Smoothed Analysis of Renegar's Condition Number for Linear Programming

John Dunagan[*], Daniel A. Spielman [†]
Department of Mathematics
Massachusetts Institute of Technology

Shang-Hua Teng [‡]
Akamai Technologies Inc. and
Department of Computer Science
University of Illinois at Urbana-Champaign

Abstract

For any linear program, we show that a slight random relative perturbation of that linear program has small condition number with high probability. Following [ST01], we call this smoothed analysis of the condition number. Part of our main result is that the expectation of the log of the condition number of any appropriately scaled linear program subject to a Gaussian perturbation of variance σ^2 is at most $O(\log nd/\sigma)$ with high probability. Since the condition number bounds the running time of many algorithms for convex programming, this may explain their observed fast convergence.

1 Introduction

Condition numbers are ubiquitous in numerical analysis and scientific computing. For many computational tasks with matrices, the ratio of the maximum and minimum eigenvalues of the matrix is a good condition number. For other tasks, such as solving a discretized partial differential equation for given boundary conditions, a different condition number may be defined. A condition number typically has two uses,

1. to estimate the sensitivity of the problem's answer to error in the input, and
2. to bound the number of iterations required by an iterative method to achieve a given degree of accuracy.

Analysis of algorithms using condition numbers may be interpreted as a *parameterized* worst-case complexity analysis. For many iterative methods, the maximum number of iterations is bounded by some function of the condition number, although the actual number of iterations may be less. Additionally, the condition number is typically bounded by some function of the input size, where the input size includes both the number of parameters and the bit size required to represent these parameters, and so condition number bounds typically imply worst-case complexity bounds. Thus condition number is a refinement of *input size* as a measure of problem difficulty.

Another reason for the study of condition numbers is that:

[*]Supported in part by NSF Career Award CCR-9875024. `jdunagan@mit.edu`

[†]Partially supported by an Alfred P. Sloan Foundation Fellowship, and NSF grant CCR-0112487. School of Science. `spielman@math.mit.edu`

[‡]Partially suppoted by an Alfred P. Sloan Foundation Fellowship, and NSF grants CCR-9972532, and CCR-0112487. `steng@akamai.com`

Numerical analysis is the study of algorithms for the problems of continuous mathematics.[1]

For a continuous input domain, it may be unnatural to discretize the input in the problem definition. Condition numbers are well-defined for arbitrary real-valued inputs, where measuring the input size in bits may not be possible. The fields of numerical analysis and scientific computing consider such problems and inputs, and condition numbers have been a pervasive underpinning of research in these fields.

Renegar [Ren94] introduced a condition number for linear programs, which we refer to as Renegar's condition number. In this work, he suggested that the study of condition numbers for linear programming was a natural outgrowth of the central role iterative solvers, particularly interior point methods, had assumed in the study of algorithms for convex programming. A large body of further work, detailed below, has developed on bounding the number of iterations required to solve a given linear program as a function of the condition number. This body of work includes both new bounds on old methods and the development of new algorithms.

Our work addresses the question of "what are likely values for the condition number?" In particular, for a very natural model of noise in the input data, we show that the condition number is likely to be low. This addresses a question outside the scope of previous work on the condition number for linear programs. The body of work on how condition number influences running time is an extensive foundation, and we hope to build another layer underneath, on how noisy data leads to bounded condition number.

Spielman and Teng proposed the *smoothed complexity* model in [ST01]. In this work, they showed that for an arbitrary linear program, a small random relative perturbation of that program is solved by the simplex algorithm (with the shadow vertex pivot rule) in polynomial time with very high probability. Blum and Dunagan proved a similar guarantee for the perceptron algorithm in [BD02].

Spielman and Teng expressed the hope in [ST01] that their result might explain the observed good performance of the simplex algorithm in practice: if your linear program is defined by a constraint matrix drawn from noisy data, it will probably be one that is easily solved by the simplex algorithm.

The smoothed complexity model seeks to interpolate between worst-case and average-case complexity analysis. By letting the size of the random perturbation to the data (i.e., the variance of the noise) become large, one obtains the traditional average-case complexity measure. By letting the size of the random perturbation go to zero, one obtains the traditional worst-case complexity measure. In between, one obtains new theoretical results that may also be practically meaningful.

The examples given above and the work in this paper pertain to the smoothed analysis of algorithms for linear programming. In this paper, we use a two-step approach:

1. Bound the running time of an algorithm in terms of a condition number.
2. Perform a smoothed analysis of this condition number.

Step 1. has already been done (see subsection 1.1). Our main theorem accomplishes step 2.

We do not wish to give the impression that smoothed analysis is only meaningful for linear programming, or even convex optimization problems. Recall that different problems (matrix inversion, solving a partial differential equation, etc.) have different condition numbers. Typically these condition numbers are defined to be (for any given problem instance) the sensitivity of the output to change in the input. Many of these condition numbers have the property that the condition number is low if the smallest relative change to the input data necessary for the problem to be ill-posed is large. (Loosely speaking, a problem instance is ill-posed if an arbitrarily small further change to the input data may yield an arbitrarily large change in the answer. The linear programming condition number we consider here is defined to

[1]Lloyd Trefethen, November 1992 SIAM News.

be the distance to ill-posedness, and can then be shown to bound the sensitivity of the output due to change in the input.) For such condition numbers it may be the case that, from any initial instance, a small random perturbation to that instance is quite likely to yield a new instance that is not too close to ill-posedness. One exciting aspect of [ST01, BD02] is that they show the simplex algorithm and the perceptron algorithm both fit into this general framework. This paper shows the same thing for the linear programming condition number. This phenomenon may be very common (the condition number for matrix inverion is addressed in [ST]).

To give a preview of our results, we state a rough version of our main theorem without constants or logarithmic factors.

Statement 1 (Smoothed Complexity of Renegar's Condition Number) *For an arbitrary linear program defined by an appropriately scaled n-by-d constraint matrix subject to a Gaussian perturbation of variance σ^2, with probability at least $1-\delta$ over the random perturbation, Renegar's condition number C satisfies*

$$C = \tilde{O}(\frac{n^2 d^2}{\sigma^2 \delta})$$

where $f = \tilde{O}(g)$ denotes $f = O(g \log^{O(1)} g)$. A precise version of this statement is theorem 1.4.1.

As an example of what kind of conclusion we derive on the overall performance of algorithms, we mention that a particular interior point method [FM00] only requires $O(\sqrt{n+d}\ln(C/\epsilon))$ iterations to come within ϵ of the optimal solution, and each iteration requires only an approximate matrix inversion computable in $O((n+d)^{2.5})$ time. Thus the smoothed complexity is $O((n+d)^3 \ln(nd/(\sigma\epsilon)))$ for this particular method to come within ϵ of the optimal solution.

1.1 Condition number based linear programming

Since Renegar's initial papers [Ren94, Ren95a, Ren95b] on condition numbers for linear programs, there has been a large body of subsequent work. The running time of a number of algorithms for optimization has been analyzed in terms of their dependence on the condition number [FN01, FV00]. The notion of condition number has even inspired new algorithms for optimization [FE00a, FE00b]. Additionally, some variants of Renegar's original condition number have also been studied [FE01].

Part of the reason for the volume of work is that every linear programming formulation requires a separate condition number analysis. This point is made by [Ren94, Ren95a, Ren95b, Ver96, CP01] in their work developing interior point methods that have good dependence on the condition number. In addition to bounding the time necessary to optimize in terms of the condition number, there has been work on quickly estimating the condition number [FV99], a well-known question for the condition numbers of other problems.

The notion of condition number for linear programs has been extended to semi-definite programs [FN01], but we will not elaborate further on this topic here.

1.2 Notation

Throughout this paper we use the notational convention that

- lower case letters such as a and α denote scalars,
- bold lower case letters such as $\boldsymbol{a}$ and $\boldsymbol{b}$ denote vectors,

- capital letters such as A denote matrices, and
- bold capital letters such as $\boldsymbol{C}$ denote convex sets.

If $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are column vectors, we let $[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ denote the matrix whose columns are the $\boldsymbol{a}_i$s.

For a vector $\boldsymbol{a}$, we let $\|\boldsymbol{a}\|$ denote the standard Euclidean norm of the vector. We will make frequent use of the Frobenius norm of a matrix, $\|A\|_F$, which is the square root of the sum of squares of the entries in the matrix. We extend this notation to let $\|A, \boldsymbol{a}\|_F$ denote the square root of the sum of squares of the entries in A and $\boldsymbol{a}$. Different choices of norm are possible; we use the Frobenius norm throughout this chapter. The following proposition relates several common choices of norm:

Proposition 1.2.1 (Choice of norm) *For an n-by-d matrix A,*

$$\begin{aligned} \frac{\|A\|_F}{\sqrt{dn}} &\leq \|A\|_\infty \leq \|A\|_F, \text{ and} \\ \frac{\|A\|_F}{\sqrt{d}} &\leq \|A\|_{OP} \leq \|A\|_F, \end{aligned}$$

where $\|A\|_{OP}$ denotes the operator norm of A, $\max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$.

We also make use of the following definitions:

Definition 1.2.2 (Ray) *For a vector $\boldsymbol{p}$, let $\mathbf{Ray}(\boldsymbol{p})$ denote $\{\alpha \boldsymbol{p} : \alpha > 0\}$.*

Definition 1.2.3 (Open convex cone) *An open convex cone is a convex set $\boldsymbol{C}$ such that for all $\boldsymbol{x} \in \boldsymbol{C}$ and all $\alpha > 0$, $\alpha \boldsymbol{x} \in \boldsymbol{C}$, and there exists a vector $\mathbf{t}$ such that $\mathbf{t}^T \boldsymbol{x} < 0$ for all $\boldsymbol{x} \in \boldsymbol{C}$.*

Warning 1.2.4 (Open convex cone?) *An open convex cone cannot contain the origin, and is not necessarily open in the topological sense.*

Definition 1.2.5 (Positive half-space) *For a vector $\boldsymbol{a}$ we let $\mathcal{H}(\boldsymbol{a})$ denote the half-space of points with non-negative inner product with $\boldsymbol{a}$.*

For example $\mathbb{R}^d$ and $\mathcal{H}(\boldsymbol{x})$ are not open convex cones, while $\{\boldsymbol{x} : \boldsymbol{x}_0 > 0\}$ and $\mathbf{Ray}(\boldsymbol{p})$ are open convex cones.

These definitions enable us to express the feasible $\boldsymbol{x}$ for the linear program

$$A\boldsymbol{x} \geq \mathbf{0} \text{ and } \boldsymbol{x} \in \boldsymbol{C}$$

as

$$\boldsymbol{C} \cap \bigcap_{i=1}^{n} \mathcal{H}(\boldsymbol{a}_i),$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are the rows of A. Throughout this chapter, we will call a set *feasible* if it is non-empty, and *infeasible* if it is empty. Thus, we say that $\boldsymbol{C} \cap \bigcap_{i=1}^{n} \mathcal{H}(\boldsymbol{a}_i)$ is feasible if the corresponding linear program is feasible.

1.3 Definition of condition number for linear programming

For a feasible linear program of the form,

$$\max\ \boldsymbol{c}^T\boldsymbol{x} \quad s.t. \quad A\boldsymbol{x} \leq \boldsymbol{b} \tag{1}$$

we follow Renegar [Ren94, Ren95a, Ren95b] in defining the primal condition number, C_P, of the program to be the normalized reciprocal of the distance to ill-posedness. A program is ill-posed if the program can be made both feasible and infeasible by arbitrarily small changes to the pair (A, b). The distance to ill-posedness of the pair (A, b) is the distance to the set of ill-posed programs under the Frobenius norm. We similarly define the dual condition number, C_D, to be the normalized reciprocal of the distance to ill-posedness of the dual program. The condition number, C_{PD}, is the maximum of C_P and C_D.

We can equivalently define the condition number without introducing the concept of ill-posedness. For programs of form (1), define $C_P^{(1)}(A, \boldsymbol{b})$ by

Definition 1.3.1 (Primal Distance to Ill-Posedness)

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

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

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

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

The dual of a program of form (1) is

$$\min\ \boldsymbol{b}^T\boldsymbol{y} \quad s.t. \quad A^T\boldsymbol{y} = \boldsymbol{c}, \quad \boldsymbol{y} \geq \boldsymbol{0},$$

and we define the dual condition number, $C_D^{(1)}(A, \boldsymbol{c})$, analogously.

Any linear program may be expressed in form (1); however, transformations among linear programming formulations do not in general (and commonly do not) preserve condition number [Ren95a]. We will therefore have to define different condition numbers for each normal form we consider. For linear programs with canonical forms:

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

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

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

we define their condition numbers, $C_{PD}^{(2)}$, $C_{PD}^{(3)}$ and $C_{PD}^{(4)}$, analogously. We follow the convention that $\boldsymbol{0}$ is not considered a feasible solution to (4).

As we mentioned previously, the condition numbers for numerous other problems (i.e., matrix inversion) are defined as the sensitivity of the output to perturbations in the input, and then shown to be equivalent to the distance to ill-posedness. Renegar inverts this scheme by defining the condition number for linear programming to be distance to ill-posedness, and then showing that the condition number does bound the sensitivity of the output to perturbations in the input [Ren94, Ren95a].

1.4 Smoothed analysis of the condition number

Following [ST01], we perform a smoothed analysis of these condition numbers. That is, we bound the distributions of these condition numbers for arbitrary programs under slight perturbations. We then derive bounds on the expectations of the logarithms of the condition numbers in terms of the size of the program and the magnitude of the perturbation.

For a linear program specified by $(\bar{A}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}})$, we consider the condition number of the program specified by $(A, \boldsymbol{b}, \boldsymbol{c})$, where A, $\boldsymbol{b}$, and $\boldsymbol{c}$ are a Gaussian random matrix and vectors of variance σ^2 centered at $\bar{A}$, $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{c}}$ respectively. As the condition numbers are unchanged by multiplying all the data by a constant, we assume without loss of generality that in each input form the Frobenius norm of the data is at most 1. This also provides a scaling of the program so that σ measures the relative size of the random perturbation. For completeness, we recall needed facts about Gaussian random variables in section A.

The following is the principal theorem of this chapter:

Theorem 1.4.1 (Smoothed Complexity of Renegar's Condition Number) *For every $\bar{A}$, $\bar{\boldsymbol{b}}$, and $\bar{\boldsymbol{c}}$ such that $\left\|\bar{A}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{c}}\right\|_F \leq 1$, and for all $i \in \{1, 2, 3, 4\}$,*

$$\mathbf{Pr}_{A,\boldsymbol{b},\boldsymbol{c}} \left[C_{PD}^{(i)}(A, \boldsymbol{b}, \boldsymbol{c}) > \frac{2^{14}\ n^2 d^{3/2}}{\delta\sigma^2} \left(\log^2 \frac{2^{10}\ n^2 d^{3/2}}{\delta\sigma^2} \right) \right] \quad < \quad \delta$$

and hence

$$\mathbf{E}_{A,\boldsymbol{b},\boldsymbol{c}} \left[\log C_{PD}^{(i)}(A, \boldsymbol{b}, \boldsymbol{c}) \right] \quad \leq \quad 21 + 3\log(nd/\sigma)$$

where A is a matrix and $\boldsymbol{b}$ and $\boldsymbol{c}$ are vectors of independent Gaussian random variables of variance σ^2, $\sigma^2 \leq 1/(nd)$, centered at $\bar{A}$, $\bar{\boldsymbol{b}}$, and $\bar{\boldsymbol{c}}$, respectively.

2 Primal Condition Number

In this section, we consider problems in conic form

$$\max\ \boldsymbol{c}^T \boldsymbol{x} \text{ such that } A\boldsymbol{x} \geq \boldsymbol{0}, \boldsymbol{x} \in \boldsymbol{C},$$

where $\boldsymbol{C}$ is an open convex cone. Because $\boldsymbol{C}$ is an open convex cone, $\boldsymbol{0}$ cannot be a feasible solution of this program. The primal program of form (1) can be put into conic form with the introduction of the homogenizing variable x_0. Letting $\boldsymbol{C} = \{(\boldsymbol{x}, x_0) : x_0 > 0\}$, the primal program of form (1) is feasible if and only if

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

is feasible. Similarly, the primal and dual programs of form (2) and the dual program of form (3) can also be put into conic form. In each case the transformation into conic form leaves the Frobenius norm unchanged. Also, a random Gaussian perturbation in the original form maps to a random Gaussian perturbation in the conic form.

The following is a generalization of the distance to ill-posedness that we will use throughout this section.

Definition 2.0.2 (Generalized primal distance to ill-posedness) *For an open convex cone, $\boldsymbol{C}$, and a matrix, A, we define $\rho(A, \boldsymbol{C})$ by* *a. if* $A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in C$ *is feasible, then*

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

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

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

We note that this definition makes sense even when A is a column vector. In this case, $\rho(\boldsymbol{a}, \boldsymbol{C})$ measures the distance to ill-posedness when we only allow perturbation to $\boldsymbol{a}$.

The primal program of form (4) is not quite in conic form; to handle it, we need

Definition 2.0.3 (Alternate generalized primal distance to ill-posedness) *For a non-open convex cone,* $\boldsymbol{C}$*, and a matrix,* A*, we define* $\rho(A, \boldsymbol{C})$ *by*

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

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

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

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

This definition would allow us to prove the analog of lemma 2.0.4 for primal programs of form (4). We omit the details of this variation on the arguments in the interest of simplicity.

The main result of this section is:

Lemma 2.0.4 (Primal condition number is likely low) *For any open convex cone* $\boldsymbol{C}$ *and a Gaussian random matrix* A *of variance* σ^2 *centered at a matrix* $\bar{A}$ *satisfying* $\|\bar{A}\|_F \leq 1$*, for* $\sigma \leq 1/\sqrt{nd}$*, we have*

$$\mathbf{Pr}\left[\frac{\|A\|_F}{\rho(A, \boldsymbol{C})} \geq \frac{2^{12} n^2 d^{3/2}}{\delta \sigma^2} \log^2\left(\frac{2^9 n^2 d^{3/2}}{\delta \sigma^2}\right)\right] \leq \delta.$$

The analysis of C_P will proceed as follows: we consider the cases that the program is feasible and infeasible separately. In section 2.1, we show that it is unlikely that a program is feasible and yet can be made infeasible by a small change to its constraints (lemma 2.1.1). In section 2.2, we show that it is unlikely that a program is infeasible and yet can be made feasible by a small change to its constraints (lemma 2.2.1). In section 2.3, we combine these results to show that the primal condition number is low with high probability.

The thread of argument in these sections consists of a geometric characterization of those programs with poor condition number, and then a probabilistic argument demonstrating that this characterization is rarely satisfied. Throughout the proofs in this section, $\boldsymbol{C}$ will always refer to the original open cone, and a subscripted $\boldsymbol{C}$ (i.e., $\boldsymbol{C}_0$) will refer to a modification of this cone.

The key probabilistic tool used in the analysis is lemma 2.0.5, which was proved in [Bal93]. A slightly weaker version of this lemma was proved in [BD02], and also in [BR76].

Lemma 2.0.5 (ϵ-Boundaries are likely to be missed) *Let $\boldsymbol{K}$ be an arbitrary convex body, and let $\mathbf{bdry}(\boldsymbol{K},\epsilon)$ denote the ϵ-boundary of $\boldsymbol{K}$; that is,*

$$\mathbf{bdry}(\boldsymbol{K},\epsilon) = \{\boldsymbol{x} : \exists \boldsymbol{x}' \in \boldsymbol{K}, |\boldsymbol{x} - \boldsymbol{x}'| \leq \epsilon\} \setminus \boldsymbol{K}.$$

Let $\boldsymbol{x}$ be a d-dimensional Gaussian random vector with variance σ^2. Then,

$$\mathbf{Pr}\left[\boldsymbol{x} \in \mathbf{bdry}(\boldsymbol{K},\epsilon)\right] \leq \frac{4\epsilon d^{1/4}}{\sigma}$$

In this section and the next, we use the following consequence of lemma 2.0.5 repeatedly.

Lemma 2.0.6 (Feasible likely quite feasible, single constraint) *Let $\boldsymbol{C}_0$ be any convex cone in $\mathbb{R}^d$ and let $\boldsymbol{a}$ be a Gaussian random vector of variance σ^2. Then,*

$$\mathbf{Pr}_{\boldsymbol{a}}\left[\rho(\boldsymbol{a},\boldsymbol{C}_0) \leq \epsilon\right] \leq \left(\frac{4\epsilon d^{1/4}}{\sigma}\right).$$

Proof: Let $\boldsymbol{K}$ be the set of $\boldsymbol{a}$ for which $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a})$ is infeasible. 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, lemma 2.0.5 tells us that the probability that $\boldsymbol{a}$ has distance at most ϵ to the boundary of $\boldsymbol{K}$ is at most $\left(\frac{4\epsilon d^{1/4}}{\sigma}\right)$. $\square$

2.1 Primal number, feasible case

In this subsection, we analyze the primal condition number in the feasible case, and prove:

Lemma 2.1.1 (Feasible is likely quite feasible, all constraints) *Let $\boldsymbol{C}$ be an open convex cone in $\mathbb{R}^d$ and let A be an n-by-d Gaussian random matrix of variance σ^2. Then,*

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

The remaining lemmas in this subsection establish a necessary geometric condition for ρ to be small. In the proof of lemma 2.1.1 at the end of this subsection, we use lemma 2.0.6 to show that this geometric condition is unlikely to be met.

Lemma 2.1.2 (Feasibility as a dot product) *For every vector $\boldsymbol{a}$ and every unit vector $\boldsymbol{p}$,*

$$\rho(\boldsymbol{a},\mathbf{Ray}\,(\boldsymbol{p})) = \left|\boldsymbol{a}^T\boldsymbol{p}\right|$$

Proof: Since $\mathbf{Ray}\,(\boldsymbol{p}) \cap \mathcal{H}(\boldsymbol{a})$ is feasible if and only if $\mathbf{Ray}\,(-\boldsymbol{p}) \cap \mathcal{H}(\boldsymbol{a})$ is infeasible, it suffices to consider the case where $\mathbf{Ray}\,(\boldsymbol{p}) \cap \mathcal{H}(\boldsymbol{a})$ is feasible. In this case $\boldsymbol{a}^T\boldsymbol{p} \geq 0$. We first prove that $\rho(\boldsymbol{a},\mathbf{Ray}\,(\boldsymbol{p})) \geq \boldsymbol{a}^T\boldsymbol{p}$. For every vector $\Delta\boldsymbol{a}$ of norm at most $\boldsymbol{a}^T\boldsymbol{p}$, we have

$$(\boldsymbol{a}+\Delta\boldsymbol{a})^T\boldsymbol{p} = \boldsymbol{a}^T\boldsymbol{p} + \Delta\boldsymbol{a}^T\boldsymbol{p} \geq \boldsymbol{a}^T\boldsymbol{p} - \|\Delta\boldsymbol{a}\| \geq 0.$$

Thus $\boldsymbol{p} \in \mathcal{H}(\boldsymbol{a}+\Delta\boldsymbol{a})$. As this holds for every $\Delta\boldsymbol{a}$ of norm at most $\boldsymbol{a}^T\boldsymbol{p}$, we have $\rho(\boldsymbol{a},\mathbf{Ray}\,(\boldsymbol{p})) \geq \boldsymbol{a}^T\boldsymbol{p}$.

To show that $\rho(\boldsymbol{a},\mathbf{Ray}\,(\boldsymbol{p})) \leq \boldsymbol{a}^T\boldsymbol{p}$, note that for any $\epsilon > 0$, setting $\Delta\boldsymbol{a} = -(\epsilon + \boldsymbol{a}^T\boldsymbol{p})\boldsymbol{p}$ yields

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

so $\mathbf{Ray}\,(\boldsymbol{p}) \cap \mathcal{H}(\boldsymbol{a}+\Delta\boldsymbol{a})$ is infeasible. As this holds for every $\epsilon > 0$, $\rho(\boldsymbol{a},\mathbf{Ray}\,(\boldsymbol{p})) \leq \boldsymbol{a}^T\boldsymbol{p}$. $\square$

Lemma 2.1.3 (Quite feasible region implies quite feasible point, single constraint) *For every* $\boldsymbol{a}$ *and every open convex cone* $\boldsymbol{C}_0$ *for which* $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a})$ *is feasible,*

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

Proof: The "$\geq$" direction is obvious from lemma 2.1.2, so we concentrate on showing

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

As $\boldsymbol{C}_0$ is open, there exists a vector $\boldsymbol{t}$ such that $\boldsymbol{t}^T \boldsymbol{x} < 0$ for each $\boldsymbol{x} \in \boldsymbol{C}_0$. If $\boldsymbol{a} \in \boldsymbol{C}_0$, then

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

For every $\epsilon > 0$, $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a} - (\boldsymbol{a} + \epsilon \boldsymbol{t}))$ is infeasible; so $\rho(\boldsymbol{a}, \boldsymbol{C}_0) \leq \|\boldsymbol{a}\|$.

If $\boldsymbol{a} \notin \boldsymbol{C}_0$, let $\boldsymbol{q}$ be the point of $\boldsymbol{C}_0$ that is closest to $\boldsymbol{a}$. As $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a})$ is feasible, $\boldsymbol{q}$ is not the origin and we can define $\boldsymbol{p} = \boldsymbol{q}/\|\boldsymbol{q}\|$. As $\boldsymbol{C}_0$ is a cone, $\boldsymbol{q}$ is perpendicular to $\boldsymbol{a} - \boldsymbol{q}$. Thus, the distance from $\boldsymbol{a}$ to $\boldsymbol{q}$ is the square root of $\|\boldsymbol{a}\|^2 - (\boldsymbol{a}^T \boldsymbol{p})^2$, and $\boldsymbol{p}$ must be the point of unit norm maximizing $\boldsymbol{a}^T \boldsymbol{p}$.

As $\boldsymbol{C}_0$ is convex, there is a plane through $\boldsymbol{q}$ separating $\boldsymbol{C}_0$ from $\boldsymbol{a}$ and perpendicular to the line segment $\boldsymbol{a} - \boldsymbol{q}$. Thus, every point of $\boldsymbol{C}_0$ has inner product at most zero with the vector $\boldsymbol{a} - \boldsymbol{q}$; and hence, for every $\epsilon > 0$, $\boldsymbol{C}_0 \cap \mathcal{H}(\boldsymbol{a} - \boldsymbol{q} + \epsilon \boldsymbol{t})$ is infeasible. To conclude the proof, we note that $\|\boldsymbol{q}\| = \boldsymbol{a}^T \boldsymbol{p}$. □

Lemma 2.1.4 (Quite feasible point for each constraint implies quite feasible point for all constraints) *If there exist vectors* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *and unit vectors* $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n \in \boldsymbol{C}_0$, $\boldsymbol{C}_0 \subset \mathbb{R}^d$, *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 point $\boldsymbol{p}$ *of unit norm,* $\boldsymbol{p} \in \boldsymbol{C}_0$, *such that*

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

Proof: We prove this using Helly's theorem [LDK63]. Let $\boldsymbol{S}_i = \{\boldsymbol{x} \in \boldsymbol{C}_0 : \boldsymbol{a}_i^T \boldsymbol{x}/\|\boldsymbol{x}\| \geq \epsilon/d\}$. As $\boldsymbol{C}_0$ is open, there exists $\boldsymbol{t}$ such that $\boldsymbol{t}^T \boldsymbol{x} < 0$, $\forall \boldsymbol{x} \in \boldsymbol{C}_0$. Let $\boldsymbol{S}'_i = \boldsymbol{S}_i \bigcap \{\boldsymbol{x} : \boldsymbol{t}^T \boldsymbol{x} = -1\}$. The $\{\boldsymbol{S}'_i\}$ have similar intersection to the $\{\boldsymbol{S}_i\}$ in that $\boldsymbol{x} \in \boldsymbol{S}'_i \Rightarrow \boldsymbol{x} \in \boldsymbol{S}_i$ and $\boldsymbol{x} \in \boldsymbol{S}_i \Rightarrow \boldsymbol{x}/\boldsymbol{t}^T \boldsymbol{x} \in \boldsymbol{S}'_i$. However, the $\{\boldsymbol{S}'_i\}$ are convex sets in a $(d-1)$-dimensional subspace. By Helly's theorem, if every subcollection of d of the $\{\boldsymbol{S}'_i\}$ has a common point, then the entire collection has a common point. Because the $\{\boldsymbol{S}_i\}$ have similar intersection to the $\{\boldsymbol{S}'_i\}$, the same statement holds for the $\{\boldsymbol{S}_i\}$. So assume $n = d$.

Let $\boldsymbol{p} = \sum_{i=1}^d \boldsymbol{p}_i/d$. Then, for each $1 \leq j \leq d$,

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

Moreover, $\boldsymbol{p}$ has norm at most one, so $\boldsymbol{p}/\|\boldsymbol{p}\|$ is a point that lies in each of $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_d$. □

Lemma 2.1.5 (Quite feasible point for all constraints implies quite feasible program) *For every set of vectors* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *and* $\boldsymbol{p}$ *such that* $\mathbf{Ray}(\boldsymbol{p}) \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i)$ *is feasible,*

$$\rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \mathbf{Ray}(\boldsymbol{p})) = \min_i \rho(\boldsymbol{a}_i, \mathbf{Ray}(\boldsymbol{p})).$$

Proof: It suffices to observe that $\mathbf{Ray}\,(\boldsymbol{p}) \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i)$ is feasible if and only if $\boldsymbol{p}^T(\boldsymbol{a}_i + \Delta \boldsymbol{a}_i) \geq 0$ for all i. □

We now prove the main result of this section.

Proof of Lemma 2.1.1 Let $\boldsymbol{C}_0 = \boldsymbol{C} \cap \bigcap_i \mathcal{H}(\boldsymbol{a}_i)$ and $\boldsymbol{C}_i = \boldsymbol{C} \cap \bigcap_{j \neq i} \mathcal{H}(\boldsymbol{a}_j)$. Note that

$$\{\boldsymbol{x} : A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C}\} = \boldsymbol{C}_0.$$

Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the columns of A. Our first step will be to show that if $\boldsymbol{C}_0$ is feasible, then

$$\rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}) \leq \epsilon/d$$

implies that there exists an i for which

$$\rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon.$$

To show this, we prove the contrapositive. That is, we assume $\boldsymbol{C}_0$ is feasible and that $\rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \geq \epsilon$ for all i. Then, lemma 2.1.3 implies that there exist unit vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n \in \boldsymbol{C}_0$ such that $\boldsymbol{a}_i^T \boldsymbol{p}_i \geq \epsilon$. Applying lemma 2.1.4, we find a unit vector $\boldsymbol{p} \in \boldsymbol{C}_0$ such that $\boldsymbol{a}_i^T \boldsymbol{p} \geq \epsilon/d$ for all i. From lemmas 2.1.2 and 2.1.5, we then compute

$$\rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}) \geq \rho([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \mathbf{Ray}\,(\boldsymbol{p})) = \min_i \rho(\boldsymbol{a}_i, \mathbf{Ray}\,(\boldsymbol{p})) \geq \epsilon/d.$$

Thus, we now know

$$\begin{aligned}
&\mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is feasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon/d)\right] \\
\leq\ &\mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n}\left[\boldsymbol{C}_0 \text{ is feasible and } \exists i : \rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon\right].
\end{aligned}$$

To bound the latter probability, we use lemma 2.0.6, which tells us that

$$\mathbf{Pr}_{\boldsymbol{a}_i}\left[\rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon \text{ and } \boldsymbol{C}_i \text{ is feasible}\right] \leq \left(\frac{4\epsilon d^{1/4}}{\sigma}\right).$$

Applying a union bound and the fact that $\boldsymbol{C}_0$ feasible implies $\boldsymbol{C}_i$ is feasible, we compute

$$\begin{aligned}
\mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n}\left[\boldsymbol{C}_0 \text{ is feasible and } \exists i : \rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon\right] &\leq \\
\sum_i \mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n}\left[\boldsymbol{C}_0 \text{ is feasible and } \rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon\right] &\leq \\
\sum_i \mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n}\left[\boldsymbol{C}_i \text{ is feasible and } \rho(\boldsymbol{a}_i, \boldsymbol{C}_i) \leq \epsilon\right] &\leq \left(\frac{4\epsilon n d^{1/4}}{\sigma}\right).
\end{aligned} \tag{1}$$

Setting $\epsilon = d\epsilon'$ yields the lemma as stated. □

This concludes the analysis that it is unlikely that the primal program is both feasible and has small distance to ill-posedness. Next, we show that it is unlikely that the primal program is both infeasible and has small distance to ill-posedness.

2.2 Primal number, infeasible case

The main result of this subsection is:

Lemma 2.2.1 (Infeasible is likely quite infeasible) *Let $\boldsymbol{C}$ be an open convex cone in $\mathbb{R}^d$ and let A be a Gaussian random matrix of variance σ^2 centered at a matrix $\bar{A}$ satisfying $\|\bar{A}\|_F \leq 1$, where $\sigma \leq 1/\sqrt{d}$. Then,*

$$\mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C} \text{ is infeasible}) \text{ and } (\rho(A, \boldsymbol{C}) \leq \epsilon)\right] \quad \leq \quad \left(\frac{360\ \epsilon n^2 d^{3/2} \lceil \log^{1.5}(1/\epsilon) \rceil}{\sigma^2}\right).$$

To prove lemma 2.2.1, we consider adding the constraints one at a time. If the program is infeasible in the end, then there must be some next constraint that takes it from being feasible to being infeasible. Lemma 2.2.2 gives a sufficient geometric characterization for the program to be quite infeasible when the next constraint is added, and in the proof of lemma 2.2.1, we show that this characterization is met with good probability. The geometric characterization is that the program is quite feasible before the next constraint is added and every previously feasible point is far from being feasible for the next constraint.

Lemma 2.2.2 (The feasible to infeasible transition) *Let $\boldsymbol{C}$ be an open convex cone, $\boldsymbol{p}$ be a unit vector, $\boldsymbol{p} \in \boldsymbol{C}$, and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}$ be vectors such that*

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

Then,

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

Proof: We will prove this by showing that for all ϵ satisfying

$$\epsilon \leq \alpha/2, \text{ and} \tag{2}$$

$$\epsilon < \frac{\beta}{4 + 2\|\boldsymbol{a}_{k+1}\|/\alpha}, \tag{3}$$

and $\{\Delta\boldsymbol{a}_1, \ldots, \Delta\boldsymbol{a}_{k+1}\}$ satisfying $\|\Delta\boldsymbol{a}_i\| < \epsilon$ for $1 \leq i \leq k+1$, we have

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

Assume by way of contradiction that

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

Then, there exists a unit vector $\boldsymbol{x}' \in \boldsymbol{C} \cap \bigcap_{i=1}^{k+1} \mathcal{H}(\boldsymbol{a}_i + \Delta\boldsymbol{a}_i)$. We first show that

$$\boldsymbol{x}' + \frac{\epsilon}{\alpha}\boldsymbol{p} \in \boldsymbol{C} \cap \bigcap_{i=1}^{k} \mathcal{H}(\boldsymbol{a}_i).$$

To see this consider any $i \leq k$ and note that

$$(\boldsymbol{a}_i + \Delta\boldsymbol{a}_i)^T \boldsymbol{x}' \geq 0 \implies \boldsymbol{a}_i^T \boldsymbol{x}' \geq -\Delta\boldsymbol{a}_i^T \boldsymbol{x}' \geq -\|\Delta\boldsymbol{a}_i\|\|\boldsymbol{x}'\| \geq -\epsilon.$$

Thus,

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

Also,

$$\boldsymbol{x}' \in \boldsymbol{C}, \quad \boldsymbol{p} \in \boldsymbol{C} \quad \Longrightarrow \quad \boldsymbol{x}' + \frac{\epsilon}{\alpha}\boldsymbol{p} \in \boldsymbol{C}$$

Let $\boldsymbol{x} = \boldsymbol{x}' + \frac{\epsilon}{\alpha}\boldsymbol{p}$. Then $\boldsymbol{x} \in \boldsymbol{C} \cap \cap_{i=1}^k \mathcal{H}(\boldsymbol{a}_i)$ and $\boldsymbol{x}$ has norm at most $1 + \epsilon/\alpha$ and at least $1 - \epsilon/\alpha$. To derive a contradiction, we now compute

$$\begin{aligned}
(\boldsymbol{a}_{k+1} + \Delta\boldsymbol{a}_{k+1})^T\boldsymbol{x}' &= (\boldsymbol{a}_{k+1} + \Delta\boldsymbol{a}_{k+1})^T(\boldsymbol{x} - (\epsilon/\alpha)\boldsymbol{p}) \\
&= \boldsymbol{a}_{k+1}^T\boldsymbol{x} + \Delta\boldsymbol{a}_{k+1}^T\boldsymbol{x} - (\epsilon/\alpha)\boldsymbol{a}_{k+1}^T\boldsymbol{p} - (\epsilon/\alpha)\Delta\boldsymbol{a}_{k+1}^T\boldsymbol{p} \\
&\leq -\beta\|\boldsymbol{x}\| + \|\Delta\boldsymbol{a}_{k+1}\|\|\boldsymbol{x}\| + (\epsilon/\alpha)\|\boldsymbol{a}_{k+1}\| + (\epsilon/\alpha)\|\Delta\boldsymbol{a}_{k+1}\| \\
&\leq -\beta(1 - \epsilon/\alpha) + \epsilon(1 + \epsilon/\alpha) + (\epsilon/\alpha)\|\boldsymbol{a}_{k+1}\| + (\epsilon^2/\alpha) \\
&= -\beta(1 - \epsilon/\alpha) + \epsilon\left((1 + \epsilon/\alpha) + \|\boldsymbol{a}_{k+1}\|/\alpha + \epsilon/\alpha\right) \\
&\leq -\beta/2 + \epsilon\left((2 + \|\boldsymbol{a}_{k+1}\|/\alpha\right), \text{ by (2)} \\
&< 0 \text{ by (3)}.
\end{aligned}$$

□

The next two items are trivial.

Proposition 2.2.3 *For positive α, β and any vector $\boldsymbol{a}_{k+1}$,*

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

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

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

while for $\alpha \leq 1$ we have

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

□

Proposition 2.2.4 *If $\boldsymbol{C} \cap \bigcap_{i=1}^k \mathcal{H}(\boldsymbol{a}_i)$ is infeasible, then*

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

Proof: Adding constraints cannot make it easier to change the program to make it feasible. □

We now prove the main result of this section.

Proof of Lemma 2.2.1 Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the columns of A, let

$$\boldsymbol{C}_k = \boldsymbol{C} \cap \bigcap_{i=1}^k \mathcal{H}(\boldsymbol{a}_k),$$

and let $\boldsymbol{C}_n$ be the final program. Let E_k denote the event that $\boldsymbol{C}_{k-1}$ is feasible and $\boldsymbol{C}_k$ is infeasible. Using the fact that $\boldsymbol{C}_n$ infeasible implies that E_k must hold for some k and proposition 2.2.4, we obtain

$$\begin{aligned} \mathbf{Pr}\left[\boldsymbol{C}_n \text{ is infeasible and } \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}\right) \leq \epsilon\right] &\leq \\ \sum_{k=1}^{n} \mathbf{Pr}\left[E_k \text{ and } \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \boldsymbol{C}\right) \leq \epsilon\right] &\leq \\ \sum_{k=1}^{n} \mathbf{Pr}\left[E_k \text{ and } \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k], \boldsymbol{C}\right) \leq \epsilon\right]. \end{aligned} \tag{4}$$

If $\boldsymbol{C}_k$ is feasible, define

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

By equation 1 and lemma 2.1.2,

$$\mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k}\left[\boldsymbol{C}_k \text{ is feasible and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq \epsilon\right] \quad \leq \quad \frac{4\epsilon n d^{5/4}}{\sigma} \tag{5}$$

By lemma 2.2.2 and proposition 2.2.3, E_{k+1} implies

$$\begin{aligned} \rho\left([\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}], \boldsymbol{C}\right) &\geq \min\left\{\frac{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)}{2}, \frac{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)}{4 + 2\|\boldsymbol{a}_{k+1}\|}, \frac{\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)}{4 + 2\|\boldsymbol{a}_{k+1}\|}\right\} \\ &\geq \frac{\min\left\{\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k),\ \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k),\ \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k)\right\}}{4 + 2\|\boldsymbol{a}_{k+1}\|} \end{aligned} \tag{6}$$

We now proceed to bound the probability that the numerator of this fraction is small. We first note that

$$\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq \delta$$

implies that either $\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq \delta$, $\rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq \delta$, or there exists an l between 1 and $\lceil \log(1/\delta) \rceil$ for which

$$\kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq 2^{-l+1} \text{ and } \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq 2^l \delta.$$

We apply lemma 2.0.6 to bound

$$\mathbf{Pr}\left[E_{k+1} \text{ and } \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq \delta\right] \leq \frac{4\delta d^{1/4}}{\sigma},$$

and lemma 2.1.1 to bound

$$\mathbf{Pr}\left[E_{k+1} \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq \delta\right] \leq \frac{4\delta n d^{5/4}}{\sigma}.$$

For $1 \leq l \leq \lceil \log(1/\delta) \rceil$, we bound

$$\begin{aligned} &\mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}}\left[E_{k+1} \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq 2^{-l+1} \text{ and } \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq 2^l \delta\right] \\ &= \mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k}\left[\boldsymbol{C}_k \neq \emptyset \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq 2^{-l+1}\right] \cdot \\ &\qquad \mathbf{Pr}_{\boldsymbol{a}_{k+1}}\left[\ \boldsymbol{C}_{k+1} = \emptyset \text{ and } \rho(\boldsymbol{a}_{k+1}, \boldsymbol{C}_k) \leq 2^l \delta \ \middle|\ \boldsymbol{C}_k \neq \emptyset \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq 2^{-l+1}\right] \\ &\leq \mathbf{Pr}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k}\left[\boldsymbol{C}_k \neq \emptyset \text{ and } \kappa(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \leq 2^{-l+1}\right] \frac{2^l 4\delta d^{1/4}}{\sigma}\ , \text{ by lemma 2.0.6,} \\ &\leq \frac{2^{-l+1} 4 n d^{5/4}}{\sigma} \frac{2^l 4\delta d^{1/4}}{\sigma}\ , \text{ by equation 5,} \\ &= \frac{32\delta n d^{3/2}}{\sigma^2} \end{aligned}$$

Summing over the choice for l, we obtain

$$\begin{aligned}
&\mathbf{Pr}\left[E_{k+1} \text{ and } \min\{\kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k), \kappa(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)\rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k), \rho(\boldsymbol{a}_{k+1},\boldsymbol{C}_k)\} < \delta\right] \\
&\quad\leq \frac{4\delta n d^{5/4}}{\sigma} + \frac{4\delta d^{1/4}}{\sigma} + \lceil\log(1/\delta)\rceil \frac{32\delta n d^{3/2}}{\sigma^2} \\
&\quad\leq \delta\left(\frac{4nd^{3/4} + 4 + 32\lceil\log(1/\delta)\rceil n d^{3/2}}{\sigma^2}\right), \text{ by } \sigma \leq 1/\sqrt{d}, \\
&\quad\leq \delta\left(\frac{40\lceil\log(1/\delta)\rceil n d^{3/2}}{\sigma^2}\right).
\end{aligned} \tag{7}$$

On the other hand, we can bound the denominator of (6) by observing that $\boldsymbol{a}_{k+1}$ is a Gaussian centered at a point $\bar{\boldsymbol{a}}_{k+1}$ of norm at most 1; so, corollary A.0.14 implies

$$\mathbf{Pr}\left[4 + 2\left\|\boldsymbol{a}_{k+1}\right\| \geq 6 + 2\sigma\sqrt{2d\ln(e/\epsilon)}\right] \leq \epsilon.$$

By applying this bound and (7) with $\delta = \epsilon(6 + 2\sigma\sqrt{2d\ln(e/\epsilon)})$, we obtain

$$\text{(via the schema: } \Pr[\frac{num}{den} \leq \epsilon] \;\leq\; \Pr[den \geq \frac{\delta}{\epsilon}] + \Pr[num \leq \delta] \text{)}$$

$$\begin{aligned}
&\mathbf{Pr}\left[E_k \text{ and } \rho\left([\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k],\boldsymbol{C}\right) \leq \epsilon\right] \\
&\quad\leq \epsilon + \epsilon\left(6 + 2\sigma\sqrt{2d\ln(e/\epsilon)}\right)\left(\frac{40\lceil\log(1/\epsilon(3+\sigma\sqrt{d\log(e/\epsilon)}))\rceil n d^{3/2}}{\sigma^2}\right) \\
&\quad\leq \epsilon + \epsilon\left(6 + 3\sqrt{\ln(e/\epsilon)}\right)\left(\frac{40\lceil\log(1/(3\epsilon))\rceil n d^{3/2}}{\sigma^2}\right), \text{ using } \sigma \leq 1/\sqrt{d} \text{ in the first term} \\
&\quad\leq \epsilon + \epsilon\left(9\sqrt{\ln(e/\epsilon)}\right)\left(\frac{40\lceil\log(1/(3\epsilon))\rceil n d^{3/2}}{\sigma^2}\right) \\
&\quad\leq \epsilon + \epsilon\left(\frac{360\lceil\log^{1.5}(1/\epsilon)\rceil n d^{3/2}}{\sigma^2}\right), \text{ since } (\sqrt{\ln(e/\epsilon)})(\log(1/3\epsilon)) \leq \log^{1.5}(1/\epsilon) \\
&\quad\leq \epsilon\left(\frac{361\lceil\log^{1.5}(1/\epsilon)\rceil n d^{3/2}}{\sigma^2}\right)
\end{aligned}$$

Plugging this in to (4), we get

$$\mathbf{Pr}\left[\boldsymbol{C}_0 \text{ is infeasible and } \rho\left([\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n],\boldsymbol{C}\right) \leq \epsilon\right] \;\leq\; \frac{361\epsilon n^2 d^{3/2}\lceil\log^{1.5}(1/\epsilon)\rceil}{\sigma^2}.$$

□

2.3 Primal number, both cases

We combine the results of sections 2.1 and 2.2 to prove lemma 2.0.4, that the primal condition number is probably low.

Proof of Lemma 2.0.4 In lemma 2.1.1, we show that

$$\mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C}_0 \text{ is feasible}) \text{ and } (\rho(A,\boldsymbol{C}_0) \leq \epsilon)\right] \leq \left(\frac{4\epsilon n d^{5/4}}{\sigma}\right),$$

while in lemma 2.2.1, we show

$$\mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C}_0 \text{ is infeasible}) \text{ and } (\rho(A, \boldsymbol{C}_0) \leq \epsilon)\right] \leq \left(\frac{361\epsilon\lceil\log^{1.5}(1/\epsilon)\rceil n^2 d^{3/2}}{\sigma^2}\right).$$

Thus,

$$\begin{aligned}
\mathbf{Pr}\left[\rho(A, \boldsymbol{C}) \leq \epsilon\right] &= \mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C}_0 \text{ is feasible}) \text{ and } (\rho(A, \boldsymbol{C}_0) \leq \epsilon)\right] \\
&\quad + \mathbf{Pr}\left[(A\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \boldsymbol{C}_0 \text{ is infeasible}) \text{ and } (\rho(A, \boldsymbol{C}_0) \leq \epsilon)\right] \\
&\leq \left(\frac{4\epsilon n d^{5/4}}{\sigma}\right) + \left(\frac{361\epsilon\lceil\log^{1.5}(1/\epsilon)\rceil n^2 d^{3/2}}{\sigma^2}\right) \\
&\leq \left(\frac{365\epsilon\lceil\log^{1.5}(1/\epsilon)\rceil n^2 d^{3/2}}{\sigma^2}\right)
\end{aligned}$$

Letting $\epsilon = \delta/(3\alpha\log^{1.5}(\alpha/\delta))$ where $\alpha = 365\frac{n^2 d^{3/2}}{\sigma^2}$, and using that $\mathbf{Pr}\left[\rho(A, \boldsymbol{C}) < \epsilon\right] \Leftrightarrow \mathbf{Pr}\left[\frac{1}{\rho(A, \boldsymbol{C})} > \frac{1}{\epsilon}\right]$ yields

$$\begin{aligned}
\mathbf{Pr}\left[\frac{1}{\rho(A, \boldsymbol{C})} > \frac{1100\ n^2 d^{3/2}}{\delta\sigma^2}\log^{3/2}\left(\frac{370\ n^2 d^{3/2}}{\delta^2\sigma^2}\right)\right] &\leq \frac{\alpha\delta\log^{1.5}\left(\frac{3\alpha}{\delta}\log^{1.5}\left(\frac{\alpha}{\delta}\right)\right)}{3\alpha\log^{1.5}\left(\frac{\alpha}{\delta}\right)} \\
&\leq \delta/2. \qquad (8)
\end{aligned}$$

At the same time, corollary A.0.14 tells us that

$$\mathbf{Pr}\left[\|A\|_F \geq 1 + \sigma\sqrt{nd\ 2\ln(2e/\delta)}\right] \leq \delta/2.$$

The lemma now follows by applying this bound, $\sigma \leq 1/\sqrt{nd}$, and (8), to get

$$\mathbf{Pr}\left[\frac{\|A\|_F}{\rho(A, \boldsymbol{C})} > \frac{(1 + \sqrt{2\ln(2e/\delta)})1100\ n^2 d^{3/2}}{\delta\sigma^2}\log^{3/2}\left(\frac{370\ n^2 d^{3/2}}{\delta^2\sigma^2}\right)\right] < \delta$$

To derive the lemma as stated, we note

$$\frac{(1 + \sqrt{2\ln(2e/\delta)})1100\ n^2 d^{3/2}}{\delta\sigma^2}\log^{3/2}\left(\frac{370\ n^2 d^{3/2}}{\delta^2\sigma^2}\right) \leq \frac{2^{12}\ n^2 d^{3/2}}{\delta^2\sigma^2}\log^2\left(\frac{2^9\ n^2 d^{3/2}}{\delta^2\sigma^2}\right)$$

. □

3 Dual Condition Number

In this section, we consider linear programs of the form

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

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 begin by defining distance to ill-posedness appropriately for the form of linear program considered in this section:

Definition 3.0.1 (Dual distance to ill-posedness) *For a matrix, A, and a vector $\boldsymbol{c}$, we define $\rho(A, \boldsymbol{c})$ by* *a. if* $A^T \boldsymbol{y} = \boldsymbol{c},\ \boldsymbol{y} \geq \boldsymbol{0}$ *is feasible, then* $\rho(A, \boldsymbol{c}) =$

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

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

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

The main result of this section is:

Lemma 3.0.2 (Dual condition number is likely low) *Let* A *and* $\boldsymbol{c}$ *be a Gaussian random matrix and vector of variance* σ^2*,* $\sigma \leq 1/\sqrt{nd}$*, centered at* $\bar{A}$ *and* $\bar{\boldsymbol{c}}$*, respectively. If* $\left\|\bar{A}\right\|_F \leq 1$ *and* $\|\bar{\boldsymbol{c}}\| \leq 1$*, then*

$$\mathbf{Pr}\left[\frac{\|A, \boldsymbol{c}\|_F}{\rho(A, \boldsymbol{c})} > \frac{1000\ d^{1/4} n^{1/2}}{\epsilon \sigma^2} \log^{1.5}\left(\frac{200\ d^{1/4} n^{1/2}}{\epsilon \sigma^2}\right)\right] \quad \leq \quad \epsilon.$$

We begin by giving several common definitions that will be useful in our analysis of the dual condition number. We define a change of variables, and we then develop a sufficient geometric condition for the dual condition number to be low. In lemma 3.0.9 and in the proof of lemma 3.0.2, we show that this geometric condition is met with good probability.

Definition 3.0.3 (Cone) *For a set of vectors* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$*, let* $\mathbf{Cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ *denote* $\{\boldsymbol{x} : \boldsymbol{x} = \sum_i \lambda_i \boldsymbol{a}_i, \quad \lambda_i \geq 0\}$.

Definition 3.0.4 (Hull) *For a set of vectors* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$*, let* $\mathbf{Hull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ *denote* $\{\boldsymbol{x} : \boldsymbol{x} = \sum_i \lambda_i \boldsymbol{a}_i, \quad \lambda_i \geq 0, \quad \sum_i \lambda_i = 1\}$.

Definition 3.0.5 (Boundary of a set) *For a convex set* $\boldsymbol{S}$*, let* $\mathbf{bdry}(\boldsymbol{S})$ *denote the boundary of* $\boldsymbol{S}$*, i.e.,* $\{\boldsymbol{x} : \forall \epsilon > 0,\ \exists \boldsymbol{e},\ \|\boldsymbol{e}\| \leq \epsilon,\ s.t.\ \boldsymbol{x} + \boldsymbol{e} \in \boldsymbol{S},\ \boldsymbol{x} - \boldsymbol{e} \notin \boldsymbol{S}\}$.

Definition 3.0.6 (Point-to-set distance) *Let* $\mathbf{dist}(\boldsymbol{x}, \boldsymbol{S})$ *denote the distance of* $\boldsymbol{x}$ *to* $\boldsymbol{S}$*, i.e.,* $\min\{\epsilon : \exists \boldsymbol{e},\ \|\boldsymbol{e}\| \leq \epsilon,\ s.t.\ \boldsymbol{x} + \boldsymbol{e} \in \boldsymbol{S}\}$.

Note that $\mathbf{Cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is not an open cone, while $\mathbf{Hull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is the standard convex hull of $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.

We observe that there exists a solution to the system $A^T \boldsymbol{y} = \boldsymbol{c},\ \boldsymbol{y} \geq \boldsymbol{0}$ if and only if

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

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

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

The main idea we need beyond the ideas of section 2 is to perform an illuminating change of variables. We set

$$\begin{aligned} \boldsymbol{z} &= (1/n) \sum_{i=1}^{n} \boldsymbol{a}_i, \text{ and} \\ \boldsymbol{x}_i &= \boldsymbol{a}_i - \boldsymbol{z}, \text{for } i = 1 \text{ to } n - 1. \end{aligned}$$

For notational convenience, we let $\boldsymbol{x}_n = \boldsymbol{a}_n - \boldsymbol{z}$, although $\boldsymbol{x}_n$ is not independent of $\{\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\}$.

We can restate the condition for the linear progam to be ill-posed in these new variables:

Lemma 3.0.7 (Ill-posedness in new variables)

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

Proof: We observe

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

For $\boldsymbol{c} \neq \boldsymbol{0}$, $\mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is a continuous mapping from $\boldsymbol{c}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ to subsets of Euclidean space, and so for $\boldsymbol{z}$ in the set and not on the boundary, a sufficiently small change to all the variables simultaneously will always leave $\boldsymbol{z}$ in the set, and similarly for $\boldsymbol{z}$ not in the set and not on the boundary.

To establish the other direction, if $\boldsymbol{z}$ is on the boundary, we can just perturb $\boldsymbol{z}$ to bring it in or out of the set. Although $\boldsymbol{z}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are determined by the $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, we can perturb the $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ so as to change the value of $\boldsymbol{z}$ without changing the values of any of the $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ (see the proof of lemma 3.0.8 below for more detail on why the change of variables permits this).

The lemma is also true for $\boldsymbol{c} = \boldsymbol{0}$, but we will not need this. □

Note that $\mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is a convex set. The following lemma will allow us to apply lemma 2.0.5 to determine the probability that $\boldsymbol{z}$ is near the boundary of this convex set.

Lemma 3.0.8 (Independence of mean among new variables) *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random vectors of variance σ^2 lying in $\mathbb{R}^d$. Let*

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

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

Proof: As $\boldsymbol{z}$ is the average of n Gaussian random vectors of variance σ^2, it is a Gaussian random vector of variance σ^2/n. We have that $\boldsymbol{z}$ is independent of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ because the linear combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ used to obtain $\boldsymbol{z}$ is orthogonal to the linear combinations of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ used to obtain the $\boldsymbol{x}_i$s. □

We proceed to apply lemma 2.0.5.

Lemma 3.0.9 (Mean is likely far from ill-posedness) *Let $\boldsymbol{c}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random vectors of variance σ^2 lying in $\mathbb{R}^d$. Let*

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

Then,

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

Proof: Let $\boldsymbol{c}$ be arbitrary. By lemma 3.0.8, we can choose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and then choose $\boldsymbol{z}$ independently. Having chosen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, we fix the convex body $\mathbf{Ray}\,(\boldsymbol{c}) - \mathbf{Hull}\,(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and apply lemma 2.0.5. The factor of 2 arises because $\boldsymbol{z}$ must miss an ϵ boundary on either side of the convex body. □

Lemma 3.0.10 (Geometric condition to be far from ill-posedness in new variables.) *If*

$$\mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) > \alpha \tag{9}$$

and

$$\begin{aligned} \|\Delta \boldsymbol{x}_i\| &\leq \alpha/4, \\ \|\Delta \boldsymbol{z}\| &\leq \alpha/4, \\ \|\Delta \boldsymbol{c}\| &\leq \frac{\alpha \|\boldsymbol{c}\|}{2\alpha + 4(\|\boldsymbol{z}\| + \max_i \|\boldsymbol{x}_i\|)}, \end{aligned}$$

then

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

Proof: Assume for the purpose of showing a contradiction that

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

Consider the case that $\boldsymbol{z} \notin \mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. We will show that $\mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) \leq \alpha$, contradicting our lemma assumption (9). Since $\boldsymbol{z} + \Delta \boldsymbol{z} \in \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c} + \Delta \boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1 + \Delta \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta \boldsymbol{x}_n\right))$,

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

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

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

We lower bound the denominator by $\|\boldsymbol{c}\|/2$ by observing that

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

We upper bound the numerator by

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

Thus,

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

Since

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

We find that

$$\begin{aligned}\textbf{dist}\left(\boldsymbol{z}, \textbf{bdry}(\textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) &\leq \left\| \Delta\boldsymbol{z} - \lambda\Delta\boldsymbol{c} + \sum_i \gamma_i \Delta\boldsymbol{x}_i \right\| \\ &\leq \left\|\Delta\boldsymbol{z}\right\| + \lambda\left\|\Delta\boldsymbol{c}\right\| + \sum_i \gamma_i \left\|\Delta\boldsymbol{x}_i\right\|\end{aligned}$$

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

This establishes a contradiction in the case that $\boldsymbol{z} \notin \textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. Now consider the case that $\boldsymbol{z} \in \textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. Since

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

there exists a hyperplane H passing through $\boldsymbol{z} + \Delta\boldsymbol{z}$ and tangent to the convex set $\textbf{Ray}\left(\boldsymbol{c} + \Delta\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta\boldsymbol{x}_n\right)$. By the assumption that $\textbf{dist}\left(\boldsymbol{z}, \textbf{bdry}(\textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) > \alpha$, there is some $\delta_0 > 0$ such that, for every $\delta \in (0, \delta_0)$, every point within $\alpha + \delta$ of $\boldsymbol{z}$ lies within $\textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$. Choose $\delta \in (0, \delta_0)$ that also satifises $\delta \leq \|\boldsymbol{z}\| + \max_i \|\boldsymbol{x}_i\|$. Let $\boldsymbol{q}$ be a point at distance $\frac{3\alpha}{4} + \delta$ from $\boldsymbol{z} + \Delta\boldsymbol{z}$ in the direction perpendicular to H. Since $\textbf{dist}\left(\boldsymbol{z}, \boldsymbol{z} + \Delta\boldsymbol{z}\right) \leq \frac{\alpha}{4}$, and $\textbf{dist}\left(\boldsymbol{z} + \Delta\boldsymbol{z}, \boldsymbol{q}\right) \leq \frac{3\alpha}{4} + \delta$,

$$\boldsymbol{q} \in \textbf{Ray}\left(\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$$

At the same time,

$$\textbf{dist}\left(\boldsymbol{q}, \textbf{Ray}\left(\boldsymbol{c} + \Delta\boldsymbol{c}\right) - \textbf{Hull}\left(\boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n + \Delta\boldsymbol{x}_n\right)\right) > \frac{3\alpha}{4}$$

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

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

and thus

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

which is a contradiction. This concludes the proof of the lemma. □

We now derive a consequence of lemma 3.0.10 that uses both the original and the new variables.

Lemma 3.0.11 (Reciprocal of distance to ill-posedness) *Let $\boldsymbol{c}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be vectors in $\mathbb{R}^d$. Let*

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

$$k_1 = \mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right)$$

$$k_2 = \|\boldsymbol{c}\|$$

Then

$$\frac{1}{\rho(A,\boldsymbol{c})} \le \max\left\{\frac{8}{k_1}, \frac{4}{k_2}, \frac{24\max_i\|\boldsymbol{a}_i\|}{k_1k_2}\right\}.$$

Proof: By the definition of k_1 and k_2 and lemma 3.0.10, we can tolerate any change of magnitude up to $k_1/4$ in $\boldsymbol{z}, \{\boldsymbol{x}_i\}$ and any change of up to $\frac{k_1k_2}{2k_1+4(\|\boldsymbol{z}\|+\max\|\boldsymbol{x}_i\|)}$ in $\boldsymbol{c}$ without the program becoming ill-posed. We show that this means we can tolerate any change of up to $k_1/8$ in $\boldsymbol{a}_i$ without the program becoming ill-posed. Formally, we need to show that if $\|\Delta\boldsymbol{a}_i\| \le k_1/8$ for all i, then $\|\Delta\boldsymbol{z}\| \le k_1/4$ and $\|\Delta\boldsymbol{x}_i\| \le k_1/4$. Since $\Delta\boldsymbol{z} = (1/n)\sum\Delta\boldsymbol{a}_i$, $\|\Delta\boldsymbol{z}\| \le k_1/8$. Since $\Delta\boldsymbol{x}_i = \Delta\boldsymbol{a}_i - \Delta\boldsymbol{z}$, $\|\Delta\boldsymbol{x}_i\| \le k_1/8 + k_1/8 = k_1/4$. Thus

$$\rho(A,\boldsymbol{c}) \ge \min\left\{\frac{k_1}{8}, \frac{k_1k_2}{2k_1 + 4(\|\boldsymbol{z}\| + \max\|\boldsymbol{x}_i\|)}\right\}$$

which implies

$$\frac{1}{\rho(A,\boldsymbol{c})} \le \max\left\{\frac{8}{k_1}, \frac{4}{k_2}, \frac{8(\|\boldsymbol{z}\| + \max\|\boldsymbol{x}_i\|)}{k_1k_2}\right\}$$

Since $\boldsymbol{z} = (1/n)\sum\boldsymbol{a}_i \Rightarrow \|\boldsymbol{z}\| \le \max\|\boldsymbol{a}_i\|$, and $\boldsymbol{x}_i = \boldsymbol{a}_i - \boldsymbol{z} \Rightarrow \|\boldsymbol{x}_i\| \le \|\boldsymbol{a}_i\| + \|\boldsymbol{z}\| \le 2\max\|\boldsymbol{a}_i\|$, we have

$$\frac{1}{\rho(A,\boldsymbol{c})} \le \max\left\{\frac{8}{k_1}, \frac{4}{k_2}, \frac{24\max\|\boldsymbol{a}_i\|}{k_1k_2}\right\}$$

This concludes the proof. □

Proof of Lemma 3.0.2 Let

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

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

We will apply the bound of lemma 3.0.11. We first lower bound $\min\{k_1, k_2, k_1k_2\}$. We begin by noting that if

$$\min\{k_1, k_2, k_1k_2\} < \epsilon,$$

then either

$$\mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) < \epsilon, \tag{10}$$

or

$$\|\boldsymbol{c}\| < \epsilon, \tag{11}$$

or there exists some integer l, $1 \le l \le \lceil \log \frac{1}{\epsilon} \rceil$, for which

$$\mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) < 2^l \epsilon \quad \text{and} \quad \|\boldsymbol{c}\| \le 2^{-l+1}. \tag{12}$$

The probabilities of events 10 and 11 will also be bounded in our analysis of event 12. By corollary A.0.15, for $d \ge 2$, we have

$$\mathbf{Pr}\left[\|\boldsymbol{c}\| \le \epsilon\right] \le \frac{e\epsilon}{\sigma},$$

which translates to

$$\mathbf{Pr}\left[\|\boldsymbol{c}\| \le 2^{-l+1}\right] \le \frac{e2^{-l+1}}{\sigma},$$

while lemma 3.0.9 implies

$$\mathbf{Pr}\left[\mathbf{dist}\left(\boldsymbol{z}, \mathbf{bdry}(\mathbf{Ray}\left(\boldsymbol{c}\right) - \mathbf{Hull}\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right))\right) < 2^l \epsilon\right] \le \frac{8 \cdot 2^l \epsilon d^{1/4} n^{1/2}}{\sigma}.$$

Thus, we compute

$$\begin{aligned}
\mathbf{Pr}\left[\min\left\{k_1, k_2, k_1 k_2\right\} < \epsilon\right] &\le \frac{8\ \epsilon d^{1/4} n^{1/2}}{\sigma} + \frac{e\epsilon}{\sigma} + \sum_{l=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \frac{e2^{-l+1}}{\sigma} \frac{8 \cdot 2^l \epsilon d^{1/4} n^{1/2}}{\sigma} \\
&= \frac{8\ \epsilon d^{1/4} n^{1/2}}{\sigma} + \frac{e\epsilon}{\sigma} + \frac{16 e\epsilon d^{1/4} n^{1/2}}{\sigma^2} \log(\frac{1}{\epsilon}) \\
&\le \frac{55\ \epsilon d^{1/4} n^{1/2}}{\sigma^2} \log(\frac{1}{\epsilon}).
\end{aligned}$$

We re-write this as

$$\mathbf{Pr}\left[\max\left\{1/k_1, 1/k_2, 1/k_1 k_2\right\} > \frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2} \log(\frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2})\right] < \frac{\epsilon}{2}.$$

From corollary A.0.14, we know that

$$\mathbf{Pr}\left[\|A, \mathbf{c}\|_F > 3 + \sigma\sqrt{(d+1)n\ 2\ln(2e/\epsilon)}\right] < \frac{\epsilon}{2}.$$

Thus,

$$\mathbf{Pr}\left[\frac{\|A, \mathbf{c}\|_F}{\rho(A, \mathbf{c})} > \frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2} \log(\frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2})(3 + \sigma\sqrt{(d+1)n 2\ln(2e/\epsilon)})\right] \quad \le \quad \epsilon.$$

To derive the lemma as stated, we conclude with

$$\frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2} \log(\frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2})(3 + \sigma\sqrt{(d+1)n\ 2\ln(2e/\epsilon)}) \le$$

$$\frac{1000\ d^{1/4} n^{1/2}}{\epsilon\sigma^2} \log^{1.5}\left(\frac{200\ d^{1/4} n^{1/2}}{\epsilon\sigma^2}\right)$$

□

4 Combining the Primal and Dual Analyses

Our main theorem is now very easy to prove.

Proof of Theorem 1.4.1 Apply lemmas 3.0.2 and 2.0.4:

$$\frac{2^{13}n^2d^{3/2}}{\epsilon\sigma^2}\log^2\left(\frac{2^9n^2d^{3/2}}{\epsilon\sigma^2}\right)+\frac{2^{11}d^{1/4}n^{1/2}}{\epsilon\sigma^2}\log^{1.5}\left(\frac{2^8d^{1/4}n^{1/2}}{\epsilon\sigma^2}\right)$$
$$\leq \frac{2^{14}n^2d^{3/2}}{\epsilon\sigma^2}\log^2\left(\frac{2^{10}n^2d^{3/2}}{\epsilon\sigma^2}\right)$$

□

We recall the four canonical forms for linear programs that we have discussed.

$$\begin{array}{rcll} \max\ \boldsymbol{c}^T\boldsymbol{x}\quad \textit{s.t.}\quad A\boldsymbol{x}\leq\boldsymbol{b} & \text{and its dual} & \min\ \boldsymbol{b}^T\boldsymbol{y}\quad \textit{s.t.}\quad A^T\boldsymbol{y}=\boldsymbol{c},\quad \boldsymbol{y}\geq\boldsymbol{0}, & (1)\\ \max\ \boldsymbol{c}^T\boldsymbol{x}\ \text{ s.t. } A\boldsymbol{x}\leq\boldsymbol{b},\ \boldsymbol{x}\geq\boldsymbol{0} & \text{and its dual} & \min\ \boldsymbol{b}^T\boldsymbol{y}\ \text{ s.t. } A^T\boldsymbol{y}\leq\boldsymbol{c},\ \boldsymbol{y}\geq\boldsymbol{0} & (2)\\ \max\ \boldsymbol{c}^T\boldsymbol{x}\ \text{ s.t. } A\boldsymbol{x}=\boldsymbol{b},\ \boldsymbol{x}\geq\boldsymbol{0} & \text{and its dual} & \min\ \boldsymbol{b}^T\boldsymbol{y}\ \text{ s.t. } A^T\boldsymbol{y}\leq\boldsymbol{c} & (3)\\ \text{find } \boldsymbol{x}\neq\boldsymbol{0} \text{ s.t. } A\boldsymbol{x}\leq\boldsymbol{0} & \text{and its dual} & \text{find } \boldsymbol{y}\neq\boldsymbol{0} \text{ s.t. } A^T\boldsymbol{y}=\boldsymbol{0},\ \boldsymbol{y}\geq\boldsymbol{0}, & (4) \end{array}$$

Renegar developed efficient algorithms for both solving and estimating the condition number of programs in form (2) in [Ren94]. Vera [Ver96] developed efficient algorithms for forms (1) and (3). Cucker and Peña developed algorithms for form (4) in [CP01]. 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. A bound on the smoothed complexity of all of these algorithms follows from theorem 1.4.1.

5 Future Directions

We hope that smoothed analysis of algorithms provides an attractive avenue for other researchers to explore the discrepancy that is sometimes observed between the worst-case complexity and the typical performance of algorithms. We also hope that this work illuminates some of the potential shared interests of the numerical analysis, operations research, and theoretical computer science communities. One potential direction for future research is the application of smoothed analysis to other problem domains.

We do not address in this thesis the question of the actual distribution of condition numbers. We would be particularly interested to hear the results of computational experiments, like those of Freund and Ordoñez[FO02], that explore the distribution of condition numbers occurring in real-world problems.

References

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

[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. Available at `http://theory.lcs.mit.edu/~jdunagan/`.

[BR76] Bhattacharya and Rao. Normal approximation and asymptotic expansion. pages 23–38, 1976.

[CP01] F. Cucker and J. Peña. A primal-dual algorithm for solving polyhedral conic systems with a finite-precision machine. Submitted to SIAM Journal on Optimization, 2001.

[DG99] S. Dasgupta and A. Gupta. An elementary proof of the johnson-lindenstrauss lemma. *International Computer Science Institute, Technical Report 99-006*, 1999.

[FE00a] Robert Freund and Marina Epelman. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Mathematical Programming*, 88(3):451–485, 2000.

[FE00b] Robert Freund and Marina Epelman. Condition number complexity of an elementary algorithm for resolving a conic linear system. Technical Report O.R. Working Paper 319-97, MIT, 2000.

[FE01] Robert Freund and Marina Epelman. A new condition measure, pre-conditioners, and relations between different measures of conditioning for conic linear systems. *SIAM Journal on Optimization*, 2001.

[FM00] Robert Freund and Shinji Mizuno. Interior point methods: Current status and future directions. In H. Frenk et al., editor, *High Performance Optimization*, pages 441–466. Kluwer Academic Publishers, 2000.

[FN01] Robert Freund and Manuel Nunez. Condition-measure bounds on the behavior of the central trajectory of a semi-definite program. *SIAM Journal on Optimization*, 11(3):818–836, 2001.

[FO02] R. Freund and F. Ordoñez. Ipm practical performance on lps and the explanatory value of complexity measures. *Proceedings of the 7th SIAM Conference on Optimization (SIOPT)*, 2002.

[FV99] Robert Freund and Jorge Vera. On the complexity of computing estimates of condition measures of a conic linear system. Operations Research Center Working Paper, MIT, 1999, submitted to *Mathematics of Operations Research*, 1999.

[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.

[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.

[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.

[ST] Daniel A. Spielmana and Shang-Hua Teng. Models and applications of smoothed analysis. In submission.

[ST01] Daniel Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (STOC '01)*, pages 296–305, 2001. Available at `http://math.mit.edu/~spielman/SmoothedAnalysis/`.

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

A Gaussian random variables

We recall that the probability density function of a Gaussian random variable is given by

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

A Gaussian random vector of variance σ^2 is a vector where each element is a Gaussian random variable of variance σ^2. A Gaussian random matrix is defined similarly. The probability density function of a d-dimensional Gaussian random vector of variance σ^2 centered at $\bar{\boldsymbol{x}}$ is given by

$$\mu(\boldsymbol{x}) = (1/(\sigma\sqrt{2\pi})^d)e^{-\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|/(2\sigma^2)}.$$

We now derive particular versions of well-known bounds on the Chi-Squared distribution. These bounds are used in the body of the paper, and bounds of this form are well-known. We thank DasGupta and Gupta [DG99] for this particular derivation.

Fact A.0.12 (Sum of gaussians) *Let $X_1, \ldots, X_d$ be independent $N(0, \sigma)$ random variables. Then*

$$\Pr[\sum_{i=1}^{d} X_i^2 \geq \kappa^2] \leq e^{\frac{d}{2}(1-\frac{\kappa^2}{d\sigma^2}+\ln\frac{\kappa^2}{d\sigma^2})}$$

Proof: For simplicity, we begin with $Y_i \sim N(0, 1)$. A simple integration shows that if $Y \sim N(0, 1)$ then $E[e^{tY^2}] = \frac{1}{\sqrt{1-2t}} \quad (t < \frac{1}{2})$. We proceed with

$$\begin{aligned}
\Pr[\sum_{i=1}^{d} Y_i^2 \geq k] &= \\
\Pr[\sum_{i=1}^{d} Y_i^2 - k \geq 0] &= \qquad \text{(for } t > 0\text{)} \\
\Pr[e^{t(\sum_{i=1}^{d} Y_i^2 - k)} \geq 1] &\leq \qquad \text{(by Markov's Ineq.)} \\
\mathbf{E}[e^{t(\sum_{i=1}^{d} Y_i^2 - k)}] &= \\
\left(\frac{1}{1-2t}\right)^{d/2} e^{-kt} &\leq \qquad \text{(letting } t = \frac{1}{2} - \frac{d}{2k}\text{)} \\
\left(\frac{k}{d}\right)^{d/2} e^{-\frac{k}{2}+\frac{d}{2}} &= \quad e^{\frac{d}{2}(1-\frac{k}{d}+\ln\frac{k}{d})}
\end{aligned}$$

Since

$$\Pr[\sum_{i=1}^{d} Y_i^2 \geq k] = \Pr[\sum_{i=1}^{d} X_i^2 \geq \sigma^2 k]$$

we set $k = \frac{\kappa^2}{\sigma^2}$ and obtain $e^{\frac{d}{2}(1-\frac{k}{d}+\ln\frac{k}{d})} = e^{\frac{d}{2}(1-\frac{\kappa^2}{d\sigma^2}+\ln\frac{\kappa^2}{d\sigma^2})}$ which was our desired bound. □

Fact A.0.13 (Alternative sum of gaussians) *Let $X_1, \ldots, X_d$ be independent $N(0, \sigma)$ random variables. Then*

$$\Pr[\sum_{i=1}^{d} X_i^2 \geq cd\sigma^2] \leq e^{\frac{d}{2}(1-c+\ln c)} \qquad c \geq 1$$

$$\Pr[\sum_{i=1}^{d} X_i^2 \leq cd\sigma^2] \leq e^{\frac{d}{2}(1-c+\ln c)} \qquad c \leq 1$$

Proof: The first inequality is proved by setting $k = cd$ in the last line of the proof of fact A.0.12. To prove the second inequality, begin the proof of fact A.0.12 with $\Pr[\sum_{i=1}^{d} Y_i^2 \leq k]$ and continue in the obvious manner. □

Corollary A.0.14 *Let $\boldsymbol{x}$ be a d-dimensional Gaussian random vector of variance σ^2 centered at the origin. Then, for $d \geq 2$ and $\epsilon \leq 1/e^2$,*

$$\mathbf{Pr}\left[\|\boldsymbol{x}\| \geq \sigma\sqrt{d(1+2\ln(1/\epsilon)}\right] \leq \epsilon$$

Proof: Set $c = 1 + 2\ln(1/\epsilon)$ in fact A.0.13. We then compute

$$e^{\frac{d}{2}(1-c+\ln c)} \leq e^{1-c+\ln c} \leq e^{-2\ln\frac{1}{\epsilon}+\ln(1+2\ln\frac{1}{\epsilon})} = \epsilon e^{-\ln\frac{1}{\epsilon}+\ln(1+2\ln\frac{1}{\epsilon})}$$

We now seek to show

$$\begin{aligned} e^{-\ln\frac{1}{\epsilon}+\ln(1+2\ln\frac{1}{\epsilon})} &\leq 1 \\ \Leftrightarrow \quad -\ln\frac{1}{\epsilon} + \ln(1+2\ln\frac{1}{\epsilon}) &\leq 0 \\ \Leftrightarrow \quad 1+2\ln\frac{1}{\epsilon} &\leq \frac{1}{\epsilon} \end{aligned}$$

For $\epsilon = 1/e^2$, the left-hand side of the last inequality is 5, while the right-hand side is greater than 7. Taking derivatives with respect to $1/\epsilon$, we see that the right-hand side grows faster as we increase $1/\epsilon$ (decrease ϵ), and therefore will always be greater. □

Corollary A.0.15 *Let $\boldsymbol{x}$ be a d-dimensional Gaussian random vector of variance σ^2 centered at the origin. Then, for $d \geq 2$,*

$$\mathbf{Pr}\left[\|\boldsymbol{x}\| \leq \epsilon\right] \leq \frac{e\epsilon}{\sigma}$$

Proof: If $\epsilon \leq \sigma$, set $c = \frac{\epsilon^2}{d\sigma^2}$ in fact A.0.13.

$$e^{\frac{d}{2}(1-c+\ln c)} \leq e^{1-c+\ln c} \leq e^{1+\ln c} = \frac{e\epsilon^2}{d\sigma^2} \leq \frac{e\epsilon}{\sigma}$$

If $\epsilon > \sigma$, the statement is vacuously true. □