Smoothed Analysis for the Simplex Method: Nearly Tight Noise Dependence

Eleon Bach

Sophie Huiberts

December 15, 2024

Abstract

Smoothed analysis is a method for analysing the performance of algorithms, used especially for those algorithms whose running time in practice is significantly better than what can be proven through worst-case analysis. Given an arbitrary linear program with d variables and n inequality constraints, Spielman and Teng (STOC '01) proved that the simplex method runs in time $O(\sigma^{-30}d^{55}n^{86})$, where $\sigma > 0$ is the standard deviation of Gaussian distributed noise added to the original LP data. Their result was simplified and strengthened over a series of work, with the current strongest upper bound being $O(\sigma^{-3/2}d^{13/4}\log(n)^{7/4})$ pivot steps due to Huiberts, Lee and Zhang (STOC '23).

We prove that there is a simplex method with smoothed complexity upper bounded by $O(\sigma^{-1/2}d^{11/4}\log(n/\sigma)^{7/4})$ pivot steps. For the same method we prove a lower bound on its smoothed complexity of $0.03\sigma^{-1/2}d^{-1/4}\ln(n)^{-1/4}$ pivot steps for $n = (4/\sigma)^d$ inequality constraints. This nearly closes the gap between the upper and lower bounds in regards to their dependence on the noise parameter σ. We obtain our results by studying a variant of the *shadow size*, a quantity which has been key to all previous smoothed analyses of the simplex method to date.

1 Introduction

Ever since its first use in early 1948, the simplex method has been one of the primary algorithms for solving linear programming (LP) problems. For the purpose of this paper, an LP is a problem described by input data $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n, c \in \mathbb{R}^d$ and is written as

$$\begin{aligned} \text{maximize} \quad & c^\top x \\ \text{subject to} \quad & Ax \le b. \end{aligned}$$

The computational task at hand is to find if there exists any $x \in \mathbb{R}^d$ such that the system of coordinatewise inequalities $Ax \le b$ holds. If such a *feasible solution* x exists, then one must report either a feasible solution x for which additionally the inner product $c^\top x$ is maximal among all feasible solutions, or a certificate that the set of feasible solutions is *unbounded*.

The simplex method is best thought of as a class of algorithms, differing in specific details such as the choice of the *pivot rule* or the *phase 1* procedure. Navigating from one vertex of the *feasible set* to another, the pivot rule is the part of a simplex method that decides in which direction the pivot step will move. Notable examples of pivot rules include the most negative reduced cost rule [Dan51], the steepest edge rule and its approximations [Har73, Gol76, FG92], and the shadow vertex rule [GS55, Bor77].

Although there have been substantial improvements over the first simplex method as introduced by Dantzig, one thing has not changed: the total number of pivot steps required to solve an LP in practice scales roughly linear in the dimensions of the problem [Dan63, Sha87, And04, xpr]. Despite many decades of practical experience supporting this rule of thumb, it remains a major challenge to the theory of algorithms to explain this phenomenon. This is further complicated by the results from worst-case analysis. For almost every major pivot rule, there are theoretical constructions known that make the simplex method take exponentially many pivot steps before reaching an optimal solution. The majority of these constructions are based on *deformed products* [KM72, Jer73, AC78, GS79, Mur80, Gol83, AZ98] or on Markov decision processes [FHZ11, Fri11, DFH22, DM23]. The fastest provable simplex method is randomized and requires $2^{O(\sqrt{d\log(1+n/d)})}$ pivot steps [Kal92, MSW96, HZ15].

During the 70's and 80's, there were a number of investigations into the average-case complexity of the simplex method. A wide variety of models was studied, including drawing the rows of A from a spherically-symmetric distribution [Bor77, Bor82, Bor87, Bor99, BDG$^+$22], drawing the combined vector (c, b) from a spherically symmetric distribution [Sma83], having fixed A, b and every inequality constraint independently being either $a_i^\top x \leq b_i$ or $a_i^\top \geq b_i$ [Hai83], and a range of others [Meg86, AKS87, Tod86, AM85]. For an in-depth survey we refer the reader to [Bor87]. A major weakness of average-case analysis is that real-life LPs are structured in recognizable ways, and that average-case LPs demonstrate no such structure. As such, it is reasonable to question to what extent average-case analyses succeed at explaining the simplex method's performance.

Smoothed analysis is a way of going beyond worst-case analysis [Rou20], drawing on some of the advantages of average-case analysis while still preserving an amount of structure in the solved instances. One assumes that a base LP problem is adversarially constructed

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & \bar{A}x \leq \bar{b}, \end{array}$$

with the assumption that $\bar{A} \in \mathbb{R}^{n\times d}$ and $\bar{b} \in \mathbb{R}^n$ are such that the rows of the combined matrix $(\bar{A}, \bar{b})$ each have Euclidean norm at most 1. Subsequently, this input data gets randomly perturbed. For a parameter $\sigma > 0$, one samples $\hat{A} \in \mathbb{R}^{n\times d}$ and $\hat{b} \in \mathbb{R}^n$ with independent entries, each entry being drawn from a Gaussian distribution with mean 0 and variance σ^2. The *smoothed complexity* of an algorithm is the expected running time to solve the perturbed problem

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & (\bar{A} + \hat{A})x \leq \bar{b} + \hat{b}, \end{array}$$

where the running time is to be bounded as a polynomial function in n, d and σ^{-1}. This definition was introduced by [ST04] in order to demonstrate that inputs on which the simplex method performs badly are rare, and as a way to model the influence of inherently noisy input data. Their result is made up of two parts. First an analysis of the *shadow size* $D(n, d, \sigma)$, which consists of taking a (fixed) two-dimensional linear subspace W and upper bounding the expected number of vertices of the orthogonal projection $\pi_W(\{x : (\bar{A}+\hat{A})x \le 1\})$ onto W. Formally, what is bounded is the quantity

$$D(n,d,\sigma) \;=\; \max_{\bar{A},c,c'} \mathbb{E}_{\hat{A}}\left[\text{ vertices}\left(\pi_{\mathrm{span}(c,c')}(\{x : (\bar{A}+\hat{A})x \le 1\})\right)\right],$$

where $\bar{A}$ is assumed to have rows each of Euclidean norm at most 1. The second part is an algorithmic reduction, showing that there exists a simplex method whose running time can be bounded as a function of $D(n,d,\sigma)$. Their algorithm is based on the *shadow vertex pivot rule*. This pivot rule works based on having two objectives $c, c' \in \mathbb{R}^d$ and visiting all basic solutions that maximize some positive linear combination of the two. Starting from an optimal basic feasible solution to the first objective, it pivots until it finds an optimal basic feasible solution to the second objective (or finds an infinite ray certifying unboundednes). When the two objectives are chosen independently of the noise $\hat{A}$, then the number of pivot steps required by the shadow vertex rule is naturally upper bounded by the shadow size $D(n,d,\sigma)$. They proved a bound on the shadow size of

$$D(n,d,\sigma) \le \frac{10^8 n d^3}{\min(\sigma, 1/3\sqrt{d\ln n})^6},$$

and found a simplex method that requires an estimated

$$O(nd\ln(nd/\min(1,\sigma))D(n,d,\frac{\min(1,\sigma^5)}{d^{8.5}n^{14}\ln^{2.5} n}))$$

pivot steps under the smoothed analysis framework. This combines to a total of $O^*(n^{86}d^{55}\sigma^{-30})$ pivot steps, ignoring logarithmic factors and assuming that $\sigma \le 1/3\sqrt{d\ln n}$. Note that this last assumption on σ may be made without loss of generality, for we can scale down the constraints of the LP to make the assumption hold true. The result of this scaling can be captured in an additive term in the upper bound that is independent of σ.

This work was built upon by [DS05], who improved the shadow bound to

$$D(n,d,\sigma) \le \frac{10^4 n^2 d\ln n}{\sigma^2} + 10^5 n^2 d^2 \ln^2 n.$$

Vershynin later proved in [Ver09] a shadow bound of $D(n,d,\sigma) \le d^3\sigma^{-4} + d^5\ln^2 n$, and found an alternative algorithm running in time $O(D(n,d,\min(\sigma, 1/\sqrt{d\ln n}, 1/d^{3/2}\ln d)))$. This led to a situation where one shadow bound had better dependence on σ and the other had much better dependence on n. This was resolved with the work of [DH20], who proved that $D(n,d,\sigma) \le d^2\sigma^{-2}\sqrt{\ln n} + d^3\ln^{1.5} n$. They also observed that the comparatively simple *dimension-by-dimension* phase 1 algorithm of [Bor87] could be used with an expected number of pivot steps of at most $(d+1)D(n,d+1,\sigma)$.

The shadow bound with the current best dependence on σ comes from [HLZ23] and states that

$$D(n,d,\sigma) \le O\left(\frac{d^{13/4}\ln^{7/4} n}{\sigma^{3/2}} + d^{19/4}\ln^{5/2} n\right).$$

[HLZ23] also proved the first non-trivial lower bound, stating that

$$D(4d-13,d,\sigma) \ge \Omega\left(\min\left(2^d, \frac{1}{\sqrt{d\sigma\sqrt{\log d}}}\right)\right)$$

using a particular construction of LP data $\bar{A}, \bar{b}$ and a subspace W, assuming that $d \geq 5$. By a computational experiment they conjecture that their construction might have smoothed shadow sizes as large as $\sigma^{-3/4}/\operatorname{poly}(d)$.

Our results We improve the algorithmic reduction, obtaining an algorithm whose running time is $O(R(n, d, \min\{\sigma, 1/\sqrt{d \ln n}, 1/d^{3/2} \log d\}))$, where $R(n, d, \sigma)$ is what we call the *semi-random shadow size*

$$R(n,d,\sigma) \;=\; \max_{\bar{A},\bar{b},c} \mathbb{E}_{\hat{A},\hat{b},Z}\Big[\; \text{vertices}\Big(\pi_{\text{span}(c,Z)}(\{x : (\bar{A}+\hat{A})x \leq \bar{b}+\hat{b}\})\Big)\;\Big].$$

Here, $\bar{A}, \bar{b}$ are again chosen such that the rows of $(\bar{A}, \bar{b})$ each have norm at most 1, $c \in \mathbb{S}^{d-1}$ is a unit vector, $\hat{A}, \hat{b}$ have independent entries that are Gaussian distributed with mean 0 and standard deviation σ, and $Z \in \mathbb{R}^d$ is independently sampled from any spherically symmetric distribution. Note the difference compared to the definition of $D(n, d, \sigma)$, which allows both objectives to be adversarially chosen. Having our algorithm sample Z at random is the key algorithmic improvement which allows us to prove stronger bounds than was possible using $D(n, d, \sigma)$. Specifically we find

$$R(n,d,\sigma) \leq O(\sqrt{\sigma^{-1}\sqrt{d^{11}\log(n)^7}})$$

Notably, this upper bound is lower than the conjectured lower bound of [HLZ23] for the fixed-plane shadow size. In terms of the exponent on σ this is the best that a shadow size bound can be, as we demonstrate in Section 5 with a nearly-matching lower bound

$$R\left((4/\sigma)^d, d, \sigma\right) \geq \frac{1}{32\sqrt{\sigma\sqrt{d\ln n}}}.$$

More detail on the techniques used to prove the upper bound can be found in Section 1.2.

1.1 Related Work

The semi-random shadow vertex method used in this paper is adapted from [DH16]. They give a simplex method which uses an expected number of $O(\frac{d^2}{\delta}\ln(d/\delta))$ pivot steps, where δ is a parameter of the constraint matrix which measures the curvature of the feasible region. Their work improves over a similar weaker result that uses "less random" shadows [BR13].

The shadow size for polyhedra all whose vertices are integral was studied in [BDLKS21, Bla23, BC24], the last of which studies uniformly random planes. Semi-random shadow planes were also used in a different context to obtain a weakly polynomial-time "simplex-like" algorithm for LP in [KS06].

Shadow bounds for random objectives were previously used by [NSS22] in order to derive results similar to diameter bounds for smoothed polyhedra. Specifically they proved that with high probability there exists a large subset of vertices (according to some specific measure) which has small diameter. Here the random objectives were sampled from some non-uniform distribution, and the sizes of the resulting shadows were bounded using the shadow bound of [DH20].

In this paper we make use of a notion of vertices being "well-separated" from each other. That assumption was pioneered in a line of work starting with [KM11]. Different from both the current work and that of [HLZ23], in [KM11] the data is deterministic, the pivot rule is that of the most negative reduced cost, and progress is measured with respect to the objective value.

1.2 Proof Overview

The primitive operation that our algorithmic approach is built on is that of following a semi-random shadow path. That is, to run the shadow vertex method with one fixed objective vector $c \in \mathbb{R}^d$ and one random objective vector sampled from a spherically symmetric distribution. We adapt the two-phase approach laid out in [Ver09] to work with this primitive. Full detail on the algorithm, including how to deal with the cases of infeasible or unbounded LPs, can be found in Section 3.

1.2.1 Auxiliary LPs

For the smoothed objective data A, b, we start out by sampling $Z \in \mathbb{R}^d$ with independent entries from the standard Gaussian distribution, with mean 0 and standard deviation 1, and solving $\max Z^\top x$ s.t. $Ax \leq 1$. This is done by adding d artificial constraints to create a starting vertex and traversing the semi-random shadow path from that starting objective to the random objective. A number of repeated trials will be required in order to have the artificial constraints not cut off the optimal solution to the LP, see Lemma 22. At the end, this results in an optimal basic feasible solution $A_I^{-1} 1_I$ to this first auxiliary LP.

Secondly, the algorithm will operate on the feasible region of a second auxiliary LP given by constraints $Ax + (1 - b)t \leq 1$. We sample another independent standard Gaussian distributed entry Z_{d+1}. The optimal basis of the first LP will be the set of constraints I tight for an edge of this second LP, and this edge connects two vertices on the combined shadow path from $-e_{d+1}$ to (Z, Z_{d+1}) and onwards to e_{d+1}. The algorithm will follow this path until it finds, on one of its traversed edges, a feasible solution with $x = A_{I'}^{-1} b_{I'}, I' \in \binom{[n]}{d}$ and $t = 1$. This gives us an optimal basic feasible solution $A_{I'}^{-1} b_{I'}$ to the linear program $\max Z^\top x$ s.t. $Ax \leq b$ with random objective. To finish the algorithm we follow the shadow path from Z to c to find the optimal solution to the intended LP.

Hence, it suffices to study the shadow size for the specific case where one objective is random and the other is fixed, both independent from the noise on the constraint data. We require this for constraint data where either the right-hand side is identical to 1 or where the right-hand side is smoothed.

1.2.2 Upper bound

To illustrate the guiding principle behind the upper bound, we will review the upper bound on the smoothed complexity of a two-dimensional polygon of [HLZ23]. We write $\pi_{c,Z} : \mathbb{R}^d \to \mathrm{span}(c, Z)$ for the orthogonal projection onto $\mathrm{span}(c, Z)$. Let $P = \{x \in \mathbb{R}^d : Ax \leq b\}$ be the polytope that is induced by the LP $\max c^\top x$ subject to $Ax \leq b$. Suppose that $p_1, \ldots, p_k \in \mathbb{R}^2$ are points with $\|p_i\| \leq 1$ for every $i \in [k]$, that we have a polygon given as their convex hull $Q = \mathrm{conv}(p_1, \ldots, p_k)$, and suppose that there exists an $\varepsilon > 0$ such that the following holds:

For every $i \in [k]$ for which p_i is a vertex of Q, there exists a functional $y_i \in \mathbb{R}^2$ with $\|y_i\| = 1$ such that $y_i^\top p_i > y_i^\top p_j + \varepsilon$ for every $j \in [n], j \neq i$ for which p_j is a vertex of Q.

Assuming the above statement is true, consider 2 consecutive vertices of Q. Without loss of generality let us call them p_1 and p_2. For a number $\rho > 0$ to be decided later, let us make a case distinction based on whether $\|p_1 - p_2\| > \rho$ or $\|p_1 - p_2\| \leq \rho$.

In the former case, the edge $[p_1, p_2]$ "takes up a lot of perimeter", in the sense that Q has perimeter at most 2π and hence Q can have at most $2\pi/\rho$ edges of length ρ or longer. This is good for us.

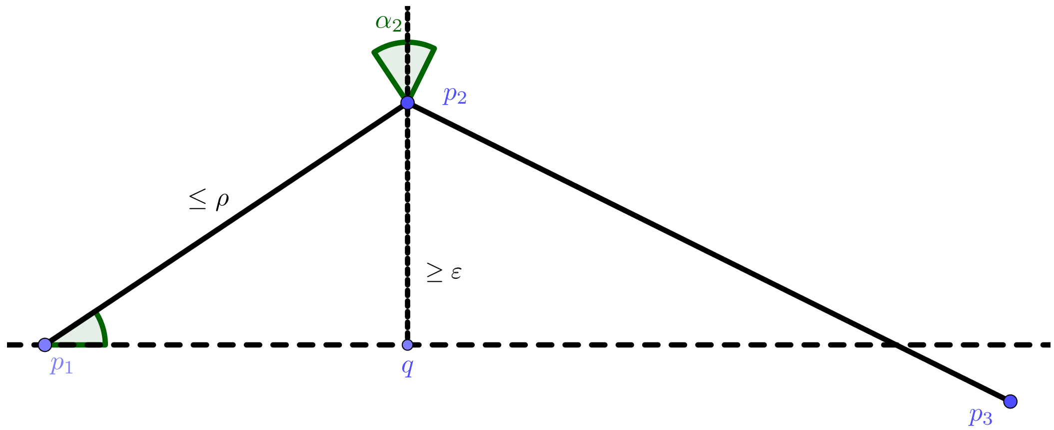

Figure 1: Lower bounding the exterior angle at p_2

In the latter case, consider the triangle with vertices p_1, p_2 and $q = p_2 - y_2^\top(p_2 - p_1) \cdot y_2$ as depicted in Figure 1. Because the next vertex on the boundary satisfies $y_2^\top p_3 < y_2^\top p_2$, the exterior angle α_2 at p_2 is at least as large as $\angle(p_2, p_1, q)$. The right-angled triangle has a hypotenuse of length $\|p_1 - p_2\| \leq \rho$ and an opposite side of length $\|p_2 - q\| \geq \varepsilon$, from which we derive a lower bound of

$$\alpha_2 \geq \angle(p_2, p_1, q) \geq \sin(\angle(p_2, p_1, q)) \geq \varepsilon/\rho.$$

Since the sum of the exterior angles of all the vertices of Q is equal to 2π, that means that there can be at most $2\pi\rho/\varepsilon$ vertices with exterior angle that large.

This argument shows that every vertex must either have a long edge or a large exterior angle, hence Q can only have at most $\min_{\rho>0} 2\pi\rho/\varepsilon + 2\pi/\rho$ vertices. Choosing $\rho = \sqrt{2\pi/\varepsilon}$ yields an upper bound of $2\sqrt{2\pi/\varepsilon}$ vertices.

The argument in [HLZ23] is essentially what is described above, but adapted to the case where the assumption of $y_i^\top p_i \geq y_i^\top p_j + \varepsilon$ only holds in expectation. For our upper bound on the semi-random shadow size in Section 4 we will use a similar concept of vertices being "well-separated" from each other. In contrast to [HLZ23] however, our notion of being "well-separated" allows us an analysis which remains completely in primal space. Hence, we will count the number of vertices of the shadow path as images of vertices of the polytope P under $\pi_{c,Z}$ and not as given by the intersection of the polar dual of P with the shadow plane. One important improvement that we make is that we homogenize the separation argument with respect to both the norm of the vertex and the norm of the functional. In the next paragraphs we will examine the building blocks of this argument.

Random objective Since the size of the shadow path from Z to c depends only on the direction $Z/\|Z\|$ and not on the norm $\|Z\|$ (see Fact 19), for the purposes of the shadow path upper bound we assume that $Z \in \mathbb{R}^d$ is exponentially distributed, i.e., such that for any measurable $S \subseteq \mathbb{R}^d$ it holds that $\Pr[Z \in S] = \frac{1}{d! \operatorname{vol}_d(\mathbb{B}_2^d)} \int_S e^{-\|z\|} \,\mathrm{d}\, z$.

For our analysis we need a certain amount of randomness in the objectives with which we traverse the shadow path. If we have a random objective $c + 1/\xi Z$ on the shadow path that is close enough to our fixed objective c, we can show that there is only a constant number of pivot steps from $c + 1/\xi Z$ to c. We parametrize this closeness by some $\xi > 0$. We observe that the path from Z to $c + 1/\xi Z$ has the same length as the path from Z to $\xi c + Z$. We construct $k = \lceil \log(\xi) \rceil$ intermediate objectives $Z + 2^i c$ for $i = -1, \dots, k$ and traverse step by step the shadow path from $Z + 2^{i-1} c$ to $Z + 2^i c$ in order to get better control over the shadow path length. In Lemma 25 we will show by an argument similar to the angle bound of [ST04], that k is of order $O(d \log(n/\sigma)$.

Separation When traversing the shadow path from Z to $2^k c + Z$ we will see in Section 4.2 that for 99% of the traversed bases I there exists an intermediate objective $y_I \in [Z, 2^k c + Z]$ such that $y_I^\top A_I^{-1} \geq 0.005/d$. This fact is independent of the noise on the constraint data and is proven using only the randomness in Z.

At the same time, using the randomness in the perturbations, for at least 80% of bases I on the path the feasible solution $x_I = A_I^{-1} b_I$ has slack at least $b_j - a_j^\top x_I \geq \frac{\|x_I\|}{5000 d^{3/2} \ln(n)^{3/2}}$ for every nonbasic constraint $j \in [n] \setminus I$. This is established in Section 4.3.

It will turn out that if the majority of traversed bases satisfy both these criteria then we have a kind of "separation", like in the proof of [HLZ23], although slightly different.

Bound We make a distinction of four cases based on numbers $R > r > 0$ and $\rho > 0$.

- The total number of bases $I \in \binom{[n]}{d}$ with $\|\pi_{c,Z}(x_I)\| > R$ is bounded per Lemma 46.
- The total number of bases $I \in \binom{[n]}{d}$ with $\|\pi_{c,Z}(x_I)\| < r$ is bounded per Lemma 48.
- The number of bases I on the path from Z to c satisfying $\|\pi_{c,Z}(x_I)\| \in [r, R]$ and which have at least one neighbour J on this path at distance $\|\pi_{c,Z}(x_I) - \pi_{c,Z}(x_J)\| \geq \rho \|\pi_{c,Z}(x_I)\|$ is at most $O(\rho^{-1} \log(R/r))$ as shown in Lemma 42.
- The bases I on the path from Z to $2^k c + Z$ with only close-by neighbours $\|\pi_{c,Z}(x_I) - \pi_{c,Z}(x_J)\| < \rho \|\pi_{c,Z}(x_I)\|$ are counted in Lemma 39 and Lemma 44 and there are at most $O(\rho d^{5/3} k \log(n)^{3/2})$ in expectation.

The last of these points hides an amount of complexity where we homogenize the norm of the functional $y_I \in [Z, 2^k c + Z]$. If I has its functional $y_I \in [2^{i-1} c + Z, 2^i c + Z]$ of large norm, $i \in [k]$, then the largest exterior angle that can be argued is only of order

$$\theta_I \geq \frac{\sigma}{\mathrm{poly}(d, \log n) \cdot \rho \cdot \|y\|} \approx \frac{\sigma}{\mathrm{poly}\,(d, \log n) \cdot \rho \cdot \|2^i c + Z\|}.$$

In order to nevertheless upper bound the total number of such vertices, we can show that the total sum of exterior angle captured by bases I on the path from $2^{i-1} c + Z$ to $2^i c + Z$ shrinks as $\frac{\|Z\|}{\|2^i c + Z\|}$, meaning the upper bound on the total number of pivot steps on every subpath from $2^{i-1} c + Z$ to $2^i c + Z$ is constant. We refer to Section 4 for full details.

1.2.3 Lower bound

In Section 5 we construct unperturbed LP data $\bar{A}, \bar{b}, c$ for which, when the LP data is perturbed and two random objective vectors are chosen, the shadow path will have at least $\frac{1}{32\sqrt{\sigma\sqrt{d \ln n}}}$ steps in expectation, assuming that the number n of constraints is permitted to be exponentially large $n \leq (4/\sigma)^d$. Although this result should not be thought of as indicative of real-world performance due to the high number of constraints, it does demonstrate that the dependence on σ in Theorem 45 cannot be further decreased without increasing its dependence on n from $\log(n)^{O(1)}$ to $n^{O(1/d)}$.

The construction involves having the rows of $\bar{A}$ be a set of unit vectors that are "well-spread-out" on the sphere. In technical terms we require this set to be a σ-net: for every $\theta \in \mathbb{S}^{d-1}$ there must be an index $i \in [n]$ such that the row i is close to θ, i.e., $\|\theta - \bar{a}_i\| \leq \sigma$. Taking $\bar{b} = 1$, the resulting feasible region after perturbation will be close to the unit ball in the sense that

$$(1 - 8\sigma\sqrt{d \ln n})\mathbb{B}_2^d \subseteq \{x : Ax \leq b\} \subseteq (1 + 16\sigma\sqrt{d \ln n})\mathbb{B}_2^d$$

with probability at least $1 - n^{-d}$. Following [HLZ23] we can use the Pythagorean theorem to quickly see that every edge of the feasible region has length at most $32\sqrt{\sigma\sqrt{d \ln n}}$. The vertices $x_I, x_{I'}$ of the feasible region respectively maximizing the objectives $c^\top x$ and $Z^\top x$ will be spaced far apart in Euclidean distance most of the time, i.e., $\|x_I - x_{I'}\| \geq 1$, which implies that any set of edges connecting the two must be of size at least $\frac{1}{32\sqrt{\sigma\sqrt{d \ln n}}}$.

2 Preliminaries

We write $[d] := \{1, \dots, d\}$. Whenever the given dimension is clear from the context, we write 1 for the all-ones vector and I for the identity matrix. The standard basis vectors are denoted by $e_1, \dots, e_d \in \mathbb{R}^d$. Let $W \subseteq \mathbb{R}^d$ be a linear subspace. Then we denote the orthogonal projection onto W by π_W.

The ℓ_2-norm is $\|x\|_2 = \sqrt{\sum_{i \in [d]} x_i^2}$ and the ℓ_∞-norm is $\|x\|_\infty = \max_{i \in [d]} |x_i|$ for a vector $x \in \mathbb{R}^d$. A norm without a subscript is always the ℓ_2-norm. Given $p \geq 1, d \in \mathbb{Z}_+$, define $\mathbb{B}_p^d = \{x \in \mathbb{R}^d : \|x\|_p \leq 1\}$ as the d-dimensional unit ball of ℓ_p norm. Further, let for $p = 2$, $\mathbb{S}^{d-1}$ denote the unit sphere in $\mathbb{R}^d$, i.e., $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \colon \|x\| = 1\}$.

For sets $A, B \subseteq \mathbb{R}^d$, the distance between the two is $\operatorname{dist}(A, B) = \inf_{a \in A, b \in B} \|a - b\|$. For a point $x \in \mathbb{R}^d$ we write $\operatorname{dist}(x, A) = \operatorname{dist}(A, x) = \operatorname{dist}(A, \{x\})$. The affine hull of d vectors $a_1, \dots, a_d$ is denoted as $\operatorname{affhull}(a_i : i \in [d])$ and their convex hull as $\operatorname{conv}(a_1, \dots, a_d) = \operatorname{conv}(a_i : i \in [d])$.

For a convex body $K \in \mathbb{R}^d$, we define $\partial K \subseteq \operatorname{span}(K)$ as the boundary of K in the linear subspace spanned by the vectors in K.

2.1 Polytopes, Cones and Fans

Definition 1 (Polyhedron)**.** *Let $A \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^n$ where $n \in \mathbb{N}$. We call a convex set $Q \subset \mathbb{R}^d$ a polyhedron if it can be written as $Q = \{x \in \mathbb{R}^d : Ax \leq b\}$.*

Definition 2. *Let $I \subseteq \binom{[n]}{d}$ index a basis, let $A_I \subseteq \mathbb{R}^{d \times d}$ and $b_I \in \mathbb{R}^d$ be the corresponding submatrix of A respectively the corresponding subset of b indexed by I and call $x_I = A_I^{-1} b_I$ the corresponding basic solution. We say that x_I and I are feasible for the LP $\max c^\top x$ subject to $Ax \leq b$ if it satisfies $Ax_I \leq b$. We denote the set of feasible bases of the system $Ax \leq b$ by $F(A, b)$.*

Definition 3. *Let $\{a_1, \dots, a_n \colon a_i \in \mathbb{R}^d\}$ be a set of vectors in $\mathbb{R}^d$. The cone $\operatorname{cone}(a_1, \dots, a_n)$ generated by $a_1, \dots, a_n$ is defined as $\operatorname{cone}(a_1, \dots, a_n) \coloneqq \{x \in \mathbb{R}^d \colon x = \sum_{i=1}^n \lambda_i a_i\}$ for $\lambda_i \in \mathbb{R}_{\geq 0}$.*

2.2 Probability Distributions

All probability distributions considered in this paper will admit a probability density function with respect to the Lebesgue measure.

First we look at a useful properties that density functions may have and which we use throughout the paper.

Definition 4 (L-log-Lipschitz random variable)**.** *Given $L > 0$, we say a random variable $x \in \mathbb{R}^d$ with probability density μ is L-log-Lipschitz (or μ is L-log-Lipschitz), if for all $x, y \in \mathbb{R}^d$, we have*

$$|\log(\mu(x)) - \log(\mu(y))| \leq L\|x - y\|,$$

or equivalently, $\mu(x)/\mu(y) \leq \exp(L\|x - y\|)$.

In the following we see an equality for the expected value of any convex function applied to any random variable.

Lemma 5 (Jensen's inequality)**.** *Let X be a random variable and f a convex function. Then we have $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.*

Definition 6. *Let $S \subseteq \mathbb{R}^d$. A random variable $X \in \mathbb{R}^d$ is exponentially distributed on $\mathbb{R}^d$ if*

$$\Pr[X \in S] = \int_S Ce^{-\|x\|} \, \mathrm{d}\, x$$

.

Lemma 7. *The normalizing constant C of the exponential distribution is $C = \frac{1}{d! \operatorname{vol}_d(\mathbb{B}_2^d)}$. For X exponentially distributed on $\mathbb{R}^d$, the k'th moment of $\|X\|$ is $\mathbb{E}[\|X\|^k] = \frac{(k+d-1)!}{(d-1)!}$.*

Proof. See Appendix B. □

For the exponential distribution we have the following tail bound.

Lemma 8. *Let X be exponentially distributed on $\mathbb{R}^d$. Then for any $t > 1$ we have*

$$\Pr[X \geq 2ed \ln t] \leq t^{-d}.$$

Proof. See Appendix B. □

The exponential distribution is 1-Lipschitz continuous.

Definition 9 (Gaussian distribution)**.** *The d-dimensional Gaussian distribution $\mathcal{N}_d(\bar{a}, \sigma^2 I)$ with support on $\mathbb{R}^d$, mean $\bar{a} \in \mathbb{R}^d$, and standard deviation σ, is defined by the probability density function*

$$\sigma^{-d} \cdot (2\pi)^{-d/2} \cdot \exp\left(-\|s - \bar{a}\|^2/(2\sigma^2)\right)$$

at every $s \in \mathbb{R}^d$.

A useful standard property of the Gaussian distribution is the following tail bound:

Lemma 10 (Gaussian tail bound)**.** *Let $x \in \mathbb{R}^d$ be a random vector sampled with independent Gaussian distributed entries of mean 0 and variance σ^2. For any $t \geq 1$ and any $\theta \in \mathbb{S}^{d-1}$ where $\mathbb{S}^{d-1}$ is the unit sphere in the d-dimensional space, we have*

$$\Pr[\|x\| \geq t\sigma\sqrt{d}] \ \leq \exp(-(d/2)(t-1)^2).$$

From this, we can upper-bound the maximum norm over n Gaussian random vectors with mean 0 and variance σ^2 by $4\sigma\sqrt{d\log n}$ with dominating probability.

Corollary 11 (Global diameter of Gaussian random variables)**.** *For any $n \geq 2$, let $x_1, \dots, x_n \in \mathbb{R}^d$ be random variables where each x_i is independent Gaussian distributed with mean 0 and standard deviation σ. Then with probability at least $1 - n^{-d}$, $\max_{i\in[n]} \|x_i\| \leq 4\sigma\sqrt{d\log n}$.*

Proof. From Lemma 10, we have for each $i \in [n]$ that

$$\Pr[\|x_i\| > 4\sigma\sqrt{d\log n}] \leq \exp(-\frac{d(4\sqrt{\log n}-1)^2}{2}) \leq \exp(-2d\log n) \leq n^{-1} \cdot n^{-d}.$$

Then the statement follows from the union bound over $i = 1, \dots, n$. □

Theorem 12 (Chernoff bound)**.** *Let $X_1, \dots, X_n \in \{0, 1\}$ be n independently distributed random variables. Let $X \coloneqq \sum_{i=1}^n X_i$. Then*

$$\Pr[X = 0] \leq e^{-\mathbb{E}[X]/2}.$$

Theorem 13 (Mass distribution of the sphere)**.** *If e_1 is a fixed and arbitrary unit vector, and if $\theta \in \mathbb{S}^{d-1}$ is sampled uniformly at random from the unit sphere, then for any $\alpha > 0$ we have*

$$\Pr[|\theta^\top e_1| \leq \alpha] \leq \alpha\sqrt{de}.$$

Moreover we also have a tail bound

$$\Pr[|\theta^\top e_1| \geq t/\sqrt{d}] \leq \sqrt{de} \cdot e^{-t^2/2}.$$

Proof. See Appendix B. □

3 Algorithms

This section will show how to adapt the algorithmic reduction of [Ver09] such that it can be used for the semi-random shadow vertex method. Proofs of the stated lemmas can be found in [Ver09]. The full procedure will output one of following scenarios

- a vector $x \in \mathbb{R}^d$ with $Ax \leq 0$, certifying *unboundedness*,
- a vector $y \in \mathbb{R}^n$ with $y^\top A < 0$, certifying *infeasibility*, or
- a basis $I \in \binom{[n]}{d}$ which is both feasible $A(A_I^{-1}b_I) \leq b$ and optimal $c^\top A_I^{-1} \geq 0$.

Note that we have a rather generous definition for unboundedness: an LP can simultaneously be unbounded and infeasible, or simultaneously be unbounded and admit an optimal basic feasible solution. This flexibility we grant ourselves out of kindness and not necessity. All that follows can be adapted to work with a more restrictive definition of unboundedness. For the sake of the clarity of our argument, we proceed with this terminology.

3.1 Shadow vertex method

The Shadow vertex method Algorithm 1 is a pivot rule for the simplex method. Let two objective vectors $y, y' \in \mathbb{R}^d$ and a feasible basis $I \in \binom{[n]}{d}$ optimal for y be given. We want to find a basis optimal for y'. The shadow vertex pivot rule will prescribe pivot steps in such a manner that, throughout the algorithm's duration, the current basis is optimal for an objective $y_\lambda := \lambda y' + (1-\lambda)y$ with $\lambda \in [0, 1]$. In the following we will describe the primal interpretation of the shadow vertex method which we summarize as Algorithm 1. In each iteration of Algorithm 1, λ is increased according to the condition of line 9. If λ is found to be at least 1, then our current basis is optimal for y' and returned. Otherwise Algorithm 1 finds in lines $13 - 20$ an $l \neq j$ such that the new basis $I \setminus \{l\} \cup \{j\}$ is optimal for y_λ and repeat with increased λ. With probability 1, l and j are uniquely determined because the system $Ax \leq b$ is non-degenerate since the bounds b are perturbed. Since $y_1 = y'$, at the end of this path the method has found an optimal basis for the objective y'. Algorithm 1 is called the shadow vertex method because, when the feasible set is orthogonally projected onto the two-dimensional linear subspace $\mathrm{span}(y, y')$, the vertices visited by the algorithm project onto the boundary of the projection ("shadow") $\pi_{\mathrm{span}(y,y')}(\{x : Ax \leq b\})$ of the feasible region.

Algorithm 1 Shadow vertex method ShadowVertex(A, b, y, y', I)

1: **Input:** non-degenerate polyhedron $= \{x \in \mathbb{R}^d \colon Ax \leq b\}$
2: objective functions $y, y' \in \mathbb{R}^d$
3: feasible basis $I \subseteq [n]$, optimal for y
4: **Output:** basis $I \subseteq [n]$ optimal for y' or *unbounded*

5: $i \leftarrow 0$ // Iteration counter
6: $\lambda_i \leftarrow 0$ // Shadow progress
7: **while** $\lambda_i \neq 1$ **do**
8: $\quad i \leftarrow i + 1$
9: $\quad \lambda_i \leftarrow$ maximal λ such that $y_\lambda^\top A_I^{-1} \geq 0$ // Maximal λ such that I is optimal for $\lambda y' + (1-\lambda)y$
10: $\quad$ **if** $\lambda_i \geq 1$ **then**
11: $\qquad$ **return** I // If basis is optimal for y, return said basis
12: $\quad$ **end if**
13: $\quad j \leftarrow j \in I$ such that $(y_\lambda^\top A_I^{-1})_j = 0$ // Pivot rule. Will be unique
14: $\quad x_I \leftarrow A_I^{-1} b_I$
15: $\quad s_i \leftarrow$ supremum over all s such that $A(x_I - sA_I^{-1}e_j) \leq b$ // Find simplex step length s
16: $\quad$ **if** $s_i = \infty$ **then**
17: $\qquad$ **return** *unbounded*
18: $\quad$ **end if**
19: $\quad l \leftarrow l \in [n] \setminus I$ such that $a_l^\top (x_I - s_i A_I^{-1} e_j) \leq b_l$ // Ratio test. Will be unique
20: $\quad I \leftarrow I \setminus \{l\} \cup \{j\}$
21: **end while**

Definition 14. *We denote by* $\pi_{c,c'} : \mathbb{R}^d \to \mathrm{span}(c, c')$ *the orthogonal projection onto the span of* c *and* c'. *We call the image* $\pi_{c,c'}(Q)$ *of* Q *under* $\pi_{c,c'}$ *shadow polygon. Note that as* Q *can be unbounded, the shadow polygon* $\pi_{c,c'}(Q)$ *might be unbounded.*

Definition 15. *Given a basis $I \in \binom{[n]}{d}$ we write the corresponding solution as $x_I = A_I^{-1} b_I$. The set $F(A,b) \subseteq \binom{[n]}{d}$ consists of all feasible bases, i.e., bases for which $Ax_I \leq b$.*

The subset $P(A,b,c,c') \subseteq F(A,b)$ is called shadow path from c to c', and consists of all bases such that x_I is maximized by some (intermediate) $y \in [c,c']$. The vertices v_1, v_2 on $P(A,b,c,c')$ maximizing c respectively c' are called endpoints.

Definition 16. *Let $I, I' \in P(A,b,c,c')$. We say that I' is a neighbor of I on the shadow path if there exists an edge on the shadow polygon between $\pi_{c,c'}(x_I)$ and $\pi_{c,c'}(x_{I'})$. Note that there can be other bases $J \in P(A,b,c,c')$ which, despite having intersection $|I \cap J| = d-1$, are not neighbors* on the shadow path*.*

Let $N(A,b,c,c',I)$ denote the set of neighbors of I on the shadow path $P(A,b,c,c')$.

Definition 17. *A shadow path $P(A,b,c,c,')\subseteq F(A,B)$ is called non-degenerate if the pre-image $\pi_{c,c'}^{-1}(\pi_{c,c'}(x_I))$ of every basic solution x_I for $I \in S$ consists of exactly one basic feasible solution x_I.*

Note that by convexity this definition implies that also the edges between the basic feasible solutions of $P(A,b,c,c,')$ induced by the graph structure on the shadow polygon have exactly one pre-image.

Fact 18 (Non-degeneracy of the shadow path)**.** *If the matrix $A \in \mathbb{R}^{n \times d}$ has independent Gaussian-distributed entries, which are also independent of b, c and Z, then the shadow path is non-degenerate with probability* 1*.*

Fact 19. *For any $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n, \lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$ and linearly independent $c, c' \in \mathbb{R}^d$ we have*

- $P(A,b,c,c') = P(\lambda_1 A, \lambda_2 b, \lambda_3 c, \lambda_4 c')$
- $|P(A,b,c,c')| = |P(A,b,c',c)|$.

Fact 20. *Let $P(A,b,c,c')$ be a non-degenerate shadow path. Then the subgraph of the shadow polygon induced by the bases $I \in P(A,b,c,c')$ is a path in the graph-theoretical sense. If $P(A,b,c,c') \geq 2$ then for any $I \in P(A,b,c,c')$, we have $|N(A,b,c,c',I)| = 2$ except for the two endpoints where it is* 1*.*

Fact 21. *Let $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$ and let $c, c' \in \mathbb{R}^d$ be linearly independent objectives. Let $P(A,b,c,c')$ be a non-degenerate shadow path. Then we have that $|P(A,b,c,y) \cap P(A,b,y,c')| \leq 1$ for all $y \in [c,c']$.*

3.2 First auxiliary LP

For the first auxiliary LP that we solve, we sample $Z \in \mathbb{R}^d$ with independent entries, each following a Gaussian distribution with mean 0 and standard deviation 1. The LP we will solve in this step of the algorithm is:

$$\begin{aligned} \max\ & Z^\top x \\ & Ax \leq 1 \end{aligned} \tag{Unit LP}$$

The all-zeroes solution is feasible. We note that, of the original LP data, only the constraint matrix A appears in (Unit LP). In order to obtain a feasible starting basis that is independent of the noise on A, we will add d artificial constraints. Let $\bar{s}_1, \ldots, \bar{s}_d \in \mathbb{R}^d$ be such that $\mathrm{conv}(\bar{s}_1, \ldots, \bar{s}_d)$ is a regular $d-1$-dimensional simplex, and furthermore satisfy $e_d^\top \bar{s}_i = 3$ and $\|e_d - \bar{s}_i\| = \frac{1}{10\sqrt{\ln d}}$ for each $i = 1, \ldots, d$. Sample independently perturbed vectors $s_1, \ldots, s_d \in \mathbb{R}^d$ with means respectively equal to $\bar{s}_1, \ldots, \bar{s}_d$ and standard deviation $\sigma > 0$. Let $R \in O(d)$ denote a uniformly random rotation matrix and construct (Unit LP') as follows:

$$\begin{aligned} \max Z^\top x & \\ Ax &\leq 1 \\ (Rs_i)^\top x &\leq 1 \qquad \forall i = 1, \ldots, d. \end{aligned} \tag{Unit LP'}$$

This construction we take from [Ver09] and has the following helpful properties:

Lemma 22. *If* (Unit LP) *admits an optimal solution* x^* *then with probability at least* 0.3 *it satisfies* $(Rs_i)^\top x^* \leq 0$ *for all* $i = 1, \ldots, d$. *This probability is independent of* A.

Lemma 23. *Let* $S \in \mathbb{R}^{d \times d}$ *denote the matrix with rows* $s_1, \ldots, s_d$. *Conditional on the rows of* A *each having norm at most* 2 *then, with probability at least* 0.9, *independent of* A, *the basic solution* $(RS)^{-1}1$ *is feasible and satisfies* $(Re_d)^\top (RS)^{-1} \geq 0$.

The outcome of these lemmas is as follows. We can construct (Unit LP') as described, take Re_d as our fixed objective and Z as our random objective, and attempt to follow the shadow path from Re_d to Z. With constant probability this succeeds and gives an optimal basic feasible solution to (Unit LP). On a failure the procedure is repeated, until success. Since the success probability is independent of A, the lengths of all attempted shadow paths are identically distributed. We prove in Theorem 49 that this path has length $O(\sqrt{\sigma^{-1}\sqrt{d^1 1 \log(n/\sigma)^7}})$.

Since the smoothening of the system should not effect the properties (i)–(iii), we need to restrict the perturbation size σ. Dadush and Huiberts [DH20] claimed that the restriction of the perturbation size σ suffices $\sigma \leq \frac{c}{\max_{\{\sqrt{d \log n}, \sqrt{d} \log d\}}}$ for some $c > 0$. As this does not chance the structure of (Unit LP), one can always increase σ by scaling up the constraint matrix A. Thus, the restriction on σ is without loss of generality.

If any attempted shadow path finds that the feasible region of (Unit LP') is unbounded, then according to our definition so is the original LP. Thus we may simply return *unbounded* whenever this occurs.

3.3 Second auxiliary LP

Sample a $(d+1)$'th standard Gaussian distributed number $Z_{d+1} \in \mathbb{R}$. We now think of an optimal basis $I \in \binom{[n]}{d}$ to (Unit LP) as indexing constraints in the *interpolation LP* given as

$$\begin{aligned} \max Z^\top x + Z_{d+1} t & \\ Ax + (1-b)t \leq 1. & \end{aligned} \tag{Int-LP}$$

The slice where $t = 0$ equals the feasible region of (Unit LP), meaning that I indexes a set of constraints that is tight for some edge. Both endpoints of this edge are part of the combined shadow path

$$P\Big((A, (1-b)), 1, -e_{d+1}, (Z, Z_{d+1})\Big) \cup P\Big((A, (1-b)), 1, (Z, Z_{d+1}), e_{d+1}\Big).$$

As such, the algorithm will be dropped somewhere on this path and we are able to use the shadow vertex method to increase t. The slice where $t = 1$ has a feasible region equal to the original LP, meaning that, as soon as we find a point satisfying this, we have obtained a basic feasible solution for phase 1. Again by Theorem 49 we know that this path has length $O(\sqrt{\sigma^{-1}\sqrt{d^11\log(n/\sigma)^7}})$.

If the shadow vertex method stops early, finding that the optimal solution to

$$\max t$$
$$Ax + (1-b)t \leq 1.$$

has value strictly less than 1, then this basis gives a certificate that the feasible set $\{x : Ax \leq b\}$ is empty. In that case the algorithm may return said certificate.

3.4 Phase 2

Having obtained a basic feasible solution $A_I^{-1}b_I$ that maximizes the random objective $Z^\top x$, all that remains is to follow the semi-random shadow path from Z to c. We prove in Theorem 50 that this can be done in $O(\sqrt{\sigma^{-1}\sqrt{d^11\log(n/\sigma)^7}})$ pivot steps. This finishes the algorithmic reduction.

4 Semi-random shadow bound

We will prove a semi-random shadow bound in two cases: either when b is perturbed as is prescribed for smoothed analysis, or when b is fixed to be the all-ones vector.

Although for algorithmic purposes we were satisfied with any rotationally symmetric distribution for Z, the proofs in this section will have the norm $\|Z\|$ require a specific distribution as well. For that purpose, recall from Fact 19 that for any $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n, \lambda_1, \lambda_2 > 0$ and linearly independent $c, c' \in \mathbb{R}^d$ we have

$$P(A, b, c, c') = P(A, b, \lambda_1 c, \lambda_2 c').$$

As such, changing the norm of the random vector Z has no algorithmic consequences. We will sample Z to be a 1-log-Lipschitz random variable as per Definition 4.

4.1 Pivot steps close to the fixed objective

As the algorithm traverses the shadow path, the main analysis requires there to be a "large amount" of randomness in the objectives that are visited. This is true for the majority of the path, except when the angle between the "current objective" and the LP's true objective is small. For that reason, we must treat this part of the shadow path separately first. The following statement is inspired by the angle bound of [ST04], but to keep this present document self-contained we give a simple proof of a similar but much weaker result. For our purposes this weaker version suffices.

Definition 24. *(Angle) Given two nonzero vectors $s, s' \in \mathbb{R}^d$, the angle $\theta(s, s') \in [0, \pi]$ between s and s' is defined to be the unique number such that* $\cos(\theta(s, s')) \cdot \|s\| \cdot \|s'\| = s^\top s'$.

For two sets $S, S' \subset \mathbb{R}^d \setminus \{0\}$ we define $\theta(S, S') = \inf_{s\in S, s'\in S'} \theta(s, s')$.

Lemma 25 (Angle bound)**.** *Let $c \in \mathbb{R}^d \setminus \{0\}$ be an objective vector. Assume that $a_1, \ldots, a_n \in \mathbb{R}^d$ are independent Gaussian distributed random vectors, each with standard deviation $\sigma \leq 1/4\sqrt{d \ln n}$ and $\|\mathbb{E}[a_i]\| \leq 1$. Let $0 < \varepsilon \leq \pi/10$. Then*

$$\Pr\left[\exists J \in \binom{[n]}{d-1} : \theta\big(c, \operatorname{cone}(a_j : j \in J)\big) < \varepsilon\right] \leq 4d \cdot n^d \cdot \frac{\varepsilon}{\sigma\sqrt{2\pi}} + n^{-d}.$$

Proof. Consider the event E that, for every $J \in \binom{[n]}{d-1}$ and every $j \in J$, we have $\operatorname{dist}\big(a_j, \operatorname{span}(\{c\} \cup \{a_i : i \in J \setminus \{j\}\})\big) \geq 2d\varepsilon$. Moreover, consider the event D that, for every $j \in [n]$, we have $\|a_j\| \leq 2$. We first show that $E \wedge D$ implies that for all $J \in \binom{[n]}{d-1}$ we have $\theta(c, \operatorname{cone}(a_i : i \in J)) > \varepsilon$. After that we will show that $\Pr[\neg(E \wedge D)] \leq n^d \cdot \frac{\varepsilon}{\sigma\sqrt{2\pi}} + n^{-d}$.

Assume that E and D hold. Let $J \in \binom{[n]}{d-1}$ be arbitrary. By our assumption of E, for each $j \in J$ there exists some separator $y_j \in \operatorname{span}(\{c\} \cup \{a_i : i \in J \setminus \{j\}\})^\perp$ with $\|y_j\| = 1$ that certifies this distance through the inequalities $y_i^\top a_j \geq 2d\varepsilon$ and $y_j^\top a_i = 0$ for each $i \in J \setminus \{j\}$. For their sum $y = \sum_{j \in J} y_j$ we know that $y^\top a_j \geq 2d\varepsilon$ for all $j \in J$, as well as that $y^\top c = 0$. Now consider any $p \in \operatorname{cone}(a_j : j \in J)$ that achieves $\theta(c, p) = \theta(c, \operatorname{cone}(a_j : j \in J))$. Without loss of generality we assume $p \in \operatorname{conv}(a_j : j \in J)$. In particular we know from the above that $y^\top p \geq 2d\varepsilon$. The triangle inequality gives us that $\|y\| \leq d$. We further deduce since D holds that $\|p\| \leq \max_{j \in J}\|a_j\| \leq 2$. From the definition of θ we get $\cos(\theta(y, p)) = y^\top p \cdot \|y\|^{-1} \cdot \|p\|^{-1} \geq \varepsilon$. In particular, we find that $\theta(y, p) \leq \pi/2 - \varepsilon$. We know that $\theta(c, y) = \pi/2$ due to $y^\top c = 0$, and hence the triangle inequality on the sphere gives us $\theta(c, p) \geq \varepsilon$.

It remains to show that $\Pr[\neg(E \wedge D)] \geq n^d \cdot \frac{\varepsilon}{\sigma\sqrt{2\pi}} + n^{-d}$. We use the union bound:

$$\begin{aligned}\Pr[\neg(E \wedge D)] &\leq \Pr[\neg E] + \Pr[\neg D] \\ &\leq \Pr[\neg D] + \sum_{J \in \binom{[n]}{d-1}} \sum_{j \in J} \Pr[\operatorname{dist}(a_j, \operatorname{span}(a_i : i \in J \setminus \{j\})) \geq 2d\varepsilon].\end{aligned}$$

Since $\sigma \leq 1/\sqrt{d \ln n}$ we know that $\|a_i\| > 2$ implies $\|a_i - \mathbb{E}[a_i]\| > 4\sigma\sqrt{d \ln n}$, so Corollary 11 gives that $\Pr[\neg D] \leq n^{-d}$. The double summation has $(d-1) \cdot \binom{n}{d-1} \leq n^d$ terms in total, so in the remainder we need only upper bound the summand uniformly. For that purpose, let $J \in \binom{[n]}{d-1}$ and $j \in J$ be arbitrary. We may as well consider $V := \operatorname{span}(\{c\} \cup \{a_i : i \in J \setminus \{j\}\})$ to be fixed. Write $y_j \in \mathbb{S}^{d-1}$ to be one of the two unit normal vectors to this linear subspace V. We are interested in the distance $\operatorname{dist}(a_j, V) = |y_j^\top a_j|$.

Note that V depends only on the values of a_i for $i \in J \setminus \{j\}$, and as such y_j is independent of a_j. That makes the inner product $y_j^\top a_j$ follow a Gaussian distribution with mean $y_j^\top \mathbb{E}[a_j]$ and standard deviation σ. The probability density function of this random variable is uniformly upper bounded by $\frac{1}{\sigma\sqrt{2\pi}}$, and hence the probability that it is contained in an interval of length $2d\varepsilon$ is at most

$$\Pr[\operatorname{dist}(a_j, V) < 2d\varepsilon] = \Pr\left[y_j^\top a_j \in (-2d\varepsilon, 2d\varepsilon)\right] \leq \frac{4d\varepsilon}{\sigma\sqrt{2\pi}}.$$

The union bound over all choices for $J \in \binom{[n]}{d-1}$ and $j \in J$ closes out the proof. $\square$

For upper bounding the number of pivot steps between objectives with small angle between them on the total shadow path we need a slightly different characterization, captured by the following lemma.

Corollary 26. *Let $c \in \mathbb{S}^{d-1}$ be a fixed objective, and let $Z \in \mathbb{R}^d$ be a random objective that is linearly independent of c and satisfies $\Pr[\|Z\| \geq t] \leq n^{-d}$ for some $t > 0$. Assume $b \in \mathbb{R}^n$ is arbitrary, and that $a_1, \ldots, a_n$ are independent Gaussian distributed random vectors each with standard deviation $\sigma \leq 1/4\sqrt{d \ln n}$ and $\|\mathbb{E}[a_i]\| \leq 1$. Write $k = \left\lceil \log_2 \left(\frac{2dt \cdot n^{2d}}{\|c\| \sigma \sqrt{2\pi}} \right) \right\rceil$. The expected length of the shadow path between the objective and a perturbed objective satisfies*

$$\mathbb{E}\big[|P(A, b, 2^k c + Z, c)|\big] \leq 7.$$

Proof. In the event that $\|Z\| \geq t$ we count at most $\binom{n}{d}$ distinct bases. Since $\Pr[\|Z\| > t] \leq n^{-d}$, the expected number of pivot steps incurred by this situation is at most 1. For that reason, we will for the remainder of this proof only consider the case $\|Z\| < t$.

Note that $\theta(2^k c + Z, c) = \theta(2^k c + Z, 2^k c)$ and consider the triangle $\triangle(0, 2^k c + Z, 2^k c)$. Let us abbreviate its vertices by $a = 2^k c + Z$ and $b = 2^k c$, so that our triangle is $\triangle(0, a, b)$. The assumption on k gives that $\|2^k c\| \geq \frac{2dtn^{2d}}{\sigma\sqrt{2\pi}}$ and with the triangle inequality we find that the edge $[a, b]$ has the shortest length of the three.

We recall the law of sines to derive

$$\frac{\sin(\theta(a, b))}{\|a - b\|} = \frac{\sin(\theta(0 - a, b - a))}{\|b\|} \leq \frac{1}{\|b\|} = \frac{1}{2^k \|c\|} \leq \frac{\sigma\sqrt{2\pi}}{2dt \cdot n^{2d}}.$$

This gives an upper bound on the sine of our desired angle. To relate this to the angle itself, note that the shortest edge of a triangle is opposite of the smallest angle, which gives us that $\theta(a, b) \leq \pi/3$. For any $\theta \in [0, \pi/3]$ one has $\sin(\theta) > 0.8\theta$, so in particular

$$\theta(a, b) \leq \frac{1}{0.8} \sin(\theta(a, b)) \|Z\| \leq \frac{5\sigma\sqrt{2\pi}}{4 \cdot 2d \cdot n^{2d}}.$$

As such, our two objectives $2^k c + Z$ and c have an angle at most $\frac{5\sigma\sqrt{2\pi}}{4 \cdot 2d \cdot n^{2d}}$ between them. If $|P(A, b, 2^k c + Z, c)| \geq 2$, i.e., if there was a pivot step taken between them, then that implies there is a basis $I \in P(A, b, 2^k c + Z, c)$ such that $A_I^{-\top} c \geq 0$ but $A_I^{-\top}(2^k c + Z) \ngeq 0$. This implies that there is a subset $J \subset I, |J| = d - 1$ and a point $p \in [2^k c + Z, c] \cap \operatorname{span}(a_j : j \in J)$. This point must satisfy $\theta(p, c) \leq \theta(2^k c + Z, c) \leq \frac{5\sigma\sqrt{2\pi}}{8d \cdot n^{2d}}$, implying that in fact $\theta(c, \operatorname{span}(a_j : j \in J)) \leq \frac{5\sigma\sqrt{2\pi}}{8d \cdot n^{2d}}$.

Lemma 25 shows us that the probability of this happening is at most $6n^{-d}$. Counting at most $\binom{n}{d}$ pivot steps in this case, we may conclude

$$\mathbb{E}[|P(A, b, 2^k c + Z, c)|] \leq 1 + \binom{n}{d} \Pr[|P(A, b, 2^k c + Z, c)| > 1] \leq 7.$$

□

4.2 Multipliers

We will give a bound on the expected number of pivot steps for most of the shadow path, i.e., the path segment $P(A, b, Z, 2^k c + Z)$. To start, we require the following theorem proven by [BBHK25]. At the moment of writing their manuscript is yet unpublished. For that reason we reproduce a verbatim proof in the appendix (Appendix A).

Theorem 27. *Let $B \in \mathbb{R}^{d\times d}$ be an invertible matrix, every whose column has Euclidean norm at most 2, and define, for any $m > 0$, $C_m = \{x \in \mathbb{R}^d : B^{-1}x \geq m\}$. Suppose $c, c' \in \mathbb{R}^d$ are fixed. Let $Z \in \mathbb{R}^d$ be a random vector with 1-log-Lipschitz probability density μ. Then*

$$\Pr\left[[c+Z, c'+Z] \cap C_m \neq \emptyset\right] \geq 0.99 \Pr\left[[c+Z, c'+Z] \cap C_0 \neq \emptyset\right]$$

for $m = \ln(1/0.99)/2d$.

The elements of the shadow path satisfying the property described above form a set that we will keep track of.

Definition 28. *Given $A \in \mathbb{R}^{n\times d}$ and $c, c' \in \mathbb{R}^d$, and a threshold $m > 0$, the set of bases with* good multipliers *is*

$$M(A, c, c', m) = \left\{ I \in \binom{[n]}{d} \;\middle|\; \exists y \in [c, c'] \quad s.t. \quad yA_I^{-1} \geq m \right\}.$$

In the language of this definition, the above theorem says most bases with all nonnegative multipliers $I \in M(A, c+Z, c'+Z, 0)$ will have good multipliers $I \in M(A, c+Z, c'+Z, \ln(1/0.99)/2d)$.

Corollary 29. *For any fixed $A \in \mathbb{R}^{n\times d}$ and any fixed $c, c' \in \mathbb{R}^d$, if $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function then for $m = \ln(1/0.99)/2d$ we have*

$$\Pr\big[I \in M(A, c+Z, c'+Z, m)\big] \geq 0.99 \cdot \Pr\big[I \in M(A, c+Z, c'+Z, 0)\big].$$

Proof. Write $C_m = \{y \in \mathbb{R}^d : A^{-\top} y \geq m\}$. We observe that

$$\Pr\big[I \in M(A, c+Z, c'+Z, m)\big] = \Pr\big[\exists y \in [c+Z, c'+Z] \colon A_I^{-1} y \geq m\big] = \Pr\big[[c+Z, c'+Z] \cap C_m \neq \emptyset\big],$$

and similarly for $M(A, c+Z, c'+Z, 0)$. At this point we can directly apply Theorem 27 to the invertible matrix $A_I^\top$ and get

$$\begin{aligned}
\Pr\big[I \in M(A, c+Z, c'+Z, m)\big] &= \Pr\big[[c+Z, c'+Z] \cap C_m \neq \emptyset\big] \\
&\leq 0.99 \Pr\big[[c+Z, c'+Z] \cap C_0 \neq \emptyset\big] \\
&= \Pr\big[I \in M(A, c+Z, c'+Z, 0)\big].
\end{aligned}$$

□

4.3 Slack

Having good multipliers alone is not sufficient as we want every vertex on the shadow-path to be "well-separated" from the others. For that purpose we will over the course of this subsection prove that all bases which have non-negligible probability of being feasible also have a good probability of being feasible by a good margin, i.e., the minimum non-zero slack is bounded away from 0. This subsection is loosely based on Section 5.3 (Randomized lower bound for δ) in [HLZ23]. We require a few facts about the Gaussian distribution. First a technical lemma about the range in which we may treat the Gaussian distribution as having a log-Lipschitz probability density function.

Lemma 30 (Gaussian as log-Lipschitz)**.** *Assume $s \in \mathbb{R}$ is Gaussian distributed with variance σ^2 and denote its probability density function by $f(\cdot)$. If $t \in \mathbb{R}, p \in (0, 1/e]$ satisfy* $\Pr[s \geq t - \varepsilon] \geq p$ *and* $\Pr[s \leq t] \geq p$ *then for any $x_1, x_2 \in [t - 4\sigma\sqrt{\ln p^{-1}}, t + 4\sigma\sqrt{\ln p^{-1}}]$ we have*

$$\frac{f(x_1)}{f(x_2)} \leq \exp\Big(8\sigma^{-1}\sqrt{\ln p^{-1}} \cdot |x_1 - x_2|\Big).$$

Proof. We first prove that $t \leq \mathbb{E}[s] + 4\sigma\sqrt{\ln p^{-1}}$. Suppose not, then we bound

$$\Pr[s \geq t - \varepsilon] = \Pr\left[s - \mathbb{E}[s] \geq t - \mathbb{E}[s] - \varepsilon\right] \leq \Pr\left[|s - \mathbb{E}[s]| \geq 3\sigma\sqrt{\ln p^{-1}}\right],$$

Since $3\sigma\sqrt{\ln p^{-1}} \geq 2\sigma\sqrt{\ln p^{-1}} + 1$, we conclude from Lemma 10 that this last probability is strictly less than p, giving a contradiction. Similarly, we may prove that $t \geq \mathbb{E}[s] - 4\sigma\sqrt{\ln p^{-1}}$ by assuming its opposite and computing

$$\Pr[s \leq t] = \Pr\left[\mathbb{E}[s] - s \geq \mathbb{E}[s] - t\right] \leq \Pr\left[|s - \mathbb{E}[s]| \geq 4\sigma\sqrt{\ln p^{-1}}\right],$$

once again leading to a contradiction by way of Lemma 10.

Recall that the probability density function of s is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mathbb{E}[s])^2}{2\sigma^2}}$, which means that on the interval $x_1, x_2 \in \left[t - 4\sigma\sqrt{\ln p^{-1}}, t + 4\sigma\sqrt{\ln p^{-1}}\right] \subseteq \left[\mathbb{E}[s] - 8\sigma\sqrt{\ln p^{-1}}, \mathbb{E}[s] + 8\sigma\sqrt{\ln p^{-1}}\right]$ it satisfies

$$\begin{aligned}
\log(f(x_1)) - \log(f(x_2)) &= \frac{1}{2\sigma^2}\left((x_2 - \mathbb{E}[s])^2 - (x_1 - \mathbb{E}[s])^2\right) \\
&= \frac{1}{2\sigma^2}\left(x_2^2 - 2x_2\mathbb{E}[s] - x_1^2 + 2x_1\mathbb{E}[s]\right) \\
&= \frac{1}{2\sigma^2} \cdot (x_1 + x_2 - 2\mathbb{E}[s]) \cdot (x_2 - x_1) \\
&\leq 8\sigma^{-1}\sqrt{\ln p^{-1}} \cdot |x_1 - x_2|.
\end{aligned}$$

This is equivalent to our desired statement. □

With the above lemma in place, we can set out to prove the main non-trivial fact that we require of the Gaussian distribution.

Lemma 31 (Condition-reversing interval lemma)**.** *Suppose $s \in \mathbb{R}$ is Gaussian distributed with variance σ^2. For $t \in \mathbb{R}, p \in (0, 1/e]$, write $L = 8\sigma^{-1}\sqrt{\ln p^{-1}}$ and pick any $0 \leq \varepsilon \leq 1/L$. Assuming that $\Pr[s \geq t - \varepsilon] \geq p$ and $\Pr[s \leq t] \geq p$ we have*

$$\Pr[s \geq t - \varepsilon \mid s \leq t] \leq e^3 \varepsilon L \cdot \Pr[s \geq t].$$

Proof. We start by proving that $\Pr\left[s \in [t - \varepsilon, t]\right] \leq e^2 \varepsilon L \cdot \Pr\left[s \in [t, t + 1/L]\right]$. In the final paragraphs of this proof we will extend this statement into the desired conclusion.

From Lemma 30 we find that $f(x_1)/f(x_2) \leq L \cdot |x_1 - x_2|$ for any two points $x_1, x_2 \in [t - 4\sigma\sqrt{\ln p^{-1}}, t + 4\sigma\sqrt{\ln p^{-1}}]$. From here we can upper bound the probability in our intended left-hand side as

$$\Pr\left[s \in [t - \varepsilon, t]\right] = \int_{t-\varepsilon}^{t} f(x)\,\mathrm{d}\,x \leq \int_{t-\varepsilon}^{t} f(t) e^{L|x-t|}\,\mathrm{d}\,x \leq e\varepsilon f(t).$$

Similarly, we may use this log-Lipschitzness property to lower bound the probability in our intended right-hand side and find

$$\Pr\left[s \in [t, t + 1/L]\right] = \int_{t}^{t+1/L} f(x)\,\mathrm{d}\,x$$

$$
\begin{aligned}
&\geq \int_t^{t+1/L} f(t)e^{-L\cdot|x-t|}\,\mathrm{d}\,x \\
&\geq e^{-1}f(t)L^{-1}.
\end{aligned}
$$

Putting these two inequalities together, we find $\Pr\big[s \in [t-\varepsilon, t]\big] \leq e^2\varepsilon L \Pr\big[s \in [t, t+1/L]\big]$. This is the initial statement mentioned at the start of this proof. Still using the log-Lipschitzness of Lemma 30, we may observe too that

$$
\begin{aligned}
\Pr[s \leq t] &= \int_{-\infty}^{t} f(x)\,\mathrm{d}\,x \\
&= \int_{-\infty}^{t+1/L} f(x-1/L)\,\mathrm{d}\,x \\
&\geq \int_{-\infty}^{t+1/L} f(x)e^{-1}\,\mathrm{d}\,x \\
&= e^{-1}\Pr[s \leq t+1/L].
\end{aligned}
$$

Using the above two inequalities in order to bound the numerator and the denominator, we can now prove the lemma as follows

$$
\begin{aligned}
\Pr[s \geq t-\varepsilon \mid s \leq t] &= \frac{\Pr\big[s \in [t-\varepsilon, t]\big]}{\Pr[s \leq t]} \\
&\leq \frac{e\Pr\big[s \in [t-\varepsilon, t]\big]}{\Pr[s \leq t+1/L]} \\
&\leq \frac{e^3\varepsilon L\cdot \Pr\big[s \in [t, t+1/L]\big]}{\Pr[s \leq t+1/L]} \\
&= e^3\varepsilon L\cdot \Pr\big[s \geq t \mid s \leq t+1/L]\big] \\
&\leq e^3\varepsilon L\cdot \Pr\big[s \geq t\big],
\end{aligned}
$$

where the last inequality follows due to $s > t+1/L$ implying $s \geq t$. This, finally, proves the lemma. ☐

This all leads up to a kind of anti-concentration result first described in [HLZ23] which allowed them (and will allow us) to substantially improve over what a naive union bound argument can achieve. Whereas [HLZ23] proved this for log-Lipschitz probability distributions, we obtain a similar result for the Gaussian distribution. It will be the primary tool used to establish that the non-zero slack values are bounded away from 0.

Lemma 32 (Conditional Anti-concentration)**.** *Suppose $s_1, \ldots, s_k \in \mathbb{R}$ are independently Gaussian distributed, each with standard deviation $\sigma > 0$, and suppose $t_1, \ldots, t_k \in \mathbb{R}$ are fixed. If $q \in (0, 1/e)$ is such that $\Pr[s \leq t] \geq q$ then for any $\varepsilon > 0$ we have*

$$
\Pr\Big[\exists j \in [k] : s_j \geq t_j - \varepsilon \;\Big|\; s \leq t\Big] \leq q + 16e^3\varepsilon\sigma^{-1}\ln^{3/2}(k/q).
$$

Proof. Since we are bounding a probability, without loss of generality we assume $\varepsilon < 16e^3\varepsilon\ln^{3/2}(k/q)$. Define $B = \Big\{j \in [k] : \Pr\big[s_j \geq t_j - \varepsilon\big] \geq q/k\Big\}$. We proceed by independence of the random variables to find

$$\begin{aligned}
\Pr\left[\exists j\in[k]\ :\ s_j\geq t_j-\varepsilon\middle|s\leq t\right] &\leq \sum_{j\in[k]}\Pr\left[s_j\geq t_j-\varepsilon\middle|s_j\leq t_j\right]\\
&\leq \sum_{j\in[k]\setminus B}\Pr\left[s_j\geq t_j-\varepsilon\right]+\sum_{j\in B}\Pr\left[s_j\geq t_j-\varepsilon\middle|s_j\leq t_j\right]\\
&\leq q+\sum_{j\in B}\Pr\left[s_j\geq t_j-\varepsilon\middle|s_j\leq t_j\right].
\end{aligned}$$

For any $j\in B$ we know that $\Pr[s_j\geq t_j-\varepsilon]\geq q/k$. The assumption of $\Pr[s\leq t]\geq q$ implies that $\Pr[s_j\leq t_j]\geq q\geq q/k$, and so we satisfy the conditions of Lemma 31 and conclude

$$\begin{aligned}
\sum_{j\in B}\Pr\left[s_j\geq t_j-\varepsilon\middle|s_j\leq t_j\right] &\leq \sum_{j\in B}8e^3\varepsilon\sigma^{-1}\sqrt{\ln(k/q)}\Pr[s_j\geq t_j]\\
&= 8e^3\varepsilon\sigma^{-1}\sqrt{\ln(k/q)}\cdot\mathbb{E}\Big[|\{j\in B\ :\ s_j\geq t_j\}|\Big].
\end{aligned}$$

Denote this last random set as $V=\{j\in B\ :\ s_j\geq t_j\}$. Now recall the Chernoff bound (Theorem 12) which establishes that $q\leq\Pr[s\leq t]=\Pr\big[|V|=0\big]\leq\exp(-\mathbb{E}[|V|]/2)$. Taking all of the above together we find

$$\begin{aligned}
\Pr\left[\exists j\in[k]\ :\ s_j\geq t_j-\varepsilon\middle|s\leq t\right] &\leq q+\sum_{j\in B}\Pr\left[s_j\geq t_j-\varepsilon\middle|s_j\leq t_j\right]\\
&\leq q+8e^3\varepsilon\cdot\sigma^{-1}\sqrt{\ln(k/q)}\cdot\mathbb{E}[|V|]\\
&\leq q+16e^3\varepsilon\ln^{3/2}(k/q),
\end{aligned}$$

finishing the proof. □

With these technical prerequisites in place, we can now prove the main result of this subsection. Let us define the main properties of interest.

Definition 33. *For a matrix $A\in\mathbb{R}^{n\times d}$ and vector $b\in\mathbb{R}^n$, define the set of feasible bases as*

$$F(A,b)=\{I\in\binom{[n]}{d}:A_I\text{ invertible and }Ax_I\leq b\}.$$

Following that, define the set of feasible bases with relative gap $g>0$ *as*

$$G(A,b,g)=\{I\in F(A,b):A_{[n]\setminus I}x_I\leq b_{[n]\setminus I}-g\cdot\|x_I\|\}.$$

For an appropriate choice of g, we prove that the set $G(A,b,g)$ contains most of the set $F(A,b)$ on average.

Theorem 34. *(Slacks are large) Let the matrix $A\in\mathbb{R}^{n\times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma>0$. Take $b\in\mathbb{R}^n$ to be fixed. If $I\in\binom{[n]}{d}$ satisfies $\Pr[I\in F(A,b)]\geq 10n^{-d}$ then*

$$\Pr\left[I\in G(A,b,\frac{\sigma}{5000d^{3/2}\ln(n)^{3/2}})\ \middle|\ I\in F(A,b)\right]\geq 0.8.$$

Proof. Let $I \in \binom{[n]}{d}$ be arbitrary, assuming $\Pr[I \in F(A,b)] \geq 10n^{-d}$. For the sake of simplicity we assume that $I = \{n-d+1, \ldots, n\}$, and thus that $[n] \setminus I = [n-d]$.

We first sample the submatrix A_I. With probability 1 it is invertible and we may compute $x_I = A_I^{-1} b_I$. Introducing a case distinction, we first assume that conditional on the value of the sample, say $B \in \mathbb{R}^{d \times d}$, that we have $\Pr[I \in F(A,b) \mid A_I = B] \geq n^{-d}$. In this case, for each $j \in [n] \setminus I$ define $t_j = b_j / \|x_I\|$ and $s_j = a_j^\top x_I / \|x_I\|$, thus defining two vectors $s, t \in \mathbb{R}^{n-d}$. Taking $\varepsilon = \frac{\sigma}{5000 d^{3/2} \ln(n)^{3/2}}$, we find that $I \in F(A,b)$ is equivalent to the system of inequalities $s \leq t$. Moreover, we observe that $I \in G(A,b,\varepsilon)$ is equivalent to the system of inequalities $s \leq t - \varepsilon$. Plugging in the conditional anti-concentration Lemma 32 with $q = n^{-d}$ gives that

$$\Pr[I \notin G(A,b,\varepsilon) \mid I \in F(A,b) \wedge A_I = B] \leq n^{-d} + 16e^3 \varepsilon \sigma^{-1} \ln(n^{d+1})^{3/2}$$

for any invertible $B \in \mathbb{R}^{d \times d}$. We deduce from the previous inequality the hopefully more friendly seeming

$$\Pr[I \notin G(A,b,\varepsilon) \wedge I \in F(A,b) \mid A_I = B] \leq 0.1 \cdot \Pr[I \in F(A,b) \mid A_I = B] + n^{-d}. \tag{1}$$

For the other side of the case distinction, the one where $\Pr[I \in F(A,b) \mid A_I = B] < n^{-d}$, we get a stronger conclusion in only two steps:

$$\Pr[I \notin G(A,b,\varepsilon) \wedge I \in F(A,b) \mid A_I = B] \leq \Pr[I \in F(A,b) \mid A_I = B] < n^{-d}.$$

We have thus learned that inequality (1) holds pointwise for any invertible value that A_I might take, which implies that the same inequality holds when A_I is random:

$$\Pr[I \notin G(A,b,\varepsilon) \wedge I \in F(A,b)] \leq 0.1 \cdot \Pr[I \in F(A,b)] + n^{-d}.$$

This ends the case distinction. Using the knowledge that $\Pr[I \in F(A,b)] \geq 10n^{-d}$ we immediately derive $\Pr[I \notin G(A,b,\varepsilon) \wedge I \in F(A,b)] \leq 0.2 \cdot \Pr[I \in F(A,b)]$. Elementary rewriting turns this into

$$\Pr[I \in G(A,b,\varepsilon) \wedge I \in F(A,b)] \geq 0.8 \Pr[I \in F(A,b)],$$

which is equivalent to the statement of the lemma. □

4.4 Triples

In order to get an upper bound on the size of the shadow path, we will need to reason about the case where the shadow path contains a basis in $M(A, c+Z, c'+Z, m)$ and whose neighbors on the shadow path are in $G(A,b,g)$. In order to do this effectively without having to worry about these being correlated in non-trivial ways, we will consider sequences of three bases on the shadow path contained in $M(A, c+Z, c'+Z, m) \cap G(A,b,g)$.

Definition 35. *For a graph $G = (V,E)$ and $S \subseteq V$, write $T^S \subseteq G$ for the vertices $v \in S$ who have at least 2 neighbors in S.*

As long as the bases on the shadow path are all in this set, and we can bound the number of such triples, then this leads to an upper bound on the shadow path length.

Lemma 36. *Let $S \subseteq V \subseteq [k]$ be random sets such that $\Pr[i \in S \mid i \in V] \geq p > 2/3$ for all $i \in [k]$. Suppose that $P = (V,E)$ is a graph on vertex set V that is either a cycle or a path. Then we have*

$$\mathbb{E}[|V|] \leq \frac{2 + \mathbb{E}[|T^S|]}{3p-2}.$$

Proof. We count the sum of the degrees $\delta(i)$ of vertices i in S. Counting per vertex, we find a lower bound of $2|S|-2$ since all vertices have degree at most 2, except possibly the endpoints of the path.

$$2|S| - 2 \leq \sum_{i \in S} \delta(i).$$

This sum counts every edge in the induced subgraph $G[S]$ twice. Every remaining edge connects to a vertex in $V \setminus S$, contributing 1 to that vertex's degree. This implies that

$$\sum_{i \in S} \delta(i) \leq 2|E(G[S])| + \sum_{i \in V \setminus S} \delta(i) \leq 2|E(G[S])| + 2|V \setminus S|.$$

To further upper bound this last quantity, observe that every edge in the subgraph $G[S]$ connects to two vertices in S. Every vertex $i \in T^S$ has degree 2 in the subgraph, while every vertex $i \in S \setminus T^S$ has degree 0 or 1 in the subgraph. From this we find that $|E(G[S])| \leq \frac{1}{2}(2|T^S| + |S \setminus T^S|) = \frac{1}{2}(|S| + |T^S|)$. Taking all of the above together, we find

$$2|S| - 2 \leq |S| + |T^S| + 2|V \setminus S| \leq 2|T^S| - |S| + 2|V|.$$

Simplifying, we get $3|S| \leq 2 + |T^S| + 2|V|$. Now it is time to remember that everything is random to conclude

$$3p\mathbb{E}[|V|] \leq 3\mathbb{E}[|S|] \leq 2 + \mathbb{E}[|T^S|] + 2\mathbb{E}[|V|],$$

and hence we may rearrange to $\mathbb{E}[|V|] \leq \frac{2+\mathbb{E}[|T^S|]}{3p-2}$. □

This subsection and the previous two all lead up to the following theorem.

Theorem 37. *Let the matrix $A \in \mathbb{R}^{n \times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma > 0$. Take $b \in \mathbb{R}^n$ and $c, c' \in \mathbb{R}^d$ to be fixed. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function then*

$$\mathbb{E}[|P(A, b, c + Z, c' + Z)|] \leq 8 + 4\mathbb{E}[|T^{G(A,b,\frac{\sigma}{5000 d^{3/2} \ln(n)^{3/2}}) \cap M(A, c+Z, c'+Z, \frac{\ln(1/0.99)}{2d})}|].$$

Proof. Regard the shadow path as a random graph on the nodes $P(A, b, c + Z, c' + Z) \subset \binom{[n]}{d}$, with an edge between two bases on the path if and only if they are adjacent in the shadow path. Write $m = \ln(1/0.99)/2d$ and $g = \frac{\sigma}{5000 d^{3/2} \ln(n)^{3/2}}$. For any $I \in \binom{[n]}{d}$ we have by Corollary 29 that $\Pr\left[I \in M(A, c + Z, c' + Z, m)\right] \geq 0.99 \Pr\left[I \in M(A, c + Z, c' + Z, 0)\right]$ using the randomness over Z, and by Theorem 34 we have $\Pr\left[I \in G(A, b, g) \;\middle|\; I \in F(A, b)\right] \geq 0.8$ using the randomness over A. Recall that $I \in P(A, b, c + Z, c' + Z)$ is equivalent to $I \in M(A, c + Z, c' + Z, 0) \cap G(A, b, 0)$ and that $I \in G(A, b, 0)$ is equivalent to $I \in F(A, b)$. Since we are integrating non-negative functions we can change the order of integration and find that

$$\begin{aligned}
&\Pr_{A,Z}\left[I \in M(A, c + Z, c' + Z, m) \cap G(A, b, g) \;\middle|\; I \in P(A, b, c + Z, c' + Z)\right] \\
&\quad = \mathbb{E}_A\left[1[I \in G(A, b, g)] \cdot \Pr_Z\left[I \in M(A, c + Z, c' + Z, m) \mid I \in M(A, c + Z, c' + Z, 0)\right] \mid I \in G(A, b, 0)\right] \\
&\quad \geq \mathbb{E}_A\left[0.99 \cdot 1[I \in G(A, b, g)] \mid I \in G(A, b, 0)\right] \\
&\quad \geq 0.99 \cdot 0.8 \geq 3/4.
\end{aligned}$$

We apply Lemma 36 to the above-mentioned subset of the graph vertices and find

$$\mathbb{E}[|P(A,b,c{+}Z,c'{+}Z)|] \leq \frac{2+\mathbb{E}[|T^{M(A,c+Z,c'+Z,m)\cap G(A,b,g)}|]}{3\cdot\frac{3}{4}-2} = 8{+}4\mathbb{E}[|T^{M(A,c+Z,c'+Z,m)\cap G(A,b,g)}|].$$

This is what was needed. □

4.5 Close and far neighbors

In order to make use of Theorem 37, the remainder of this section is dedicated to giving an upper bound on $T^{M(A,c,c',m)\cap G(A,b,g)}$. From here on we use thinking of the shadow path as lying on the boundary of the shadow polygon.

Any basis in $T^{M(A,c,c',m)\cap G(A,b,g)}$ will be counted in one of two ways, depending on the distance to its neighbors as measured in the projection. Recall the definition of neighbor from Definition 16.

Definition 38. *For a given matrix $A \in \mathbb{R}^{n\times d}$, a right-hand side $b \in \mathbb{R}^n$, a pair of objectives $c, c' \in \mathbb{R}^d$ and some threshold $0 < \rho \leq 1/2$, we denote the set of shadow path elements at far distance from their neighbors by*

$$H(A,b,c,c',\rho) = \Big\{I \in P(A,b,c,c') : \ \forall I' \in N(A,b,c,c',I),\ \|\pi_{c,c'}(x_I) - \pi_{c,c'}(x_{I'})\| \geq \rho\|\pi_{c,c'}(x_I)\|\Big\}.$$

Lemma 39. *For any $A \in \mathbb{R}^{n\times d}$, any $b \in \mathbb{R}^n$, any $c, c' \in \mathbb{R}^d$ and any $g, m > 0$ the set of bases on the shadow path with large separation satisfies*

$$|T^{G(A,b,g)\cap M(A,c,c',m)}| \leq |H(A,b,c,c',\rho)| + \frac{\rho\cdot\theta(c,c')\cdot\max(\|c\|,\|c'\|)}{(1-\rho)\cdot gm} + 2$$

Proof. The total number of $I \in T^{G(A,b,g)\cap M(A,c,c',m)}$ that satisfy $I \in H(A,b,c,c',\rho)$ is at most $|H(A,b,c,c',\rho)|$. For that reason, we need only prove that the total number of $I \in T^{G(A,b,g)\cap M(A,c,c',m)}$ satisfying $I \notin H(A,b,c,c',\rho)$ is at most $\frac{\rho\cdot\theta(c,c')\cdot\max(\|c\|,\|c'\|)}{(1-\rho)\cdot gm}$.

Let $I \in T^{G(A,b,g)\cap M(A,c,c',m)} \setminus H(A,b,c,c',\rho)$ and $J, J' \in N(A,b,c,c',\rho)$ be arbitrary. This implies that $I \in M(A,c,c',m)$ and $J, J' \in G(A,b,g)$. Take $y \in [c,c']$ such that $yA_I^{-1} \geq m$. Write j for the unique element $I \cap J = \{j\}$. We directly compute

$$\begin{aligned} y^\top(x_I - x_J) &= (y^\top A_I^{-1})(A_I x_I - A_I x_J) \\ &\geq m\cdot(b_I - (A_I x_J))_j \\ &\geq mg\|x_J\| \\ &\geq mg\|\pi_{c,c'}(x_J)\|. \end{aligned}$$

Since $I \notin H(A,b,c,c',\rho)$ we must have at least one close-by neighbor. Without loss of generality assume it is J satisfying $\|\pi_{c,c'}(x_I - x_J)\| \leq \rho\|\pi_{c,c'}(x_I)\|$. For J we can now express that x_I and x_J are "far-apart" as measured by the inner product with y, for we must have

$$\|\pi_{c,c'}(x_J)\| \geq \|\pi_{c,c'}(x_I)\| - \|\pi_{c,c'}(x_I - x_J)\| \geq (1-\rho)\|\pi_{c,c'}(x_I)\|,$$

implying that $y^\top x_I \geq y^\top x_J + (1-\rho)\cdot mg\cdot\|\pi_{c,c'}(x_I)\|$.

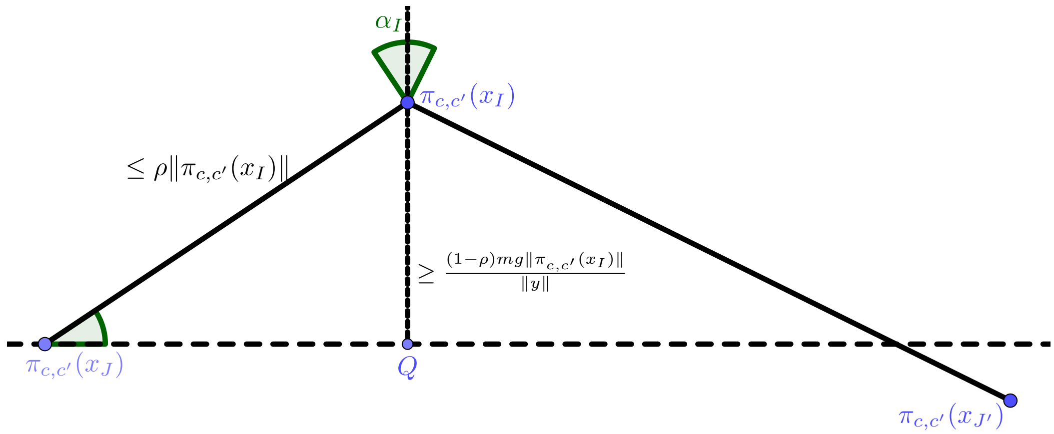

Figure 2: The plane $\text{span}(c, c')$ in Lemma 39. The vector y points straight up.

In Figure 2, we draw the points $\pi_{c,c'}(x_I), \pi_{c,c'}(x_J)$ and $\pi_{c,c'}(x_{J'})$. We find the line orthogonal to y which passes through $\pi_{c,c'}(x_J)$, and draw a single additional point Q, which is the orthogonal projection of $\pi_{c,c'}(x_I)$ onto said line. The triangle $\triangle(\pi_{c,c'}(x_J), Q, \pi_{c,c'}(x_I))$ has a right angle at Q.

Let α_I denote the exterior angle of the shadow polygon at the vertex $\pi_{c,c'}(x_I)$, also drawn in the figure. Chasing angles we find that $\alpha_I \geq \angle(\pi_{c,c'}(x_I), \pi_{c,c'}(x_J), Q)$. This latter angle we can lower bound with its sine

$$\angle(\pi_{c,c'}(x_I), \pi_{c,c'}(x_J), Q) \geq \sin(\angle(\pi_{c,c'}(x_I), \pi_{c,c'}(x_J), Q)) = \frac{\|\pi_{c,c'}(x_I) - Q\|}{\|\pi_{c,c'}(x_J) - \pi_{c,c'}(x_I)\|} \geq \frac{(1-\rho)mg}{\rho\|y\|},$$

using the lower bound on the length of the opposite edge and the upper bound on the length of the hypothenuse described in the text above, canceling the two factors of $\|\pi_{c,c'}(x_I)\|$. What we have found is that for every $I \in T^{G(A,b,g)\cap M(A,c,c',m)} \setminus H(A,b,c,c',\rho)$ we have $\alpha_I \geq \frac{(1-\rho)mg}{\rho\|y\|} \geq \frac{(1-\rho)mg}{\rho\max(\|c\|,\|c'\|)}$.

It is now that we note that, for all $I \in P(A,b,c,c')$ except the two endpoints, the exterior angles pack into the angle between the objectives. In particular, for the sum over I's currently under consideration, except the endpoints, we have

$$\sum_{\substack{I \in T^{G(A,b,g)\cap M(A,c,c',m)} \setminus H(A,b,c,c',\rho) \\ \text{not an endpoint}}} \frac{(1-\rho)mg}{\rho\max(\|c\|,\|c'\|)} \leq \sum_{\substack{I\in P(A,b,c,c') \\ \text{not an endpoint}}} \alpha_I \leq \theta(c,c').$$

Observe that the first summand does not depend on I. Therefore we can divide through $\frac{(1-\rho)mg}{\rho\max(\|c\|,\|c'\|)}$ on all sides. Accounting for the possible contributions by the endpoints, we find

$$|T^{G(A,b,g)\cap M(A,c,c',M)} \setminus H(A,b,c,c',\rho)| \leq \frac{\rho\max(\|c\|,\|c'\|)\cdot\theta(c,c')}{(1-\rho)mg} + 2.$$

$\square$

Our next immediate concern is to bound the number of bases in $H(A, b, c, c', \rho)$. For that purpose we will integrate a potential function over part of the boundary of the shadow polygon.

Definition 40. *We define the ring with inner radius r and outer radius R as $D(R, r) = R\mathbb{B}_2^2 \setminus r\mathbb{B}_2^2$.*

Lemma 41. *Let $T \subseteq \mathbb{R}^2$ be a closed convex set, and let $R > r > 0$. Then we can upper bound the following integral as follows*

$$\int_{D(R,r)\cap\partial T} \|x\|^{-1} \,\mathrm{d}\, x \leq 4\pi\lceil\log_2(R/r)\rceil.$$

Proof. To start, we define for $i = 1, \dots, l = \lceil\log_2(R/r)\rceil$ the ring $D_i := D(2^{i-1}r, 2^i r)$. Note that $\bigcup_{i=1}^{l} D_i = D(R, r)$. We break up the large integral into smaller parts as

$$\int_{D(R,r)\cap\partial T} \|x\|^{-1} \,\mathrm{d}\, x = \sum_{i=1}^{l} \int_{D_i\cap\partial T} \|x\|^{-1} \,\mathrm{d}\, x. \tag{2}$$

For each $i = 1, \dots, l$ we know that $x \in D_i$ implies an upper bound on the integrand $\|x\|^{-1} \leq \frac{1}{r2^{i-1}}$. We will now upper bound the size of the integration domain. Take any point $x \in D_i \cap \partial T$. Since x is on the boundary of the convex set T there exists a nonzero vector $y \in \mathbb{R}^2$ such that $y^\top x = \max_{x'\in T} y^\top x'$. This same vector y demonstrates that $y^\top x = \max_{x'\in D_i\cap\partial T} y^\top x'$, and hence our point is also on the boundary of the restricted set $\mathrm{conv}(D_i \cap \partial T)$. It follows that our integration domain satisfies the inclusion

$$D_i \cap \partial T \subseteq \partial\,\mathrm{conv}(D_i \cap \partial T).$$

We know that $\mathrm{conv}(D_i \cap \partial T) \subseteq 2^i r\mathbb{B}_2^2$ is convex. By the monotonicity of surface area for inclusions of convex sets we find that $\int_{\partial\,\mathrm{conv}(D_i\cap\partial T)} \mathrm{d}\, x \leq \int_{\partial 2^i r\mathbb{B}_2^2} \mathrm{d}\, x \leq 2\pi \cdot 2^i r$. Taken together, we have found that

$$\int_{D_i\cap\partial T} \|x\|^{-1} \,\mathrm{d}\, x \leq \int_{D_i\cap\partial T} r/2^{i-1} \,\mathrm{d}\, x \leq 4\pi.$$

Summing over all values of $i = 1, \dots, \lceil\log_2(R/r)\rceil$ in (2) gives the result. $\square$

With this we find an upper bound on the size of $H(A, b, c, c', \rho)$ that is independent of the objectives c, c'.

Lemma 42. *For given constraint data $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$, objectives $c, c' \in \mathbb{R}^d$ and threshold $\rho > 0$ we have for any $R > 2r > 0$ that*

$$|H(A, b, c, c', \rho)| \leq \frac{25\ln(R/r)}{\rho} + |\{I \in H(A, b, c, c', \rho) : \|\pi_{c,c'}(x_I)\| \notin [r, R]\}|.$$

Proof. Abbreviate $S = \{I \in H(A, b, c, c', \rho) : \|\pi_{c,c'}(x_I)\| \in [r, R]\}$. It is our goal to show that $|S| \leq \frac{8\pi\log_2(R/r)}{\rho}$.

For every $I \in P(A, b, c, c')$ and every $I' \in N(A, b, c, c', I)$ we abbreviate the norm $\ell_I = \|\pi_{c,c'}(x_I)\|$ and the line segment $L_{I,I'} = [\pi_{c,c'}(x_I), \pi_{c,c'}(x_{I'})]$.

Every such line segment is an edge of the shadow polygon $\pi_{c,c'}(\{x : Ax \leq b\})$. Notably, every edge is found at most twice in this manner. With that knowledge, we can upper bound the sum

$$\sum_{I\in S}\ \sum_{I'\in N(A,b,c,c',I)} \int_{D(2R,r/2)\cap L_{I,I'}} \frac{1}{\|t\|} \,\mathrm{d}\, t \leq 2\int_{D(2R,r/2)\cap\partial\pi_{c,c'}(\{x:Ax\leq b\})} \frac{1}{\|t\|} \,\mathrm{d}\, t \leq 4\pi\lceil\log_2(4R/r)\rceil.$$

For our last observation, notice that the intersection $L_{I,I'} \cap D(2\ell_I, \ell_I/2)$ contains a line segment of length at least $\ell_I/2$, and on this line segment the integrand is at least $2/\ell_I$. This implies that the integral on that line segment is at least $\int_{D(2R,r/2)\cap L_{I,I'}} \|t\|^{-1} \,\mathrm{d}\, t \geq 1$ and we can lower bound

$$\sum_{I\in S} \sum_{I'\in N(A,b,c,c',I)} \int_{D(2R,r/2)\cap L_{I,I'}} \frac{1}{\|t\|} \,\mathrm{d}\, t \geq \sum_{I\in S} \sum_{I'\in N(A,b,c,c',I)} 1 \geq |S|.$$

We thus learn that $|S| \leq 4\pi\lceil \log_2(4R/r)\rceil \leq 25\ln(R/r)$. $\square$

4.6 Theorem

We are almost ready to prove our key theorem for upper bounding the smoothed shadow size. The last remaining issue that needs resolving is that Lemma 39 depends linearly on the norm of the longest objective vector. We can build on top of it to remedy this fact.

Lemma 43. *Let $c, Z \in \mathbb{R}^d$. If $i \geq \lceil \log(\|Z\|/\|c\|)\rceil + 2$ then $\theta(2^{i-1}c + Z, 2^i c + Z) \leq \frac{5\|Z\|}{2^{i+1}\|c\|}$.*

Proof. Note that $\theta(2^{i-1}c + Z, 2^i c + Z) = \theta(2^i c + 2Z, 2^i c + Z)$ and consider the triangle $\triangle(0, 2^i c + 2Z, 2^i c + Z)$. Let us abbreviate its vertices by $a = 2^i c + 2Z$ and $b = 2^i c + Z$, so that our triangle is $\triangle(0, a, b)$. The assumption on i gives that $\|2^{i-1}c\| \geq 2\|Z\|$. By the triangle inequality we find that the edge $[a, b]$ has the shortest length of the three.

We recall the law of sines to derive

$$\frac{\sin(\theta(a,b))}{\|a-b\|} = \frac{\sin(\theta(0-a, b-a)}{\|b\|} \leq \frac{1}{\|b\|} \leq \frac{1}{2^{i-1}\|c\|}$$

This gives an upper bound on the sine of our desired angle. To relate this to the angle itself, note that the shortest edge of a triangle is opposite of the smallest angle, which gives us that $\theta(a, b) \leq \pi/3$. For any $\theta \in [0, \pi/3]$ one has $\sin(\theta) > 0.8\theta$, so in particular $\theta(2^{i-1}c + Z, 2^i c + Z) \leq \frac{5\|Z\|}{2^{i+1}\|c\|}$. $\square$

Lemma 44. *Consider constraint data $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$, linearly independent objectives $c, z \in \mathbb{R}^d$, any $g, m > 0$, and analysis parameters $R > 2r > 0$ and $k \in \mathbb{N}$. Then we find*

$$|T^{G(A,b,g)\cap M(A,z,c+2^k z,m)}| \leq 40\sqrt{\frac{\|z\|}{gm} k \log(R/r)} + |\{I \in F(A,b) : \|\pi_{c,Z}(x_I)\| \notin [r,R]\}| + 3k + 2.$$

Proof. For $y, y' \in \mathbb{R}^d$ we abbreviate $S(y, y') = T^{G(A,b,g)\cap M(A,y,y',m)}$ and we cover the set studied by smaller segments as follows

$$S(z, 2^k c + z) \subset S(z, c + z) \cup \bigcup_{i=1}^{k} S(2^{i-1}c + z, 2^i c + z).$$

For the sake of succinctness write $c^{-1} = Z$ and for each $i = 0, \dots, k$ write $c^i = 2^i c + Z$. With this notation, our task is to bound $\sum_{i=0}^{k} |S(c^{i-1}, c^i)|$. For each $i = 0, \dots, k$ we apply Lemma 39 and find, for some $\rho \in (0, 1/2]$ to be decided later, that

$$\sum_{i=0}^{k} |S(c^{i-1}, c^i)| \leq 2(k+1) + \sum_{i=0}^{k} |H(A, b, c^{i-1}, c^i, \rho)| + \frac{\rho \cdot \theta(c^{i-1}, c^i) \cdot \max(\|c^{i-1}\|, \|c^i\|)}{(1-\rho)\cdot gm} \tag{3}$$

$$\leq 2(k+1) + \sum_{i=0}^{k} |H(A, b, c^{i-1}, c^i, \rho)| + \frac{\rho \cdot \theta(c^{i-1}, c^i) \cdot (\|z\| + 2^i \|c\|)}{(1-\rho) \cdot gm}. \tag{4}$$

Observe that, as $\rho \leq 1/2$, the term $(1-\rho)$ is lower bounded by $1/2$. The sets $H(A, b, c^{i-1}, c^i, \rho)$ have pairwise overlap at most 1 by Fact 21, which gives with Lemma 42 that

$$\sum_{i=0}^{k} |H(A, b, c^{i-1}, c^i, \rho)| \leq |H(A, b, c^{-1}, c^k, \rho)| + k \tag{5}$$

$$\leq 25\rho^{-1} \ln(R/r) + |\{I \in F(A, b) : \|\pi_{z,c}(x_I)\| \notin [r, R]\}| + k. \tag{6}$$

Now for the angle terms we want to bound $\sum_{i=0}^{k} \theta(c^{i-1}, c^i) \cdot (\|z\| + 2^i \|c\|)$. Observe that by construction the angles sum as $\sum_{i=0}^{k} \theta(c^{i-1}, c^i) = \theta(c^{-1}, c^k)$ and subdivide into three parts based on a threshold $h = \lceil \log_2(\|z\|/\|c\|) \rceil + 2$ as

$$\sum_{i=0}^{k} \theta(c^{i-1}, c^i) \cdot (\|z\| + 2^i \|c\|) = \theta(c^{-1}, c^k) \cdot \|z\| + \sum_{i=1}^{k} 2^i \theta(c^{i-1}, c^i) \cdot \|c\| \tag{7}$$

$$\leq \pi \cdot \|z\| + \sum_{i=1}^{h} 2^i \theta(c^{i-1}, c^i) \cdot \|c\| + \sum_{i=h}^{k} 2^i \theta(c^{i-1}, c^i) \cdot \|c\|. \tag{8}$$

For the middle term, $i \leq h$ implies $2^i \|c\| \leq 8\|z\|$ and hence $\sum_{i=0}^{h} 2^i \theta(c^{i-1}, c^i) \cdot \|c\| \leq 8\pi k \|z\|$. For the third term we use Lemma 43 to find $\sum_{i=h}^{k} 2^i \theta(c^{i-1}, c^i) \cdot \|c\| \leq 3k\|z\|$. Taken all together we have found

$$S(z, 2^k c + z) \leq 3k + 2 + 25\rho^{-1} \ln(R/r) + |\{I \in F(A, b) : \|\pi_{z,c}(x_I)\| \notin [r, R]\}| + \frac{64\rho k \|z\|}{gm}.$$

It remains to choose $\rho \in (0, 1/2]$ so as to find the strongest upper bound, which is attained at $\rho = \sqrt{\frac{25 \ln(R/r) \cdot gm}{100k\|z\|}}$. □

We are now ready to state our key technical theorem, which we will apply with two different choices for b.

Theorem 45. *Let the matrix constraint $A \in \mathbb{R}^{n \times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma > 0$, and such that the rows of $\mathbb{E}[A]$ each have norm at most 1. Let $b \in \mathbb{R}^n, c \in \mathbb{R}^d$ be arbitrary and fixed, as well as the analysis parameters $R > 2r > 0$ and $t > 0$. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function that satisfies $\Pr[\|Z\| \geq t] \leq n^{-d}$ and is independent of A then*

$$\mathbb{E}[|P(A, b, Z, c)|] \leq O\left(\sqrt{\frac{d}{\sigma} \log(R/r) \sqrt{d^7 \log^5\left(\frac{nt}{\sigma\|c\|}\right)}}\right) + 4\mathbb{E}[|\{I \in F(A, b) : \pi_{c,Z}(x_I) \notin [r, R]\}|]$$

Proof. Abbreviating a number of expressions, we write $k = \left\lceil \log_2\left(\frac{2dt \cdot n^{2d}}{\|c\|\sigma\sqrt{2\pi}}\right)\right\rceil$ to help break up the shadow path into shorter pieces, take the bound on the multipliers as $m = \ln(1/0.99)/2d$ and on the relative slacks as $g = \frac{\sigma}{5000 d^{3/2} \ln(n)^{3/2}}$.

We have $\mathbb{E}[|P(A,b,Z,c)|] \le \mathbb{E}[|P(A,b,Z,2^k c+Z)|] + \mathbb{E}[|P(A,b,2^k c+Z,c)|]$. The second term is at most 7 by Corollary 26. For the first term we apply Theorem 37 to find

$$\mathbb{E}[|P(A,b,Z,2^k c+Z)|] \le 8 + 4\mathbb{E}[|T^{G(A,b,g)\cap M(A,Z,2^k c+Z,m)}|].$$

Through applying Lemma 44 we find that

$$\mathbb{E}[|T^{G(A,b,g)\cap M(A,Z,2^k c+Z,m)}|] \le \mathbb{E}\Big[40\sqrt{k\frac{\|Z\|}{gm}\log(R/r)}\Big] + \mathbb{E}[|\{I\in F(A,b) : \|\pi_{c,Z}(x_I)\| \notin [r,R]\}|] + 3k + 2,$$

and Jensen's inequality gives that $\mathbb{E}[\sqrt{\|Z\|}] \le \sqrt{\mathbb{E}[\|Z\|]}$. By Lemma 7, the expected norm of Z is $\mathbb{E}[\|Z\|] = d$. Taken all together we find our conclusion. ☐

4.7 Norms

In this subsection we will look at basic solutions with "very small" and "very large" norms and show that they are very unlikely to occur. Our parameter for "very large" norms is R and our parameter for "very small" norms is r. For the former we will deduce guarantees in Lemma 46 and for the latter in Lemma 47 and Lemma 48.

Lemma 46. *Let each row a_i, $i \in [n]$ of A be an independent σ^2-Gaussian random variable, and let $b \in \mathbb{R}^n$ be fixed. Then we have for any $R > 0$ that*

$$\Pr[\max_{I\in\binom{[n]}{d}} \|x_I\| \ge R\|b\|_\infty] \le \frac{2\cdot d^2 n^d}{\sigma R\sqrt{2\pi}}.$$

Proof. Let $I \subseteq [n]$ denote the index set of a subset of rows of cardinality d. Let E_I denote the event that $\operatorname{dist}\big(a_j, \operatorname{span}(a_i : i \in I\setminus\{j\})\big) \ge d/R$ holds for each $j \in I$. Note that if the matrix A_I is invertible then the column of A_I^{-1} corresponding to index $j \in I$ has norm exactly equal to $1/\operatorname{dist}\big(a_j, \operatorname{span}(a_i : i\in I\setminus\{j\})\big)$. It follows by the triangle inequality that E_I implies that x_I has norm at most $\|x_I\| \le \sum_{i\in I}\|(A_I^{-\top})_i\|\cdot|b_i| \le R\|b_I\|_\infty$. Using this implication along with a union bound we find

$$\Pr\Big[\max_{I\in\binom{[n]}{d}} \|x_I\| \ge R\|b\|_\infty\Big] \le \Pr\Big[\bigvee_{I\in\binom{[n]}{d}} \neg E_I\Big] \le \sum_{I\in\binom{[n]}{d}} \Pr[\neg E_I]. \tag{9}$$

It remains to show that $\Pr[\neg E_I] \le \frac{2d^2}{\sigma R\sqrt{2\pi}}$ for all $I \in \binom{[n]}{d}$. Using another union bound, it suffices if we show for each $j \in I$ that

$$\Pr\Big[\operatorname{dist}\big(a_j, \operatorname{span}(a_i : i\in I\setminus\{j\})\big) \le d/R\Big] \le \frac{2d}{\sigma R\sqrt{2\pi}}.$$

We take the linear subspace $V = \operatorname{span}(a_i : i \in I\setminus\{j\})$ to be fixed, and take $y \in V^\perp \cap \mathbb{S}^{d-1}$ to be any fixed unit normal vector. Using this notation we can write

$$\operatorname{dist}\big(a_j, V\big) = |y^\top a_j|.$$

Note that V is defined only using a_i with $i \in I\setminus\{j\}$, and in particular that y is independent of a_j. That means that, after conditioning on the values of a_i for $i \in I\setminus\{j\}$, the signed distance $y^\top a_j$ is

Gaussian distributed with mean $y^\top \mathbb{E}[a_j]$ and standard deviation σ. The distance can only be small if $y^\top a_j \in (-d/R, d/R)$ and hence we find

$$\begin{aligned}
\Pr\left[\operatorname{dist}\left(a_j, V\right) \le d/R\right] &= \Pr\left[|y^\top a_j| \le d/R\right] \\
&= \Pr\left[y^\top a_j \in [-d/R, d/R]\right] \\
&\le \frac{2d}{R} \cdot \frac{1}{\sigma\sqrt{2\pi}}, \qquad (10)
\end{aligned}$$

using the fact that the probability density function of $y^\top a_j$ is uniformly upper bounded by $1/\sigma\sqrt{2\pi}$. Combining (9) with the union bound over all $j \in I$ and (10) proves the lemma. $\square$

Lemma 47 (No small norms)**.** *Let the rows a_i, $i \in [n]$ of A have independent Gaussian distributed entries with expectations of norm $\|\mathbb{E}[a_i]\| \le 1$ for $i = 1, \dots, n$, each with standard deviation σ. Let $b \in \mathbb{R}^n$ be arbitrary subject to $|b_i| \ge \varepsilon$ for all $i \in [n]$. Then we have*

$$\Pr[\min_{I \in \binom{[n]}{d}} \|x_I\| < \varepsilon/2] \le n^{-d}.$$

Proof. We first show that if $\|a_i\| \le 2$ for all $i \in [n]$ then it follows that $\|x_I\| \ge \varepsilon/2$ for every basis $I \in \binom{[n]}{d}$.

Assume that $\|a_i\| \le 2$ for some $i \in [n]$. Then for any $x \in \mathbb{R}^d$ with $\|x\| < \varepsilon/2$ it follows that $a_i^\top x \le \|a_i\| \cdot \|x\| < \varepsilon \ne b_i$. In particular this implies that x cannot be obtained as $A_I^{-1} b_I$ for any $I \in \binom{[n]}{d}$ with $i \in I$. Thus if $\|a_i\| \le 2$ then any basic solution x_I for $I \in \binom{[n]}{d}$ must satisfy $\|x_I\| \ge \varepsilon/2$.

For $a_1, \dots, a_n$ we call on Corollary 11 to find that

$$\Pr[\exists i \in [n] : \|a_i\| > 2] \le \Pr[\exists i \in [n] : \|a_i - \mathbb{E}[a_i]\| > 4\sigma\sqrt{d \log n}] \le n^{-d}.$$

We have thus found that $\Pr[\min_{I \in \binom{[n]}{d}} \|x_I\| < \varepsilon/2] \le \Pr[\exists i \in [n] : \|a_i\| > 2] \le n^{-d}$ as required. $\square$

Lemma 48. *Let the rows a_i, $i \in [n]$ of A have independent Gaussian distributed entries with expectations of norm $\|\mathbb{E}[a_i]\| \le 1$ for $i = 1, \dots, n$, each with standard deviation σ. Let $c \in \mathbb{R}^d \setminus \{0\}$ and $b \in \mathbb{R}^n$ be fixed subject to $|b_i| > \varepsilon$ for every $i \in [n]$ and let $Z \in \mathbb{R}^d$ be distributed independently from A, b and rotationally symmetric. Then we have, for $\alpha, \varepsilon > 0$, that*

$$\Pr\left[\min_{I \in \binom{[n]}{d}} \|\pi_{\operatorname{span}(c,Z)}(x_I)\| < \frac{\alpha \cdot \varepsilon}{2}\right] \le n^{-d} + \alpha n^d \cdot \sqrt{de}.$$

Proof. We start with a simple bound, writing

$$\min_{I \in \binom{[n]}{d}} \|\pi_{\operatorname{span}(c,Z)}(x_I)\| \ge \min_{I \in \binom{[n]}{d}} \|x_I\| \cdot \min_{I' \in \binom{[n]}{d}} \frac{\|\pi_{\operatorname{span}(c,Z)}(x_{I'})\|}{\|x_{I'}\|}.$$

Thus, if $\min_{I \in \binom{[n]}{d}} \|\pi_{\operatorname{span}(c,Z)}(x_I)\| < \alpha \cdot \varepsilon/2$ is small then necessarily we need at least one of $\min_{I \in \binom{[n]}{d}} \|x_I\| < \varepsilon/2$ or $\min_{I' \in \binom{[n]}{d}} \frac{\|\pi_{\operatorname{span}(c,Z)}(x_{I'})\|}{\|x_{I'}\|} < \alpha$ to be small. A union bound over these two events gives us

$$\Pr[\min_{I \in \binom{[n]}{d}} \|\pi_{\operatorname{span}(c,Z)}(x_I)\| \le \alpha \cdot \varepsilon] \le \Pr[\min_{I \in \binom{[n]}{d}} \|x_I\| \le \varepsilon] + \Pr[\min_{I' \in \binom{[n]}{d}} \frac{\|\pi_{\operatorname{span}(c,Z)}(x_{I'})\|}{\|x_{I'}\|} \le \alpha].$$

As we have proven in Lemma 47, we have $\Pr[\min_{I\in\binom{[n]}{d}}\|x_I\| \le \varepsilon/2] \le n^{-d}$ for the first summand. It remains to upper bound the second summand. For this, we start by observing that for any $I \in \binom{[n]}{d}$ we have $\frac{\|\pi_{\mathrm{span}(c,Z)}(x_I)\|}{\|x_I\|} \ge \frac{|Z^\top x_I|}{\|Z\|\cdot\|x_I\|}$. This inequality implies that if the former quantity is small then the second quantity must be small. This in turn results in the inequality $\Pr[\min_{I\in\binom{[n]}{d}} \frac{\|\pi_{\mathrm{span}(c,Z)}(x_I)\|}{\|x_I\|} \le \alpha] \le \Pr[\min_{I\in\binom{[n]}{d}} \frac{|Z^\top x_I|}{\|Z\|\cdot\|x_I\|} \le \alpha]$. To upper bound this last probability, we observe that for each $I \in \binom{[n]}{d}$ the fraction $\frac{Z^\top x_I}{\|Z\|\cdot\|x_I\|}$ has a distribution identical to the inner product $\theta^\top e_1$ between a uniformly random unit vector $\theta \in \mathbb{S}^{d-1}$ and an arbitrarily chosen standard basis vector. Taking a union bound over all $|\binom{[n]}{d}| \le n^d$ choices of I, we bound

$$\begin{aligned}\Pr[\min_{I\in\binom{[n]}{d}} \frac{|Z^\top x_I|}{\|Z\|\cdot\|x_I\|} \le \alpha] &\le \sum_{I\in\binom{[n]}{d}} \Pr[\frac{|Z^\top x_I|}{\|Z\|\cdot\|x_I\|} \le \alpha] \\ &\le n^d \cdot \Pr[|\theta^\top e_1| \le \alpha].\end{aligned}$$

Using Theorem 13 to upper bound $\Pr[|\theta^\top e_1| \le \alpha] \le \alpha\sqrt{de}$ we obtain the result. □

4.8 Conclusion

We require the semi-random shadow bound for two cases, either when the entries of $b \in \mathbb{R}^n$ are all fixed to 1, or when the entries of b are Gaussian distributed. For the former we will present Theorem 49 and for the latter Theorem 50.

Theorem 49. *Let the constraint matrix $A \in \mathbb{R}^{n\times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma > 0$ and such that the rows of $\mathbb{E}[A]$ each have norm at most 1. Let the right hand side vector b be fixed to be 1. Let $c \in \mathbb{R}^d$ be arbitrary and fixed, as well as the analysis parameters $R > 2r > 0$. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function that satisfies $\Pr[\|Z\| \ge 2ed\ln(n)] \le n^{-d}$ and is independent of A, then the semi-random shadow path on $\{x : Ax \le 1\}$ has length bounded as*

$$|P(A, 1, Z, c)| \le O\left(\sqrt{\frac{d}{\sigma}\sqrt{d^9 \log^7\left(\frac{nd\ln(n)}{\sigma\|c\|}\right)}}\right).$$

Proof. Choose $R = 2d^2n^{2d}/\sigma\sqrt{2\pi}$ and $r = (2n^d \cdot \sqrt{de})^{-1}$. We apply Theorem 45 and with these values, we obtain $\log(R/r) \le O(d\log(n/\sigma)) \le O(d\log(nd\ln(n)/\sigma\|c\|))$. It remains to upper bound

$$\mathbb{E}[|\{I \in F(A, b) : \|\pi_{c,Z}(x_I)\| > R\}|] + \mathbb{E}[|\{I \in F(A, b) : \|\pi_{c,Z}(x_I)\| < r\}|].$$

Using Lemma 46 we get $\mathbb{E}[|\{I \in F(A, b) : \|\pi_{c,Z}(x_I)\| > R\}|] \le n^d \Pr[\max_{I\in\binom{[n]}{d}}\|x_I\| > R] \le 1$. To bound the expected number of bases with small projected norms, we start similarly by

$$\mathbb{E}[|\{I \in F(A, b) : \|\pi_{c,Z}(x_I)\| < r\}|] \le n^d \Pr[\max_{I\in\binom{[n]}{d}} \|\pi_{c,Z}(x_I)\| < r].$$

We apply Lemma 48 with $\varepsilon = 1$ and $\alpha = (n^{2d} \cdot \sqrt{de})^{-1}$ to get $\Pr[\min_{I\in\binom{[n]}{d}}\|\pi_{c,Z}(x_I)\| < r] \le 2n^{-d}$, which finishes the argument since now $\mathbb{E}[|\{I \in F(A, b) : \pi_{c,Z}(x_I) \notin [r, R]\}|] \le 3$. □

Theorem 50. *Let the constraint matrix $A \in \mathbb{R}^{n\times d}$ have independent Gaussian distributed entries, as well as the vector $b \in \mathbb{R}^n$, each with standard deviation $\sigma > 0$, and such that the rows of $\mathbb{E}[(A,b)]$ each have norm at most 1. Let $c \in \mathbb{R}^d$ be arbitrary and fixed, as well as the analysis parameters $R > 2r > 0$. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function that satisfies $\Pr[\|Z\| \geq 2ed\ln(n)] \leq n^{-d}$ and is independent of A, then the semi-random shadow path on $\{x : Ax \leq b\}$ has length bounded as*

$$|P(A,b,Z,c)| \leq O\left(\sqrt{\frac{d}{\sigma}\sqrt{d^9 \log^7\left(\frac{nd\ln(n)}{\sigma\|c\|}\right)}}\right).$$

Proof. Choose $R = 2d^2n^{2d}/\sigma\sqrt{2\pi}$ as before, giving

$$\mathbb{E}[|\{I \in F(A,b) : \|\pi_{c,Z}(x_I)\| > R\}|] \leq n^d \Pr[\max_{I\in\binom{[n]}{d}} \|x_I\| > R] \leq 1.$$

For this proof we pick $r = \frac{\sigma\sqrt{2\pi}}{n^{3d}\sqrt{de}}$. We apply Theorem 45 and with these values, getting $\log(R/r) \leq O(d\log(n/\sigma)) \leq O(d\log(nd\ln(n)/\sigma\|c\|))$. We are thus left with bounding $\mathbb{E}[|\{I \in F(A,b) : \|\pi_{c,Z}(x_I)\| < r\}|]$. Take $\varepsilon = \sigma\sqrt{2\pi}n^{-d}$ and $\alpha = (n^{2d}\cdot\sqrt{de})^{-1}$. We separately treat the scenario where there exists $i \in [n]$ with $|b_i| < \varepsilon$ and the scenario where for all $i \in [n]$ it holds that $|b_i| \geq \varepsilon$.

In the first scenario we count at most $\binom{n}{d}$ bases I with $\|\pi_{c,Z}(x_I)\| < r$, and this scenario occurs with probability $\Pr[\exists i \in [n], |b_i| < \varepsilon] \leq \frac{2\varepsilon}{\sigma\sqrt{2\pi}} < n^{-d}$. Thus this scenario contributes at most 1 to the expectation.

For the second scenario we apply Lemma 48 to learn that $\Pr[\min_{I\in\binom{[n]}{d}} \|\pi_{c,Z}(x_I)\| < r] \leq 2n^{-d}$, finding that this contributes at most 2 to the expectation. This suffices for the theorem. $\square$

5 Lower bound

In this section we will demonstrate that the exponent for σ in the shadow bound proved in the previous section cannot be further improved without significantly worsening the dependence on n.

Definition 51. *For $\eta > 0$ and $d \in \mathbb{N}$, a set $S \subset \mathbb{S}^{d-1}$ is called η-dense if for any $x \in \mathbb{S}^{d-1}$ there exists $s \in S$ such that $\|x - s\| \leq \eta$.*

Dense sets have been previously studied, and in particular there are known bounds on their size for greedy constructions.

Lemma 52 (See, e.g., [Mat02] p.314)**.** *There exists an η-dense set $S \subset \mathbb{S}^{d-1}$ with cardinality $|S| \leq (4/\eta)^d$.*

For our unperturbed constraint data we will use a matrix whose rows form an η-dense set. This will result in a feasible set which is "close to the unit ball".

Lemma 53. *Let $S \subset \mathbb{S}^{d-1}$ be η-dense, $\eta \leq 1/8$, and let $A \in \mathbb{R}^{n\times d}$ be a matrix with rows $a_1, \ldots, a_n$. Assume that for every $i \in [n]$ there exists an $s \in S$ with $\|a_i - s\| \leq \eta$. Given a vector $b \in [1-\eta, 1+\eta]^n$, the set $Q := \{x \in \mathbb{R}^d : Ax \leq b\}$ satisfies*

$$(1-2\eta)\mathbb{B}_2^d \subseteq Q \subseteq (1+4\eta)\mathbb{B}_2^d.$$

Proof. Suppose $x \in \mathbb{R}^d$ satisfies $\|x\| \le 1 - 2\eta$. Consider any $i \in [n]$ and find some $s \in S$ such that $\|a_i - s\| \le \eta$. By the triangle inequality we find $\|a_i\| \le \|s\| + \|a_i - s\| \le 1 + \eta$. By the Cauchy-Schwarz inequality we find

$$a_i^\top x \le \|a_i\| \cdot \|x\| \le (1+\eta)(1-2\eta) = 1 - \eta - 2\eta^2 \le b_i.$$

Since this inequality $a_i^\top x \le b_i$ holds for all $i \in [n]$ we conclude that any $x \in \mathbb{R}^d$ with $\|x\| \le 1 - 2\eta$ satisfies $x \in Q$.

Now suppose $x \in \mathbb{R}^d$ satisfies $\|x\| > 1 + 4\eta$. By construction of S there exists an $s \in S$ such that $\|\frac{x}{\|x\|} - s\| \le \eta$, and by assumption on A there exists $i \in [n]$ such that $\|a_i - s\| \le \eta$. By the triangle inequality we know that $\|\frac{x}{\|x\|} - a_i\| \le 2\eta$. We use the Cauchy-Schwartz inequality to find

$$\begin{aligned}
a_i^\top x &= \|x\| - \left(\frac{x}{\|x\|} - a_i\right)^\top x \\
&\ge (1-2\eta)\|x\| \\
&> (1-2\eta)(1+4\eta) \\
&= 1 + 2\eta - 8\eta^2.
\end{aligned}$$

Assuming that $\eta \le 1/8$ gives us $a_i^\top x > 1 + \eta \ge b_i$, implying $x \notin Q$. □

With these lemmas in place, we can prove our lower bound on the length of the shadow path between two random objectives.

Theorem 54. *Given $d \ge 2$ and $0 < \sigma \le \frac{1}{32\sqrt{d \ln n}}$, there exist $n \le (4/\sigma)^d$, $\bar{A} \in \mathbb{R}^{n \times d}$, and $\bar{b} \in \mathbb{R}^n$ such that the following holds. The rows of the combined matrix (A, b) each have norm at most 1. If A, b have their entries independently Gaussian distributed with variance σ^2 and expectation $\mathbb{E}[A] = \bar{A}, \mathbb{E}[b] = \bar{b}$, and if $Z, Z' \in \mathbb{R}^d$ are independently distributed uniformly on the unit sphere, then the expected shadow path length satisfies*

$$\mathbb{E}[|P(A, b, Z, Z')|] \ge \frac{1}{32\sqrt{\sigma\sqrt{d \ln n}}}.$$

Proof. We pick $S \subset \mathbb{S}^{d-1}$ to be σ-dense with $|S| \le (4/\sigma)^d$ as demonstrated by Lemma 52. We set $n = |S|$ and let $\bar{A} \in \mathbb{R}^{n \times d}$ be formed by having the elements of S as its rows. We set $\bar{b} \in \mathbb{R}^n$ to be the all-ones vector, sample $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$ with Gaussian distributed entries as specified in this theorem's statement and write

$$Q = \{x \in \mathbb{R}^d : Ax \le b\}.$$

Let us consider the 4 paths of the form $P(A, b, \pm Z, \pm Z')$. Note that these 4 sets are all identically distributed. Their union gives the set of all feasible bases $I \in F(A, b)$ such that $yA_I^{-1} \ge 0$ for some $y \in \mathrm{span}(Z, Z') \setminus \{0\}$. Thus their combined size is at least the number of vertices of $\pi_{Z,Z'}(Q)$, where $\pi_{Z,Z'}$ denotes the orthogonal projection onto the linear subspace $\mathrm{span}(Z, Z')$. As such, the expectation $\mathbb{E}[|P(A, b, Z, Z')|]$ is at least 1/4 times the expected number of vertices of $\pi_{Z,Z'}(Q)$.

We will show that $\pi_{Z,Z'}(Q)$ has at least $\frac{\pi}{16\sqrt{\sigma\sqrt{d \ln n}}}$ vertices with probability $1 - n^{-d} \ge 2/3$.

Using the Gaussian tail bound Lemma 10 we get that, with probability at least $1 - n^{-d}$, the rows of $\bar{A} - A$ all have norm at most $4\sigma\sqrt{d \ln n}$ and also $\|\bar{b} - b\|_\infty \le 4\sigma\sqrt{d \ln n}$. In the remainder of this proof we will show that these conditions imply the desired lower bound on the number of vertices of $\pi_{Z,Z'}(Q)$.

Note that the fact that S is σ-dense implies that it is $4\sigma\sqrt{d\ln n}$-dense, where $4\sigma\sqrt{d\ln n} \leq 1/8$ by assumption on σ. Abbreviate $\eta = 4\sigma\sqrt{d\ln n}$. We apply Lemma 53 to the perturbed data A, b and find

$$(1-2\eta)\mathbb{B}_2^d \subseteq Q \subseteq (1+4\eta)\mathbb{B}_2^d.$$

In particular, it follows that the projection satisfies

$$(1-2\eta)\pi_{Z,Z'}(\mathbb{B}_2^d) \subseteq \pi_{Z,Z'}(Q) \subseteq (1+4\eta)\pi_{Z,Z'}(\mathbb{B}_2^d). \tag{11}$$

In words, this says that the polygon $\pi_{Z,Z'}(Q)$ contains a circular disc of radius $1-2\eta$ and is contained in a circular disc of radius $1+4\eta$. We now show that $\pi_{Z,Z'}(Q)$ must have many edges.

Let $[p,q] \subset \pi_{Z,Z'}(Q)$ be any edge of the polygon, and let $y \in [p,q]$ be the minimum-norm point inside this edge. The optimality condition implies that $y^\top p \geq \|y\|^2$, which gives us

$$\|y-p\|^2 = (y-p)^\top(y-p) = \|y\|^2 + \|p\|^2 - 2y^\top p \leq \|p\|^2 - \|y\|^2.$$

From (11) we know that $\|p\| \leq 1+4\eta$ and $\|y\| \geq 1-2\eta$, resulting in

$$\|p\|^2 - \|y\|^2 \leq (1+8\eta+16\eta^2) - (1-4\eta+4\eta^2) \leq 12\eta + 12\eta^2 \leq 16\eta.$$

We thus found that $\|y-p\| \leq 4\sqrt{\eta}$. A nearly identical calculation gives $\|y-q\| \leq 4\sqrt{\eta}$, and hence the edge has length $\|p-q\| \leq 8\sqrt{\eta}$. This holds for every edge of $\pi_{Z,Z'}(Q)$. The perimeter of $\pi_{Z,Z'}(Q)$ is at least $2\pi \cdot (1-2\eta) \geq \pi$. We conclude that the polygon $\pi_{Z,Z'}(Q)$ must have at least $\frac{\pi}{8\sqrt{\eta}}$ edges, assuming that the rows of $\bar{A} - A$ all have norm at most $4\sigma\sqrt{d\ln n}$ and also $\|\bar{b} - b\|_\infty \leq 4\sigma\sqrt{d\ln n}$. Since this happens with probability at least $1 - n^{-d} \geq 1/2$ we find that the expected number of edges of $\pi_{Z,Z'}(Q)$ is at least $\frac{\pi}{16\sqrt{4\sigma\sqrt{d\ln n}}}$.

Wrapping up, we find that the expected cardinality of $P(A,b,Z,Z')$ is at least $\frac{\pi\cdot(1-n^{-d})}{64\sqrt{\sigma\sqrt{d\ln n}}}$. We get the lemma's statement by noting that $\pi \cdot (1-n^{-d}) \geq 2$. $\square$

Appendices

A Proofs From Elsewhere

The following is a verbatim reproduction of a necessary theorem due to [BBHK25]. It is reproduced here in its entirety for the ease of reviewing, since the manuscript [BBHK25] is not yet publically available at the time of this manuscript's submission.

Theorem 55 (Repeated Theorem 27)**.** *Let $B \in \mathbb{R}^{d\times d}$ be an invertible matrix, every whose column has Euclidean norm at most 2, and define, for any $m > 0$, $C_m = \{x \in \mathbb{R}^d : B^{-1}x \geq m\}$. Suppose $c, c' \in \mathbb{R}^d$ are fixed. Let $Z \in \mathbb{R}^d$ be a random vector with 1-log-Lipschitz probability density μ. Then*

$$\Pr\left[[c+Z, c'+Z] \cap C_m \neq \emptyset\right] \geq 0.99 \Pr\left[[c+Z, c'+Z] \cap C_0 \neq \emptyset\right]$$

for $m = \ln(1/0.99)/2d$.

Proof. Write g_i for the i'th column of B and $p = \sum_{i=1}^{d} g_i$. We have $\|p\| \leq 2d$. Suppose we have a vector $z \in \mathbb{R}^d$ such that $[c+z, c'+z] \cap C_0 \neq \emptyset$. Let $\lambda \in [0,1]$ be the minimum number such that $z + c + \lambda(c' - c) \in C_0$. Then we define, for $\alpha = -\frac{\ln(0.99)}{2d}$, the function

$$f(z) := z + \alpha p.$$

Reusing the previous multiplier, we find the coordinatewise inequality of vectors

$$B^{-1}(f(z) + c + \lambda(c' - c)) = B^{-1}(z + c + \lambda(c' - c)) + \alpha B^{-1} p \geq \alpha. \tag{12}$$

Hence we find that $[c + f(z), c' + f(z)] \cap C_\alpha \neq \emptyset$. Note that neither f nor the non-emptiness of this intersection depends on the value of λ.

Thus we have, for a constant γ depending only on d and and the distribution μ, that

$$\begin{aligned} \Pr[[c+Z, c'+Z] \cap C_0 \neq \emptyset] &= \Pr[Z \in -[c,c'] + C_0] \\ &= \gamma \int_{-[c,c']+C_0} \mu(z) \,\mathrm{d}\, z. \end{aligned}$$

Since the map f is a translation independent of Z, it is volume-preserving and because of (12), it follows that

$$\begin{aligned} \Pr[[c+Z, c'+Z] \cap C_0 \neq \emptyset] &= \gamma \int_{-[c,c']+C_0} \mu(z) \,\mathrm{d}\, z \\ &= \gamma \int_{-[c,c']+C_\alpha} \mu(f(z)) \,\mathrm{d}\, z \\ &\leq \gamma \int_{-[c,c']+C_\alpha} \mu(z) \cdot e^{\|\alpha p\|} \,\mathrm{d}\, z \\ &\leq \gamma e^{2\alpha d} \int_{-[c,c']+C_\alpha} \mu(z) \,\mathrm{d}\, z \\ &= e^{2\alpha d} \Pr[Z \in -[c,c'] + C_\alpha] \\ &= e^{2\alpha d} \Pr[[c+Z, c'+Z] \cap C_\alpha \neq \emptyset] \end{aligned}$$

Thus, it follows that

$$\Pr[[c+Z, c'+Z] \cap C_\alpha \neq \emptyset] \geq e^{-2\alpha d} \Pr\left[[c+Z, c'+Z] \cap C_0 \neq \emptyset\right].$$

$\square$

B Additional Proofs

Proof of Lemma 7. Integrating in polar coordinates gives us the normalizing constant

$$\begin{aligned} \int_{\mathbb{R}^d} e^{-\|x\|} \,\mathrm{d}\, x &= \int_0^\infty \mathrm{vol}_{d-1}(r\mathbb{S}^{d-1}) e^{-r} \,\mathrm{d}\, r \\ &= \mathrm{vol}_{d-1}(\mathbb{S}^{d-1}) \int_0^\infty r^{d-1} e^{-r} \,\mathrm{d}\, r \end{aligned}$$

$$= \mathrm{vol}_{d-1}(\mathbb{S}^{d-1}) \cdot (d-1)!,$$

using the Gamma function. We can obtain the moments of $\|X\|$ using a similar calculation:

$$\begin{aligned}\int_{\mathbb{R}^d} \|x\|^k e^{-\|x\|} \,\mathrm{d}\,x &= \int_0^\infty \mathrm{vol}_{d-1}(r\mathbb{S}^{d-1}) r^k e^{-r} \,\mathrm{d}\,r \\ &= \mathrm{vol}_{d-1}(\mathbb{S}^{d-1}) \int_0^\infty r^{k+d-1} e^{-r} \,\mathrm{d}\,r \\ &= \mathrm{vol}_{d-1}(\mathbb{S}^{d-1}) \cdot (k+d-1)!.\end{aligned}$$

Dividing $\frac{\mathrm{vol}_{d-1}(\mathbb{S}^{d-1})\cdot(k+d-1)!}{\mathrm{vol}_{d-1}(\mathbb{S}^{d-1})\cdot(d-1)!} = \frac{(k+d-1)!}{(d-1)!}$ gives the result. $\square$

Proof of Lemma 8. Using Markov's inequality we know for $k = d\ln n$ and $t = 2ed\ln n$ that

$$\begin{aligned}\Pr[\|X\| > t] &= \Pr[\|X\|^k > t^k] \\ &= \frac{\mathbb{E}[\|X\|^k]}{t^k} \\ &\leq \frac{(k+d)^k}{t^k} \\ &\leq \frac{(2d\ln n)^{d\ln n}}{(2ed\ln n)^{d\ln n}} = n^{-d}.\end{aligned}$$

$\square$

Proof of Theorem 13. Assume that without loss of generality the fixed and arbitrary unit vector is e_1. We notice that for any $\alpha > 0$ the probability $\Pr[|\theta^\top e_1| \leq \alpha]$ is given as the ratio between the volume of the unit sphere $\mathbb{S}^{d-1}$ intersected with the half-spaces $H \coloneqq \{x \in \mathbb{R}^d \colon x_1 \leq \alpha\}$ and $\{x \in \mathbb{R}^d \colon x_1 \geq -\alpha\}$ and the volume of the unit sphere $\mathbb{S}^{d-1}$ itself. Further, we notice that the volume of $\mathbb{S}^{d-1}$ can be computed as

$$\mathrm{vol}(\mathbb{S}^{d-1}) = \int_{-1}^1 \mathrm{vol}_{d-2}((\sqrt{1-s^2})\mathbb{S}^{d-2})\sqrt{1+\left(\frac{-2s}{2\sqrt{1-s^2}}\right)^2}\,\mathrm{d}\,s = \mathrm{vol}_{d-2}(\mathbb{S}^{d-2})\int_{-1}^1 \sqrt{1-s^2}^{d-3}\,\mathrm{d}\,s$$

We notice that the factor part $\left(\frac{-2s}{2\sqrt{1-s^2}}\right)^2$ in the first equality is the derivative of the radius of the sphere $(\sqrt{1-s^2})\mathbb{S}^{d-2}$. Hence, we can write the probability that the first coordinate of θ is at most α as

$$\Pr[|\theta^\top e_1| \leq \alpha] = \frac{\mathrm{vol}_{d-2}(\mathbb{S}^{d-2})\int_{-\alpha}^{\alpha}\sqrt{1-s^2}^{d-3}\,\mathrm{d}\,s}{\mathrm{vol}_{d-2}(\mathbb{S}^{d-2})\int_{-1}^{1}\sqrt{1-s^2}^{d-3}\,\mathrm{d}\,s} \leq \frac{\int_{-\alpha}^{\alpha}\sqrt{1-s^2}^{d-3}\,\mathrm{d}\,s}{\int_{-1/\sqrt{d}}^{1/\sqrt{d}}.\sqrt{1-s^2}^{d-3}\,\mathrm{d}\,s}.$$

We will upper bound the integrant of the nominator by 1. If $s \in [-1/\sqrt{d}, 1/\sqrt{d}]$, then one calculates that $\sqrt{1-s^2}^{d-3} \in [1/\sqrt{e}, 1]$. We use this for upper bounding the denominator as

$$\Pr[|\theta^\top e_1| \leq \alpha] \leq \frac{\int_{-\alpha}^{\alpha} 1\,\mathrm{d}\,s}{(2/\sqrt{d})\cdot(1/\sqrt{e})} = \alpha\sqrt{de}$$

and find the desired bound. $\square$

Notation Index

$I \in \binom{[n]}{d}$	indexes a basis
A_I	submatrix of A induced by basis I
b_I	rows of b indexed by I
x_I	basic solution $x_I = A_I^{-1} b_I$
$y \in [c, c']$	intermediate objective
$F(A, b) \subseteq \binom{[n]}{d}$	set of feasible bases
$P(A, b, c, c') \subseteq F(A, b)$	bases on the shadow path from c to c'
$G(A, b, g) \subseteq F(A, b)$	set of bases I with $a_j^\top x_I \leq b_j - g\|x_I\|$ for all $j \notin I$
$M(A, c, c', m) \subseteq \binom{[n]}{d}$	bases I s.t. $\exists\ y \in [c, c']$ with $A_I^{-1} y \geq m$
$N(A, b, c, c', I) \subset P(A, b, c, c')$	neighbours of $I \in P(A, b, c, c')$ on the shadow path
$L(A, b, c, c', \rho) \subseteq P(A, b, c, c')$	bases I s.t. $\frac{\|\pi_{c,c'}(x_I - x_{I'})\|}{\|\pi_{c,c'}(x_I)\|} \geq \rho$ for $I' \in N(A, b, c, c', I)$
$T^S \subseteq V$	graph nodes in $S \subseteq V$ who have 2 neighbours in S
$\theta(s, s') \in [0, \pi]$	angle, the unique number such that $\cos(\theta(s, s')) \cdot \|s\| \cdot \|s'\| = s^\top s'$
$\theta(S, S') := \inf_{s \in S, s' \in S'} \theta(s, s')$	minimum angle between two sets of points $S, S' \subseteq \mathbb{R}^d$

References

[AC78] David Avis and Vasek Chvátal. Notes on bland's pivoting rule. In *Polyhedral Combinatorics*, pages 24–34. Springer, 1978.

[AKS87] Ilan Adler, Richard M. Karp, and Ron Shamir. A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps. *J. Complexity*, 3(4):372–387, 1987.

[AM85] Ilan Adler and Nimrod Megiddo. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM (JACM)*, 32(4):871–895, 1985.

[And04] Neculai Andrei. On the complexity of MINOS package for linear programming. *Studies in Informatics and Control*, 13(1):35–46, 2004.

[AZ98] Nina Amenta and Günter M. Ziegler. Deformed products and maximal shadows. *Contemporary Math.*, 223:57–90, 1998.

[BBHK25] Eleon Phil Bach, Alexander Black, Sophie Huiberts, and Sean Kafer. Analyzing the simplex method with assumptions from the user manual, 2025+.

[BC24] Alexander E. Black and Francisco Criado. Random shadows of fixed polytopes, 2024.

[BDG+22] Gilles Bonnet, Daniel Dadush, Uri Grupel, Sophie Huiberts, and Galyna Livshyts. Asymptotic bounds on the combinatorial diameter of random polytopes. In Xavier Goaoc and Michael Kerber, editors, *38th International Symposium on Computational Geometry, SoCG 2022, June 7-10, 2022, Berlin, Germany*, volume 224 of *LIPIcs*, pages 18:1–18:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[BDLKS21] Alexander Black, Jesús De Loera, Sean Kafer, and Laura Sanità. On the simplex method for 0/1 polytopes, 2021.

[Bla23] Alexander E. Black. Small shadows of lattice polytopes, January 2023.

[Bor77] Karl Heinz Borgwardt. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung.* PhD thesis, Universität Kaiserslautern, 1977.

[Bor82] Karl-Heinz Borgwardt. The average number of pivot steps required by the simplex-method is polynomial. *Zeitschrift für Operations Research*, 26(1):157–177, 1982.

[Bor87] Karl Heinz Borgwardt. *The Simplex Method: A Probabilistic Analysis*, volume 1. Springer-Verlag, Berlin, 1987.

[Bor99] Karl Heinz Borgwardt. A sharp upper bound for the expected number of shadow vertices in lp-polyhedra under orthogonal projection on two-dimensional planes. *Mathematics of Operations Research*, 24(3):544–603, 1999.

[BR13] Tobias Brunsch and Heiko Röglin. Finding short paths on polytopes by the shadow vertex algorithm. In *Automata, Languages, and Programming*, ICALP '13, pages 279–290, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[Dan51] George B Dantzig. Maximization of a linear function of variables subject to linear inequalities. *Cowles Commission Monograph 13: Activity analysis of production and allocation*, pages 339–347, 1951.

[Dan63] George Dantzig. *Linear Programming and Extensions.* Princeton University Press, December 1963.

[DFH22] Yann Disser, Oliver Friedmann, and Alexander V. Hopp. An exponential lower bound for Zadeh's pivot rule. *Mathematical Programming*, July 2022.

[DH16] Daniel Dadush and Nicolai Hähnle. On the shadow simplex method for curved polyhedra. *Discrete Comput. Geom.*, 56(4):882–909, 2016.

[DH20] Daniel Dadush and Sophie Huiberts. A friendly smoothed analysis of the simplex method. *SIAM Journal on Computing*, 49(5):STOC18–449–499, January 2020.

[DM23] Yann Disser and Nils Mosis. A Unified Worst Case for Classical Simplex and Policy Iteration Pivot Rules. In Satoru Iwata and Naonori Kakimura, editors, *34th International Symposium on Algorithms and Computation (ISAAC 2023)*, volume 283 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 27:1–27:17, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[DS05] Amit Deshpande and Daniel A Spielman. Improved smoothed analysis of the shadow vertex simplex method. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 349–356, 2005.

[FG92] John J. Forrest and Donald Goldfarb. Steepest-edge simplex algorithms for linear programming. *Mathematical Programming*, 57(3):341–374, 1992.

[FHZ11] Oliver Friedmann, Thomas Dueholm Hansen, and Uri Zwick. Subexponential lower bounds for randomized pivoting rules for the simplex algorithm. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 283–292, 2011.

[Fri11] Oliver Friedmann. A subexponential lower bound for Zadeh's pivoting rule for solving linear programs and games. In Oktay Günlük and Gerhard J. Woeginger, editors, *Proceedings of 15th International Conference on Integer Programming and Combinatorial Optimization*, pages 192–206. Springer, 2011.

[Gol76] Donald Goldfarb. Using the steepest-edge simplex algorithm to solve sparse linear programs. In *Sparse matrix computations*, pages 227–240. Elsevier, 1976.

[Gol83] Donald Goldfarb. Worst case complexity of the shadow vertex simplex algorithm. *preprint, Columbia University*, 1983.

[GS55] Saul Gass and Thomas Saaty. The computational algorithm for the parametric objective function. *Naval research logistics quarterly*, 2(1-2):39–45, 1955.

[GS79] Donald Goldfarb and William Y Sit. Worst case behavior of the steepest edge simplex method. *Discrete Applied Mathematics*, 1(4):277–285, 1979.

[Hai83] M. Haimovich. The simplex method is very good! On the expected number of pivot steps and related properties of random linear programs. Technical report, Columbia University, 1983.

[Har73] Paula M. J. Harris. Pivot selection methods of the devex lp code. *Mathematical Programming*, 5(1):1–28, December 1973.

[HLZ23] Sophie Huiberts, Yin Tat Lee, and Xinzhi Zhang. Upper and lower bounds on the smoothed complexity of the simplex method. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*. ACM, June 2023.

[HZ15] Thomas Dueholm Hansen and Uri Zwick. An improved version of the random-facet pivoting rule for the simplex algorithm. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 209–218, 2015.

[Jer73] Robert G Jeroslow. The simplex algorithm with the pivot rule of maximizing criterion improvement. *Discrete Mathematics*, 4(4):367–377, 1973.

[Kal92] Gil Kalai. A subexponential randomized simplex algorithm. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 475–482, 1992.

[KM72] Victor Klee and George J Minty. How good is the simplex algorithm. *Inequalities : III : proceedings of the 3rd Symposium on inequalities*, pages 159–175, 1972.

[KM11] Tomonari Kitahara and Shinji Mizuno. A bound for the number of different basic solutions generated by the simplex method. *Mathematical Programming*, 137(1-2):579–586, August 2011.

[KS06] Jonathan A. Kelner and Daniel A. Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 51–60, 2006.

[Mat02] Jiří Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.

[Meg86] Nimrod Megiddo. Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Math. Programming*, 35(2):140–172, 1986.

[MSW96] Jiří Matoušek, Micha Sharir, and Emo Welzl. A subexponential bound for linear programming. *Algorithmica*, 16(4-5):498–516, 1996.

[Mur80] Katta G Murty. Computational complexity of parametric linear programming. *Mathematical programming*, 19(1):213–219, 1980.

[NSS22] Hariharan Narayanan, Rikhav Shah, and Nikhil Srivastava. A Spectral Approach to Polytope Diameter. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 108:1–108:22, 2022.

[Rou20] Tim Roughgarden, editor. *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press, December 2020.

[Sha87] Ron Shamir. The efficiency of the simplex method: a survey. *Management science*, 33(3):301–334, 1987.

[Sma83] Steve Smale. On the average number of steps of the simplex method of linear programming. *Mathematical programming*, 27(3):241–262, 1983.

[ST04] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[Tod86] Michael J. Todd. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Mathematical Programming*, 35(2):173–192, 1986.

[Ver09] Roman Vershynin. Beyond hirsch conjecture: walks on random polytopes and smoothed complexity of the simplex method. *SIAM Journal on Computing*, 39(2):646–678, 2009.

[xpr] FICO Xpress Optimizer reference manual. https://www.fico.com/fico-xpress-optimization/docs/latest/solver/optimizer/HTML/chapter4.html, archived at https://web.archive.org/web/20230328100937/https://www.fico.com/fico-xpress-optimization/docs/latest/solver/optimizer/HTML/chapter4.html. Retrieved September 18th, 2023.