced the smoothed analysis of algorithms as an alternative to worst-case and average-case analyses in the hope that it could provide a measure of the

*This work was supported in part by NSF Grants CCR-9875024, CCR-9972532, CCR-0112487, CCR-0325630, and CCF-0707522, and felloships from the Alfred P. Sloan Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

† jdunagan@microsoft.com. Part of this work was done while the author was at MIT.

‡ spielman@cs.yale.edu. Part of this work was done while the author was at MIT.

§ steng@cs.bu.edu

complexity of algorithms that better agrees with practical experience. The smoothed complexity of an algorithm is the maximum over its inputs of the expected running time of the algorithm under slight perturbations of that input. In particular, they consider the linear programming problem of the normal form

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \text{ s.t. } \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}$$

in which the input data $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ are subject to slight Gaussian perturbations. Recall that the probability density function of the univariate Gaussian random variable with mean 0 and variance σ^2 is given by

$$\mu(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}.$$

Definition 1.0.1 (Gaussian Perturbation). *For any $\bar{\mathbf{A}} \in \mathbb{R}^{n\times d}$ and $\sigma \geq 0$, the Gaussian perturbation of variance σ^2 of $\bar{\mathbf{A}}$ is $\mathbf{A} = \bar{\mathbf{A}} + \mathbf{G}$, where each entry of $\mathbf{G}$ is an independent univariate Gaussian variable with mean 0 and variance σ^2.*

Spielman and Teng [ST04] proved that the smoothed complexity of a two-phase shadow vertex simplex method was polynomial in n, d and $1/\sigma$. That is, for any $(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})$ such that $\left\|(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})\right\|_F \leq 1$ and $\sigma \leq 1$, if $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ is the Gaussian perturbation of $(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})$ of variance σ^2, then the linear program defined by $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ can be solved by a simplex algorithm in expected time polynomial in nd/σ, where $\left\|(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})\right\|_F$ is the square root of the sum of squares of the entries in $\bar{\mathbf{A}}$, $\bar{\boldsymbol{b}}$, and $\bar{\boldsymbol{c}}$.

In this paper, we perform a smoothed analysis of condition numbers for linear programs, and thereby obtain a smoothed analysis of interior-point algorithms. Interior point algorithms for linear programming are exciting both because they are known to run in polynomial time in the worst case [Kar84] and because they have been used to efficiently solve linear programs in practice [LMS90].

The worst-case complexity of solving linear programs has traditionally been stated in terms of n, d, and L, where L is commonly called the "bit-length" of the input linear program, which could also be a parameter measuring the precision needed to perform the arithmetic operations exactly. The definition of L varies in the literature: Khachiyan [Kha79], Karmarkar [Kar84], and Vaidya [Vai90] define L for integer matrices A to be some constant times

$$\log(\text{largest absolute value of the determinant of any square sub-matrix of } \mathbf{A}) \\ + \log(\|\boldsymbol{c}\|_\infty) + \log(\|\boldsymbol{b}\|_\infty) + \log(n+d).$$

Under this definition, L is not efficiently computable, and unless $\mathbf{A}$ comes from a very special class of matrices, it is difficult to find L below $\Omega(n)$. Others use cruder upper bounds on L such as the total number of bits in a row of the matrix or the total number of bits in the entire matrix [Wri97].

Without using fast matrix multiplication[1], the best bound known on the worst-case complexity of any linear programming algorithm is Anstreicher's [Ans99] bound of $O\left((n^3/\log n)L\right)$. This bound improves slightly upon the bound of $O(n^3L)$, first independently obtained by

[1] Vaidya [Vai89] showed that fast matrix multiplication can be used to further improve the complexity of interior-point algorithms. Also see Chapter 8 of Nesterov and Nemirovskii [NN94].

Vaidya's [Vai90][2] and Gonzaga [Gon88]. Several other variants of interior-point algorithms have since been shown to have complexity $O(n^3L)$. These algorithms are iterative algorithms that use $O(\sqrt{n}L)$ iterations with an average of $O(n^{2.5})$ operations per iteration. Note that all these bounds assume $n \geq d$.

The speed of interior point methods in practice is much better than that proved in their worst-case analyses [IL94, LMS90, EA96]. It has been observed that the number of iterations of various interior-point algorithms is much smaller than $\sqrt{n}L$. This discrepancy between worst-case analysis and practical experience is our main motivation for studying the smoothed complexity of interior point methods.

As a corollary of the main result of this paper, we show that there exist interior-point algorithms that return the exact answer to linear programs and have smoothed iteration complexity $O(\sqrt{n}\log(nd/\sigma))$ and smoothed complexity $O(n^3\log(nd/\sigma))$ operations, for $\sigma \leq 1$. In comparison, L could be much larger than $\log(nd/\sigma)$ in the $O(n^3L)$ worse-case bound. We thus partially close the gap between the theoretical worst-case bound and practical observations.

Recently, Deza, Nematollahi and Terlaky [DET08] have proved a lower bound of $\Omega(\sqrt{n/\log^3 n})$ on the complexity of path-following interior point methods. This lower bound holds even when they only require the algorithm to produce a solution of small duality gap. The precision to which numbers are specified in their construction is sub-quadratic in n, which suggests that it is unlikely that one could greatly reduce the number of iterations by perturbing their examples by perturbations of standard deviation $O(1/n^3)$. Other lower bounds on the complexity of interior point algorithms have been obtained by Todd [Tod94] and Todd and Ye [TY96]. However, the programs for which these lower bounds hold are very ill-conditioned.

Before stating our main result, we first review Renegar's condition number for linear programs.

1.1 Condition Numbers of a Linear Program

A linear program is typically specified by a matrix $\mathbf{A} \in \mathbb{R}^{n\times d}$ together with two vectors $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{c} \in \mathbb{R}^d$. There are several canonical forms of linear programs specified by $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$. The following are four commonly used canonical forms:

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \text{ s.t. } \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b} \quad \text{and its dual} \quad \min\ \boldsymbol{b}^T\boldsymbol{y} \text{ s.t } \mathbf{A}^T\boldsymbol{y} = \boldsymbol{c},\ \ \boldsymbol{y} \geq \mathbf{0} \tag{1}$$

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \text{ s.t. } \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{x} \geq \mathbf{0} \quad \text{and its dual} \quad \min\ \boldsymbol{b}^T\boldsymbol{y} \text{ s.t. } \mathbf{A}^T\boldsymbol{y} \geq \boldsymbol{c},\ \boldsymbol{y} \geq \mathbf{0} \tag{2}$$

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \text{ s.t. } \mathbf{A}^T\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \mathbf{0} \quad \text{and its dual} \quad \min\ \boldsymbol{b}^T\boldsymbol{y} \text{ s.t. } \mathbf{A}\boldsymbol{y} \geq \boldsymbol{c} \tag{3}$$

$$\text{find } \boldsymbol{x} \neq \mathbf{0} \text{ s.t. } \mathbf{A}\boldsymbol{x} \leq \mathbf{0} \quad \text{and its dual} \quad \text{find } \boldsymbol{y} \neq \mathbf{0} \text{ s.t. } \mathbf{A}^T\boldsymbol{y} = \mathbf{0},\ \boldsymbol{y} \geq \mathbf{0} \tag{4}$$

In his pioneering work [Ren95b, Ren95a, Ren94], Renegar defined the condition number $C(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ of a linear program as the scale-invariant reciprocal of the distance of that program to "ill-posedness". A linear program is ill-posed if the program can be made both feasible and infeasible by arbitrarily small changes to its data. Any linear program may be expressed in each of the first three canonical forms. However, transformations among linear programming formulations

[2]Vaidya's bound is explicit in its dependence on both n and d: $O((n+d)d^2 + (n+d)^{1.5}d)L)$.

do not in general preserve their condition numbers [Ren95a]. We therefore define a condition number for each normal form considered.

Definition 1.1.1 (Primal Condition Number: Form (1)). *For* $\mathbf{A} \in \mathbb{R}^{n\times d}$ *and* $\boldsymbol{b} \in \mathbb{R}^n$,

(a) if $\mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}$ *is feasible, then*

$$C_P^{(1)}(\mathbf{A}, \boldsymbol{b}) = \frac{\|\mathbf{A}, \boldsymbol{b}\|_F}{\sup\{\delta : \|\Delta\mathbf{A}, \Delta\boldsymbol{b}\|_F \leq \delta \textit{ implies } (\mathbf{A}+\Delta\mathbf{A})\boldsymbol{x} \leq (\boldsymbol{b}+\Delta\boldsymbol{b}) \textit{ is feasible}\}},$$

(b) if $\mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}$ *is infeasible, then*

$$C_P^{(1)}(\mathbf{A}, \boldsymbol{b}) = \frac{\|\mathbf{A}, \boldsymbol{b}\|_F}{\sup\{\delta : \|\Delta\mathbf{A}, \Delta\boldsymbol{b}\|_F \leq \delta \textit{ implies } (\mathbf{A}+\Delta\mathbf{A})\boldsymbol{x} \leq (\boldsymbol{b}+\Delta\boldsymbol{b}) \textit{ is infeasible}\}}.$$

It follows from the definition above that $C_P^{(1)}(\mathbf{A}, \boldsymbol{b}) \geq 1$. We define the dual condition number $C_D^{(1)}(\mathbf{A}, \boldsymbol{c})$ analogously. The condition number $C^{(1)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ is then defined as

$$C^{(1)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}) = \max\left(C_P^{(1)}(\mathbf{A}, \boldsymbol{b}), C_D^{(1)}(\mathbf{A}, \boldsymbol{c})\right).$$

For linear programs with canonical forms (2), (3), and (4) we define their condition numbers, $C^{(2)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$, $C^{(3)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ and $C^{(4)}(\mathbf{A})$, analogously. Note that each condition number is always at least 1.

Working with linear programs in canonical form (2), Renegar developed an algorithm that returns a feasible point with duality gap $R \leq O(nC(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}))$ or determines that the program is infeasible or unbounded in $O(n^3 \log C(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}))$ arithmetic operations. Applying any interior-point algorithm that can reduce the duality gap from R to ϵ in $O(\sqrt{n}\log(R/\epsilon))$ iterations and $O(n^3 \log(R/\epsilon))$ operations to this feasible point, one can obtain a solution with relative accuracy ϵ.

Subsequently, algorithms with complexity logarithmic in the condition number were developed by Vera [Ver96] for forms (1) and (3) and by Cucker and Peña [CP01] for form (4). In [FV00], Freund and Vera give a unified approach which both efficiently estimates the condition number and solves the linear programs in any of these forms. The following theorem summarizes these complexity results.

Theorem 1.1.2 (Condition-Based Complexity). *For any linear program of form (i),* $i \in \{1, 2, 3\}$*, specified by* $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ *with* $n \geq d$ *and parameter* $\epsilon \leq 1$*, there is an interior-point algorithm that determines that the program is infeasible or unbounded in* $O(n^3 \log(nC^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})/\epsilon))$ *operations, or finds a feasible solution* $\boldsymbol{x}$ *with duality gap at most* $\epsilon\,\|\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}\|_F$ *in* $O(n^3 \log(nC^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})/\epsilon))$ *operations. For a linear program of form (4) given by* $\mathbf{A}$*, there is an algorithm that finds a feasible solution* $\boldsymbol{x}$ *or determines that the program is infeasible in* $O(n^3 \log(nC^{(4)}(\mathbf{A})))$ *operations.*

1.2 Smoothed Analysis of Condition Numbers: Our Results

The condition-based complexity of linear programming provides not only a natural framework for evaluating the performance of linear-programming algorithms in the Real Turing machine model or a finite-precision model, but also an instance-based characterization of a non-trivial class of linear programs that can be solved quickly. If $\log C^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ is much smaller than L, then $O(n^3 L)$ is an overly pessimistic upper bound on the complexity of solving the linear program defined by $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$.

Our main result is an upper bound on the smoothed value of $\log C^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$. In Theorem 4.0.3, we show for any $(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})$, $\sigma \leq 1$, and each i

$$\mathop{\mathbf{E}}_{\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}} \left[\log C^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}) \right] = O(\log(nd/\sigma)), \qquad (5)$$

where $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are Gaussian perturbations of $\bar{\mathbf{A}}$, $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{c}}$ of variance σ^2.

The bound on the smoothed complexity of interior point methods follows immediately from Theorem 1.1.2, and (5). Note that in (5) and in the corollary below, we abuse notation by writing $C^{(4)}(A, \boldsymbol{b}, \boldsymbol{c})$ instead of $C^{(4)}(A)$.

Corollary 1.2.1 (Smoothed Complexity of IPM). *For any $n \geq d$, $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$, $\bar{\boldsymbol{b}} \in \mathbb{R}^n$ and $\bar{\boldsymbol{c}} \in \mathbb{R}^d$ such that $\left\| \bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}} \right\|_F \leq 1$ and $\sigma \leq 1$, let $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ be the Gaussian perturbation of $(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})$ of variance σ^2. For any $\epsilon \leq 1$, let $T^{(i)}((\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}), \epsilon)$ be the time complexity of the interior point algorithms of Theorem 1.1.2 for finding ϵ-accurate solutions of the linear program defined by $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ in form (i) or determining that the program is infeasible or unbounded. Then,*

$$\mathop{\mathbf{E}}_{(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})} \left[T^{(i)}((\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}), \epsilon) \right] = O\left(n^3 \log\left(\frac{nd}{\sigma\epsilon} \right) \right).$$

Remark 1 (Assumption on σ and Big-O notation). *In the remainder of this paper, we will explicitly keep track of constants in our analysis and will not use "big-O" notation to hide them. We will, however, assume $\sigma \leq 1/\sqrt{(n+1)(d+1)}$ and use this assumption to simplify our expressions. To apply our analysis to the case when $1/\sqrt{(n+1)(d+1)} \leq \sigma \leq 1$, we can simply scale the space by a factor of $1/(\sigma\sqrt{(n+1)(d+1)})$.*

Note also that Corollary 1.2.1 is stated with the assumption $n \geq d$. In the analysis of condition numbers of this paper, we will explicitly keep track of the contributions of n and d.

As explained in [ST03b], when one combines this analysis with the smoothed analysis of the finite termination procedure in that paper, one obtains an interior point algorithm that returns the exact answer to the linear program and that has smoothed complexity $O(n^3 \log(nd/\sigma))$, for $\sigma \leq 1$.

Our work is partially motivated by a recent result of Blum and Dunagan [BD02] on the smoothed analysis of the perceptron algorithm for linear programming. We build upon and extend their analysis to Renegar's condition number for the four standard forms discussed in Section 1.1. As already noted, one can apply the smoothed condition number bounds of this paper directly to many other linear programming algorithms whose complexity can be bounded in terms of the condition number of its input. These algorithms include the ellipsoid algorithm [Kha79, FV00], von Neumann's algorithm [EF00], and the recent perceptron algorithm with rescaling [DV04].

There have been many average-case analyses of condition numbers and interior point algorithms. Anstreicher, Ji, Potra and Ye [AJPY93, AJPY99], have shown that under Todd's degenerate model for random linear programs [Tod91], a homogeneous self-dual interior point method runs in $O(\sqrt{n}\log n)$ expected iterations. Borgwardt and Huhn [HB02] have obtained similar results under any spherically symmetric distribution. The performance of other interior point methods on random inputs has been heuristically analyzed through "one-step analyses", but it is not clear that these analyses can be made rigorous [Nem88, GT92, MTY93]. Cucker and Wschebor [CW03], Cheung and Cucker [CC02], and Cheung, Cucker, and Hauser [CCH03] studied the distribution of condition numbers of random linear programs drawn from various distributions, and their bounds also imply that the average-case complexity of the interior-point method is $O(n^3 \log n)$. If one specializes our results to perturbations of the all-zero matrix $\bar{\mathbf{A}}$ and the all-zero vectors $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{c}}$, then one obtains a similar average-case analysis of the distribution of condition numbers under the Gaussian distribution.

In contrast with the average-case analysis, our smoothed analysis can be interpreted as demonstrating that if there is a little bit of imprecision or noise in the input data, then the linear program is unlikely to be poorly-conditioned, and hence can be solved quickly by interior-point algorithms. We refer interested readers to [ST04] for discussions of worst-case, average-case, and smoothed analyses of algorithms.

1.3 Organization of the Paper

In our analysis, we divide the eight condition numbers $C_P^{(i)}$ and $C_D^{(i)}$, for $i \in \{1,2,3,4\}$, into two groups. The first group includes $C_P^{(1)}$, $C_P^{(2)}$, $C_D^{(2)}$, $C_D^{(3)}$, and with some additional work, $C_P^{(4)}$. The remaining condition numbers belong to the second group. We will refer to a condition number from the first group as a *primal condition number* and a condition number from the second group as a *dual condition number*.

Section 2 is devoted to the smoothed analysis of primal condition numbers. We remark that the techniques used in Section 2 do not critically depend upon $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ being Gaussian perturbations, and similar theorems could be proved using slight modifications of our techniques if these were smoothly distributed within spheres or cubes. In Section 3, we consider dual condition numbers. Our analysis in this section does make critical use of the Gaussian distribution of $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. In Section 4, we prove our main result on condition numbers, Theorem 4.0.3, using the smoothed bounds of the previous two sections. We conclude the paper in Section 5 with some open questions.

2 Primal Condition Number

In this section we consider the the primal condition numbers. Instead of analyzing each primal condition number separately, we perform a unified analysis, motivated by the work of Peña [Peñ00], by transforming each canonical form to conic form. We then obtain the stronger result that the smoothed value of the logarithm of the primal condition number of conic programs in form (6) below is $O(\log nd/\sigma)$.

2.1 Primal Condition Number of Conic Programming

As the condition numbers are independent of the objective functions, we focus on the feasibility problem for a conic linear program which can be written as:

$$\text{find } \boldsymbol{x} \text{ such that } \mathbf{A}\boldsymbol{x} \geq \mathbf{0}, \boldsymbol{x} \in \boldsymbol{C}, \tag{6}$$

where $\mathbf{A}$ is an n-by-d matrix and $\boldsymbol{C}$ is a *strictly-supported convex cone* in $\mathbb{R}^d$.

Definition 2.1.1 (Strictly-supported convex cone). *A strictly-supported convex cone is a non-empty convex set $\boldsymbol{C}$ such that for all $\boldsymbol{x} \in \boldsymbol{C}$ and all $\alpha > 0$, $\alpha\boldsymbol{x} \in \boldsymbol{C}$, and such that there exists a vector $\mathbf{t}$ for which $\mathbf{t}^T\boldsymbol{x} < 0$ for all $\boldsymbol{x} \in \boldsymbol{C}$.*

For a column vector $\boldsymbol{a}$, let $\mathbf{Ray}\,(\boldsymbol{a})$ denote $\{\alpha\boldsymbol{a} : \alpha > 0\}$ and let $\mathcal{H}(\boldsymbol{a}) = \left\{\boldsymbol{x} : \boldsymbol{a}^T\boldsymbol{x} \geq 0\right\}$ denote the half-space of points with non-negative inner product with $\boldsymbol{a}$. Note that

$$\mathbf{A}\boldsymbol{x} \geq \mathbf{0} \text{ and } \boldsymbol{x} \in \boldsymbol{C} \quad \Longleftrightarrow \quad \boldsymbol{x} \in \boldsymbol{C} \cap \bigcap_{i=1}^{n} \mathcal{H}(\boldsymbol{a}_i),$$

where $\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_n^T$ are the rows of $\mathbf{A}$. Thus, the set $\boldsymbol{C} \cap \bigcap_{i=1}^{n} \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset$ if and only if the corresponding conic program is feasible. Throughout this paper, if $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are column vectors in $\mathbb{R}^d$, we let $[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ denote the matrix whose rows are the $\boldsymbol{a}_i^T$s.

Note that a strictly-supported convex cone cannot contain the origin. For example, $\mathbb{R}^d$ and $\mathcal{H}(\boldsymbol{a})$ are not strictly-supported convex cones, while $\{\boldsymbol{x} : \boldsymbol{x}_0 > 0\}$ and $\mathbf{Ray}\,(\boldsymbol{a})$ are. Thus, $\mathbf{0}$ cannot be a feasible solution of program (6).

The following definition generalizes distance to ill-posedness from linear programing to conic programing by explicitly taking into account the strictly-supported convex cone $\boldsymbol{C}$.

Definition 2.1.2 (Distance to ill-posed). *For a strictly-supported convex cone $\boldsymbol{C}$ (not subject to perturbation) and a matrix, $\mathbf{A}$, we define $\rho(\mathbf{A}, \boldsymbol{C})$ by*

a. if $\mathbf{A}\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \in C$ is feasible, then

$$\rho(\mathbf{A}, \boldsymbol{C}) = \sup\left\{\epsilon : \|\Delta\mathbf{A}\|_F < \epsilon \text{ implies } (\mathbf{A} + \Delta\mathbf{A})\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C} \text{ is feasible}\right\};$$

b. if $\mathbf{A}\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C}$ is infeasible, then

$$\rho(\mathbf{A}, \boldsymbol{C}) = \sup\left\{\epsilon : \|\Delta\mathbf{A}\|_F < \epsilon \text{ implies } (\mathbf{A} + \Delta\mathbf{A})\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C} \text{ is infeasible}\right\}.$$

We note that this definition makes sense even when $\mathbf{A}$ is a row vector $[\boldsymbol{a}]$. In this case, $\rho(\boldsymbol{a}^T, \boldsymbol{C})$ measures the distance to ill-posedness when we only allow modification of $\boldsymbol{a}$.

2.2 Transformation to Conic Programming

The primal program of form (1) can be put into conic form with the introduction of a homogenizing variable x_0. Setting $\boldsymbol{C} = \{(\boldsymbol{x}, x_0) : x_0 > 0\}$, the homogenized primal program of form (1) is

$$[-\mathbf{A}, \boldsymbol{b}](\boldsymbol{x}, x_0) \geq \mathbf{0},\ (\boldsymbol{x}, x_0) \in \boldsymbol{C}.$$

By setting $\boldsymbol{C} = \{(\boldsymbol{x}, x_0) : x_0 > 0 \textit{ and } \boldsymbol{x} \geq \mathbf{0}\}$, one can similarly homogenize the primal program of form (2). The dual programs of form (2) and form (3) can be homogenized by setting $\boldsymbol{C} = \{(\boldsymbol{y}, y_0) : y_0 > 0 \textit{ and } \boldsymbol{y} \geq \mathbf{0}\}$ and $\boldsymbol{C} = \{(\boldsymbol{y}, y_0) : y_0 > 0\}$, respectively, and considering the program

$$[\mathbf{A}^T, -\boldsymbol{c}](\boldsymbol{y}, y_0) \geq \mathbf{0},\ (\boldsymbol{y}, y_0) \in \boldsymbol{C}.$$

Note that in each of these homogenized programs, the variables lie in a strictly-supported convex cone.

Proposition 2.2.1 (Preserving feasibility). *Each of the homogenized programs is feasible if and only if its original program is feasible.*

Even though transformations among linear programming formulations in general do not preserve condition numbers, Peña [Peñ00] has proved that homogenization does not alter the distance to ill-posedness. For convenience, we will state the lemma for form (1), and note that similar statements hold for $C_P^{(2)}$, $C_D^{(2)}$, and $C_D^{(3)}$.

Lemma 2.2.2 (Preserving the condition number). *Let*

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \quad s.t. \quad \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}$$

be a linear program. Let $\boldsymbol{C} = \{(\boldsymbol{x}, x_0) : x_0 > 0\}$*. Then,* $C_P^{(1)}(\mathbf{A}, \boldsymbol{b}) = \|\mathbf{A}, \boldsymbol{b}\|_F / \rho([-\mathbf{A}, \boldsymbol{b}], \boldsymbol{C})$.

The primal program of form (4) does not fit into form (6), as $\boldsymbol{x}$ can be any non-zero vector. To handle it, we need the following definition.

Definition 2.2.3 (Alternative distance to ill-posed). *For a convex cone that is* not strictly-supported, $\boldsymbol{C}$*, and a matrix,* $\mathbf{A}$*, we define* $\rho(\mathbf{A}, \boldsymbol{C})$ *by*

a. if $\mathbf{A}\boldsymbol{x} \geq 0,\ \boldsymbol{x} \neq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C}$ *is feasible, then*

$$\rho(\mathbf{A}, \boldsymbol{C}) = \sup\{\epsilon : \|\Delta\mathbf{A}\|_F < \epsilon \textit{ implies } (\mathbf{A} + \Delta\mathbf{A})\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \neq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C} \textit{ is feasible}\}$$

b. if $\mathbf{A}\boldsymbol{x} \geq 0,\ \boldsymbol{x} \neq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C}$ *is infeasible, then*

$$\rho(\mathbf{A}, \boldsymbol{C}) = \sup\{\epsilon : \|\Delta\mathbf{A}\|_F < \epsilon \textit{ implies } (\mathbf{A} + \Delta\mathbf{A})\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \neq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C} \textit{ is infeasible}\}$$

This definition would allow us to prove the analogues of Lemmas 2.3.15 and 2.3.1 for primal programs of form (4). We omit the details of this variation on the arguments in the interest of simplicity.

2.3 Smoothed Analysis of Primal Condition Number of Conic Programs

Lemma 2.3.1 below states the main result of this section. Note that a simple union bound over $C_P^{(2)}$ and $C_D^{(2)}$ using Lemma 2.3.1 together with Theorem 1.1.2 yields Corollary 1.2.1 for form (2).

Lemma 2.3.1 (Logarithm of primal conic condition number). *For any $n \geq d \geq 2$, any strictly-supported convex cone $\boldsymbol{C}$, $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$ satisfying $\left\|\bar{\mathbf{A}}\right\|_F \leq 1$, and $\sigma \leq 1/\sqrt{nd}$,*

$$\mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \frac{\|\mathbf{A}\|_F}{\rho(\mathbf{A}, \boldsymbol{C})}\right] \leq 4 \log_2\left(\frac{nd}{\sigma}\right) + 17,$$

where $\mathbf{A}$ is the Gaussian perturbation of $\bar{\mathbf{A}}$ of variance σ^2.

In a moment, we will derive Lemma 2.3.1 from the following lemma.

Lemma 2.3.2 (Smoothed Analysis of Distance to ill-posed). *For any $n \geq d \geq 2$, any strictly-supported convex cone $\boldsymbol{C}$, $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$ satisfying $\left\|\bar{\mathbf{A}}\right\|_F \leq 1$, $\sigma \leq 1/\sqrt{nd}$, and $0 < \epsilon \leq 1/2$,*

$$\mathop{\mathbf{Pr}}_{\mathbf{A}}\left[\rho(\mathbf{A}, \boldsymbol{C}) \leq \epsilon\right] \leq \left(\frac{98 n^2 d^{1.5}}{\sigma^2}\right) \epsilon \ln^{1.5}\left(\frac{1}{\epsilon}\right), \tag{7}$$

where $\mathbf{A}$ is the Gaussian perturbation of $\bar{\mathbf{A}}$ of variance σ^2.

We postpone the proof of Lemma 2.3.2 to Section 2.3.3. We now apply it to prove Lemma 2.3.1.

Proof of Lemma 2.3.1. First notice that

$$\mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \frac{\|\mathbf{A}\|_F}{\rho(\mathbf{A}, \boldsymbol{C})}\right] = \mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \|\mathbf{A}\|_F\right] + \mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \frac{1}{\rho(\mathbf{A}, \boldsymbol{C})}\right].$$

Because $\left\|\bar{\mathbf{A}}\right\|_F \leq 1$ and $\sigma \leq 1/\sqrt{nd}$, Lemma A.1.5 implies

$$\mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \|\mathbf{A}\|_F\right] \leq 1/2.$$

By Lemma 2.3.2,

$$\mathop{\mathbf{Pr}}_{\mathbf{A}}\left[\frac{1}{\rho(\mathbf{A}, \boldsymbol{C})} \geq x\right] \leq \frac{\alpha \log_2^{1.5} x}{x}, \quad \text{where} \quad \alpha = \frac{98 n^2 d^{1.5}}{\sigma^2}.$$

As $n, d \geq 1$, $\sigma \leq 1$, and $\alpha \geq 98 > 10$, we may apply Lemma A.2.2 to bound

$$\mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \frac{1}{\rho(\mathbf{A}, \boldsymbol{C})}\right] \leq 2 \log_2(e\alpha).$$

Thus,

$$\mathop{\mathbf{E}}_{\mathbf{A}}\left[\log_2 \frac{\|\mathbf{A}\|_F}{\rho(\mathbf{A}, \boldsymbol{C})}\right] \leq 1/2 + 2\log_2(e\alpha) \leq 1/2 + 2\log_2(98e) + 2\log_2 \frac{n^2 d^{1.5}}{\sigma^2} < 4 \log_2 \frac{nd}{\sigma} + 17.$$

□

We will prove Lemma 2.3.2 by separately considering the cases in which the program is feasible and infeasible. We will handle the feasible case in Section 2.3.1, and deal with the infeasible case in Section 2.3.2. We then combine both cases and complete the analysis in Section 2.3.3.

The thread of argument in both Sections 2.3.1 and 2.3.2 consists of a geometric characterization of the conic programs with poor primal condition number, followed by a probabilistic argument demonstrating that this characterization is rarely satisfied in the smoothed model. Throughout the proofs in this section, $\boldsymbol{C}$ will always refer to the original strictly-supported convex cone, and a subscripted $\boldsymbol{C}$, such as $\boldsymbol{C}_1$, will refer to a modification of this cone.

For a convex body $\boldsymbol{K}$ in $\mathbb{R}^d$, let $\partial \boldsymbol{K}$ be its boundary. For any $\epsilon \geq 0$, let

$$\partial(\boldsymbol{K}, \epsilon) = \left\{\boldsymbol{x} : \exists \boldsymbol{x}' \in \partial \boldsymbol{K}, \left\|\boldsymbol{x} - \boldsymbol{x}'\right\|_2 \leq \epsilon\right\}$$

The key probabilistic tool used in the analysis is Lemma 2.3.4, which we will derive from the following result of Ball [Bal93]. A slightly weaker version of this lemma was proved in [BD02] and also in [BR76, pages 23-38].

Theorem 2.3.3 (Ball [Bal93]). *Let μ be the density function of a d-dimensional vector of independent Gaussian random variables of mean 0 and variance 1. Then for any convex body $\boldsymbol{K}$ in $\mathbb{R}^d$,*

$$\int_{\partial \boldsymbol{K}} \mu \leq 4d^{1/4}.$$

Lemma 2.3.4 (Hitting $\partial(\boldsymbol{K}, \epsilon)$). *Let $\boldsymbol{K}$ be a convex body in $\mathbb{R}^d$, let $\bar{\boldsymbol{x}} \in \mathbb{R}^d$, and let $\boldsymbol{x}$ be the Gaussian perturbation of $\bar{\boldsymbol{x}}$ of variance σ^2. Then,*

$$\Pr_{\boldsymbol{x}}[\boldsymbol{x} \in \partial(\boldsymbol{K}, \epsilon) \setminus \boldsymbol{K}] \leq \frac{4d^{1/4}\epsilon}{\sigma}, \text{ and} \qquad \textit{(outside boundary)}$$

$$\Pr_{\boldsymbol{x}}[\boldsymbol{x} \in \partial(\boldsymbol{K}, \epsilon) \cap \boldsymbol{K}] \leq \frac{4d^{1/4}\epsilon}{\sigma}. \qquad \textit{(inside boundary)}$$

Proof. We derive the result assuming $\sigma = 1$. The result for general σ follows by scaling.

Let μ denote the density according to which $\boldsymbol{x}$ is distributed. To derive the first inequality, we let $\boldsymbol{K}_\epsilon$ denote the set of points of distance at most ϵ from $\boldsymbol{K}$, and observe that $\boldsymbol{K}_\epsilon$ is convex.

Integrating by shells, we obtain

$$\begin{aligned} \Pr[\boldsymbol{x} \in \partial(\boldsymbol{K}, \epsilon) \setminus \boldsymbol{K}] &\leq \int_{t=0}^{\epsilon} \int_{\partial \boldsymbol{K}_t} \mu \\ &\leq 4d^{1/4}\epsilon, \end{aligned}$$

by Theorem 2.3.3.

We similarly derive the second inequality by defining $\boldsymbol{K}^\epsilon$ to be the set of points inside $\boldsymbol{K}$ of distance at least ϵ from the boundary of $\boldsymbol{K}$ and observing that $\boldsymbol{K}^\epsilon$ is convex for any ϵ. To see that $\boldsymbol{K}^\epsilon$ is convex, note that $\boldsymbol{x} \in \boldsymbol{K}^\epsilon$ if and only if the ball of radius ϵ around $\boldsymbol{x}$ is contained in $\boldsymbol{K}$, and that the ball of radius ϵ around any convex combination of points $\boldsymbol{x}$ and $\boldsymbol{y}$ is contained in the convex hull of the balls of radius ϵ around these points. ☐

In this section and the next, we use the following consequence of Lemma 2.3.4 repeatedly.

Corollary 2.3.5 (Distance to ill-posed, single constraint). *For a vector $\bar{\boldsymbol{a}} \in \mathbb{R}^d$, let $\boldsymbol{a}$ be the Gaussian perturbation of $\bar{\boldsymbol{a}}$ of variance σ^2. Then, for any strongly supported convex cone $\boldsymbol{C}_0$ in $\mathbb{R}^d$*

$$\Pr_{\boldsymbol{a}}\left[\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a}) \text{ is non-empty and } \rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \epsilon\right] \leq \frac{4d^{1/4}\epsilon}{\sigma}, \text{ and}$$
$$\Pr_{\boldsymbol{a}}\left[\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a}) \text{ is empty and } \rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \epsilon\right] \leq \frac{4d^{1/4}\epsilon}{\sigma}.$$

Proof. Let $\boldsymbol{K}$ be the set of $\boldsymbol{a}$ for which $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a}) = \emptyset$, i.e., let $\boldsymbol{K}$ be the interior of the polar cone of $\boldsymbol{C}_0$. Observe that $\rho(\boldsymbol{a}, \boldsymbol{C}_0)$ is exactly the distance from $\boldsymbol{a}$ to the boundary of $\boldsymbol{K}$. Since $\boldsymbol{K}$ is a convex cone, these two inequalities follow directly from the two inequalities of Lemma 2.3.4. □

2.3.1 Feasible Case

In this subsection, we will prove:

Lemma 2.3.6 (Distance to ill-posed: feasible case). *Let $\boldsymbol{C}$ be a strictly-supported convex cone in $\mathbb{R}^d$, let $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$, and let $\mathbf{A}$ be the Gaussian perturbation of $\bar{\mathbf{A}}$ of variance σ^2. Then for any $\sigma \geq 0$ and $\epsilon > 0$,*

$$\Pr_{\mathbf{A}}\left[(\mathbf{A}\boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{x} \in \boldsymbol{C} \text{ is feasible}) \text{ and } (\rho(\mathbf{A}, \boldsymbol{C}) \leq \epsilon)\right] \leq \frac{4nd^{5/4}\epsilon}{\sigma}.$$

To prove Lemma 2.3.6, we first establish a necessary geometric condition for ρ to be small. This condition is stated and proved in Lemma 2.3.7. In Lemma 2.3.8, we apply Helly's Theorem [Hel23, LDK63] to simplify this geometric condition, expressing it in terms of the minimum of ρ over individual constraints. This allows us to use Lemma 2.3.5 to show that this geometric condition is unlikely to be met. Lemma 2.3.6 then follows immediately from Lemmas 2.3.7 and 2.3.11 below.

We remark that a result similar to Lemma 2.3.7 appears in [CC01].

Lemma 2.3.7 (Geometric condition that ρ is small). *For any strictly-supported convex cone $\boldsymbol{C}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^d$ such that $\boldsymbol{C} \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset$,*

$$\rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}) \geq \sup_{\boldsymbol{p} \in \boldsymbol{C}, \|\boldsymbol{p}\|_2 = 1} \min_i \boldsymbol{a}_i^T \boldsymbol{p}.$$

Proof. Let $\boldsymbol{p}$ be any unit vector in $\boldsymbol{C}$, and let $\Delta\boldsymbol{a}_1, \ldots, \Delta\boldsymbol{a}_n$ be such that $\boldsymbol{C} \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i + \Delta\boldsymbol{a}_i) = \emptyset$. As $\boldsymbol{p} \notin \bigcap_i \mathcal{H}(\boldsymbol{a}_i + \Delta\boldsymbol{a}_i)$, there exists an i for which $(\boldsymbol{a}_i + \Delta\boldsymbol{a}_i)^T \boldsymbol{p} < 0$, which implies

$$\min_i \boldsymbol{a}_i^T \boldsymbol{p} \leq \boldsymbol{a}_i^T \boldsymbol{p} < -\Delta\boldsymbol{a}_i^T \boldsymbol{p} \leq \|\Delta\boldsymbol{a}_i\|_2 \|\boldsymbol{p}\|_2 = \|\Delta\boldsymbol{a}_i\|_2 \leq \rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}).$$

□

Lemma 2.3.8 (Consequnce of Helly's Theorem). *For any strictly-supported convex cone* $\boldsymbol{C} \subseteq \mathbb{R}^d$ *and* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^d$ *such that* $\boldsymbol{C} \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset$,

$$\sup_{\boldsymbol{p} \in \boldsymbol{C}, \|\boldsymbol{p}\|_2 = 1} \min_i \boldsymbol{a}_i^T \boldsymbol{p} \geq \min_i \rho \left(\boldsymbol{a}_i, \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j) \right) \Big/ d.$$

We will derive Lemma 2.3.8 from Lemmas 2.3.9 and 2.3.10, which we now state and prove.

Lemma 2.3.9 (Cone to unit vector, single constraint). *For any strictly-supported convex cone* $\boldsymbol{C}_0 \subset \mathbb{R}^d$ *and* $\boldsymbol{a} \in \mathbb{R}^d$ *such that* $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a}) \neq \emptyset$,

$$\rho(\boldsymbol{a}, \boldsymbol{C}_0) = \sup_{\boldsymbol{p} \in \boldsymbol{C}_0, \|\boldsymbol{p}\|_2 = 1} \boldsymbol{a}^T \boldsymbol{p}.$$

Proof. The "$\geq$" direction follows from Lemma 2.3.7. So it suffices to show

$$\rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \sup_{\boldsymbol{p} \in \boldsymbol{C}_0, \|\boldsymbol{p}\|_2 = 1} \boldsymbol{a}^T \boldsymbol{p}.$$

As $\boldsymbol{C}_0$ is strictly-supported, there exists a vector $\boldsymbol{t}$ such that $\boldsymbol{t}^T \boldsymbol{x} < 0$ for all $\boldsymbol{x} \in \boldsymbol{C}_0$. We now divide the proof into two cases depending on whether $\boldsymbol{a} \in \boldsymbol{C}_0$.

If $\boldsymbol{a} \in \boldsymbol{C}_0$, then we let $\boldsymbol{p} = \boldsymbol{a} / \|\boldsymbol{a}\|_2$. It is easy to verify that

$$\boldsymbol{a}^T \boldsymbol{p} = \|\boldsymbol{a}\|_2 = \max_{\|\boldsymbol{p}\|_2 = 1} \boldsymbol{a}^T \boldsymbol{p} = \sup_{\boldsymbol{p} \in \boldsymbol{C}_0, \|\boldsymbol{p}\|_2 = 1} \boldsymbol{a}^T \boldsymbol{p}.$$

Moreover, $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a} - (\boldsymbol{a} - \epsilon \boldsymbol{t})) = \emptyset$ for every $\epsilon > 0$. So, $\rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \|\boldsymbol{a}\|_2 = \boldsymbol{a}^T \boldsymbol{p}$.

If $\boldsymbol{a} \notin \boldsymbol{C}_0$, we consider $\mathrm{cl}\,(\boldsymbol{C}_0)$, the closure of $\boldsymbol{C}_0$. There are two cases depending on whether

$$\sup_{\boldsymbol{p} \in \boldsymbol{C}_0, \|\boldsymbol{p}\|_2 = 1} \boldsymbol{a}^T \boldsymbol{p} \tag{8}$$

is equal to 0. If this quantity is equal to 0, then $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a} + \epsilon \boldsymbol{t}) = \emptyset$ for every ϵ, because for any $\boldsymbol{q} \in \boldsymbol{C}_0$, $\boldsymbol{a} + \epsilon \boldsymbol{t}^T \boldsymbol{q} = \boldsymbol{a}^T \boldsymbol{q} + \epsilon \boldsymbol{t}^T \boldsymbol{q} < 0$. Thus, $\rho(\boldsymbol{a}, \boldsymbol{C}_0) = 0$. If the quantity (8) is greater than 0, let $\boldsymbol{q}$ be the point of $\mathrm{cl}\,(\boldsymbol{C}_0)$ that is closest to $\boldsymbol{a}$. As the quantity (8) is strictly positive, $\boldsymbol{q}$ lies inside $\mathcal{H}(\boldsymbol{a})$ and is not the origin. Let $\boldsymbol{p} = \boldsymbol{q} / \|\boldsymbol{q}\|_2$. As $\boldsymbol{C}_0$ is a cone, $\boldsymbol{q}$ is perpendicular to $\boldsymbol{a} - \boldsymbol{q}$, and so $\boldsymbol{q}^T \boldsymbol{a} = \boldsymbol{q}^T \boldsymbol{q}$. Thus, the distance from $\boldsymbol{a}$ to $\boldsymbol{q}$ is $\sqrt{\|\boldsymbol{a}\|_2^2 - \|\boldsymbol{q}\|_2^2} = \sqrt{\|\boldsymbol{a}\|_2^2 - (\boldsymbol{a}^T \boldsymbol{p})^2}$, as $\boldsymbol{a}^T \boldsymbol{p} = \|\boldsymbol{q}\|_2$. Conversely, for any unit vector $\boldsymbol{r} \in \mathrm{cl}\,(\boldsymbol{C}_0)$, the distance from $\mathbf{Ray}\,(\boldsymbol{r})$ to $\boldsymbol{a}$ is $\sqrt{\|\boldsymbol{a}\|_2^2 - (\boldsymbol{a}^T \boldsymbol{r})^2}$. Thus, the unit vector $\boldsymbol{r} \in \mathrm{cl}\,(\boldsymbol{C}_0)$ maximizing $\boldsymbol{a}^T \boldsymbol{r}$ must be $\boldsymbol{p}$.

As $\boldsymbol{C}_0$ is convex, the hyperplane through the origin and $\boldsymbol{q}$ that is orthogonal to $\boldsymbol{a} - \boldsymbol{q}$ separates $\boldsymbol{C}_0$ from $\boldsymbol{a}$. So, $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a} - \boldsymbol{q} + \epsilon \boldsymbol{t}) = \emptyset$ for any $\epsilon > 0$, and so $\rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \|\boldsymbol{q}\|_2 = \boldsymbol{a}^T \boldsymbol{p}$.

□

Lemma 2.3.10 (Helly Reduction). *For a strictly-supported convex cone* $\boldsymbol{C}_0 \subset \mathbb{R}^d$, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^d$ *and* $\epsilon > 0$, *if there exist unit vectors* $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n \in \boldsymbol{C}_0$, *such that*

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{p}_i &\geq \epsilon, \text{ for all } i, \text{ and} \\ \boldsymbol{a}_i^T \boldsymbol{p}_j &\geq 0, \text{ for all } i \text{ and } j, \end{aligned}$$

then there exists a unit vector $\boldsymbol{p} \in \boldsymbol{C}_0$ *such that*

$$\boldsymbol{a}_i^T \boldsymbol{p} \geq \epsilon/d, \text{ for all } i.$$

Proof. We prove this lemma using Helly's Theorem [Hel23, LDK63] which states that if a collection of convex sets in $\mathbb{R}^d$ has the property that every sub-collection of $d+1$ of the sets has a common point, then the entire collection has a common point. Let

$$\mathbf{S}_i = \{\boldsymbol{x} \in \boldsymbol{C}_0 : \boldsymbol{a}_i^T \boldsymbol{x} / \|\boldsymbol{x}\|_2 \geq \epsilon/d\}.$$

We begin by proving that every d of the $\mathbf{S}_i$s contain a point in common. Without loss of generality, we consider $\mathbf{S}_1, \dots, \mathbf{S}_d$. Let $\boldsymbol{q} = \sum_{i=1}^d \boldsymbol{p}_i/d$. Then, for each $1 \leq j \leq d$,

$$\boldsymbol{a}_j^T \boldsymbol{q} = \boldsymbol{a}_j^T \left(\sum_{i=1}^d \boldsymbol{p}_i/d \right) \geq \boldsymbol{a}_j^T \left(\boldsymbol{p}_j/d \right) \geq \epsilon/d.$$

As $\|\boldsymbol{q}\|_2 = \left\| \sum_{i=1}^d \boldsymbol{p}_i/d \right\|_2 \leq \left(\sum_i^d \|\boldsymbol{p}_i\|_2 \right)/d \leq 1$, $\boldsymbol{a}_j^T \boldsymbol{q} / \|\boldsymbol{q}\|_2 \geq \boldsymbol{a}_j^T \boldsymbol{q}$, so $\boldsymbol{q}$ is contained in each of $\mathbf{S}_1, \dots, \mathbf{S}_d$.

As $\boldsymbol{C}_0$ is strictly-supported, there exists $\boldsymbol{t}$ such that $\boldsymbol{t}^T \boldsymbol{x} < 0$, $\forall \boldsymbol{x} \in \boldsymbol{C}_0$. Let

$$\mathbf{S}_i' = \mathbf{S}_i \bigcap \{\boldsymbol{x} : \boldsymbol{t}^T \boldsymbol{x} = -1\}.$$

Then, $\boldsymbol{x} \in \mathbf{S}_i$ implies $-\boldsymbol{x}/\boldsymbol{t}^T \boldsymbol{x} \in \mathbf{S}_i'$. So, every d of the $\mathbf{S}_i'$ have a point in common. As these are convex sets lying in a $d-1$ dimensional space, by Helly's Theorem there exists a point $\boldsymbol{q}$ that lies within all of the $\mathbf{S}_i'$s. As $\mathbf{S}_i' \subset \mathbf{S}_i$, this point lies inside all the $\mathbf{S}_i$s. So, $\boldsymbol{p} = \boldsymbol{q}/\|\boldsymbol{q}\|_2$ is the desired unit vector. □

Proof of Lemma 2.3.8. For each i, we apply Lemma 2.3.9 to the vector $\boldsymbol{a}_i$ and the cone $\mathrm{cl}(\boldsymbol{C}) \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)$ to find a unit vector $\boldsymbol{p}_i \in \mathrm{cl}(\boldsymbol{C}) \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)$ such that

$$\boldsymbol{a}_i^T \boldsymbol{p}_i = \rho \left(\boldsymbol{a}_i, \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j) \right).$$

As $\boldsymbol{C} \cap \bigcap_{j=1}^n \mathcal{H}(\boldsymbol{a}_j)$ is not empty, $\boldsymbol{a}_i^T \boldsymbol{p}_i \geq 0$. As $\boldsymbol{p}_i \in \mathrm{cl}(\boldsymbol{C}) \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)$, we also have

$$\boldsymbol{a}_j^T \boldsymbol{p}_i \geq 0,$$

for all j. Applying Lemma 2.3.10 with $\boldsymbol{C}_0 = \mathrm{cl}(\boldsymbol{C}) \cap \bigcap_{j=1}^n \mathcal{H}(\boldsymbol{a}_j) - \{\boldsymbol{0}\}$, we find a unit vector $\boldsymbol{p} \in \boldsymbol{C}_0$ satisfying

$$\boldsymbol{a}_i^T \boldsymbol{p} \geq \min_i \rho \left(\boldsymbol{a}_i, \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j) \right) \Bigg/ d,$$

for all i. □

Lemma 2.3.11 (Probability of the geometric condition). *Let $\boldsymbol{C}$ be a strictly-supported convex cone in $\mathbb{R}^d$, let $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n \in \mathbb{R}^d$, and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the Gaussian perturbations of $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n$ of variance σ^2. Then,*

$$\mathbf{Pr}\left[C \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset \quad \text{and} \quad \sup_{\boldsymbol{p} \in \boldsymbol{C}, \|\boldsymbol{p}\|_2 = 1} \min_i \boldsymbol{a}_i^T \boldsymbol{p} < \epsilon\right] \leq \frac{4nd^{5/4}\epsilon}{\sigma}.$$

Proof. By Lemma 2.3.8,

$$\mathbf{Pr}\left[C \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset \quad \text{and} \quad \sup_{\boldsymbol{p} \in \boldsymbol{C}, \|\boldsymbol{p}\|_2 = 1} \min_i \boldsymbol{a}_i^T \boldsymbol{p} < \epsilon\right]$$
$$\leq \mathbf{Pr}\left[C \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i) \neq \emptyset \quad \text{and} \ \min_i \rho\left(\boldsymbol{a}_i, \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)\right) < d\epsilon\right].$$

Applying a union bound over i and then Corollary 2.3.5, we find this probability is at most

$$\sum_{i=1}^n \mathbf{Pr}\left[C \cap \bigcap_j \mathcal{H}(\boldsymbol{a}_j) \neq \emptyset \quad \text{and} \ \rho\left(\boldsymbol{a}_i, \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)\right) < d\epsilon\right] \leq \sum_{i=1}^n \frac{4d^{1/4}(d\epsilon)}{\sigma} = \frac{4nd^{5/4}\epsilon}{\sigma}.$$

□

2.3.2 Infeasible Case

In this subsection, we prove:

Lemma 2.3.12 (Distance to ill-posed: infeasible case). *Let $\boldsymbol{C}$ be a strictly-supported convex cone $\mathbb{R}^d$, let $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$, and let $\mathbf{A}$ be the Gaussian perturbation of $\bar{\mathbf{A}}$ of variance σ^2. Then for any $0 < \sigma \leq 1/\sqrt{d}$ and $0 \leq \epsilon \leq 1/2$,*

$$\mathop{\mathbf{Pr}}_{\mathbf{A}}\left[(\mathbf{A}\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \textit{ is infeasible}) \textit{ and } (\rho(\mathbf{A}, \boldsymbol{C}) \leq \epsilon)\right] \quad \leq \quad \left(\frac{97n^2d^{1.5}}{\sigma^2}\right)\left(\epsilon \ln^{1.5}\left(\frac{1}{\epsilon}\right)\right) \quad (9)$$

To prove Lemma 2.3.12, we consider adding the constraints one at a time. If the program is infeasible in the end, then there must be some constraint, which we call the *critical* constraint, that takes it from being feasible to being infeasible. In Lemma 2.3.14, we give a geometric quantity that is a lower bound of the distance to ill-posedness when the critical constraint is added. We then bound the probability that this geometric quantity is small. We start with an extension of Lemma 2.3.9 to the infeasible case.

Lemma 2.3.13 (ρ bound on inner product). *For any strictly-supported convex cone $\boldsymbol{C} \subset \mathbb{R}^d$ and $\boldsymbol{a} \in \mathbb{R}^d$ such that $\boldsymbol{C} \cap \mathcal{H}(\boldsymbol{a}) = \emptyset$,*

$$\sup_{\boldsymbol{p} \in \boldsymbol{C}, \|\boldsymbol{p}\|_2 = 1} \boldsymbol{p}^T \boldsymbol{a} \leq -\rho(\boldsymbol{a}, \boldsymbol{C}).$$

Proof. For any unit vector $\boldsymbol{p} \in \boldsymbol{C}$ and any $\epsilon > 0$, if we set $\Delta \boldsymbol{a} = \left(\epsilon - \boldsymbol{p}^T \boldsymbol{a}\right) \boldsymbol{p}$, then

$$\boldsymbol{p}^T (\boldsymbol{a} + \Delta \boldsymbol{a}) = \boldsymbol{p}^T \boldsymbol{a} + \left(\epsilon - \boldsymbol{p}^T \boldsymbol{a}\right) \boldsymbol{p}^T \boldsymbol{p} = \boldsymbol{p}^T \boldsymbol{a} + \left(\epsilon - \boldsymbol{p}^T \boldsymbol{a}\right) = \epsilon,$$

implying $\boldsymbol{C} \cap \mathcal{H}(\boldsymbol{a} + \Delta \boldsymbol{a}) \neq \emptyset$. Thus $\rho(\boldsymbol{a}, \boldsymbol{C}) \leq \left|\boldsymbol{p}^T \boldsymbol{a}\right|$. □

Lemma 2.3.14 (Geometric bound of the feasible-to-infeasible transition). *Let $n \geq d \geq 2$, let $\boldsymbol{C}$ be a strictly-supported convex cone in $\mathbb{R}^d$, let $\boldsymbol{p} \in \boldsymbol{C}$, let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1} \in \mathbb{R}^d$, and let α and β be positive. If*

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{p} &\geq \alpha, \text{ for } 1 \leq i \leq k, \text{ and} \\ \boldsymbol{a}_{k+1}^T \boldsymbol{x} &\leq -\beta, \text{ for all } \boldsymbol{x} \in \boldsymbol{C} \cap \textstyle\bigcap_{i=1}^k \mathcal{H}(\boldsymbol{a}_i), \ \ \|\boldsymbol{x}\|_2 = 1. \end{aligned}$$

then

$$\rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}], \boldsymbol{C}) \geq \min\left\{\frac{\alpha}{2}, \frac{\alpha\beta}{4\alpha + 2\|\boldsymbol{a}_{k+1}\|_2}\right\}.$$

Proof. Let

$$\epsilon = \min\left\{\frac{\alpha}{2}, \frac{\alpha\beta}{4\alpha + 2\|\boldsymbol{a}_{k+1}\|_2}\right\}. \tag{10}$$

It sufficient to show for any $\{\Delta \boldsymbol{a}_1, \ldots, \Delta \boldsymbol{a}_{k+1}\}$ satisfying $\|\Delta \boldsymbol{a}_i\|_2 < \epsilon$ for $1 \leq i \leq k+1$,

$$\boldsymbol{C} \cap \bigcap_{i=1}^{k+1} \mathcal{H}(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i) = \emptyset.$$

Assume by way of contradiction that there exists a unit vector $\boldsymbol{x}' \in \boldsymbol{C} \cap \bigcap_{i=1}^{k+1} \mathcal{H}(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i)$. Then for any $i \leq k$, $(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i)^T \boldsymbol{x}' \geq 0$, implying

$$\boldsymbol{a}_i^T \boldsymbol{x}' \geq -\Delta \boldsymbol{a}_i^T \boldsymbol{x}' \geq -\|\Delta \boldsymbol{a}_i\|_2 \left\|\boldsymbol{x}'\right\|_2 \geq -\epsilon.$$

Let $\boldsymbol{x} = \boldsymbol{x}' + \frac{\epsilon}{\alpha} \boldsymbol{p}$. Then

$$\boldsymbol{a}_i^T \boldsymbol{x} = \boldsymbol{a}_i^T \left(\boldsymbol{x}' + \frac{\epsilon}{\alpha} \boldsymbol{p}\right) = \boldsymbol{a}_i^T \boldsymbol{x}' + \boldsymbol{a}_i^T \frac{\epsilon}{\alpha} \boldsymbol{p} \geq -\epsilon + \frac{\epsilon}{\alpha} \alpha \geq 0.$$

As $\boldsymbol{x}' \in \boldsymbol{C}$ and $\boldsymbol{p} \in \boldsymbol{C}$, and hence $\boldsymbol{x}' + \frac{\epsilon}{\alpha} \boldsymbol{p} \in \boldsymbol{C}$, we have $\boldsymbol{x} \in \boldsymbol{C} \cap \bigcap_{i=1}^k \mathcal{H}(\boldsymbol{a}_i)$. Note also that $1 - \epsilon/\alpha \leq \|\boldsymbol{x}\|_2 \leq 1 + \epsilon/\alpha$.

To derive a contradiction, we now compute

$$\begin{aligned} (\boldsymbol{a}_{k+1} + \Delta \boldsymbol{a}_{k+1})^T \boldsymbol{x}' &= (\boldsymbol{a}_{k+1} + \Delta \boldsymbol{a}_{k+1})^T (\boldsymbol{x} - (\epsilon/\alpha) \boldsymbol{p}) \\ &= \boldsymbol{a}_{k+1}^T \boldsymbol{x} + \Delta \boldsymbol{a}_{k+1}^T \boldsymbol{x} - (\epsilon/\alpha) \boldsymbol{a}_{k+1}^T \boldsymbol{p} - (\epsilon/\alpha) \Delta \boldsymbol{a}_{k+1}^T \boldsymbol{p} \\ &\leq -\beta \|\boldsymbol{x}\|_2 + \|\Delta \boldsymbol{a}_{k+1}\|_2 \|\boldsymbol{x}\|_2 + (\epsilon/\alpha) \|\boldsymbol{a}_{k+1}\|_2 + (\epsilon/\alpha) \|\Delta \boldsymbol{a}_{k+1}\|_2 \\ &< -\beta(1 - \epsilon/\alpha) + \epsilon(1 + \epsilon/\alpha) + (\epsilon/\alpha) \|\boldsymbol{a}_{k+1}\|_2 + (\epsilon^2/\alpha) \\ &= -\beta(1 - \epsilon/\alpha) + \epsilon\big((1 + \epsilon/\alpha) + \|\boldsymbol{a}_{k+1}\|_2 / \alpha + \epsilon/\alpha\big) \\ &\leq -\beta/2 + \epsilon \left(2 + \|\boldsymbol{a}_{k+1}\|_2 / \alpha\right), && \text{by } \epsilon \leq \alpha/2 \\ &\leq 0, && \text{by (10)}, \end{aligned}$$

which contradicts $\boldsymbol{x}' \in \boldsymbol{C} \cap \bigcap_{i=1}^{k+1} \mathcal{H}(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i)$ □

We now prove Lemma 2.3.12 by proving that it is reasonably likely that there exist α and β that are not too small and satisfy the conditions Lemma 2.3.14.

Proof of Lemma 2.3.12. First note that for $1/100 < \epsilon \leq 1/2$, the right-hand side (9) is at least 1 and so the lemma is trivially true. So in the remainder of the proof, we assume $\epsilon \leq 1/100$. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the rows of A, and for $k \in \{0, 1, \ldots, n\}$, let

$$\boldsymbol{C}_0 = \boldsymbol{C} \quad \text{and} \quad \boldsymbol{C}_k = \boldsymbol{C} \cap \bigcap_{i=1}^{k} \mathcal{H}(\boldsymbol{a}_k).$$

We observe the following simple monotonicity property of the distance to ill-posedness for infeasible conic programs: if $\boldsymbol{C}_k = \emptyset$ then $\boldsymbol{C}_{k+1} = \emptyset$, and so

$$\rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k], \boldsymbol{C}\right) \leq \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}\right).$$

Let E_k denote the event that $\boldsymbol{C}_{k-1} \neq \emptyset$ but $\boldsymbol{C}_k = \emptyset$. We have

$$\mathbf{Pr}\left[\boldsymbol{C}_n = \emptyset \text{ and } \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}\right) \leq \epsilon\right] \leq \sum_{k=0}^{n-1} \mathbf{Pr}\left[E_{k+1} \text{ and } \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}], \boldsymbol{C}\right) \leq \epsilon\right]. \tag{11}$$

If E_{k+1} occurs, then $\boldsymbol{C}_k \neq \emptyset$, and we may define a non-negative quantity

$$\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) = \sup_{\boldsymbol{p} \in \boldsymbol{C}_k, \|\boldsymbol{p}\|_2 = 1} \min_{1 \leq i \leq k} \boldsymbol{a}_i^T \boldsymbol{p}. \tag{12}$$

E_{k+1} and Lemma 2.3.13 imply that for all $\boldsymbol{x} \in \boldsymbol{C}_k$ such that $\|\boldsymbol{x}\|_2 = 1$, $\boldsymbol{a}_{k+1}^T \boldsymbol{x} \leq -\rho\left(\boldsymbol{a}_{k+1}, \boldsymbol{C}\right)$. If we let $\boldsymbol{p}$ be the unit vector at which the supremum in (12) is achieved, we may apply Lemma 2.3.14 and Lemma A.2.6 to show that E_{k+1} implies

$$\begin{aligned}
\rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}], \boldsymbol{C}\right) &\geq \min\left\{\frac{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)}{2}, \frac{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)}{4 + 2\left\|\boldsymbol{a}_{k+1}\right\|_2}, \frac{\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)}{4 + 2\left\|\boldsymbol{a}_{k+1}\right\|_2}\right\} \\
&\geq \frac{\min\left\{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k),\ \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k),\ \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)\right\}}{4 + 2\left\|\boldsymbol{a}_{k+1}\right\|_2}.
\end{aligned} \tag{13}$$

For any $0 < \lambda$, Corollary 2.3.5 and Lemma 2.3.11, respectively, imply

$$\begin{aligned}
\forall_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k : \boldsymbol{C}_k \neq \emptyset}, \mathop{\mathbf{Pr}}_{\boldsymbol{a}_{k+1}}\left[E_{k+1} \text{ and } \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq \lambda\right] &\leq \frac{4d^{1/4}\lambda}{\sigma}, \\
\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k}\left[\boldsymbol{C}_k \text{ is feasible and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq \lambda\right] &\leq \frac{4nd^{5/4}\lambda}{\sigma}.
\end{aligned}$$

By Proposition A.2.4, we have for $0 < \lambda \leq 1/e$

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}}\left[E_{k+1} \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq \lambda\right] &\leq \frac{16nd^{1.5}\lambda}{\sigma^2}\left(1 + \max\left(0, \ln\frac{1}{\lambda} + \ln\frac{\sigma^2}{16nd^{1.5}}\right)\right). \\
&\leq \frac{16nd^{1.5}\lambda}{\sigma^2} \ln\frac{1}{\lambda}.
\end{aligned}$$

Applying Proposition A.2.5 to add up these three bounds, we obtain

$$\begin{aligned}\mathbf{Pr}\left[E_{k+1} \text{ and } \min\left\{\kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k), \kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k), \rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k)\right\} < \lambda\right] \\ \leq \lambda\left(\frac{(16\ln(1/\lambda)+8)nd^{1.5}}{\sigma^2}\right). \qquad (14)\end{aligned}$$

To complete the proof, we now consider the denominator of (13). Because $\boldsymbol{a}_{k+1}$ is the Gaussian perturbation of variance σ^2 of $\bar{\boldsymbol{a}}_{k+1}$ and $\|\bar{\boldsymbol{a}}_{k+1}\|_2 \leq 1$,

$$\begin{aligned}\mathbf{Pr}\left[4+2\|\boldsymbol{a}_{k+1}\|_2 \geq 6\ln^{1/2}1/\epsilon\right] &\leq \mathbf{Pr}\left[4+2\|\boldsymbol{a}_{k+1}\|_2 \geq 6+3\ln^{1/2}1/\epsilon\right] \\ &\leq \mathbf{Pr}\left[\|\boldsymbol{a}_{k+1}\|_2 \geq 1.5\ln^{1/2}1/\epsilon\right] \leq \epsilon, \qquad (15)\end{aligned}$$

where the first inequality follows from $6 \leq 3\ln^{1/2}(1/\epsilon)$ when $\epsilon \leq 1/100$ and the last inequality follows from Proposition A.1.3.

We now set $\lambda = 6\epsilon\ln^{1/2}(1/\epsilon)$. We note that $\epsilon \leq 1/100$ implies $\lambda \leq 1/11 \leq 1/e$. Observe that the event

$$\frac{\min\left\{\kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k),\ \kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k),\ \rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k)\right\}}{4+2\|\boldsymbol{a}_{k+1}\|_2} \leq \epsilon$$

would imply

$$\min\left\{\kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k),\ \kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k),\ \rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k)\right\} < \lambda, \text{ or } 4+2\|\boldsymbol{a}_{k+1}\|_2 \geq 6\ln^{1/2}(1/\epsilon).$$

So, we may apply (13), (14) and (15) to obtain

$$\begin{aligned}&\mathbf{Pr}\left[E_{k+1} \text{ and } \rho\left([\boldsymbol{a}_1,\ldots,\boldsymbol{a}_{k+1}],\boldsymbol{C}\right) \leq \epsilon\right] \\ &\leq \epsilon + \left(6\epsilon\ln^{1/2}\left(\frac{1}{\epsilon}\right)\right)\left(\frac{16nd^{1.5}}{\sigma^2}\right)\left(\left(\ln\left(\frac{1}{6\epsilon\ln^{1/2}(1/\epsilon)}\right)+\frac{1}{2}\right)\right) \\ &\leq \epsilon + \left(\frac{96nd^{1.5}}{\sigma^2}\epsilon\ln^{1/2}\left(\frac{1}{\epsilon}\right)\right)\left(\ln\left(\frac{1}{\epsilon}\right)-\ln 6+1/2\right) \\ &\leq \epsilon + \frac{96nd^{1.5}}{\sigma^2}\epsilon\ln^{1.5}\left(\frac{1}{\epsilon}\right) \leq \frac{97nd^{1.5}}{\sigma^2}\epsilon\ln^{1.5}\left(\frac{1}{\epsilon}\right).\end{aligned}$$

Plugging this in to (11), we get

$$\mathbf{Pr}\left[\boldsymbol{C}_0 \text{ is infeasible and } \rho\left([\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n],\boldsymbol{C}\right) \leq \epsilon\right] \leq \frac{97n^2d^{1.5}\epsilon\ln^{1.5}(1/\epsilon)}{\sigma^2}.$$

□

2.3.3 Primal condition number, putting the feasible and infeasible cases together

We now combine the results of Sections 2.3.1 and 2.3.2 to prove Lemma 2.3.2.

Proof of Lemma 2.3.2. Note that for $1/100 < \epsilon \leq 1/2$ the right-hand side of (7) is at least 1 and so the lemma is trivially true. So we assume $\epsilon \leq 1/100$. By Lemma 2.3.6 and Lemma 2.3.12, respectively

$$\mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is feasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon)\right] \leq \frac{4nd^{5/4}\epsilon}{\sigma},$$

$$\mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is infeasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon)\right] \leq \frac{97n^2d^{1.5}\epsilon\ln^{1.5}(1/\epsilon)}{\sigma^2}.$$

Thus, as $0 < \epsilon \leq 1/100$

$$\begin{aligned}\mathbf{Pr}\left[\rho(A, \boldsymbol{C}) \leq \epsilon\right] &= \mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is feasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon)\right] \\ &\quad + \mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is infeasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon)\right] \\ &\leq \frac{98n^2d^{1.5}\epsilon\ln^{1.5}(1/\epsilon)}{\sigma^2}.\end{aligned}$$

□

Lemma 2.3.15 (Primal conic condition number). *Let $n \geq d \geq 2$, let $\boldsymbol{C}$ be a strictly-supported convex cone in $\mathbb{R}^d$, let $\bar{\mathbf{A}} \in \mathbb{R}^{n\times d}$ satisfy $\left\|\bar{\mathbf{A}}\right\|_F \leq 1$, and let $\mathbf{A}$ be the Gaussian perturbation of $\bar{\mathbf{A}}$ of variance σ^2. Then for any $x \geq 1$ and $\sigma \leq 1/\sqrt{nd}$,*

$$\mathbf{Pr}\left[\frac{\|A\|_F}{\rho(A, \boldsymbol{C})} \geq x\right] \leq \frac{197n^2d^{1.5}}{\sigma^2}\left(\frac{\ln^2 x}{x}\right).$$

Proof. The lemma is trivially true for any $1 \leq x \leq 100$ as, in this case, the right-hand side is at least 1. So in the rest of the proof, we will assume $x \geq 100$. For any $\epsilon < 1/100$, $0.5\ln^{1/2}(1/\epsilon) \geq 1$, so

$$\begin{aligned}\mathbf{Pr}\left[\|\mathbf{A}\|_F \geq 2\ln^{1/2}(1/\epsilon)\right] &\leq \mathbf{Pr}\left[\|\mathbf{A}\|_F \geq 1 + 1.5\ln^{1/2}(1/\epsilon)\right] \\ &\leq \mathbf{Pr}\left[\left\|\mathbf{A} - \bar{\mathbf{A}}\right\|_F \geq 1.5\ln^{1/2}(1/\epsilon)\right] \\ &\leq \epsilon \text{ (by Proposition A.1.3 and } \sigma \leq 1/\sqrt{nd}).\end{aligned}$$

By Lemma 2.3.2 for any $\lambda \leq 1/2$

$$\mathbf{Pr}\left[\rho(A, \boldsymbol{C}) \leq \lambda\right] \leq \frac{98n^2d^{1.5}\lambda\ln^{1.5}(1/\lambda)}{\sigma^2}.$$

Let $\epsilon = 1/x$ and $\lambda = 2\epsilon\ln^{1/2}(1/\epsilon)$. As $x \geq 100$ implies $\lambda \leq 1/2$, we have

$$\begin{aligned}\mathbf{Pr}\left[\frac{\|\mathbf{A}\|_F}{\rho(\mathbf{A}, \boldsymbol{C})} \geq x\right] &\leq \mathbf{Pr}\left[\|\mathbf{A}\|_F \geq 2\ln^{1/2}(1/\epsilon)\right] + \mathbf{Pr}\left[\rho(\mathbf{A}, \boldsymbol{C}) \leq \lambda\right] \\ &\leq \epsilon + \left(\frac{98n^2d^{1.5}}{\sigma^2}\right)\left(2\epsilon\ln^{1/2}(1/\epsilon)\right)\ln^{1.5}\left(\frac{1}{2\epsilon\ln^{1/2}(1/\epsilon))}\right) \\ &\leq \epsilon + \left(\frac{196n^2d^{1.5}}{\sigma^2}\right)\left(\epsilon\ln^2(1/\epsilon)\right) \leq \left(\frac{197n^2d^{1.5}}{\sigma^2}\right)\left(\frac{\ln^2 x}{x}\right).\end{aligned}$$

□

3 Dual Condition Number

In this section, we consider linear programs of the form

$$\mathbf{A}^T \boldsymbol{y} = \boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0}.$$

The dual program of form (1) and the primal program of form (3) are both of this type. The dual program of form (4) can be handled using a slightly different argument than the one we present. As in section 2, we omit the details of the modifications necessary for form (4). We consider the following dual distance to ill-posedness.

Definition 3.0.1 (Dual distance to ill-posed). *For* $\mathbf{A} \in \mathbb{R}^{n \times d}$ *and* $\boldsymbol{c} \in \mathbb{R}^d$,

a. if $\mathbf{A}^T \boldsymbol{y} = \boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0}$ *is feasible, then*

$$\rho(\mathbf{A}, \boldsymbol{c}) = \sup \left\{ \epsilon : \|\Delta\mathbf{A}, \Delta\boldsymbol{c}\|_F < \epsilon \textit{ implies } (\mathbf{A} + \Delta\mathbf{A})^T \boldsymbol{y} = \boldsymbol{c} + \Delta\boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0} \textit{ is feasible} \right\}$$

b. if $\mathbf{A}^T \boldsymbol{y} = \boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0}$ *is infeasible, then*

$$\rho(\mathbf{A}, \boldsymbol{c}) = \sup \left\{ \epsilon : \|\Delta\mathbf{A}, \Delta\boldsymbol{c}\|_F < \epsilon \textit{ implies } (\mathbf{A} + \Delta\mathbf{A})^T \boldsymbol{y} = \boldsymbol{c} + \Delta\boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0} \textit{ is infeasible} \right\}$$

The main result of this section is:

Lemma 3.0.2 (Logarithm of dual condition number). *For any* $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$ *and* $\bar{\mathbf{c}} \in \mathbb{R}^d$ *such that* $\left\|\bar{\mathbf{A}}, \bar{\mathbf{c}}\right\|_F \leq 1$, *and for any* $\sigma \leq 1/\sqrt{(n+1)d}$, *if* $\mathbf{A}$ *and* $\boldsymbol{c}$ *are the Gaussian perturbations of variance* σ^2 *of* $\bar{\mathbf{A}}$ *and* $\bar{\mathbf{c}}$, *respectively, then*

$$\mathop{\mathbf{E}}_{(\mathbf{A}, \boldsymbol{c})} \left[\log_2 \frac{\|\mathbf{A}, \boldsymbol{c}\|_F}{\rho(\mathbf{A}, \boldsymbol{c})} \right] \leq 15 + 4 \log_2 \frac{nd}{\sigma}.$$

As with the primal conic condition number, we first develop a geometric characterization of linear programs with large dual condition numbers. We then give a smoothed analysis of this geometric condition.

3.1 Geometric conditions of dual condition numbers

For a set of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, let $\mathbf{Cone}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)$ denote $\{\boldsymbol{x} : \boldsymbol{x} = \sum_i \lambda_i \boldsymbol{a}_i, \;\; \lambda_i \geq 0\}$, and let $\mathbf{Hull}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)$ denote $\{\boldsymbol{x} : \boldsymbol{x} = \sum_i \lambda_i \boldsymbol{a}_i, \;\; \lambda_i \geq 0, \;\; \sum_i \lambda_i = 1\}$.

We observe that $A^T \boldsymbol{y} = \boldsymbol{c}, \;\; \boldsymbol{y} \geq \mathbf{0}$ is feasible if and only if

$$\boldsymbol{c} \in \mathbf{Cone}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right),$$

and that for $\boldsymbol{c} \neq \mathbf{0}$, this holds if and only if

$$\mathbf{Ray}\left(\boldsymbol{c}\right) \text{ intersects } \mathbf{Hull}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right).$$

For a convex set $\mathbf{S}$, let $\partial(\mathbf{S})$ denote the boundary of $\mathbf{S}$. For $S \subset \mathbb{R}^d$ and $\boldsymbol{x} \in \mathbb{R}^d$, let $\mathbf{dist}(\boldsymbol{x}, \mathbf{S})$ denote the distance of $\boldsymbol{x}$ to $\mathbf{S}$, that is, $\inf\{\epsilon : \exists \boldsymbol{e}, \ \|\boldsymbol{e}\|_2 \leq \epsilon, \ s.t. \ \boldsymbol{x} + \boldsymbol{e} \in \mathbf{S}\}$.

We consider the following change of variables from $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ to $(\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1})$ which will help to simplify our analysis.

$$\boldsymbol{z} = \left(\sum_{i=1}^{n} \boldsymbol{a}_i\right)/n, \text{ and } \boldsymbol{x}_i = \boldsymbol{a}_i - \boldsymbol{z}, \text{ for } i = 1 \text{ to } n-1. \tag{16}$$

For notational convenience, we let $\boldsymbol{x}_n = \boldsymbol{a}_n - \boldsymbol{z}$, although $\boldsymbol{x}_n$ is not independent of $\{\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\}$. In this set of new variables, the ill-posed linear programs are characterized by the following lemma.

Lemma 3.1.1 (Ill-posed in new variables). *For any* $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{n \times d}$ *and* $\boldsymbol{c} \in \mathbb{R}^d$, *let* $\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *be defined by* (16). *Then*

$$\mathbf{A}^T \boldsymbol{y} = \boldsymbol{c}, \ \boldsymbol{y} \geq \mathbf{0}, \ \boldsymbol{c} \neq \mathbf{0} \textit{ is ill-posed if and only if } \boldsymbol{z} \in \partial\left(\mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\right).$$

Proof. First, observe that

$$\begin{aligned} A^T \boldsymbol{y} = \boldsymbol{c}, \ \boldsymbol{y} \geq \mathbf{0} \text{ is feasible} \quad &\Longleftrightarrow \quad \mathbf{Ray}(\boldsymbol{c}) \text{ intersects } \mathbf{Hull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \\ &\Longleftrightarrow \quad \mathbf{Ray}(\boldsymbol{c}) \text{ intersects } \boldsymbol{z} + \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \\ &\Longleftrightarrow \quad \boldsymbol{z} \in \mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n). \end{aligned}$$

If $\boldsymbol{z} \in \partial(\mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n))$, then the system is feasible, but an arbitrarily small change in $\boldsymbol{z}$ can make the system infeasible, so the instance is ill-posed. Conversely, if $\boldsymbol{z} \notin \partial(\mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n))$, then sufficiently small changes in $\boldsymbol{x}$, $\boldsymbol{c}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ will not cause $\boldsymbol{z}$ to cross the boundary, and so the system is well-posed. For a quantitative version of this later statement, see Lemma A.3.1. □

We now establish the following geometric condition.

Lemma 3.1.2 (Distance to ill-posed). *For any* $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{n \times d}$ *and* $\boldsymbol{c} \in \mathbb{R}^d$, *let* $\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *be defined by* (16). *Let*

$$k_1 = \mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\right)\right) \quad \textit{and} \quad k_2 = \|\boldsymbol{c}\|_2.$$

Then

$$\rho(\mathbf{A}, \boldsymbol{c}) \geq \min\left\{\frac{k_1}{8}, \frac{k_2}{4}, \frac{k_1 k_2}{24 \max_i \|\boldsymbol{a}_i\|_2}\right\}.$$

Proof. For any $\Delta\boldsymbol{c}$ and $\Delta\boldsymbol{a}_1, \ldots, \Delta\boldsymbol{a}_n \in \mathbb{R}^d$, let

$$\Delta\boldsymbol{z} = \left(\sum_{i=1}^{n} \Delta\boldsymbol{a}_i\right)/n, \text{ and } \Delta\boldsymbol{x}_i = \Delta\boldsymbol{a}_i - \Delta\boldsymbol{z}, \text{ for } i = 1 \text{ to } n.$$

If $\max_i \|\Delta\boldsymbol{a}_i\|_2 \leq k_1/8$, then clearly $\|\Delta\boldsymbol{z}\|_2 \leq k_1/8$, and $\|\Delta\boldsymbol{x}_i\|_2 \leq k_1/4$. If further

$$\|\Delta\boldsymbol{c}\|_2 \leq \frac{k_1 k_2}{2k_1 + 4(\|\boldsymbol{z}\|_2 + \max_i \|\boldsymbol{x}_i\|_2)},$$

then by Lemma A.3.1, $\mathbf{A}^T \boldsymbol{y} = \boldsymbol{c},\ \boldsymbol{y} \geq \mathbf{0}$ is feasible if and only if $(\mathbf{A} + \Delta\mathbf{A})^T \boldsymbol{y} = \boldsymbol{c} + \Delta\boldsymbol{c},\ \boldsymbol{y} \geq \mathbf{0}$ is feasible, where $\Delta\mathbf{A} = [\Delta\boldsymbol{a}_1, \ldots, \Delta\boldsymbol{a}_n]$. Thus

$$\rho(A, \boldsymbol{c}) \geq \min \left\{ \frac{k_1}{8}, \frac{k_1 k_2}{2k_1 + 4(\|\boldsymbol{z}\|_2 + \max \|\boldsymbol{x}_i\|_2)} \right\} \geq \min \left\{ \frac{k_1}{8}, \frac{k_2}{4}, \frac{k_1 k_2}{8(\|\boldsymbol{z}\|_2 + \max \|\boldsymbol{x}_i\|_2)} \right\},$$

where the last inequality follows from Lemma A.2.7. The lemma then follows from $\|\boldsymbol{z}\|_2 = \|(1/n) \sum \boldsymbol{a}_i\|_2 \leq \max \|\boldsymbol{a}_i\|_2$, and $\|\boldsymbol{x}_i\|_2 \leq \|\boldsymbol{a}_i\|_2 + \|\boldsymbol{z}\|_2 \leq 2 \max_i \|\boldsymbol{a}_i\|_2$. $\square$

3.2 Probabilistic analysis of dual condition numbers

We start with a simple probability lemma which will be useful for our analysis.

Lemma 3.2.1 (Unlikely ill-posed in new variables). *Let $\boldsymbol{c}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$. For any $\bar{\boldsymbol{z}} \in \mathbb{R}^d$ and for any positive σ, if $\boldsymbol{z}$ is the Gaussian perturbation of $\bar{\boldsymbol{z}}$ of variance σ^2/n, then*

$$\Pr_{\boldsymbol{z}} \left[\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) \leq \epsilon\right] \leq \frac{8 d^{1/4} n^{1/2} \epsilon}{\sigma}.$$

Proof. The lemma follows directly from Lemma 2.3.4 because $\mathbf{Ray}(\boldsymbol{c}) - \mathbf{Hull}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is a convex set. $\square$

Lemma 3.2.2 (Dual condition number). *Let $n \geq d \geq 2$, $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$ and $\bar{\mathbf{c}} \in \mathbb{R}^d$ be such that $\left\|\bar{\mathbf{A}}, \bar{\mathbf{c}}\right\|_F \leq 1$, and let $\sigma \leq 1/\sqrt{(n+1)d}$. If $\mathbf{A}$ and $\boldsymbol{c}$ are the Gaussian perturbations of variance σ^2 of $\bar{\mathbf{A}}$ and $\bar{\mathbf{c}}$, respectively, then*

$$\Pr\left[\frac{\|A, \boldsymbol{c}\|_F}{\rho(A, \boldsymbol{c})} > x\right] \leq \left(\frac{2184 d^{1/4} n^{1/2}}{\sigma^2}\right) \frac{\ln^2 x}{x}.$$

Proof. The lemma is trivially true for any $1 \leq x \leq 200$, as in this case the right-hand side is at least 1. So in the remainder of the proof, we will assume $x > 200$. Let $\epsilon = 1/x$, so $0 < \epsilon \leq 1/200$.

Let $\bar{\mathbf{A}} = [\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n]$ and $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$. Define $\boldsymbol{z}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ by (16). We consider the change of variables from $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ to $\boldsymbol{z}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$. By Proposition A.1.4, $\boldsymbol{z}$ is a Gaussian perturbation of $\bar{\boldsymbol{z}} = \frac{1}{n} \sum_i \bar{\boldsymbol{a}}_i$ of variance σ^2/n, and moreover, $\boldsymbol{z}$ is independent of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$. Let

$$k_1 = \mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) \quad \text{and} \quad k_2 = \|\boldsymbol{c}\|_2.$$

By Lemma 3.2.1, for any $\lambda > 0$ and for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$,

$$\Pr_{\boldsymbol{z}}\left[k_1 \leq \lambda\right] = \Pr_{\boldsymbol{z}}\left[\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) \leq \lambda\right] \leq \frac{8 \cdot d^{1/4} n^{1/2} \lambda}{\sigma}.$$

By Proposition A.1.1, for $\lambda > 0$

$$\Pr_{\boldsymbol{c}}\left[k_2 \leq \lambda\right] = \Pr_{\boldsymbol{c}}\left[\|\boldsymbol{c}\|_2 \leq \lambda\right] \leq \Pr_{\boldsymbol{c}}\left[|c_1| \leq \lambda\right] \leq \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sigma}.$$

Because $\boldsymbol{z}$ and $\boldsymbol{c}$ are independent, by Proposition A.2.4, we have for $0 < \lambda \leq 1/e$ and for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$

$$\begin{aligned}\Pr_{\boldsymbol{z},\boldsymbol{c}}[k_1 k_2 < \lambda] &\leq \frac{8\sqrt{2/\pi}d^{1/4}n^{1/2}\lambda}{\sigma^2}\left(1 + \max\left(0, \ln\frac{1}{\lambda} + \ln\frac{\sigma^2}{8\sqrt{2/\pi}d^{1/4}n^{1/2}}\right)\right) \\ &\leq \frac{7d^{1/4}n^{1/2}\lambda}{\sigma^2}\ln\frac{1}{\lambda}.\end{aligned}$$

Applying Proposition A.2.5 to add up these three bounds, we obtain for $\lambda \leq 1/e$ and for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$

$$\Pr_{\boldsymbol{z},\boldsymbol{c}}[\min\{k_1, k_2, k_1 k_2\} < \lambda] \leq \frac{8 \cdot d^{1/4}n^{1/2}\lambda}{\sigma} + \frac{e\lambda}{\sigma} + \frac{7d^{1/4}n^{1/2}\lambda}{\sigma^2}\ln\frac{1}{\lambda} \leq \frac{18d^{1/4}n^{1/2}\lambda}{\sigma^2}\ln\frac{1}{\lambda}. \quad (17)$$

By Lemma 3.1.2, we have

$$\begin{aligned}\frac{\|\mathbf{A}, \boldsymbol{c}\|_F}{\rho(A, \boldsymbol{c})} &\leq \|\mathbf{A}, \boldsymbol{c}\|_F\left(\max\left\{\frac{8}{k_1}, \frac{4}{k_2}, \frac{24\max_i\|\boldsymbol{a}_i\|_2}{k_1 k_2}\right\}\right) &(18)\\ &\leq 24\left(\|\mathbf{A}, \boldsymbol{c}\|_F\max(\max_i\|\boldsymbol{a}_i\|_2, 1)\right)\left(\max\left\{\frac{1}{k_1}, \frac{1}{k_2}, \frac{1}{k_1 k_2}\right\}\right). &(19)\end{aligned}$$

As $\|\mathbf{A}, \boldsymbol{c}\|_F + 1 \geq \max(\max_i\|\boldsymbol{a}_i\|_2, 1)$, we have

$$\begin{aligned}\Pr\left[\max(\max_i\|\boldsymbol{a}_i\|_2, 1) \geq 2.5\ln^{1/2}(1/\epsilon)\right] &\leq \Pr\left[\|\mathbf{A}, \boldsymbol{c}\|_F \geq 2\ln^{1/2}(1/\epsilon)\right], \quad \text{as } 1 \leq 0.5\ln^{1/2}(1/\epsilon),\\ &\leq \Pr\left[\|\mathbf{A}, \boldsymbol{c}\|_F > 1 + 1.5\ln^{1/2}(1/\epsilon)\right]\\ &\leq \epsilon, \quad \text{by Proposition A.1.3, using } \sigma \leq \sqrt{1/(n+1)d}.\end{aligned}$$

Let $\lambda = (2\ln^{1/2}(1/\epsilon))(2.5\ln^{1/2}(1/\epsilon))\epsilon = 5\epsilon\ln(1/\epsilon)$. As $\epsilon \leq 1/200$ implies $\lambda \leq 1/7 \leq 1/e$, we have

$$\begin{aligned}\Pr\left[\frac{\|\mathbf{A}, \boldsymbol{c}\|_F}{24\rho(A, \boldsymbol{c})} > x\right] &\leq \Pr\left[\|\mathbf{A}, \boldsymbol{c}\|_F \geq 2\ln^{1/2}(1/\epsilon)\right] + \Pr\left[\max\left(\max_i\|\boldsymbol{a}_i\|_2, 1\right) \geq 2.5\ln^{1/2}(1/\epsilon)\right]\\ &\quad + \Pr[\min(k_1, k_2, k_1 k_2) \leq \lambda]\\ &\leq \epsilon + \epsilon + \left(\frac{18d^{1/4}n^{1/2}}{\sigma^2}\right)\left(5\epsilon\ln\left(\frac{1}{\epsilon}\right)\ln\left(\frac{1}{5\epsilon\ln(1/\epsilon)}\right)\right)\\ &\leq 2\epsilon + \left(\frac{90d^{1/4}n^{1/2}}{\sigma^2}\right)\left(\epsilon\ln^2\left(\frac{1}{\epsilon}\right)\right) \leq \left(\frac{91d^{1/4}n^{1/2}}{\sigma^2}\right)\frac{\ln^2 x}{x}.\end{aligned}$$

So,

$$\Pr\left[\frac{\|\mathbf{A}, \boldsymbol{c}\|_F}{\rho(A, \boldsymbol{c})} > x\right] \leq \left(\frac{91d^{1/4}n^{1/2}}{\sigma^2}\right)\frac{\ln^2(x/24)}{x/24} \leq \left(\frac{2184d^{1/4}n^{1/2}}{\sigma^2}\right)\frac{\ln^2(x)}{x}.$$

□

We conclude this section with the proof of Lemma 3.0.2.

Proof of Lemma 3.0.2. By (18), we have

$$\begin{aligned}
\mathbf{E}\left[\log_2 \frac{\|A, \boldsymbol{c}\|_F}{\rho(A, \boldsymbol{c})}\right] &\leq \log_2(24) + \mathbf{E}\left[\log_2 \|A, \boldsymbol{c}\|_F\right] + \mathbf{E}\left[\log_2(\max(\max_i \|\boldsymbol{a}_i\|_2, 1))\right] + \\
&\quad + \mathbf{E}\left[\log_2 \max\left\{\frac{1}{k_1}, \frac{1}{k_2}, \frac{1}{k_1 k_2}\right\}\right] \\
&\leq 5 + 2\,\mathbf{E}\left[\log_2 \|A, \boldsymbol{c}\|_F\right] + \mathbf{E}\left[\log_2 \max\left\{\frac{1}{k_1}, \frac{1}{k_2}, \frac{1}{k_1 k_2}\right\}\right] \\
&\leq 5 + 1 + 2\log_2\left(\frac{18 d^{1/4} n^{1/2}}{\sigma^2}\right), \quad \text{by Lemma A.1.5, Lemma A.2.2 and (17),} \\
&\leq 15 + 4\log_2 \frac{nd}{\sigma}.
\end{aligned}$$

□

4 Combining the Primal and Dual Analyses

Theorem 4.0.3 (Smoothed Analysis of $C(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$). *Let $n \geq d \geq 2$ and let $\bar{\mathbf{A}} \in \mathbb{R}^{n \times d}$, $\bar{\boldsymbol{b}} \in \mathbb{R}^n$ and $\bar{\mathbf{c}} \in \mathbb{R}^d$ satisfy $\left\|\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\mathbf{c}}\right\|_F \leq 1$. If $\sigma \leq 1/\sqrt{(n+1)(d+1)}$ and $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ the Gaussian perturbation of $(\bar{\mathbf{A}}, \bar{\boldsymbol{b}}, \bar{\mathbf{c}})$ of variance σ^2, then for each $i \in \{1, 2, 3, 4\}$,*

$$\begin{aligned}
\Pr_{\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}}\left[C^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}) > x\right] &\leq \left(\frac{522 n^2 d^{1.5}}{\sigma^2}\right) \frac{\ln^2 x}{x}, \quad \textit{and} \\
\mathop{\mathbf{E}}_{(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})}\left[\log C^{(i)}(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})\right] &\leq 18 + 4\log \frac{nd}{\sigma}.
\end{aligned}$$

Proof. Note that the transformation of each canonical form into the conic form leaves the Frobenius norm unchanged. Also, a Gaussian perturbation in the original becomes a Gaussian perturbation in the conic form. Therefore, by Lemma 2.2.2, the smoothed bounds on the primal condition number of the conic form imply smoothed bounds on each of the condition numbers $C_P^{(1)}, C_P^{(2)}, C_D^{(2)}, C_D^{(3)}$.

By Lemma 2.3.15, Lemma 3.2.2, and Proposition A.2.5 we have

$$\begin{aligned}
\Pr_{\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}}\left[C^{(i)}(A, \boldsymbol{b}, \boldsymbol{c}) > x\right] &\leq \left(\frac{197 n^2 d^{1.5}}{\sigma^2}\right) \frac{\ln^2 x}{x} + \left(\frac{2184 d^{1/4} n^{1/2}}{\sigma^2}\right) \frac{\ln^2 x}{x}. \\
&\leq \left(\frac{522 n^2 d^{1.5}}{\sigma^2}\right) \frac{\ln^2 x}{x},
\end{aligned}$$

as $n \geq d \geq 2$.

To bound the log of the condition number, we use Lemmas 2.3.1 and Lemma 3.0.2 to show

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{A},\boldsymbol{b},\boldsymbol{c}}\left[\log C^{(i)}(\mathbf{A},\boldsymbol{b},\boldsymbol{c})\right] \\
&\leq \mathop{\mathbf{E}}_{\mathbf{A},\boldsymbol{b},\boldsymbol{c}}\left[\log\left(C_P^{(i)}(\mathbf{A},\boldsymbol{b})+C_D^{(i)}(\mathbf{A},\boldsymbol{c})\right)\right] \\
&\leq \max\left(\mathop{\mathbf{E}}_{\mathbf{A},\boldsymbol{b},\boldsymbol{c}}\left[\log\left(2C_P^{(i)}(\mathbf{A},\boldsymbol{b})\right)\right], \mathop{\mathbf{E}}_{\mathbf{A},\boldsymbol{b},\boldsymbol{c}}\left[\log\left(2C_D^{(i)}(\mathbf{A},\boldsymbol{c})\right)\right]\right) \\
&\leq 18 + 4\log\left(\frac{nd}{\sigma}\right),
\end{aligned}$$

where in the second-to-last inequality we used that fact that for positive random variables β and γ,

$$\log(\beta+\gamma) \leq \max\left(\log(2\beta), \log(2\gamma)\right).$$

□

5 Open Problems and Conclusion

One way to strengthen the results in this paper would be to prove that they hold under more restrictive models of perturbation. For example, we ask whether similar results can be proved if one perturbs the linear program subject to maintaining feasibility or infeasibility. This would be an example of a property-preserving perturbation, as defined in [ST03a].

A related question is whether these results can be proved under zero-preserving perturbations in which only non-zero entries of $\mathbf{A}$ are subject to perturbations. Unfortunately, the following example shows that in this model of zero-preserving perturbations, it is not possible to bound the condition number by *poly* $\left(n, d, \frac{1}{\sigma}\right)$ with probability at least 1/2. Thus, any attempt to prove a bound better than $O(n^3 L)$ on the smoothed complexity of interior point algorithms under zero-preserving perturbations would have to use an analysis that does not merely apply a bound on the condition number.

Define the matrix

$$\bar{\mathbf{A}} = \begin{bmatrix} -1 & \epsilon & & \\ & -1 & \epsilon & \\ & & \cdots & \\ & & -1 & \epsilon \end{bmatrix}$$

where ϵ is a parameter to be chosen later. For ease of exposition, we will normalize $\left\|\bar{\mathbf{A}}\right\|_F$ to be 1 at the end of formulation. Let $\mathbf{A}$ be a zero preserving Gaussian perturbation of $\bar{\mathbf{A}}$ with variance σ^2, and consider the linear program $\mathbf{A}\boldsymbol{x} \geq \mathbf{0}, \boldsymbol{x} \in \boldsymbol{C}$ where $\boldsymbol{C} = \{\boldsymbol{x} : \boldsymbol{x} > \mathbf{0}\}$. The i^{th} constraint of $\bar{\mathbf{A}}\boldsymbol{x} \geq \mathbf{0}$ is exactly

$$\epsilon x_{i+1} \geq x_i$$

We now set $\sigma = \delta/\sqrt{4\ln n}$, and defer our choice of δ to later. Applying Proposition A.1.2, we find

$$\Pr[|a_{i,i} - (-1)| \geq \delta] \leq \frac{1}{2n\sqrt{\ln n}}, \text{and} \tag{20}$$

$$\Pr[|a_{i,i+1} - \epsilon| \geq \delta] \leq \frac{1}{2n\sqrt{\ln n}}. \tag{21}$$

Thus, for n sufficiently large, with probability at least 1/2 no entry of $\mathbf{A}$ differs from the corresponding entry in $\bar{\mathbf{A}}$ by more than δ. For the rest of the proof, we assume this to be the case.

If $\epsilon > \delta$ (which we will arrange later), we then have that $\mathbf{A}\boldsymbol{x} \geq \mathbf{0}, \boldsymbol{x} \in \boldsymbol{C}$ is feasible, and

$$\boldsymbol{x} = \left[\left(\frac{\epsilon - \delta}{1+\delta}\right)^n, \left(\frac{\epsilon - \delta}{1+\delta}\right)^{n-1}, \ \ldots, 1 \right]$$

is a feasible solution. We also have that for every feasible $\boldsymbol{x}$, $(\epsilon + \delta)x_{i+1} \geq (1-\delta)x_i$, for all i. Define

$$\Delta\mathbf{A} = \begin{bmatrix} 0 & \ldots & 0 & -(\frac{\epsilon+\delta}{1-\delta})^{n-2} \\ 0 & \ldots & 0 & 0 \\ & \ldots & & \\ 0 & \ldots & 0 & 0 \end{bmatrix}.$$

We now show that $(\mathbf{A} + \Delta\mathbf{A})\boldsymbol{x} \geq \mathbf{0}, \boldsymbol{x} \in \boldsymbol{C}$ is infeasible, and hence $\rho(\mathbf{A}, \boldsymbol{C}) \leq \|\Delta\mathbf{A}\|_F = (\frac{\epsilon+\delta}{1-\delta})^{n-2}$. To see this, note that the constraint given by the top row of $(\mathbf{A} + \Delta\mathbf{A})$ is

$$-x_1 + \epsilon x_2 - \left(\frac{\epsilon+\delta}{1-\delta}\right)^{n-2} x_n \geq 0,$$

while the other rows of $\mathbf{A} + \Delta\mathbf{A}$ constrain $x_2 \leq (\frac{\epsilon+\delta}{1-\delta})^{n-2} x_n$. Assuming $\epsilon \leq 1$ (which we arrange later), this constraint is impossible to satisfy for $\boldsymbol{x} \in \boldsymbol{C}$.

Setting $\epsilon = \frac{1}{n}$ and $\sigma = \frac{1}{n^2}$ (and hence $\delta = \frac{\sqrt{4 \ln n}}{n^2}$) yields $\rho(\mathbf{A}, \boldsymbol{C}) = (\frac{\epsilon+\delta}{1-\delta})^{n-2} = (\frac{O(1)}{n})^{n-2}$, which is exponentially small and also satisfies the requirements on ϵ. We can upper bound $\|\mathbf{A}\|_F$ by $\|\mathbf{A}\|_F \leq \sqrt{n(1+\delta)^2 + n(\epsilon+\delta)^2} \leq 2\sqrt{n}$. Thus the condition number, which is equal to $\|\mathbf{A}\|_F / \rho(\mathbf{A}, \boldsymbol{C})$, is at least $\Omega(n)^{n-3}$.

If we had normalized $\|\bar{\mathbf{A}}\|_F = 1$ at the beginning of the proof, the corresponding variables would have been set to $\epsilon \approx \frac{1}{n\sqrt{n}}$, $\sigma \approx \frac{1}{n^2\sqrt{n}}$, which still proves the negative result. This concludes our discussion of impossibility results for smoothed analysis.

We would like to point out that condition numbers appear throughout Numerical Analysis and that condition numbers may be defined for many non-linear problems. The speed of algorithms for optimizing linear functions over convex bodies (including semidefinite programming) has been related to their condition numbers [Fre02, FV00], and it seems that one might be able to extend our results to these algorithms as well. Condition numbers have also been defined for non-linear programming problems, and one could attempt to perform a smoothed analysis of non-linear optimization algorithms by relating their performance to the condition numbers of their inputs, and then performing a smoothed analysis of their condition numbers.

This paper also made the first step towards these analyses. Our results in Section 2 showed that for any strictly-supported convex cone $\boldsymbol{C}$, the smoothed value of the logarithm of the primal condition number of conic program

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}, \quad \boldsymbol{x} \in \boldsymbol{C}$$

is $O(\log(nd/\sigma))$. We conjecture similar result holds for the dual condition number of the conic program.

The approach of proving smoothed complexity bounds by relating the performance of an algorithm to some property of its input, such as a condition number, and then performing a smoothed analysis of this quantity has also been recently used in [ST03a, SST06]. Finally, we hope that this work illuminates some of the shared interests of the Numerical Analysis, Operations Research, and Theoretical Computer Science communities.

References

[Ans99] Anstreicher, K. M. Linear programming in $O((n^3/\ln n)L)$ operations. *SIAM Journal of Optimization*. 9. 803-812, 1999.

[AJPY93] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Average performance of a self–dual interior–point algorithm for linear programming. In P. M. Pardalos, editor, *Complexity in Numerical Optimization*, pages 1–15. World Scientific Publishing Co., London, United Kingdom, 1993.

[AJPY99] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Probabilistic analysis of an infeasible-interior-point algorithm for linear programming. *Mathematics of Operations Research*, 24(1):176–192, 1999.

[AS70] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, 9 edition, 1970.

[Bal93] K. Ball. The reverse isoperimetric problem for Gaussian measure. *Discrete and Computational Geometry*, 10(4):411–420, 1993.

[BR76] Bhattacharya and Rao. Normal approximation and asymptotic expansions. Wiley, 1976.

[BD02] Avrim Blum and John Dunagan. Smoothed analysis of the perceptron algorithm for linear programming. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms (SODA '02)*, 2002.

[CC01] D. Cheung and F. Cucker. A new condition number for linear programming. *Math. Programming*, 91(1 (Ser. A)):163–174, 2001.

[CC02] C. W. Cheung and F. Cucker Probabilistic analysis of condition numbers for linear programming, *Journal of Optimization Theory and Applications*, 114(1), July 2002, pp 55-67.

[CCH03] Dennis Cheung, Felipe Cucker and Raphael Hauser. Tail decay and moment estimates of a condition number for random linear conic systems. *SIAM Journal on Optimization*, 2005, 15 (4) pp. 1237-1261.

[CP01] F. Cucker and J. Peña. A primal-dual algorithm for solving polyhedral conic systems with a finite-precision machine. *SIAM Journal on Optimization*, 2001, 12 (2) pp. 522-554.

[CW03] F. Cucker and M. Wschebor. On the Expected Condition Number of Linear Programming Problems, *Numerische Mathematik*, Vol 94, 2003, pp 419-478.

[DET08] A. Deza, E. Nematollahi and T. Terlaky. How good are interior point methods? Klee-Minty cubes tighten iteration-complexity bounds, *Mathematical Programming* 113(1):1– 14, 2008.

[DV04] John Dunagan and Santosh Vempala. A simple polynomial-time rescaling algorithm for solving linear programs, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, 2004, pp. 315–320.

[EF00] M. Epelman and R.M. Freund. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Mathematical Programming* 88(3):451– 485, 2000.

[Fe68] W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. 1. John Wiley & Sons. 1968.

[Fre02] R. Freund. Complexity of convex optimization using geometry-based measures and a reference point. *Mathematical Programming* 99:197–221, 2004.

[FV00] Robert Freund and Jorge Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization*, 10(1):155–176, 2000.

[Gon88] Clovis C. Gonzaga. *An Algorithm for Solving Linear Programming Problems in $O(n^3L)$ Operations*, pages 1–28. Springer-Verlag, 1988. Progress in Mathematical Programming, N. Megiddo ed..

[GT92] C. C. Gonzaga and M. J. Todd. An $O(\sqrt{n}L)$–iteration large–step primal–dual affine algorithm for linear programming. *SIAM Journal on Optimization*, 2:349–359, 1992

[Kha79] L L. Khachian. A polynomial algorithm in linear programming. *Doklady Adademiia Nauk SSSR*, 244:1093–96, 1979. (In Russian)

[LDK63] B. Grunbaum, L. Danzer and V. Klee. Helly's theorem and its relatives. In *Convexity (Proceedings of the Symposia on Pure Mathematics 7)*, pages 101–180. American Mathematical Society, 1963.

[Hel23] E. Helly. Über Mengen konvexen Körper mit gemeinschaflichen Punkten. In *Jber. Deutsch. Math. Verein.* 32 (1923) pages 175-176.

[HB02] Petra Huhn and Karl Heinz Borgwardt. Interior-point methods: Worst case and average case analysis of a phase-I algorithm and a termination procedure. *Journal of Complexity*, 18:833–910, 2002.

[Kar84] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[Kha79] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademia Nauk SSSR*, pages 1093–1096, 1979.

[LMS90] I. J. Lustig, R. E. Marsten, and D. F. Shanno. The primal–dual interior point method on the Cray supercomputer. In T. F. Coleman and Y. Li, editors, *Large–Scale Numerical Optimization, Papers from the Workshop held at Cornell University, Ithaca, NY, USA, October 1989*, volume 46 of *SIAM Proceedings in Applied Mathematics*, pages 70–80. Society of Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 1990.

[EA96] C. Mezaros, X. Xu E. Andersen, J. Gondzio. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior point methods in mathematical programming.* Kluwer Academic Publisher, 1996.

[MTY93] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive–step primal–dual interior–point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[NN94] J.E. Nesterov and A. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming* SIAM, Philadelphia, 1994.

[Nem88] A. S. Nemirovskii. An new polynomial algorithm for linear programming. *Doklady Akademii Nauk SSSR*, 298(6):1321–1325, 1988. Translated in : *Soviet Mathematics Doklady 37(1): 264–269, 1988.*

[Peñ00] J. Peña. Understanding the geometry of infeasible perturbations of a conic linear system. *SIAM Journal on Optimization*, 10:534–550, 2000.

[Ren94] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[Ren95a] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.

[Ren95b] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser. A):279–351, 1995.

[SST06] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446-476, 2006.

[IL94] D. Shanno, I. Lustig, R. Marsten. Interior point methods: computational state of the art. *ORSA Journal on Computing*, 6(1):1–14, 1994.

[ST03a] Daniel Spielman and Shang-Hua Teng. Smoothed analysis: Motivation and discrete models. In *Proceedings of WADS 2003, Lecture Notes in Computer Science 2748*, pages 256–270, 2003.

[ST03b] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of termination of linear programming algorithms. *Math. Program., Ser. B*, 97:375–404, 2003.

[ST04] Daniel Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51 (3) pp 385–436, May 2004.

[Tod91] M. J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16(4):671–693, 1991.

[Tod94] M. J. Todd. A lower bound on the number of iterations of primal-dual interior-point methods for linear programming. In G. A. Watson and D. F. Griffiths, editors, *Numerical Analysis 1993*, pages 237–259. Longman Press, Harlow, 1994.

[TY96] M. J. Todd and Y. Ye. A lower bound on the number of iterations of long-step and polynomial interior-point methods for linear programming. *Annals of Operations Research*, 62:233–252, 1996.

[Vai90] P. M. Vaidya. An algorithm for linear programming which requires $O((m+n)n^2 + (m+n)^{1.5}nL)$ arithmetic operations. *Mathematical Programming*, 47:175–201, 1990. Condensed version in : *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 29–38, 1987.*

[Vai89] Pravin M. Vaidya. Speeding-Up Linear Programming Using Fast Matrix Multiplication. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*: 332-337, 1989.

[Ver96] Jorge Vera. Ill-posedness and the complexity of deciding existence of solutions to linear programs. *SIAM Journal on Optimization*, 6(3), 1996.

[Wri97] Stephen Wright. *Primal-Dual Interior-Point Method.* SIAM Publications, 1997.

A Mathematical Background and Useful Inequalities

A.1 Gaussian random variables

Proposition A.1.1. *Let x be a Gaussian random variable of variance $\sigma^2 \leq 1$ and arbitrary mean. For all $\epsilon \geq 0$,*

$$\mathbf{Pr}\left[|x| \leq \epsilon\right] \leq \sqrt{\frac{2}{\pi}}\epsilon.$$

Proof. The density function of x is never more than $1/\sqrt{2\pi}$. □

The following bound on the tails of Gaussian distributions is standard (see, for example [Fe68, Section VII.1])

Proposition A.1.2 (Gaussian tail bound). *Let x be a Gaussian random variable of variance σ^2 and mean 0. For all $\delta \geq 0$,*

$$\mathbf{Pr}\left[|x| \geq \delta\right] \leq \sqrt{\frac{2}{\pi}} \frac{\sigma}{\delta} e^{-\delta^2/2\sigma^2}.$$

Proposition A.1.3 (Chi-Square bound). *For any d-dimensional Gaussian random vector $\boldsymbol{x}$ of variance σ^2 centered at the origin, and for $d \geq 2$, $\epsilon \leq 1/100$ and $\sigma \leq 1/\sqrt{d}$*

$$\mathbf{Pr}\left[\|\boldsymbol{x}\|_2 \geq 1.5 \ln^{1/2}(1/\epsilon)\right] \leq \epsilon$$

Proof. This may be derived by substituting these values into Equality (26.4.8) of [AS70], which says that for all $k \geq 0$,

$$\mathbf{Pr}\left[\|\boldsymbol{x}\|_2 \geq k\sigma\right] \leq \frac{\left(k^2\right)^{d/2-1} e^{-k^2/2}}{2^{d/2-1}\Gamma(\frac{d}{2})}.$$

For example, Spielman and Teng [ST04, Corollary 2.19] show that this equality implies that for every $\lambda \geq 3$,

$$\mathbf{Pr}\left[\|\boldsymbol{x}\|_2 \geq 3\sqrt{d \ln \lambda}\sigma\right] \leq \lambda^{-2.9d}.$$

Substituting $\lambda = \epsilon^{-1/4}$, we obtain

$$\mathbf{Pr}\left[\|\boldsymbol{x}\|_2 \geq 1.5\sqrt{d \ln(1/\epsilon)}\sigma\right] \leq \epsilon^{2.9d/4} \leq \epsilon,$$

for $d \geq 2$. □

Proposition A.1.4 (Independence of mean and displacements). *Let $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n \in \mathbb{R}^d$ and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the Gaussian perturbations of $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n$ of variance σ^2. Let*

$$\bar{\boldsymbol{z}} = \frac{1}{n}\sum_i \bar{\boldsymbol{a}}_i, \qquad \boldsymbol{z} = \frac{1}{n}\sum_i \boldsymbol{a}_i, \qquad \text{and} \qquad \boldsymbol{x}_i = \boldsymbol{a}_i - \boldsymbol{z}, \text{ for } 1 \leq i \leq n.$$

Then, $\boldsymbol{z}$ is the Gaussian perturbation of $\bar{\boldsymbol{z}}$ of variance σ^2/n and is independent of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

Lemma A.1.5. *Let $\bar{\boldsymbol{a}} \in \mathbb{R}^m$ satisfy $\|\bar{\boldsymbol{a}}\|_2 \leq 1$ and let $\sigma \leq 1/\sqrt{m}$. If $\boldsymbol{a}$ is the Gaussian perturbation of $\bar{\boldsymbol{a}}$ of variance σ^2, then*

$$\mathop{\mathbf{E}}_{\boldsymbol{a}}\left[\log_2 \|\boldsymbol{a}\|_2\right] \leq 1/2.$$

Proof. Because logarithm is a concave function, we have

$$\mathop{\mathbf{E}}_{\boldsymbol{a}}\left[\log_2 \|\boldsymbol{a}\|_2\right] \leq \log_2(\mathop{\mathbf{E}}_{\boldsymbol{a}}\left[\|\boldsymbol{a}\|_2\right]) \leq \log_2 \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{a}}\left[\|\boldsymbol{a}\|_2^2\right]} = \log_2 \sqrt{m\sigma^2 + \|\bar{\boldsymbol{a}}\|_2^2},$$

as $\|\boldsymbol{a}\|_2^2$ is a non-central χ^2 random variable with dn degrees of freedom and non-centrality parameter $\|\bar{\boldsymbol{a}}\|_2$ [AS70, 26.4.37]. Therefore, from the assumption $\sigma \leq 1/\sqrt{m}$ and $\|\boldsymbol{a}\|_2 \leq 1$,

$$\mathop{\mathbf{E}}_{\boldsymbol{a}}\left[\log_2 \|\boldsymbol{a}\|_2\right] \leq \log_2 \sqrt{2} = 1/2.$$

□

A.2 Useful Inequalities

Proposition A.2.1. *Let $\mu(x, y)$ be a non-negative integrable function, and let x and y be distributed according to $\mu(x, y)$. Then for any event $A(x, y)$,*

$$\mathop{\mathbf{Pr}}_{x,y}[A(x,y)] \quad \le \quad \max_x \mathop{\mathbf{Pr}}_y [A(x,y)],$$

where, in the right-hand term, y is distributed according to the induced distribution $\mu(x, y)$.

Lemma A.2.2 (Almost linear to log). *Let A be a positive random variable such that for all $x \ge e$,*

$$\mathop{\mathbf{Pr}}_A [A \ge x] \le \frac{\alpha \ln^{1.5} x}{x},$$

for some $\alpha \ge 10$. Then,

$$\mathop{\mathbf{E}}_A [\log_2 A] \le 2\log_2(e\alpha)$$

Proof. We will use the following inequality, which is easy to verify numerically:

Fact A.2.3. *For all $\alpha \ge 10$ and $x \ge 2\ln\alpha$, $x - 1.5\ln x \ge x/2$.*

Then,

$$\begin{aligned}
\mathop{\mathbf{E}}_A [\ln A] &= \int_0^\infty \mathop{\mathbf{Pr}}_A [\ln A > x]\, dx = \int_0^\infty \mathop{\mathbf{Pr}}_A [A > e^x]\, dx \\
&\le \int_0^\infty \min\left(1, \frac{\alpha x^{1.5}}{e^x}\right) dx \le \int_0^{2\ln\alpha} dx + \int_{2\ln\alpha}^\infty \frac{\alpha x^{1.5}}{e^x}\, dx \\
&= 2\ln\alpha + \alpha \int_{2\ln\alpha}^\infty e^{-x+1.5\ln x}\, dx \\
&\le 2\ln\alpha + \alpha \int_{2\ln\alpha}^\infty e^{-x/2}\, dx = 2\ln\alpha + 2 = 2\ln(e\alpha),
\end{aligned}$$

where the last inequality follows from Fact A.2.3.

Thus $\mathbf{E}_A[\log_2 A] = (\log_2 e)\,\mathbf{E}_A[\ln A] \le 2(\log_2 e)\ln(e\alpha) = 2\log_2(e\alpha)$. □

The following proposition is due to Sankar, Spielman and Teng [SST06, Corollary C.2].

Proposition A.2.4 (Linear combination). *Let A and B be two positive random variables. Assume $\mathbf{Pr}[A \ge x] \le \frac{\alpha}{x}$ and $\mathbf{Pr}[B \ge x | A] \le \frac{\beta}{x}$, for some $\alpha, \beta > 0$. Then,*

$$\mathbf{Pr}[AB \ge x] \le \frac{\alpha\beta}{x}\left(1 + \max\left(0, \ln\left(\frac{x}{\alpha\beta}\right)\right)\right)$$

Proposition A.2.5 (Min by sum). *For any positive random variables A and B and for any $\lambda > 0$,*

$$\mathbf{Pr}[\min(A,B) < \lambda] \le \mathbf{Pr}[A < \lambda] + \mathbf{Pr}[B < \lambda] = \mathbf{Pr}\left[\frac{1}{A} > \frac{1}{\lambda}\right] + \mathbf{Pr}\left[\frac{1}{B} > \frac{1}{\lambda}\right].$$

Lemma A.2.6. *For positive α, β and any vector $\boldsymbol{a}_{k+1}$,*

$$\frac{\alpha\beta}{2\alpha + \|\boldsymbol{a}_{k+1}\|_2} \geq \min\left\{\frac{\alpha\beta}{2 + \|\boldsymbol{a}_{k+1}\|_2}, \frac{\beta}{2 + \|\boldsymbol{a}_{k+1}\|_2}\right\}.$$

Proof. For $\alpha \geq 1$, we have

$$\frac{\alpha\beta}{2\alpha + \|\boldsymbol{a}_{k+1}\|_2} = \frac{\beta}{2 + \|\boldsymbol{a}_{k+1}\|_2 / \alpha} \geq \frac{\beta}{2 + \|\boldsymbol{a}_{k+1}\|_2},$$

while for $\alpha \leq 1$ we have

$$\frac{\alpha\beta}{2\alpha + \|\boldsymbol{a}_{k+1}\|_2} \geq \frac{\alpha\beta}{2 + \|\boldsymbol{a}_{k+1}\|_2}.$$

□

Lemma A.2.7. *For any positive A, k_1 and k_2,*

$$\frac{k_1 k_2}{2k_1 + A} \geq \min\left(\frac{k_2}{4}, \frac{k_1 k_2}{2A}\right).$$

Proof. For $2k_1 \geq A$, we have $k_1k_2/(2k_1 + A) \geq k_2/4$; and for $2k_1 \leq A$, we have $k_1k_2/(2k_1+A) \geq k_2k_2/(2A)$. □

A.3 Convex Geometry

Lemma A.3.1 (Perturbation of new variables). *For any $\alpha > 0$, let $\boldsymbol{c}, \boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ be a set of vectors such that*

$$\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) > \alpha. \tag{22}$$

Let

$$\beta = \frac{\alpha \|\boldsymbol{c}\|_2}{2\alpha + 4(\|\boldsymbol{z}\|_2 + \max_i \|\boldsymbol{x}_i\|_2)}.$$

Then for any $\|\Delta\boldsymbol{x}_i\|_2 \leq \alpha/4$, $\|\Delta\boldsymbol{z}\|_2 \leq \alpha/4$, and $\|\Delta\boldsymbol{c}\|_2 \leq \beta$,

$$\boldsymbol{z} + \Delta\boldsymbol{z} \notin \partial\left(\mathbf{Ray}\left(\boldsymbol{c} + \Delta\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta\boldsymbol{x}_n\right)\right)$$

Proof. Assume by way of contradiction that

$$\boldsymbol{z} + \Delta\boldsymbol{z} \in \partial\left(\mathbf{Ray}\left(\boldsymbol{c} + \Delta\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta\boldsymbol{x}_n\right)\right).$$

We first consider the case that $\boldsymbol{z} \notin \mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. In this case, we will show that $\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) \leq \alpha$, contradicting assumption (22). Since $\boldsymbol{z} + \Delta\boldsymbol{z} \in \partial\left(\mathbf{Ray}\left(\boldsymbol{c} + \Delta\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta\boldsymbol{x}_n\right)\right)$,

$$\boldsymbol{z} + \Delta\boldsymbol{z} = \lambda(\boldsymbol{c} + \Delta\boldsymbol{c}) - \sum_i \gamma_i(\boldsymbol{x}_i + \Delta\boldsymbol{x}_i),$$

for some $\lambda \geq 0$ and $\gamma_1, \ldots, \gamma_n \geq 0$ such that $\sum_i \gamma_i = 1$. We establish an upper bound on λ by noting that

$$\lambda = \frac{\left\| \boldsymbol{z} + \Delta \boldsymbol{z} + \sum_i \gamma_i (\boldsymbol{x}_i + \Delta \boldsymbol{x}_i) \right\|}{\left\| \boldsymbol{c} + \Delta \boldsymbol{c} \right\|}. \tag{23}$$

We may lower bound the denominator of (23) by $\left\| \boldsymbol{c} \right\| /2$, as

$$\left\| \Delta \boldsymbol{c} \right\| \leq \frac{\alpha \left\| \boldsymbol{c} \right\|}{2\alpha + 4(\left\| \boldsymbol{z} \right\| + \max_i \left\| \boldsymbol{x}_i \right\|)} \leq \left\| \boldsymbol{c} \right\| /2.$$

We upper bound the numerator of (23) by

$$\begin{aligned} \left\| \boldsymbol{z} + \Delta \boldsymbol{z} + \sum_i \gamma_i (\boldsymbol{x}_i + \Delta \boldsymbol{x}_i) \right\| &\leq \left\| \boldsymbol{z} \right\| + \alpha/4 + \sum_i \gamma_i (\left\| \boldsymbol{x}_i \right\| + \left\| \Delta \boldsymbol{x}_i \right\|) \\ &\leq \left\| \boldsymbol{z} \right\| + \alpha/4 + \max_i \left\| \boldsymbol{x}_i \right\| + \alpha/4 \\ &= \left\| \boldsymbol{z} \right\| + \max_i \left\| \boldsymbol{x}_i \right\| + \alpha/2. \end{aligned}$$

Thus,

$$\lambda \leq \frac{\left\| \boldsymbol{z} \right\| + \max_i \left\| \boldsymbol{x}_i \right\| + \alpha/2}{\left\| \boldsymbol{c} \right\| /2}.$$

Since

$$\left(\boldsymbol{z} + \Delta \boldsymbol{z} - \lambda \Delta \boldsymbol{c} + \sum_i \gamma_i \Delta \boldsymbol{x}_i \right) = \left(\lambda \boldsymbol{c} - \sum_i \gamma_i \boldsymbol{x}_i \right) \in \mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right),$$

we find that

$$\begin{aligned} &\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) \\ &\leq \left\| \Delta \boldsymbol{z} - \lambda \Delta \boldsymbol{c} + \sum_i \gamma_i \Delta \boldsymbol{x}_i \right\| \\ &\leq \left\| \Delta \boldsymbol{z} \right\| + \lambda \left\| \Delta \boldsymbol{c} \right\| + \sum_i \gamma_i \left\| \Delta \boldsymbol{x}_i \right\| \\ &\leq \frac{\alpha}{4} + \left(\frac{\left\| \boldsymbol{z} \right\| + \max_i \left\| \boldsymbol{x}_i \right\| + \alpha/2}{\left\| \boldsymbol{c} \right\| /2} \right) \left(\frac{\alpha \left\| \boldsymbol{c} \right\|}{2\alpha + 4(\left\| \boldsymbol{z} \right\| + \max_i \left\| \boldsymbol{x}_i \right\|)} \right) + \frac{\alpha}{4} \\ &= \alpha, \end{aligned}$$

contradicting assumption (22).

We now consider the case that $\boldsymbol{z} \in \mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. Since

$$\boldsymbol{z} + \Delta \boldsymbol{z} \in \partial\left(\mathbf{Ray}\left(\boldsymbol{c} + \Delta \boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n\right)\right),$$

there exists a hyperplane H passing through $\boldsymbol{z} + \Delta \boldsymbol{z}$ and tangent to the convex set $\mathbf{Ray}\left(\boldsymbol{c} + \Delta \boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n\right)$. By the assumption that $\mathbf{dist}\left(\boldsymbol{z}, \partial\left(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)\right)\right) > \alpha$, there is some $\delta_0 > 0$ such that, for every $\delta \in$

$(0, \delta_0)$, every point within distance $\alpha + \delta$ of $\boldsymbol{z}$ lies within $\mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Choose $\delta \in (0, \delta_0)$ that also satisfies $\delta \leq \|\boldsymbol{z}\| + \max_i \|\boldsymbol{x}_i\|$. Let $\boldsymbol{q}$ be the point at distance $\frac{3\alpha}{4} + \delta$ from $\boldsymbol{z} + \Delta \boldsymbol{z}$ in the direction perpendicular to H and away from the set. Clearly,

$$\mathbf{dist}\,(\boldsymbol{q}, \mathbf{Ray}\,(\boldsymbol{c} + \Delta \boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n)) = \frac{3\alpha}{4} + \delta. \tag{24}$$

Since $\mathbf{dist}\,(\boldsymbol{z}, \boldsymbol{z} + \Delta \boldsymbol{z}) \leq \frac{\alpha}{4}$ and $\mathbf{dist}\,(\boldsymbol{z} + \Delta \boldsymbol{z}, \boldsymbol{q}) = \frac{3\alpha}{4} + \delta$, $\mathbf{dist}\,(\boldsymbol{z}, \boldsymbol{q}) \leq \alpha + \delta$, and

$$\boldsymbol{q} \in \mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\,.$$

Because $\boldsymbol{q} \in \mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, there exist $\lambda \geq 0$ and $\gamma_1, \ldots, \gamma_n \geq 0$ satisfying $\sum_i \gamma_i = 1$ such that

$$\boldsymbol{q} = \lambda \boldsymbol{c} - \sum_i \gamma_i \boldsymbol{x}_i.$$

We upper bound λ as before:

$$\lambda = \frac{\|\boldsymbol{q} + \sum_i \gamma_i \boldsymbol{x}_i\|}{\|\boldsymbol{c}\|} \leq \frac{\|\boldsymbol{z}\| + \alpha + \delta + \max_i \|\boldsymbol{x}_i\|}{\|\boldsymbol{c}\|} \leq \frac{\|\boldsymbol{z}\| + \max_i \|\boldsymbol{x}_i\| + \alpha/2}{\|\boldsymbol{c}\|/2}.$$

Hence,

$$\begin{aligned} \boldsymbol{q} + \lambda \Delta \boldsymbol{c} - \sum_i \gamma_i \Delta \boldsymbol{x}_i &= \lambda(\boldsymbol{c} + \Delta \boldsymbol{c}) - \sum_i \gamma_i (\boldsymbol{x}_i + \Delta \boldsymbol{x}_i) \\ &\in \quad \mathbf{Ray}\,(\boldsymbol{c} + \Delta \boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n)\,, \end{aligned}$$

and thus

$$\begin{aligned} \mathbf{dist}\,(\boldsymbol{q}, \mathbf{Ray}\,(\boldsymbol{c} + \Delta \boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n)) &\leq \left\| \lambda \Delta \boldsymbol{c} - \sum_i \gamma_i \Delta \boldsymbol{x}_i \right\| \\ &\leq \lambda \|\Delta \boldsymbol{c}\| + \max_i \|\Delta \boldsymbol{x}_i\| \\ &\leq \alpha/2 + \alpha/4 \\ &\leq 3\alpha/4, \end{aligned}$$

which contradicts (24). □