BEYOND HIRSCH CONJECTURE: WALKS ON RANDOM POLYTOPES AND SMOOTHED COMPLEXITY OF THE SIMPLEX METHOD

ROMAN VERSHYNIN

ABSTRACT. The smoothed analysis of algorithms is concerned with the expected running time of an algorithm under slight random perturbations of arbitrary inputs. Spielman and Teng proved that the shadow-vertex simplex method has polynomial smoothed complexity. On a slight random perturbation of an arbitrary linear program, the simplex method finds the solution after a walk on polytope(s) with expected length polynomial in the number of constraints n, the number of variables d and the inverse standard deviation of the perturbation $1/\sigma$.

We show that the length of walk in the simplex method is actually *polylogarithmic* in the number of constraints n. Spielman-Teng's bound on the walk was $O^*(n^{86}d^{55}\sigma^{-30})$, up to logarithmic factors. We improve this to $O(\log^7 n(d^9 + d^3\sigma^{-4}))$. This shows that the tight Hirsch conjecture $n - d$ on the length of walk on polytopes is not a limitation for the smoothed Linear Programming. Random perturbations create short paths between vertices.

We propose a randomized phase-I for solving arbitrary linear programs, which is of independent interest. Instead of finding a vertex of a feasible set, we add a vertex at random to the feasible set. This does not affect the solution of the linear program with constant probability. So, in expectation it takes a constant number of independent trials until a correct solution is found. This overcomes one of the major difficulties of smoothed analysis of the simplex method – one can now statistically decouple the walk from the smoothed linear program. This yields a much better reduction of the smoothed complexity to a geometric quantity – the size of planar sections of random polytopes. We also improve upon the known estimates for that size, showing that it is polylogarithmic in the number of vertices.

CONTENTS

1. Introduction

The simplex method is "the classic example of an algorithm that is known to perform well in practice but which takes exponential time in the worst case" [6]. In an attempt to explain this behavior, Spielman and Teng [6] introduced the concept of *smoothed analysis* of algorithms, in which one measures the expected complexity of an algorithm under slight random perturbations of arbitrary inputs. They proved that a variant of the simplex method has polynomial smoothed complexity.

Consider a linear program of the form

$$\text{maximize } \langle z, x \rangle$$
$$\text{subject to } Ax \leq b, \tag{LP}$$

where A is an $n \times d$ matrix, representing n constraints, and x is a vector representing d variables.

A simplex method starts at some vertex x_0 of the polytope $Ax \leq b$, found by a phase-I method, and then walks on the vertices of the polytope toward the solution of (LP). A pivot rule dictates how to choose the next vertex in this walk. The complexity of the simplex method is then determined by the length of the walk – the number of pivot steps.

So far, smoothed analysis has only been done for the shadow-vertex pivot rule introduced by Gaas and Saaty [7]. The shadow-vertex simplex method first chooses an *initial objective function* z_0 optimized by the initial vertex x_0. Then it interpolates between z_0 and the actual objective function z. Namely, it rotates z_0 toward z and computes the vertices that optimize all the objective functions between z_0 and z.

A *smoothed linear program* is a linear program of the form (LP), where the rows a_i of A, called the *constraint vectors*, and b are independent Gaussian random vectors, with arbitrary centers $\bar{a}_i$ and $\bar{b}$ respectively, and whose coordinates are independent normal random variables with standard deviations $\sigma \max_i \|(\bar{a}_i, \bar{b}_i)\|$. Spielman and Teng proved

Theorem 1.1. [6] *For arbitrary linear program with $d > 3$ variables and $n > d$ constraints, the expected number of pivot steps in a two-phase shadow-vertex simplex method for the smoothed program is at most a polynomial $\mathcal{P}(n, d, \sigma^{-1})$.*

Spielman-Teng's analysis yields the following estimate on expected number of pivot steps:

$$\mathcal{P}(n, d, \sigma^{-1}) \leq O^*(n^{86} d^{55} \sigma^{-30})$$

where the logarithmic factors are disregarded. The subsequent work of Deshpande and Spielman [5] improved on the exponents of d and σ; however, it doubled the exponent of n.

Another model of randomness, in which the directions of the inequalities in (LP) are chosen independently at random, was studied in the eighties [8, 1, 15, 3, 2]. It was shown that a shadow-vertex simplex method solves such problems in $O(d^2)$ expected number of pivot steps. Note that this bound does not depend on the number of constraints n. However, it is not clear whether reversion of the inequalities can be justified as a good model for typical linear problems.

These results lead to the question – how does the smoothed complexity of the simplex method depend on the number of constraints n? Unlike in the model with randomly reversed inequalities, the number of pivots steps in the smoothed analysis must be at least logarithmic in n (see below).

In this paper, we prove the first polylogarithmic upper bound:

Theorem 1.2 (Main)**.** *The expected number of pivot steps in Theorem 1.1 is at most*

$$\mathcal{P}(n, d, \sigma^{-1}) \leq O(\log^7 n(d^9 + d^3 \sigma^{-4})).$$

See Theorem 6.1 for a more precise estimate.

So, the number of pivot steps is *polylogarithmic in the number of constraints n*, while the previous known bounds were polynomial in n.

This bound goes in some sense beyond the classical Hirsch conjecture on the diameter of polytopes, the maximal number of steps in the shortest walk between any pair of vertices (see e.g. [10]). Hirsch conjecture states that the diameter of a polytope with n faces in d dimensions, and in particular of the feasible polytope of (LP), is at most $n - d$.

Hirsch conjecture is tight, so it would be natural to think of $n - d$ as a lower bound on the worst case complexity of *any* variant of the simplex method. Theorem 1.2 (and Theorem 6.2 below) claim that a random perturbation creates a much shorter path between a given pair of vertices. Moreover, while Hirsch conjecture does not suggest any algorithm for finding a short walk, the shadow-vertex simplex method already finds a much shorter walk!

One can wonder whether a random perturbation creates a short path between the vertices by destroying most of them. This is not so even in the average case, when A is a matrix with i.i.d. Gaussian entries. Indeed, the expected number of vertices of the random polytope is asymptotic to $2^d d^{-1/2}(d-1)^{-1}(\pi \log n)^{(d-1)/2}$ ([12], see [9]). This bound is exponential in d and sublinear but not polylogarithmic in n, as Theorem 1.2 claims.

For $d = 2$, the random polygon will have $\Omega(\sqrt{\log n})$ vertices, which is thus a lower bound on the number of pivot steps of (any) simplex algorithm.

The smoothed complexity (expected running time) of the simplex method is $O(\mathcal{P}(n, d, \sigma^{-1})\, t_{\mathrm{pivot}})$, where t_{pivot} is the time to make one pivot step under the shadow-vertex pivot rule. The dependence of t_{pivot} on n is at most linear, for one only needs to find an appropriate vector a_i among the n vectors to update the running vertex. However, for many well structured linear problems the exhaustive search over all a_i is not necessary, which makes t_{pivot} much smaller. In these cases, Theorem 1.2 shows that *the shadow-vertex simplex method can solve very large scale problems* (up to exponentially many constraints).

Acknowledgements. The author is grateful to the referees for careful reading of the manuscript and many useful remarks and suggestions, which greatly improved the paper.

2. Outline of the approach

Our smoothed analysis of the simplex method is largely inspired by that of Spielman and Teng [6]. We resolve a few conceptual difficulties that arise in [6], which eventually simplifies and improves the overall picture.

2.1. **Interpolation: reduction to unit linear programs.** First, we reduce an arbitrary linear program (LP) to a *unit linear program* – the one for which $b = \mathbf{1}$. This is done by a simple interpolation. This observation is independent of any particular algorithm to solve linear programs.

An interpolation variable will be introduced, and (LP) will reduce to a unit program in dimension $d+1$ with constraint vectors of type (a_i, b_i). A simple but very useful consequence is that this reduction preserves the Gaussian distribution of the constraints – if (LP) has independent Gaussian constraints (as the smoothed program does), then so does the reduced unit program.

2.2. **Duality: reduction to planar sections of random polytopes.** Now that we have a unit linear program, it is best viewed in the polar perspective. The polar of the feasible set $Ax \le \mathbf{1}$ is the polytope

$$P = \mathrm{conv}(0, a_1, \ldots, a_n).$$

The unit linear problem is then equivalent to finding facet(z), the facet of P pierced by the ray $\{tz :\ t \ge 0\}$. In the shadow-vertex simplex method, we assume that phase-I provides us with an initial objective vector z_0 and the initial facet(z_0). Then phase-II of the simplex method computes facet(q) for all vectors q in the plane $E = \mathrm{span}(z_0, z)$ between z_0 and z. Specifically, it rotates q from z_0 toward z and updates facet(q) by removing and adding one vertex to its basis, as it becomes necessary. At the end, it outputs facet(z).

The number of pivot steps in the simplex method is bounded by the number of facets of P the plane E intersects. This is the *size of the planar section of the random polytope* P, the number of the edges of the polygon $P \cap E$. Under a hypothetical assumption that E is fixed or is statistically independent of P, the average size of $P \cap E$ is polynomial in the dimension d and the reciprocal of the standard deviation σ and polylogarithmic in the number of vertices n, see Theorem 2.1 below.

The main complication of the analysis in [6] was that plane $E = \mathrm{span}(z_0, z)$ was also random, and moreover correlated with the random polytope P. It is not clear how to find the initial vector z_0 independent of the polytope P and, at the same time, in such a way that we know the facet of P it pierces. Thus the main problem rests in phase-I. None of the previously available phase-I methods in linear programming seem to effectively overcome this difficulty.

The randomized phase-I proposed in [6] exposed a random facet of P by multiplying a random d-subset of the vectors a_i by an appropriately big constant to ensure that these vectors do span a facet. Then a random convex linear combination of these vectors formed the initial vector z_0. This approach brings about two complications:

(a) the vertices of the new random polytope are no longer Gaussian;

(b) the initial objective vector z_0 (thus also the plane E) is correlated with the random polytope.

Our new approach will overcome both these difficulties.

2.3. **Phase-I for arbitrary linear programs.** We propose the following randomized phase-I for arbitrary unit linear programs. It is of independent interest, regardless of its applications to smoothed analysis and to the simplex method.

Instead of finding or exposing a facet of P, we *add a facet* to P in a random direction. We need to ensure that this facet falls into the *numb set* of the linear program, which consists of the points that do not change the solution when added to the set of constraint vectors (a_i). Since the solution of the linear program is $\mathrm{facet}(z)$, the affine half-space below the affine span of $\mathrm{facet}(z)$ (on the same side as the origin) is contained in the numb set. Thus *the numb set always contains a half-space.*

A random vector z_0 drawn from the uniform distribution on the sphere S^{d-1} is then in the numb half-space with probability at least $1/2$. Moreover, a standard concentration of measure argument shows that such a random point is at distance $\Omega(d^{-1/2})$ from the boundary of the numb half-space, with constant probability. (This distance is the *observable diameter* of the sphere, see [11] Section 1.4). Thus a small regular simplex with center z_0 is also in the numb set with constant probability. Similarly, one can smooth the vertices of the simplex (make them Gaussian) without leaving the numb set. Finally, to ensure that

such simplex will form a facet of the new polytope, it suffices to dilate it by the factor $M = \max_{i=1,\ldots,n} \|a_i\|$.

Summarizing, *we can add d linear constraints to any linear program at random, without changing its solution with constant probability.*

Note that it is easy to check whether the solution is correct, i.e. that the added constraints do not affect the solution. The latter happens if and only if none of the added constraints turn into equalities on the solution x. Therefore, one can repeatedly solve the linear program with different sets of added constraints generated independently, until the solution is correct. Because of a constant probability of success at every step, this phase-I terminates after an expected constant number of steps, and it always produces a correct initial solution.

When applied for the smoothed analysis of the simplex method, this phase-I resolves one of the main difficulties of the approach in [6]. The initial objective vector z_0 and thus the plane E become independent of the random polytope P. Thus *the smoothed complexity of the simplex method gets bounded by the number of edges of a planar section of a random polytope P*, whose vertices have standard deviation of order $\Omega^*(\min(\sigma, d^{-3/2}))$, see (5.1). In the previous approach [6], such reduction was made with the standard deviation of order $\Omega^*(n^{-14}d^{-8.5}\sigma^5)$.

A *deterministic phase-I* is also possible, along the same lines. We have used that a random point in S^{d-1} is at distance $\Omega(d^{-1/2})$ from a half-space. The same property is clearly satisfied by at least one element of the set $\{\pm e_1, \ldots, \pm e_d\}$, where $\{e_1, \ldots, e_d\}$ is the canonical basis of $\mathbb{R}^d$. Therefore, at least one of d regular simplices of radius $\frac{1}{2}d^{-1/2}$ centered at points e_i, lies in the numb half-space. One can try them all for added constraints; at least one will give a correct solution. This however will increase the running time by a factor of d – the number of trials in this deterministic phase-I may be as large as d, while the expected number of trials in the randomized phase-I is constant. The smoothed analysis with such phase-I will also become more difficult due to having d non-random vertices.

2.4. **Sections of randomly perturbed polytopes.** To complete the smoothed analysis of the simplex algorithm, it remains to bound the size (i.e. the number of edges) of the section $P \cap E$ of a randomly perturbed polytope P with a fixed plane E.

Spielman and Teng proved the first polynomial bound on this size. We improve it to a polylogarithmic bound in the number of vertices n:

Theorem 2.1 (Sections of random polytopes). *Let $a_1, \ldots, a_n$ be independent Gaussian vectors in $\mathbb{R}^d$ with centers of norm at most 1, and with standard deviation σ. Let E be a plane in $\mathbb{R}^d$. Then the random polytope $P = \mathrm{conv}(a_1, \ldots, a_n)$ satisfies*

$$\mathbb{E} \,|\, \mathrm{edges}(P \cap E)| = O(d^5 \log^2 n + d^3 \sigma^{-4}).$$

See Theorem 6.2 for a slightly more precise result.

Spielman and Teng obtained a weaker estimate $O(nd^3\sigma^{-6})$ for this size ([6] Theorem 4.0.1). A loss of the factor of n in their argument occurs in estimating the angle of incidence ([6] Lemma 4.2.1), the angle at which a fixed ray in E emitted from the origin meets the facet of P it pierces.

Instead of estimating the angle of incidence from one viewpoint determined by the origin 0, we will *view the polytope P from three different points* 0_1, 0_2, 0_3 on E. Rays will be emitted from each of these points, and from at least one of them the angle of incidence will be good (more precisely, the angle to the edge of $P \cap E$, which is the intersection of the corresponding facet with E).

3. Preliminaries

3.1. **Notation.** The *direction* of a vector x in a vector space is the ray $\{tx : t \geq 0\}$. The non-negative *cone* of a set K, denoted $\mathrm{cone}(K)$, is the union of the directions of all the vectors x in K. The closed *convex hull* of K is denoted by $\mathrm{conv}(K)$ and $\triangle(K)$. (We prefer the latter notation when K consists of d points in $\mathbb{R}^d$).

The standard inner product in $\mathbb{R}^d$ is denoted by $\langle x, y \rangle$, and the standard Euclidean norm in $\mathbb{R}^d$ is denoted by $\|x\|$. The unit Euclidean sphere in $\mathbb{R}^d$ is denoted by S^{d-1}.

The *polar* of a set K in $\mathbb{R}^d$ is defined as $K^\circ = \{x \in \mathbb{R}^d : \langle x, y \rangle \leq 1 \ \forall y \in K\}$.

A *half-space* in $\mathbb{R}^d$ is a set of the form $\{x : \langle z, x \rangle \leq 0\}$ for some vector z. An *affine half-space* takes the form $\{x : \langle z, x \rangle \leq a\}$ for some vector z and a number a. The definitions of a *hyperplane* and *affine hyperplane* are similar, with equalities in place of the inequalities. The *normal to an affine hyperplane H* that is not a hyperplane is the vector h such that $H = \{x : \langle h, x \rangle = 1\}$. A point x is said to be *below H* if $\langle h, x \rangle \leq 1$. We say that a direction z *pierces* the hyperplane H (or a subset H_0 thereof) if the ray $\{tz : t \geq 0\}$ intersects H (respectively, H_0).

The probability will be denoted by $\mathbb{P}$, and the expectation by $\mathbb{E}$. The *conditional probability* on a measurable subset B of a probability space is denoted by $\mathbb{P}\{\cdot|B\}$ and defined as $\mathbb{P}\{A|B\} = \mathbb{P}\{A \cap B\}/\mathbb{P}\{B\}$.

A *Gaussian random vector* $g = (g_1, \ldots, g_d)$ in $\mathbb{R}^d$ with center $\bar{g} = (\bar{g}_1, \ldots, \bar{g}_d)$ and variance σ is a vector whose coordinates g_i are independent Gaussian random variables with centers $\bar{g}_i$ and variance σ.

Throughout the paper, we will assume that the vectors (a_i, b_i) that define the linear program (LP) are in general position. This assumption simplifies our analysis and it holds with probability 1 for a smoothed program. One can remove this assumption with appropriate modifications of the results.

A solution x of (LP) is determined by a d-set I of the indices of the constraints $\langle a_i, x \rangle \le b_i$ that turn into equations on x. One can compute x from I by solving these equations. So we sometimes call the index set I a solution of (LP).

For a polytope $P = \text{conv}(0, a_1, \ldots, a_n)$ and a vector z in $\mathbb{R}^d$, $\text{facet}(z) = \text{facet}_P(z)$ will denote the set of all faces of P which the direction z pierces. The point where it pierces them (which is unique) is denoted by z_P. More precisely, $\text{facet}(z)$ is the family of all d-sets I such that $\triangle(a_i)_{i \in I}$ is a facet of the polytope P and the direction z pierces it.

If z is in general position, $\text{facet}(z)$ is an empty set or contains exactly one set I. The corresponding geometric facet $\triangle(a_i)_{i \in I}$ will be denoted by $\text{Facet}_P(z)$ or $\text{Facet}_P(I)$.

Positive absolute constants will be denoted by $C, C_1, c, c_1, \ldots$. The natural logarithms will be denoted by log.

3.2. Vertices at infinity.
For convenience in describing the interpolation method, we will assume that one of the constraint vectors a_i can be *at infinity*, in a specified direction $u \in \mathbb{R}^d$. The definitions of the positive cone and the convex hull are then modified in a straightforward way. If, say, a_j is such an infinite vector and $j \in I$, then one defines $\triangle(a_i)_{i \in I} = \triangle(a_i)_{i \in I - \{j\}} + \{tu : t \ge 0\}$, where the addition is the Minkowski sum of two sets, $A + B = \{a + b : a \in A, b \in B\}$.

Although having infinite vectors is convenient in theory, all computations can be performed with numbers bounded by the magnitude of the input (e.g., checking $I \in \text{facet}(z)$ for given z and I has the same complexity whether or not some vertex of P is at infinity).

3.3. Polar shadow vertex simplex method.
This is the only variant of the simplex method whose smoothed complexity has been analyzed. We shall describe this method now; for more information see [6] Section 3.2.

The polar shadow vertex simplex method works for unit linear programs, i.e. programs (LP) with $b = \mathbf{1}$. A solution of such program is a member of $\text{facet}(z)$ of the polytope $P = \text{conv}(0, a_1, \ldots, a_n)$. The program is unbounded iff $\text{facet}(z) = \emptyset$. (See [6] Section 3.2).

The input of the polar shadow vertex simplex method is:

(1) objective vector z;
(2) initial objective vector z_0;
(3) $\text{facet}(z_0)$, provided that it consists of only one set of indices.

The simplex method rotates z_0 toward z and computes $\text{facet}(q)$ for all vectors q between z_0 and z. At the end, it outputs the limit of $\text{facet}(q)$ as q approaches z. This is the last running $\text{facet}(q)$ before q reaches z.

If $\text{facet}(z_0)$ contains more than one index set, one can use the limit of $\text{facet}(q)$ as q approaches z_0 as the input of the simplex method. This will be the first running $\text{facet}(q)$ when q departs from z_0.

If z and z_0 are linearly dependent, $z_0 = -cz$ for some $c > 0$, one can specify an arbitrary *direction of rotation* $u \in \mathbb{R}^n$, which is linearly independent of z, so that the simplex method rotates q in $\mathrm{span}(z, u)$ in the direction of u, i.e. one can always write $q = c_1 z + c_2 u$ with $c_2 \geq 0$.

4. Reduction to unit programs: Interpolation

We will show how to reduce an arbitrary linear program (LP) to a *unit linear program*

$$\text{maximize } \langle z, x \rangle$$
$$\text{subject to } Ax \leq \mathbf{1}. \tag{Unit LP}$$

This reduction, whose idea originates from [6], is quite general and is independent from any particular method to solve linear programs.

The idea is to interpolate between (Unit LP) and (LP). To this end, we introduce an additional (interpolation) variable t and a multiplier λ, and consider the *interpolated linear program* with variables x, t:

$$\text{maximize } \langle z, x \rangle + \lambda t$$
$$\text{subject to } Ax \leq tb + (1 - t)\mathbf{1}, \quad 0 \leq t \leq 1. \tag{Int LP}$$

The interpolated linear program becomes (Unit LP) for $t = 0$ and (LP) for $t = 1$. We can give bias to $t = 0$ by choosing the multiplier $\lambda \to -\infty$ and to $t = 1$ by choosing $\lambda \to +\infty$. Furthermore, (Int LP) can be written as a unit linear program in $\mathbb{R}^{d+1}$:

$$\text{maximize } \langle (z, \lambda), (x, t) \rangle$$
$$\text{subject to } \begin{cases} \langle (a_i, 1 - b_i), (x, t) \rangle \leq 1, \\ \langle (0, 1), (x, t) \rangle \leq 1, \quad \langle (0, -\infty), (x, t) \rangle \leq 1. \end{cases} \tag{Int LP'}$$

The constraint vectors are $(a_i, 1 - b_i)$, $(0, 1)$ and $(0, -\infty)$. (see Section 3.2 about vertices at infinity). This has a very useful consequence: if the constraints of the original (LP) are Gaussian, then so are the constraints of (Int LP'), except the two last ones. In other words, *the reduction to a unit program preserves the Gaussian distribution of the constraints.*

The properties of interpolation are summarized in the following elementary fact.

Proposition 4.1 (Interpolation)**.**
(i) **(LP) is unbounded** *iff (Unit LP) is unbounded iff (Int LP) is unbounded for all sufficiently big λ iff (Int LP) is unbounded for some λ.*

(ii) *Assume (LP) is not unbounded. Then the solution of (Unit LP) equals the solution of (Int LP) for all sufficiently small λ; in this solution, $t = 0$.*

(iii) *Assume (LP) is not unbounded. Then* **(LP) is feasible** *iff $t = 1$ in the solution of (Int LP) for all sufficiently big λ.*

*(iv) Assume (LP) is feasible and bounded. Then the **solution of (LP)** equals the solution of (Int LP) for all sufficiently big λ.*

Proof. See Appendix A.   ∎

Now assuming that we know how to solve unit linear programs, we will be able to solve arbitrary linear programs. The correctness of this two-phase algorithm follows immediately from Proposition 4.1.

SOLVER FOR (LP)

> **Phase-I:** Solve (Unit LP) using SOLVER FOR (UNIT LP) of Section 5. If this program is unbounded, then (LP) is also unbounded. Otherwise, the solution of (Unit LP) and $t = 0$ is a limit solution of (Int LP) as $\lambda \to -\infty$. Use this solution as the input for the next step.
>
> **Phase-II:** Use the polar shadow-vertex simplex method to find a limit solution of (Int LP) with $\lambda \to +\infty$. If $t \neq 1$ in this solution, then the (LP) is infeasible. Otherwise, this is a correct solution of (LP).

While this algorithm is stated in terms of limit solutions, one does not need to take actual limits when computing them. This follows from the properties of the polar shadow-vertex simplex method described in Section 3.3. Indeed, in phase-II of SOLVER FOR (LP) we can write (Int LP) as (Int LP') and use the initial objective vector $\bar{z}_0 = (0, -1)$, the actual objective vector $\bar{z} = (0, 1)$, and the direction of rotation $\bar{u} = (z, 0)$. Phase-I provides us with a limit solution for the objective vectors $(\varepsilon z, -1) = \bar{z}_0 + \varepsilon \bar{u}$ as $\varepsilon \to 0^+$. These vectors approach z_0 as we rotate from z toward z_0 in $\operatorname{span}(z, u)$. Similarly, we are looking for a limit solution for the objective vectors $(\varepsilon z, 1) = \bar{z} + \varepsilon \bar{u}$ as $\varepsilon \to 0^+$. These vectors approach z as we rotate from z_0 toward z in $\operatorname{span}(z, u)$. By Section 3.3, the polar shadow-vertex simplex method applied with vectors $\bar{z}_0$, $\bar{z}$, $\bar{u}$ and the initial limit solution found in Phase-I, finds the correct limit solution in Phase-II.

5. SOLVING UNIT PROGRAMS: ADDING CONSTRAINTS IN PHASE-I

We describe a randomized phase-I for solving arbitrary unit linear problems of type (Unit LP). Rather than finding an initial feasible vertex, we shall add a random vertex to the feasible set. We thus add d constraints to (Unit LP), forming

$$
\begin{aligned}
&\text{maximize } \langle z, x \rangle \\
&\text{subject to } A^+ x \leq \mathbf{1},
\end{aligned}
\qquad \text{(Unit LP}^+\text{)}
$$

where A^+ has the rows $a_1, \ldots, a_n, a_{n+1}, \ldots, a_{n+d}$ with some new constraint vectors $a_{n+1}, \ldots, a_{n+d}$.

The first big question is whether the problems (Unit LP) and (Unit LP$^+$) are *equivalent*, i.e. whether (Unit LP$^+$) is bounded if and only if (Unit LP) is bounded, and if they are bounded, the solution of (Unit LP$^+$) equals the solution of (Unit LP). This motivates:

Definition 5.1. *A* numb set *of a unit linear program is the set of all vectors a so that adding the constraint $\langle a, x \rangle \leq 1$ to the set of the constraints produces an equivalent linear program.*

We make two crucial observations – that the numb set is always big, and that one can always check if the problems (Unit LP) and (Unit LP$^+$) are equivalent. As mentioned in Section 3.1, we will assume that the constraint vectors a_i are in general position.

Proposition 5.2. *The numb set of a unit linear program contains a half-space (called a* numb half-space*).*

Proof. Given a convex set K containing the origin in a vector space, Minkowski functional $\|z\|_K$ is defined for vectors z as $\|z\|_K = \inf\{\lambda > 0 : \frac{1}{\lambda} z \in K\}$ if the infimum exists, and infinity if it does not exist. Then the duality shows that the solution $\max_{Ax \leq \mathbf{1}} \langle z, x \rangle$ of (Unit LP) equals $\|z\|_P$. (It is infinity iff the problem is unbounded; we will use the convention $1/\infty = 0$ in the sequel). By Hahn-Banach (Separation) Theorem, there exists a vector z^* such that

$$\langle z^*, x \rangle \leq \langle z^*, \tfrac{1}{\|z\|_P} z \rangle := h \;\; \text{for all } x \in P.$$

$0 \in P$ implies that $h \geq 0$. We define the affine half-space

$$H^- = \{x : \langle z^*, x \rangle \leq h\}$$

and claim that H^- lies in the numb set of (Unit LP). To prove this, let $a \in H^-$. Since $P \subset H^-$, we have $\operatorname{conv}(P \cup a) \subset H^-$, thus

$$\|z\|_P \geq \|z\|_{\operatorname{conv}(P \cup a)} \geq \|z\|_{H^-} = \|z\|_P$$

where the first two inequalities follow from the inclusion $P \subset \operatorname{conv}(P \cup a) \subset H^-$, and the last equality follows from the definition of H^-. So, we have shown that $\|z\|_{\operatorname{conv}(P \cup a)} = \|z\|_P$, which says that a and thus the affine half-space H^- is in the numb set of (Unit LP). Since $h \geq 0$, H^- contains the origin, thus contains a half-space. ∎

In particular, if (Unit LP) is bounded, then its numb set is the affine half-space below facet(z). Then a similar duality argument proves:

Proposition 5.3 (Equivalence)**.**
(i) If the added constraint vectors $a_{n+1}, \ldots, a_{n+d}$ lie in some numb half-space of (Unit LP), then (Unit LP$^+$) is equivalent to (Unit LP).

(ii) (Unit LP$^+$) is equivalent to (Unit LP) if and only if either (Unit LP$^+$) is unbounded or its solution does not satisfy any of the added constraints $\langle a_i, x \rangle \leq 1$, $i = n+1, \ldots, n+d$.

∎

Proposition 5.2 implies that a constraint vector z_0 whose direction is chosen at random in the unit sphere S^{d-1}, is in the numb set with probability at least $1/2$. By a standard concentration of measure argument, a similar statement will be true about a small simplex centered at z_0. It is then natural to take the vertices of this simplex as added constraint vectors $a_{n+1}, \ldots, a_{n+d}$ for (Unit LP$^+$). To this end, we define the size ℓ of the simplex and the standard deviation σ_1 for smoothing its vertices as

$$\ell = \frac{c_1}{\sqrt{\log d}}, \quad \sigma_1 = \min\Big(\frac{1}{6\sqrt{d \log n}}, \ \frac{c_1}{d^{3/2} \log d}\Big), \tag{5.1}$$

where $c_1 = \frac{1}{300}$ and $c_2 = \frac{c_1^2}{100}$. Then we form (Unit LP$^+$) as follows:

ADDING CONSTRAINTS

> **Input:** Size $M_0 > 0$ and rotation $U \in O(d)$.
>
> **Output:** "Failure" or vectors $a_{n+1}, \ldots, a_{n+d}$ and $z_0 \in \mathrm{cone}(a_{n+1}, \ldots, a_{n+d})$.
>
> (1) *Form a regular simplex:* let z_0' be a fixed unit vector in $\mathbb{R}^d$ and $\bar{a}'_{n+1}, \ldots, \bar{a}'_{n+d}$ be the vertices of a fixed regular simplex in $\mathbb{R}^d$ with center and normal z_0', and radius $\|z_0' - \bar{a}'_i\| = \ell$.
>
> (2) *Rotate and dilate:* let $z_0 = 2M_0 U z_0'$, $\bar{a}_i = 2M_0 U \bar{a}'_i$ for $i = n+1, \ldots, n+d$.
>
> (3) *Smooth:* let a_i be independent Gaussian random variables with mean $\bar{a}_i$ and standard deviation $2M_0\sigma_1$, for $i = n+1, \ldots, n+d$.
>
> (4) *Check if the constraints added correctly:* check if
>   (a) $z_0 \in \mathrm{cone}(a_{n+1}, \ldots, a_{n+d})$ and
>   (b) distance from 0 to $\mathrm{aff}(a_{n+1}, \ldots, a_{n+d})$ is at least M.
>   If not, return "Failure".

Remark. Steps (3) and (4) are, strictly speaking, not necessary. They facilitate the theoretical smoothed analysis of the simplex method. However, they can be skipped in practical implementations.

The crucial property of ADDING CONSTRAINTS is the following. (Recall that we regard a solution of a linear program as the index set of the inequalities that become equalities on the solution point, see Section 3.1).

Theorem 5.4. *Let (Unit LP) be a unit linear program with a numb half-space H, and let $M_0 \geq M$ where $M = \max_{i=1,\ldots,n} \|a_i\|$. Then:*

1. Let $U \in O(d)$ be arbitrary. If the algorithm ADDING CONSTRAINTS *does not return "Failure", then a solution of (Unit LP$^+$) with the objective function $\langle z_0, x \rangle$ is the index set $\{n+1, \ldots, n+d\}$.*

2. With probability at least $1/4$ in the choice of a random rotation $U \in O(d)$ and random vectors $a_{n+1}, \ldots, a_{n+d}$, the algorithm ADDING CONSTRAINTS *does not return "Failure" and the vectors $a_{n+1}, \ldots, a_{n+d}$ lie in the numb half-space H.*

Proof. See Appendix B. ∎

By Proposition 5.3, the conclusion of Theorem 5.4 is that:

(a) with constant probability the problems (Unit LP$^+$) and (Unit LP) are equivalent;

(b) we can check whether they are equivalent or not (by part (ii) of Proposition 5.3);

(c) we always know a solution of (Unit LP$^+$) for some objective function (if "Failure" is not returned).

Thus we can solve (Unit LP) by repeatedly solving (Unit LP$^+$) with independently added constraints until no "Failure" is returned and until the solution is correct. This forms a two-phase solver for unit linear programs.

SOLVER FOR (UNIT LP)

Do the following until no "Failure" is returned and the solution I^+ contains none of the indices $n+1, \ldots, n+d$:

> **Phase-I:** Apply ADDING CONSTRAINTS with $M_0 = e^{\lceil \log M \rceil}$ where $M = \max_{i=1,\ldots,n} \|a_i\|$ and with the rotation U chosen randomly and independently in the orthogonal group $O(d)$ according to the Haar measure. If no "Failure" returned, then $\{n+1, \ldots, n+d\}$ is a solution of (Unit LP$^+$) with the objective function $\langle z_0, x \rangle$. Use this solution as the input for the next step.

> **Phase-II:** Use the polar shadow-vertex simplex method to find a solution I^+ of (Unit LP$^+$) with the actual objective function $\langle z, x \rangle$.

Return I^+.

Remark. The discretized maximum M_0 is introduced in this algorithm only to simplify its smoothed analysis. In practical implementations of the algorithm, one can use $M_0 = M$.

6. BOUNDING THE COMPLEXITY VIA SECTIONS OF RANDOM POLYTOPES

We do here the smoothed analysis of SOLVER FOR (LP), and prove the following more precise version of Theorem 1.2:

Theorem 6.1. *For an arbitrary linear program with $d > 3$ variables and $n > d$ constraints, the expected number of pivot steps in a two-phase shadow-vertex simplex method for the*

smoothed program is

$$O\big(\log^2 n \cdot \log\log n \cdot (d^3\sigma^{-4} + d^5\log^2 n + d^9\log^4 d)\big) = O^*(d^9 + d^3\sigma^{-4}).$$

Remark. A more careful analysis may allow one to remove the factor $\log^2 n \cdot \log\log n$. To do this, one uses the version of the algorithm with $M_0 = M$ and bounds the sections of a polytope with an added facet. This makes the analysis a bit harder, because the magnitudes of the added vertices correlate with M and thus with the polytope.

To prove Theorem 6.1, let us recall how many calls SOLVER FOR (LP) makes to the polar shadow-vertex method. One call is made to solve (Int LP) (in the second phase), and several calls are made to solve (Unit LP$^+$) (in the subroutine SOLVER FOR (UNIT LP) in the first phase.)

The expected number of calls (iterations) in SOLVER FOR (UNIT LP) is 4. This follows from part 2 of Theorem 5.4 and Proposition 5.3. Thus:

> *The running time of* SOLVER FOR (LP) *is bounded by the total number of pivot steps made in the polar shadow-vertex simplex method, when we apply it:*
> (1) *once for (Int LP);*
> (2) *on average, four times for (Unit LP$^+$).*

Furthermore, as explained in Section 2.2, the number of pivot steps in the polar shadow-vertex simplex method on a unit linear program is bounded by the *number of edges of the polygon $P \cap E$*, where P is the convex hull of the origin and the constraint vectors, and E is the span of the initial and the actual objective vectors.

We shall now first estimate the size of the section of the polytope for (1) and (2) separately, and then combine them by a simple stopping time argument.

Recall that the vectors a_i and b are Gaussian random vectors with centers $\bar{a}_i$ and $\bar{b}$ respectively, and with standard deviation σ. We can assume without loss of generality that

$$\|(\bar{a}_i, \bar{b}_i)\| \le 1 \ \text{ for all } i = 1, \ldots, n, \quad \sigma \le \frac{1}{6\sqrt{d\log n}}. \tag{6.1}$$

(To achieve these bounds, we scale down the vectors if necessary – first to achieve $\max_i \|(a_i, b_i)\| = 1$, then further to make σ as required).

6.1. **Sections of random polytopes.** When we apply the polar shadow-vertex simplex method for (Int LP) (in phase-II of SOLVER FOR (LP)), the plane

$$E = \text{span}((z, 0), (0, 1))$$

is fixed, and the constraint vectors are

$$(0, 1), \ (0, -\infty), \ \text{and } (a_i, 1 - b_i)_{i=1}^n.$$

The vertices $(0,1)$, $(0,-\infty)$ and the origin can be removed from the definition of P using the elementary observation that if $a \in E$ then the number of edges of $\mathrm{conv}(P \cup a) \cap E$ is at most the number of edges of $P \cap E$ plus 2. Since $(0,1)$, $(0,-\infty)$ and 0 do lie in E, they can be ignored at the cost of increasing the number of edges by 6. Thus we can assume that

$$P = \mathrm{conv}(a_i, 1 - b_i)_{i=1}^n,$$

where the vectors $(a_i, 1 - b_i)$ are independent Gaussian vectors with centers of norm at most 2, and with standard deviation σ.

Scaling these vectors down so that their norms become at most 1, we deduce the desired size of the section $P \cap E$ from the following theorem. It gives a desired bound for the number of pivots when we solve (Int LP).

Theorem 6.2 (Sections of random polytopes). *Let $a_1, \ldots, a_n$ be independent Gaussian vectors in $\mathbb{R}^d$ with centers of norm at most 1, and whose standard deviation σ satisfies (6.1). Let E be a plane in $\mathbb{R}^d$. Then the random polytope $P = \mathrm{conv}(a_1, \ldots, a_n)$ satisfies*

$$\mathbb{E} \, |\operatorname{edges}(P \cap E)| \leq C d^3 \sigma^{-4}, \tag{6.2}$$

where C is an absolute constant.

Proof. See Section 7. ∎

Remark. Spielman and Teng obtained a weaker estimate $O(nd^3\sigma^{-6})$ for this size ([6] Theorem 4.0.1). Because of the polynomial, rather than a polylogarithmic, dependence on n, their bound is not sufficient for us.

Summarizing, we have shown that:

$$\textit{(Int LP) makes in expectation } D(d, \sigma) + 6 \textit{ pivot steps}, \tag{6.3}$$

where $D(d, \sigma)$ denotes the right hand side in (6.2).

6.2. **Sections of a random polytope with an added facet.** When we repeatedly apply the polar shadow-vertex simplex method for (Unit LP$^+$) (in SOLVER FOR (UNIT LP)), each time we do so with U chosen randomly and independently of everything else. Let us condition on a choice of U. Then the plane E is fixed:

$$E = \mathrm{span}\left(\frac{z_0}{\|z_0\|}, z\right) = \mathrm{span}(U z_0', z).$$

The constraint vectors are

$$a_1, \ldots, a_n, \ a_{n+1}, \ldots, a_{n+d}.$$

The first n of these are independent Gaussian vectors with centers of norm at most 1 and whose standard deviation σ satisfies (6.1).

The last d are also Gaussian vectors chosen independently with centers $2M_0\tilde{a}_i$ and variance $2M_0\sigma_1$, where $\tilde{a}_i(= U\bar{a}'_i)$ are fixed vectors of norm

$$\|\tilde{a}\| = \|\bar{a}'_i\| = \sqrt{\|z'_0\|^2 + \|z'_0 - \bar{a}'_i\|^2} = \sqrt{1 + \ell^2} \le 1.01.$$

(Here we used the orthogonality of z'_0 and $z'_0 - a'_i$, which holds by the construction of these vectors in algorithm ADDING CONSTRAINTS).

Recall that $M_0 = e^{\lceil \log M \rceil}$, where $M = \max_{i=1,\ldots,n} \|a_i\|$ and σ_1 is as in (5.1).

Let $\Phi(a_1,\ldots,a_{n+d})$ denote the density of such vectors as above. One should note that the last d of vectors correlate with the first n vectors through the random variable M_0. This difficulty will be resolved by an argument similar to that in [6]. We will show that with high probability, M_0 takes values in a small finite set. For each fixed M_0 in this set, all the vectors $a_1,\ldots,a_{n+d}$ are independent, so we can use Theorem 6.2 to get the desired size of the section $P \cap E$:

Corollary 6.3 (Sections of random polytopes with an added facet). *Let $a_1,\ldots,a_{n+d}$ be random vectors in $\mathbb{R}^d$ with joint distribution $\Phi(a_1,\ldots,a_{n+d})$. Let E be a plane in $\mathbb{R}^d$. Then the random polytope $P = \mathrm{conv}(a_1,\ldots,a_{n+d})$ satisfies*

$$\mathbb{E}\,|\,\mathrm{edges}(P \cap E)| \le C\log\log n \cdot D(d,\sigma_0)$$

where $D(d,\sigma)$ denotes the right hand side of (6.2), where

$$\sigma_0 = c\log^{-1/2} n \cdot \min(\sigma,\sigma_1),$$

and where c is an absolute constant.

Proof. See Section 8. ∎

Thus, similarly to the previous section, we have shown that:

$$(\text{Unit } LP^+) \text{ makes in expectation } C\log\log n \cdot D(d,\sigma_0) + 6 \text{ pivot steps.} \qquad (6.4)$$

6.3. **The total number of pivot steps.** As we mentioned in the beginning of this section, the total number of

$$\text{pivot steps in SOLVER FOR (LP)} = Y + Z, \qquad (6.5)$$

where Y is the number of pivot steps to solve (Int LP), and Z is the total number of pivot steps to solve (Unit LP^+) over all iterations SOLVER FOR (UNIT LP) makes.

Then (6.3) states that

$$\mathbb{E}Y \le D(d,\sigma) + 6. \qquad (6.6)$$

Furthermore, the expected number of iterations in SOLVER FOR (UNIT LP) is at most four. Then (6.4) yields a good bound for Z. This is rigorously proved by the following simple stopping time argument.

Consider a variant of SOLVER FOR (UNIT LP), from which the stopping condition is removed, i. e. which repeatedly applies phase-I and phase-II in an infinite loop. Let Z_k denote the number of pivot steps in phase-II of this algorithm in k-th iteration, and F_k denote the random variable which is 1 if k-th iteration in this algorithm results in failure, and 0 otherwise. Then the expected total number of pivot steps made in the actual SOLVER FOR (UNIT LP), over all iterations, is distributed identically with

$$Z \equiv \sum_{k=1}^{\infty} Z_k \prod_{j=1}^{k-1} F_j.$$

To bound the expectation of Z, we denote by $\mathbb{E}_0$ the expectation with respect to random (smoothed) vectors $(a_1, \ldots, a_n)$, and by $\mathbb{E}_j$ the expectation with respect to the random choice made in j-th iteration of SOLVER FOR (UNIT LP), i. e. the choice of U and of $(a_{n+1}, \ldots, a_{n+d})$.

Let us first condition on the choice of $(a_1, \ldots, a_n)$. This fixes the numb set, which makes each F_j depend only on the random choice made in j-th iteration, while Z_k will only depend on the random choice made in k-th iteration. Therefore

$$\mathbb{E}Z = \mathbb{E}_0 \sum_{k=1}^{\infty} (\mathbb{E}_k Z_k) \prod_{j=1}^{k-1} \mathbb{E}_j F_j. \tag{6.7}$$

As observed above,

$$\mathbb{E}_j F_j = \mathbb{P}(F_j = 1) \leq 3/4,$$

which bounds the product in (6.7) by $(3/4)^k$. Moreover, $\mathbb{E}_0 \mathbb{E}_k Z_k$ are equal to the same value for all k because of the identical distribution. Thus

$$\mathbb{E}_0 \mathbb{E}_k Z_k \leq \max_U \mathbb{E}_\Phi Z_1,$$

where E_Φ is the expectation with respect to the random vectors $(a_1, \ldots, a_{n+d})$ conditioned on a choice of U in the first iteration. Thus

$$\mathbb{E}Z \leq 4 \max_U \mathbb{E}_\Phi Z_1.$$

The random vectors $(a_1, \ldots, a_{n+d})$ have joint density Φ, and we can apply (6.4) to bound

$$\max_U \mathbb{E}_\Phi Z_1 \leq 2C \log \log n \cdot D(d, \sigma_0).$$

Summarizing,

$$\mathbb{E}Z \leq 8C \log \log n \cdot D(d, \sigma_0).$$

This, (6.6), and (6.5) imply that the expected number of pivot steps in SOLVER FOR (LP) is at most

$$D(d, \sigma) + 6 + 8C \log \log n \cdot D(d, \sigma_0) \leq C_1 \log^2 n \cdot \log \log n \cdot d^3 (\sigma^{-4} + d^2 \log^2 n + d^6 \log^4 d).$$

This proves Theorem 6.1 and completes the smoothed analysis of the simplex method.

7. SECTIONS OF RANDOM POLYTOPES WITH I.I.D. VERTICES

The goal of this section is to prove Theorem 6.2 about the size of planar sections of random polytopes.

Our argument improves upon the part of the argument of [6] where it looses a factor of n. Recall that we need a polylogarithmic dependence on n.

In Section 7.1, we will outline the counting argument of [6], the crucial idea of which is to reduce the counting problem to the geometric problem of bounding a fixed point away from the boundary of a random simplex. In Section 7.2, we will give a simple "three viewpoints" argument that allows us to not loose a factor of n in contrast to the original argument of [6]. In further sections, we revisit the counting argument and complete the proof.

7.1. Spielman-Teng's counting argument.

We start by outlining the counting argument of [6]. This argument leads to a loss of a factor linear in n. We shall show then how to improve this to a polylogarithmic factor.

We consider the one-dimensional torus in the plane E defined as $\mathbb{T} = E \cap S^{d-1}$. We parametrize it by

$$q = q(\theta) = z\sin(\theta) + t\cos(\theta), \qquad \theta \in [0, 2\pi),$$

where z, t are some fixed orthonormal vectors in E.

Assume for now that the origin always lies in the interior of the polytope P; we will get rid of this minor assumption later. Then clearly

$$\text{Exp} := \mathbb{E}|\operatorname{edges}(P \cap E)| = \mathbb{E}|\{\operatorname{facet}_P(q) : \ q \in \mathbb{T}\}|.$$

In order to turn this into finite counting, we quantize the torus $\mathbb{T}$ by considering

$$\mathbb{T}_m := \{m \text{ equispaced points in } \mathbb{T}\}.$$

A simple discretization argument (see [6] Lemma 4.0.6) gives that

$$\mathbb{E}|\{\operatorname{facet}_P(q) : \ q \in \mathbb{T}\}| = \lim_{m \to \infty} \mathbb{E}|\{\operatorname{facet}_P(q) : \ q \in \mathbb{T}_m\}| \qquad (7.1)$$
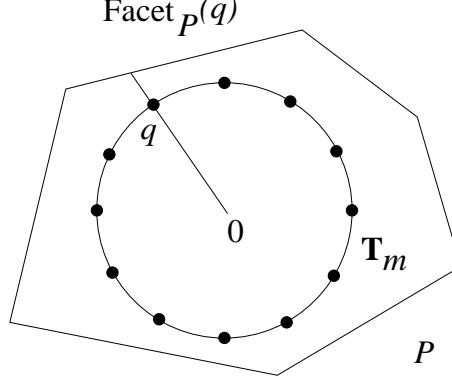
For the proper counting, one would prefer to keep one point q per facet, for example the one closest to the boundary of the facet. The closeness here is measured with respect to the angular distance:

Definition 7.1 (Angular distance). *Let x and y be two vectors in $\mathbb{R}^d$. The* angular distance

$$\operatorname{ang}(x, y) := \cos^{-1}\left(\frac{|\langle x, y\rangle|}{\|x\|\|y\|}\right)$$

is the angle formed by the vectors x and y. The angular length $\operatorname{ang}(\mathcal{I})$ *of an interval $\mathcal{I}$ in $\mathbb{R}^d$ is the angular distance between its endpoints.*

So in the counting formula (7.1), we can leave one q per facet, namely the $q = q(\theta)$ with the maximal θ (according to the parametrization of the torus in the beginning of the argument). Therefore, the angular distance of such q to the boundary of $\operatorname{Facet}_{P\cap E}(q)$ (one of the endpoints of this interval), and hence also to the boundary of $\operatorname{Facet}_P(q)$, is at most $2\pi/m$.

We have proved that

$$\operatorname{Exp} \leq \lim_{m\to\infty} \mathbb{E}|\{\operatorname{facet}_P(q): \operatorname{ang}\left(q, \partial\operatorname{Facet}_P(q)\right) \leq \frac{2\pi}{m}, \ q \in \mathbb{T}_m\}|.$$

With a slight abuse of notation, we shall denote the set of all d-subsets of $\{1, \ldots, n\}$ by $\binom{n}{d}$. Then

$$\operatorname{Exp} \leq \lim_{m\to\infty} \sum_{q\in\mathbb{T}_m} \sum_{I\in\binom{n}{d}} \mathbb{P}\{\operatorname{facet}_P(q) = I \text{ and } \operatorname{ang}\left(q, \partial\triangle(a_i)_{i\in I}\right) \leq \frac{2\pi}{m}\}. \qquad (7.2)$$

For every q, there exists at most one $I \in \binom{n}{m}$ such that $\operatorname{facet}_P(q) = I$. Hence

$$\sum_{I\in\binom{n}{d}} \mathbb{P}\{\operatorname{facet}_P(q) = I\} \leq 1$$

Using this bound in (7.2) and estimating $\sum_{q\in\mathbb{T}_m}$ above by $m \cdot \max_{q\in\mathbb{T}_m}$, we obtain

$$\operatorname{Exp} \leq \lim_{m\to\infty} m \cdot p(m), \qquad (7.3)$$

where

$$p(m) = \max_{q \in \mathbb{T}_m, I \in \binom{n}{d}} \mathbb{P}\{\operatorname{ang}\left(q, \partial \triangle(a_i)_{i \in I}\right) \le \frac{2\pi}{m} \mid \operatorname{facet}_P(q) = I\}.$$

Furthermore, one can get rid of the polytope P by analyzing the event $\operatorname{facet}_P(q) = I$. We condition on the realization on the points $(a_i)_{i \in I^c}$, as well as on the subspace $E_I :=$ $\operatorname{span}(a_i)_{i \in I}$. The randomness remains in the random points a_i, $i \in I$ inside the (fixed) subspace E_I. Then the event $\operatorname{facet}_P(q) = I$ holds if and only if the (fixed) point q_P where the direction of q pierces E_I lies in the the simplex $\triangle(a_i)_{i \in I}$.

We have thus reduced the problem to *estimating the distance of a fixed point to the boundary of a random simplex, conditioned on the event that the point lies in that simplex.*

The main difficulty is that the distance is angular rather than Euclidean; the latter is easier to estimate. Unfortunately, the two distances may be very different. This happens if the *angle of incidence* – the angle at which the direction of q meets the subspace E_I – is too small. So Spielman and Teng needed to show that the angle of incidence is at least of order $1/n$ with high probability (Section 4.2 in [6]); consequently, the angular and Euclidean distances are within a factor $O(n)$ from each other.

In this paper, we can not tolerate the loss of a factor of n since we are proving a complexity estimate that is polylogarithmic rather than polynomial in n. We will now present a simple way to avoid such a loss.

7.2. **Three viewpoints.** Instead of estimating the angle of incidence from one viewpoint determined by the origin 0, we will *view the polytope P_0 from three different points 0_1, 0_2, 0_3 on E.* Vectors q will be emitted from each of these points, and from at least one of them the angle of incidence will be good (more precisely, the angle of q to the intersection of its facet with E will be good). This is formalized in the following two elementary observations on the plane.

Lemma 7.2 (Three viewpoints). *Let $K = \operatorname{conv}(b_1, \ldots, b_N)$ be a planar polygon, where points b_i are in general position and have norms at most 1. Let 0_1, 0_2, 0_3 be the vertices of an equilateral triangle centered at the origin and with norm 4. Denote $K_i = \operatorname{conv}(0_i, K)$. Then, for every edge (b_k, b_m) of K, there exists $i \in \{1, 2, 3\}$ such that (b_k, b_m) is an edge of K_i, and*

$$\operatorname{dist}(0_i, \operatorname{aff}(b_k, b_m)) \ge 1.$$
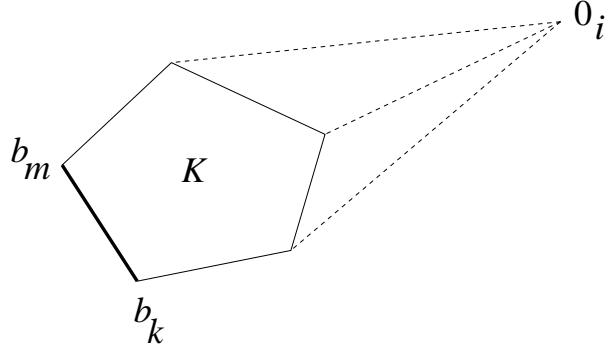
Proof. Let L be any line passing through the origin. Then, for every equilateral triangle centered at the origin and whose vertices have norms R, there exist two vertices separated by the line L and whose distances to L are at least $R/2$. (The bound $R/2$ is attained if L is parallel to one of the sides of the triangle).

Let L be the line passing through the origin and parallel to the edge (b_k, b_m). It follows that among the three points 0_1, 0_2, 0_3 there exists at least two (say, 0_1 and 0_2) separated by the line L and such that

$$\operatorname{dist}(0_i, L) \geq 4/2 = 2, \qquad i = 1, 2.$$

Moreover, since all the points b_i have norms at most 1, we have

$$\operatorname{dist}(L, \operatorname{aff}(b_k, b_m)) \leq \operatorname{dist}(0, b_k) \leq 1.$$

Then by the triangle inequality, 0_1 and 0_2 are separated by the line $\operatorname{aff}(b_k, b_m)$ and

$$\operatorname{dist}(0_i, \operatorname{aff}(b_k, b_m)) > 1, \qquad i = 1, 2.$$

Since 0_1 and 0_2 are separated by the affine span of the edge (b_k, b_m) of the polygon K, one of these points (say, 0_1) lies on the same side from $\operatorname{aff}(b_k, b_m)$ as the polygon K. It follows that (b_k, b_m) is an edge of $\operatorname{conv}(0_1, K)$. This completes the proof. ∎

Lemma 7.3 (Angular and Euclidean distances)**.** *Let L be a line in the plane such that*

$$\operatorname{dist}(0, L) \geq 1.$$

Then, for every pair of points x_1, x_2 on L of norm at most 10, one has

$$c \operatorname{dist}(x_1, x_2) \leq \operatorname{ang}(x_1, x_2) \leq \operatorname{dist}(x_1, x_2)$$

where $c = (10^2 + 1)^{-1}$.

Proof. Without loss of generality, we may assume that $\operatorname{dist}(0, L) = 1$. Choose unit vectors u and v in the plane such that $\langle u, v \rangle = 0$ and $L = \{u + tv : t \in \mathbb{R}\}$. Then

$$x_1 = u + t_1 v, \quad x_2 = u + t_2 v$$

for some $t_1, t_2 \in \mathbb{R}$.
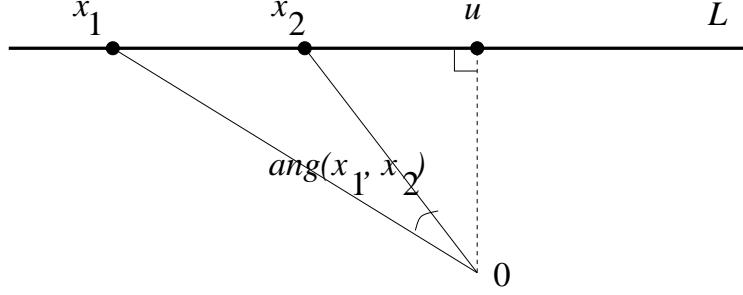
Hence we have

$$|t_1| \leq \|x_1\| \leq 10, \quad |t_2| \leq \|x_2\| \leq 10.$$

Without loss of generality, $t_1 \leq t_2$. Then one easily checks that

$$\operatorname{dist}(x_1, x_2) = t_2 - t_1, \qquad \operatorname{ang}(x_1, x_2) = \tan^{-1} t_2 - \tan^{-1} t_1.$$

Thus

$$\operatorname{ang}(x_1, x_2) = \int_{t_1}^{t_2} \frac{d}{dt}(\tan^{-1} t) \, dt = \int_{t_1}^{t_2} \frac{dt}{1 + t^2}.$$

We can estimate the integrand from both sides using the inequalities $1 \leq 1 + t^2 \leq 1 + t_2^2 \leq 1 + 10^2$. Thus

$$\frac{t_2 - t_1}{1 + 10^2} \leq \operatorname{ang}(x_1, x_2) \leq t_2 - t_1.$$

This completes the proof.   ∎

In view of Lemma 7.3, we can rephrase Lemma 7.2 as follows: every edge (facet) of K can be viewed from one of the three viewpoints 0_1, 0_2, 0_3 at a nontrivial angle, and yet remain an edge of the corresponding polygon $\operatorname{conv}(0_i, K)$.

7.3. **Boundedness of the random polytope.** In Lemmas 7.2 and 7.3, the boundedness requirements (for the points x_i and b_i respectively) are clearly essential. To make sure that these requirements are satisfied in our setting, we recall that the vertices of the polytope P are i.i.d. Gaussian vectors. We shall therefore use the following well-known estimate on the size of a Gaussian vector, see e.g. Proposition 2.4.7 in [6].

Lemma 7.4. *Let g be a Gaussian vector in $\mathbb{R}^n$ ($n \geq 3$) with center $\bar{g}$ of norm at most 1, and with variance σ. Then:*
   1. We have $\mathbb{P}\{\|g - \bar{g}\| \geq 3\sigma\sqrt{d \log n}\} \leq n^{-2.9d}$.
   2. We have $\mathbb{P}\{\|g\| \leq c\sigma\sqrt{d}\} \leq e^{-d}$, where $c = e^{-3/2}$.

We consider the event

$$\mathcal{E} := \{\|a_i\| \leq 2, \ i = 1, \ldots, n\}.$$

By Lemma 7.4 and using our assumption (6.1), we have

$$\mathbb{P}\{\mathcal{E}^c\} \leq \mathbb{P}\{\max_{i=1,\dots,n} \|a_i\| > 3\sigma\sqrt{d\log n} + 1\} \leq n^{-2.9d+1} \leq 0.0015 \binom{n}{d}^{-1}. \qquad (7.4)$$

Our goal in Theorem 6.2 is to estimate

$$\mathrm{Exp} := \mathbb{E}\,|\,\mathrm{edges}(P \cap E)|.$$

The random variable $|\,\mathrm{edges}(P \cap E)|$ is bounded above by $\binom{n}{d}$, which is the maximal number of facets of P. Using (7.4), it follows that

$$\mathrm{Exp} := \mathbb{E}\,|\,\mathrm{edges}(P \cap E)| \leq \mathbb{E}\,|\,\mathrm{edges}(P \cap E)| \cdot \mathbf{1}_{\mathcal{E}} + 1.$$

We will use a similar intersection argument several times in the sequel.

7.4. **Counting argument revisited from three viewpoints.** We will apply Lemma 7.2 in combination with 7.3 for the random polygon $P \cap E$, whenever it satisfies $\mathcal{E}$. All of the points of this polygon are then bounded by 2 in norm, so we scale the result in Lemma 7.2 by the factor 2.

Let 0_1, 0_2, 0_3 be the vertices of an equilateral triangle in the plane E, centered at the origin and with

$$\|0_1\| = \|0_2\| = \|0_3\| = 8. \qquad (7.5)$$

Denote

$$P_i = \mathrm{conv}(0, -0_i + P). \qquad (7.6)$$

Lemma 7.2 states in particular that, if $\mathcal{E}$ holds, then each edge of $P \cap E$ can be seen as an edge from one of the three viewpoints. Precisely, there is a one-to-one correspondence[1] between the edges of $P \cap E$ and the set $\{\mathrm{facet}_{P_i \cap E}(q) : q \in \mathbb{T}, i = 1, 2, 3\}$. We can further replace this set by $\{\mathrm{facet}_{P_i}(q) : q \in \mathbb{T}, i = 1, 2, 3\}$, since each $\mathrm{facet}_{P_i}(q)$ uniquely determines the edge $\mathrm{facet}_{P_i \cap E}(q)$; and vice versa, each edge can belong to a unique facet. Therefore

$$\mathrm{Exp} \leq \mathbb{E}\,|\{\mathrm{facet}_{P_i}(q) : q \in \mathbb{T}, i = 1, 2, 3\}| + 1 \qquad (7.7)$$

and, by a discretization in limit as in Section 7.1,

$$= \lim_{m\to\infty} \mathbb{E}\,|\{\mathrm{facet}_{P_i}(q) : q \in \mathbb{T}_m, i = 1, 2, 3\}| + 1. \qquad (7.8)$$

Moreover, by the same discretization argument, we may ignore in (7.7) all facets whose intersection with E have angular length no bigger than, say, $2\pi/m$. After this, we replace

---

[1]Here and in the sequel, we identify $\mathrm{facet}(q)$ with the index set it contains. Since the polytope in question is almost surely in general position, $\mathrm{facet}(q)$ contains at most one index set.

P_i by $P_i \cap E$ as we mentioned above, and intersect with the event $\mathcal{E}$ again, as before. This gives

$$\text{Exp} \leq \lim_{m \to \infty} \mathbb{E} \,|\{\text{facet}_{P_i \cap E}(q) \text{ of angular length} > \frac{2\pi}{m}; \; q \in \mathbb{T}_m, \; i = 1, 2, 3\}| \cdot \mathbf{1}_{\mathcal{E}} + 2. \quad (7.9)$$

We are going to apply Lemmas 7.2 and 7.3 for a realization of P for which the event $\mathcal{E}$ holds. Consider any facet from the set in (7.9). So let

$$I = \text{facet}_{P_i \cap E}(q) \quad \text{for some } i \in \{1, 2, 3\} \text{ and some } q \in \mathbb{T}_m.$$

By Lemma 7.2, we can choose a viewpoint 0_i which realizes this facet and from which its intersection with E is seen at a good angle. Formally, among the indices $i_0 \in \{1, 2, 3\}$ such that $I = \text{facet}_{P_{i_0} \cap E}(q_0)$ for some $q_0 \in \mathbb{T}_m$, we choose the one that maximizes the distance from 0 to the affine span of the edge

$$\mathcal{I} = \text{Facet}_{P_{i_0} \cap E}(I).$$

By Lemma 7.2,

$$\text{dist}(0, \text{aff}(\mathcal{I})) \geq 1. \tag{7.10}$$

Because only facets of angular length $> 2\pi/m$ were included in the set in (7.9), we have $\text{ang}(\mathcal{I}) > 2\pi/m$. It follows that $\mathcal{I}$ contains some point q'' in $\mathbb{T}_m$.

Summarizing, we have realized every facet $I = \text{facet}_{P_i \cap E}(q)$ from (7.9) as $I = \text{facet}_{P_{i_0} \cap E}(q'')$ for some i_0 and some $q'' \in \mathbb{T}_m$.

Recall that when the event $\mathcal{E}$ holds, all points of P have norm at most 2. Thus all points of P_{i_0} have norm at most $\|0_{i_0}\| + 2 = 10$. Since $\mathcal{I} \subset P_{i_0}$, all points in $\mathcal{I}$ also have norm at most 10. Therefore bound (7.10) yields, by view of Lemma 7.3, that *the angular and Euclidean distances are equivalent on $\mathcal{I}$ up to a factor of c.*

We shall call a facet (edge) of a polygon nondegenerate if the angular and Euclidean distances are equivalent on it up to the factor c. We have shown that

$$\text{Exp} \leq \lim_{m \to \infty} \mathbb{E} \,|\{\text{nondegenerate facet}_{P_i \cap E}(q) : \; q \in \mathbb{T}_m, \; i = 1, 2, 3\}| \cdot \mathbf{1}_{\mathcal{E}} + 2.$$

Each facet may correspond to more than one q. We are going to leave only one q per facet, namely the $q = q(\theta)$ with the maximal θ (according to the parametrization of the torus in the beginning of the argument). Therefore, the angular distance of such q to one boundary of $\text{Facet}_{P_i \cap E}(q)$ (one of the endpoints of this interval) is at most $2\pi/m$. The nondegeneracy of this facet then implies that the usual distance of q_{P_i} to the boundary of $\text{Facet}_{P_i \cap E}(q)$, thus also to the boundary of $\text{Facet}_{P_i}(q)$, is at most $\frac{1}{c} \cdot 2\pi/m =: C/m$. Therefore

$$\text{Exp} \leq \lim_{m \to \infty} \mathbb{E} \,|\{\text{facet}_{P_i}(q) \text{ such that } \text{dist}(q_{P_i}, \partial \,\text{Facet}_{P_i}(q)) \leq \frac{C}{m}, \; q \in \mathbb{T}_m, \; i = 1, 2, 3\}| \cdot \mathbf{1}_{\mathcal{E}} + 2$$

$$\leq 3 \max_{i=1,2,3} \lim_{m \to \infty} \mathbb{E} \,|\{\text{facet}_{P_i}(q) \text{ such that } \text{dist}(q_{P_i}, \partial \,\text{Facet}_{P_i}(q)) \leq \frac{C}{m}, \; q \in \mathbb{T}_m\}| \cdot \mathbf{1}_{\mathcal{E}} + 2.$$

Recall that by (7.6) and (7.5), the polytope P_k is the convex hull of the origin and a translate of the polytope P by a fixed vector of norm 8. Therefore

$$\text{Exp} \leq 3 \max \text{Exp}_0 + 2, \tag{7.11}$$

where

$$\text{Exp}_0 = \lim_{m \to \infty} \mathbb{E} \left| \left\{ \text{facet}_P(q) \text{ such that } \text{dist}(q_P, \partial \text{Facet}_P(q)) \leq \frac{C}{m}, \ q \in \mathbb{T}_m \right\} \right| \cdot \mathbf{1}_{\mathcal{E}} \tag{7.12}$$

and the maximum in (7.11) is over all centers of the distributions a_i of norm at most $1 + 8 = 9$. Scaling the vectors down by 9, we can bound Exp_0 using the following lemma:

Lemma 7.5 (Discretized counting). *Let $a_1, \ldots, a_n$ be independent Gaussian vectors in $\mathbb{R}^d$ with centers of norm at most 1, and whose standard deviation σ satisfies (6.1). Then*

$$\text{Exp}_0 \leq C_0 d^3 \sigma^{-4},$$

where C_0 is an absolute constant.

This lemma and (7.11) complete the proof of Theorem 6.2.

7.5. **Proof of Lemma 7.5.** As before, $\binom{n}{d}$ will denote the set of all d-subsets of $\{1, \ldots, n\}$. We have

$$\text{Exp}_0 = \lim_{m \to \infty} \sum_{q \in \mathbb{T}_m} \sum_{I \in \binom{n}{d}} \mathbb{P} \left\{ \text{facet}_P(q) = I \text{ and } \text{dist}(q_P, \partial \triangle(a_i)_{i \in I}) \leq \frac{C}{m} \text{ and } \mathcal{E} \right\}. \tag{7.13}$$

For every q, there exists at most one $I \in \binom{n}{m}$ such that $\text{facet}_P(q) = I$. Furthermore, the last equation is equivalent to $q_P \in \triangle(a_i)_{i \in I}$. Hence

$$\sum_{I \in \binom{n}{d}} \mathbb{P}\{\text{facet}_P(q) = I\} = \sum_{I \in \binom{n}{d}} \mathbb{P}\{q_P \in \triangle(a_i)_{i \in I}\} \leq 1.$$

Using this bound in (7.13) and estimating $\sum_{q \in \mathbb{T}_m}$ above by $m \cdot \max_{q \in \mathbb{T}_m}$, we obtain

$$\text{Exp}_0 \leq \lim_{m \to \infty} m \cdot p_0(m), \tag{7.14}$$

where

$$p_0(m) = \max_{q \in \mathbb{T}_m, I \in \binom{n}{d}} \mathbb{P}\{\text{dist}(q_P, \partial \triangle(a_i)_{i \in I}) \leq C/m \text{ and } \mathcal{E} \mid q_P \in \triangle(a_i)_{i \in I}\}.$$

Thus we should be looking for an estimate of the type $p_0(m) \lesssim 1/m$.

To this end, we fix an arbitrary point $q \in \mathbb{T}^m$ and a set $I \in \binom{n}{d}$. Consider the hyperplane

$$E_I := \text{span}(a_i)_{i \in I}$$

and the point

$$q_I := \text{point where the direction of } q \text{ pierces the hyperplane } E_I.$$

Then

$$\text{facet}_P(q) = I \Leftrightarrow \begin{cases} q_I \in \triangle(a_i)_{i \in I}; \\ \text{all vectors } (a_i)_{i \in I^c} \text{ are below } E_I. \end{cases} \tag{7.15}$$

We now pass to the local coordinates $(b_i)_{i \in I}$ for the hyperplane E_I, using the change of variables

$$a_i = R_\omega b_i + r\omega, \qquad i \in I,$$

described in Appendix C.

We condition on a realization of r, ω and the vectors $(a_i)_{i \in I^c}$. This fixes the hyperplane E_I and the point q_I determined by r and ω. The density of the vectors $(b_i)_{i \in I}$ is given in Lemma C.1 in Appendix C. By (7.15), we can assume that the vectors $(a_i)_{i \in I^c}$ are below E_I.

Let $p \in \mathbb{R}^{d-1}$ be the (fixed) representation of q_I in the new variables, i.e.

$$q_I = R_\omega p + r\omega.$$

Consider the event

$$\mathcal{E}_0 := \{\|b_i\| \le 2, \, i = 1, \dots, n\}$$

By part 1 of Lemma C.1,

$$\mathcal{E} \subseteq \mathcal{E}_0.$$

By (7.15),

$$\text{facet}_P(q) = I \Leftrightarrow q_I \in \triangle(a_i)_{i \in I} \Leftrightarrow p \in \triangle(b_i)_{i \in I}.$$

Moreover, if $p \in \triangle(b_i)_{i \in I}$ then $\|p\| \le \max_{i \in I} \|b_i\|$, thus

$$\{\mathcal{E} \text{ and } \text{facet}_P(q) = I\} \subseteq \{\mathcal{E}_0 \text{ and } p \in \triangle(b_i)_{i \in I}\} \subseteq \{\|p\| \le 2\}.$$

Summarizing, we have shown that

$$p_0(m) \le \max \mathbb{P}\{\text{dist}(p, \partial \triangle(b_i)_{i \in I}) \le C/m \text{ and } \mathcal{E}_0 \mid p \in \triangle(b_i)_{i \in I}\},$$

where the maximum is over all vectors $p \in \mathbb{R}^{d-1}$ such that $\|p\| \le 2$. We may assume that $p = 0$ by translating all the vectors by $-p$. Before this translation, the densities ν_i that make up the density of $(b_i)_{i \in I}$ in (C.2) had centers of norm at most 1 by Lemma C.1. After the translation, their norms will be at most $1 + \|p\| \le 3$. Similarly, the constant 2 in the definition of $\mathcal{E}_0$ will change to $2 + \|p\| \le 4$.

It remains to use the Distance Lemma 4.1.2 of [6]:

Lemma 7.6 (Distance Lemma (Spielman-Teng)). *Let $\nu_1, \dots, \nu_d$ be densities of Gaussian vectors in $\mathbb{R}^{d-1}$ with centers of norm at most 3 and with standard deviation $\sigma < 1/3\sqrt{d \log n}$. Then, for every $\varepsilon > 0$, we have*

$$\mathbb{P}\{\text{dist}(0, \text{aff}(b_2, \dots, b_d)) < \varepsilon \text{ and all } \|b_i\| \le 4\} \le C_1 d^2 \sigma^{-4} \varepsilon,$$

where $(b_1, \ldots, b_d)$ have joint density proportional to

$$| \triangle(b_1, \ldots, b_d)| \cdot \prod_{i=1}^{d} \nu_i(b_i),$$

and C_1 is an absolute constant.

Note that $\operatorname{dist}(0, \partial \triangle(b_i)_{i \in I}) \leq C/m$ implies that there exists $i_0 \in I$ such that

$$\operatorname{dist}(0, \operatorname{aff}(b_i)_{i \in I - \{i_0\}}) \leq C/m.$$

Since there are d choices for i_0, the Distance Lemma 7.6 yields:

$$p_0(m) \leq d \cdot C_1 d^2 \sigma^{-4}(C/m) \leq C_2 d^3 \sigma^{-4}/m,$$

where $C_2 = C_1 C$. By (7.14), we conclude that

$$\operatorname{Exp}_0 = O(d^3 \sigma^{-4}).$$

This completes the proof. ∎

8. Sections of random polytopes with an added facet

In this section, we prove Corollary 6.3 about the size of planar sections of random polytopes with an added facet.

The main difficulty is that not all vectors $(a_1, \ldots, a_{n+d})$ are independent. The last d vectors correlate with the first n vectors through the random variable $M_0 = e^{\lceil \log M \rceil}$, where $M = \max_{i=1,\ldots,n} \|a_i\|$.

This difficulty will be resolved similarly to [6]. We will show that with high probability, M_0 takes values in a set of cardinality $O(\log \log n)$. For each fixed M_0 in this set, all the vectors $a_1, \ldots, a_{n+d}$ are independent, so we will be able to use Theorem 6.2 to get the desired size of the section $P \cap E$.

8.1. Boundedness of the random polytope. Consider

$$\bar{M} = \max_{i=1,\ldots,n} \|\bar{a}_i\|, \qquad M = \max_{i=1,\ldots,n} \|a_i\|.$$

Recall that by (6.1), $\bar{M} \leq 1$.

Lemma 8.1. *Consider the event*

$$\mathcal{E}_1 := \left\{ \frac{c_1}{\sqrt{\log n}} \big(\bar{M} + \sigma\sqrt{d \log n}\big) \leq M \leq \bar{M} + 3\sigma\sqrt{d \log n} \right\}, \tag{8.1}$$

where $c_1 = c/9$, and where c is the absolute constant in Lemma 7.4. Then

$$\mathbb{P}\{\mathcal{E}_1^c\} \leq \binom{n}{d}^{-1},$$

Proof.   *Upper bound.* By part 1 of Lemma 7.4,

$$\mathbb{P}\{|M - \bar{M}| > 3\sigma\sqrt{d\log n}\} \leq \{\max_{i=1,\ldots,n} \|a_i - \bar{a}_i\| > 3\sigma\sqrt{d\log n}\} \leq n^{-2.9d+1} \leq 0.0015\binom{n}{d}^{-1}.$$

Lower bound. We consider two cases. If $\bar{M} \geq 8\sigma\sqrt{d\log n}$ then, using the last estimate, we have:

$$\mathbb{P}\{M < \frac{1}{2}(\bar{M} + \sigma\sqrt{d\log n})\} \leq \mathbb{P}\{M < \bar{M} - 3\sigma\sqrt{d\log n}\} \leq 0.0015\binom{n}{d}^{-1}.$$

If $\bar{M} < 8\sigma\sqrt{d\log n}$ then, using part 2 of Lemma 7.4 and the independence of a_i, we obtain

$$\mathbb{P}\{M < \frac{c_1}{\sqrt{\log n}}(\bar{M} + \sigma\sqrt{d\log n})\} \leq \mathbb{P}\{M < 9c_1\sigma\sqrt{d}\} \leq (e^{-d})^n \leq 0.5\binom{n}{d}^{-1}.$$

This completes the proof.   ■

8.2. **Discretization of M.** Recall that $M_0 = e^{\lceil \log M \rceil}$. It follows from Lemma 8.1 that M_0 is likely to take values in a small set.

Indeed, if $\mathcal{E}_1$ holds, then M_0 takes values in a (non-random) set

$$\mathcal{M}_0 := \{e^{\lceil \log M \rceil} : \ M \text{ satisfies the inequalities in (8.1)}\}$$

of cardinality $|\mathcal{M}_0| = O(\log\log n)$.

Therefore, using as before that the number of edges of the polytope $P \cap E$ is bounded by $\binom{n}{d}$, we have:

$$\begin{aligned}
\mathbb{E}|\text{edges}(P \cap E)| &\leq \mathbb{E}|\text{edges}(P \cap E)| \cdot \mathbf{1}_{\mathcal{E}_1} + 1 \\
&= \sum_{m \in \mathcal{M}_0} \mathbb{E}|\text{edges}(P \cap E)| \cdot \mathbf{1}_{\{M_0 = m\}} + 1 \\
&\leq C\log\log n \cdot \max_{M_0 \in \mathcal{M}_0} \mathbb{E}_0|\text{edges}(P \cap E)| + 1, \quad\quad (8.2)
\end{aligned}$$

where $\mathbb{E}_0$ denotes the expectation with respect to the distribution of $(a_1, \ldots, a_{n+d})$ as in the statement of Corollary 6.3 *except that the value of $M_0 \in \mathcal{M}_0$ is fixed.* According to this new distribution, all vectors $a_1, \ldots, a_{n+d}$ are now independent.

8.3. **Vertices with different standard deviations.** Now we dilate the vectors a_i so that the centers of their distributions become at most 1. To this end, define

$$b_i := \frac{a_i}{C'M_0\sqrt{\log n}}, \qquad i = 1, \ldots, n+d,$$

where $C' = 3/c_1$, where c_1 is the constant in Lemma 8.1. Let us estimate the standard deviations and the centers of the distributions of b_i.

Vectors $b_1, \ldots, b_n$. The standard deviation of $a_1, \ldots, a_n$ is σ. Hence the standard deviation of $b_1, \ldots, b_n$ is

$$\sigma' := \frac{\sigma}{C' M_0 \sqrt{\log n}}.$$

Using the lower bound in (8.1) (which holds by the definition of $\mathcal{M}_0$) and the trivial inequality $M \leq M_0$, we conclude that the standard deviation satisfies a condition of the type (6.1), namely

$$\sigma' \leq \frac{\sigma}{C' M \sqrt{\log n}} \leq \frac{\sigma}{C' c_1 \sigma \sqrt{d \log n}} \leq \frac{1}{3 \sqrt{d \log n}}. \tag{8.3}$$

On the other hand, since $\bar{M} \leq 1$ and by the upper bound in (8.1), we have $M_0 \leq 3M \leq 3\bar{M} + 9\sigma \sqrt{d \log n} \leq 3 + 9/6 \leq 5$. Thus

$$\sigma' \geq \frac{\sigma}{5 C' \sqrt{\log n}}. \tag{8.4}$$

The centers of the distributions of $a_1, \ldots, a_n$ have norms at most $\bar{M}$. Since $M \leq M_0$ and using the lower bound in (8.1), we conclude that the centers of the distributions of $b_1, \ldots, b_n$ have norms at most

$$\frac{\bar{M}}{C' M_0 \sqrt{\log n}} \leq \frac{M}{c_1 C' M_0} \leq 1.$$

Vectors $b_{n+1}, \ldots, b_{n+d}$. The standard deviation of $a_{n+1}, \ldots, a_{n+d}$ is $2 M_0 \sigma_1$. Hence the standard deviation of $b_{n+1}, \ldots, b_{n+d}$ is

$$\sigma'' := \frac{2 \sigma_1}{C' \sqrt{\log n}}. \tag{8.5}$$

Since σ_1 satisfies (5.1) and $C' \geq 1$, the standard deviation σ'' satisfies a condition of the type (6.1), namely

$$\sigma'' \leq \frac{1}{3 \sqrt{d \log n}}. \tag{8.6}$$

The centers of the distributions of $a_{n+1}, \ldots, a_{n+d}$ have norms at most $3 M_0$. Hence the centers of the distributions of $b_{n+1}, \ldots, b_{n+d}$ have norms at most

$$\frac{3}{C' \sqrt{\log n}} \leq 1.$$

Now we apply Lemma 4.3.2 from [6] which allows us to reduce standard deviations of the vertices to one (minimum) value.

Lemma 8.2 (Gaussians free)**.** *Let $d_1, \ldots, d_n$ be independent Gaussian vectors in $\mathbb{R}^d$ with standard deviations $\sigma_1, \ldots, \sigma_n$. Let $\sigma_0 > 0$. Assume that $\sigma_0 \leq \sigma_i \leq 1/3 \sqrt{d \log n}$ for all i. Let E be a plane in $\mathbb{R}^d$. Then the random polytope $P = \operatorname{conv}(d_1, \ldots, d_n)$ satisfies*

$$\mathbb{E} \, |\operatorname{edges}(P \cap E)| \leq D(d, \sigma_0) + 1.$$

where $D(d, \sigma)$ denotes the right hand side of (6.2).

We use this lemma for the vectors $b_1, \ldots, b_{n+d}$ and for

$$\sigma_0 := \min\left(\frac{\sigma}{5C'\sqrt{\log n}}, \frac{2\sigma_1}{C'\sqrt{\log n}}\right) \geq c \log^{-1/2}(n) \cdot \min(\sigma, \sigma_1),$$

where $c > 0$ is some absolute constant.

Indeed, it follows from (8.4) and (8.5) that $\sigma_0 \leq \min(\sigma', \sigma'')$. Similarly, (8.3) and (8.6) state that $\max(\sigma', \sigma'') \leq 1/3\sqrt{d \log n}$. Thus Lemma 8.2 applies, and it yields

$$\mathbb{E}_0 |\operatorname{edges}(P \cap E)| \leq D(d, \sigma_0) + 1.$$

Using this in (8.2), we conclude that

$$\mathbb{E} |\operatorname{edges}(P \cap E)| \leq C \log \log n \cdot D(d, \sigma_0) + 1.$$

This completes the proof of Corollary 6.3.

A. Appendix A. Proof of Proposition 4.1

In this section, we prove Proposition 4.1, which allows us to interpolate between an arbitrary linear program and a unit linear program.

For a fixed $t \in [0, 1]$, let (LP_t) denote the interpolation program (Int LP) with this fixed value of t. The feasible sets (polytopes) of (LP) and of (LP_t) will be denoted by P and P_t respectively. They are subsets of $\mathbb{R}^d$.

A.1. Recession cone.

Our proof of (i) of Proposition 4.1 will be based on an analysis of the *recession cone*, defined as

$$\operatorname{Recess}(LP) = \operatorname{Recess}(P) := \{x : Mx + P \subseteq P \text{ for all } M \geq 0\}.$$

The polytope P is unbounded iff its recession cone is nonempty.

Lemma A.1 (Recession cone). *Assume (LP) is feasible. Then*

$$\operatorname{Recess}(LP) = \{x : Ax \leq 0\} = \left(\operatorname{cone}(a_i)_{i=1}^n\right)^\circ$$

Proof.   The second equation is trivial.

To prove the inclusion $\operatorname{Recess}(\mathrm{LP}) \subseteq \{x : Ax \leq 0\}$, let us fix arbitrary $x \in \operatorname{Recess}(P)$ and $x_0 \in P$. By the definition of the recession cone, $Mx + x_0 \in P$ for all $M > 0$. Since P is the feasible polytope of (LP), we have

$$A(Mx + x_0) \leq b \qquad \text{for all } M > 0.$$

Hence

$$Ax \leq \frac{1}{M}(b - Ax_0) \to 0 \qquad \text{as } M \to \infty.$$

It follows that $Ax \leq 0$.

To prove the inclusion $\text{Recess}(LP) \supseteq \{x : Ax \leq 0\}$, let us fix x such that $Ax \leq 0$ and $x_0 \in P$. Since P is the feasible polytope of (LP), we have $Ax_0 \leq b$. Then for every $M \geq 0$ we have

$$A(Mx + x_0) = M \cdot Ax + Ax_0 \leq b.$$

Thus $Mx + x_0 \in P$ for every $M \geq 0$. Hence $x \in \text{Recess}(P)$. This completes the proof. ∎

Lemma A.2 (Boundedness of LP). *Assume the linear program (LP) is feasible. Then (LP) is bounded iff*

$$z \in \big(\text{Recess}(LP)\big)^\circ = \text{cone}(a_i)_{i=1}^n.$$

Proof. *Necessity.* Assume $z \notin \big(\text{Recess}(LP)\big)^\circ$. Then there exists $x \in \text{Recess}(LP)$ such that $\langle z, x \rangle > 1$. Let us fix an arbitrary $x_0 \in P$. By the definition of the recession cone, $Mx + x_0 \in P$ for every $M \geq 0$. Therefore, the vectors $Mx + x_0$ are feasible for (LP). On the other hand, the objective function can take arbitrarily large values on such vectors:

$$\langle z, Mx + x_0 \rangle = M \langle z, x \rangle + \langle z, x_0 \rangle \to \infty \qquad \text{as } M \to \infty.$$

Thus (LP) is unbounded.

Sufficiency. Assume $z \in \text{cone}(a_i)_{i=1}^n$. Then we can write $z = \sum_{i=1}^n \lambda_i a_i$ for some $\lambda_i \geq 0$. Let x be any feasible vector for (LP), that is $Ax \leq b$. Then, the objective function is bounded as

$$\langle z, x \rangle \leq \sum_{i=1}^n \lambda_i \langle a_i, x \rangle \leq \sum_{i=1}^n \lambda_i b_i.$$

Thus (LP) is bounded. This completes the proof. ∎

Remark. The main point in both lemmas is that their conclusion does not depend on the right hand side b of (LP). Thus, *for feasible linear programs, their boundedness does not depend on the right hand side b.*

A.2. **Proof of Proposition 4.1.**

Proof of (i). Since unbounded linear programs are feasible, Lemma A.2 implies that (LP) is unbounded iff (Unit LP) is unbounded.

Next, if (Unit LP) is unbounded then (Int LP) is clearly unbounded for every λ. Indeed, the feasible set of (Int LP) contains the P_0 as a section (for $t = 0$), and P_0 is the feasible set of (Unit LP).

The remaining part of (i) is: if (Int LP) is unbounded for some λ, then (Unit LP) is unbounded. We shall now prove this.

The unboundedness of (Int LP) means that for every $M \geq 1$ there exists a feasible point (x_M, t_M) for (Int LP) such that

$$\langle z, x_M \rangle + \lambda t_M \to \infty \qquad \text{as } M \to \infty. \tag{A.1}$$

Since $t_M \in [0, 1]$, the definition of (Int LP) yields

$$Ax_M \leq t_M b + (1 - t_M)\mathbf{1} \leq \max(\|b\|_\infty, 1) \cdot \mathbf{1} =: B \cdot \mathbf{1}. \tag{A.2}$$

Writing this estimate as $A\left(\frac{1}{B}x_M\right) \leq \mathbf{1}$, we see that the vectors $\frac{1}{B}x_M$ are feasible for (Unit LP) for all $M \geq 1$. On the other hand, the objective function can take arbitrarily large values on such vectors. Indeed, since $t_M \in [0, 1]$, estimate (A.1) implies that

$$\langle z, \frac{1}{B}x_M \rangle \to \infty \qquad \text{as } M \to \infty.$$

Thus (Unit LP) is unbounded.

Proof of (iii). Assume that (LP) is not unbounded.

Sufficiency. Assume (LP) is feasible. Then the interpolation (LP$_t$) is clearly feasible for every $0 \leq t \leq 1$. Indeed, if x is feasible for (LP), then tx is feasible for (LP$_t$). Similarly to (A.2), we have that $Ax \leq B \cdot \mathbf{1}$ fr all feasible points x of (LP$_t$). In terms of the feasible polytopes, this means that $P_t \subseteq B \cdot P_0$.

Since (LP) is not unbounded by the assumption, part (i) implies that (Unit LP) is not unbounded. Being always feasible (0 is a feasible set), (Unit LP) must therefore be bounded. Hence

$$\left| \max_{x \in P_t}\langle z, x \rangle \right| \leq B \cdot \left| \max_{x \in P_0}\langle z, x \rangle \right| =: \lambda_0. \tag{A.3}$$

We can write (Int LP) as the optimization problem

$$\max_{t \in [0,1]} f(t), \quad \text{where } f(t) = \max_{x \in P_t} \frac{1}{\lambda}\langle z, x \rangle + t.$$

By (A.3), we have

$$|f(t) - t| \leq \lambda_0/\lambda \quad \text{for all } t \in [0, 1].$$

It then follows that

$$\operatorname*{argmax}_{t \in [0,1]} f(t) \geq 1 - 2\lambda_0/\lambda,$$

which converges to 1 as $\lambda \to \infty$.

We have shown that a solution (x_λ, t_λ) of (Int LP) with parameter λ satisfies $t_l \to 1$ as $\lambda \to \infty$. On the other hand, (x_λ, t_λ) is a vertex of the feasible polytope of (Int LP). This polytope does not depend on λ, thus there are finitely many choices for (x_λ, t_λ). It follows that $t_\lambda = 1$ for all sufficiently large λ. The sufficiency in (iii) is proved.

Necessity. This part is trivial. Indeed, if $(x, 1)$ is a solution of (Int LP) with some parameter λ, then x is a solution of (LP$_1$) and thus x is a solution of (LP).

Proof of (ii). Assume (LP) is not unbounded.

Note that for all sufficiently small t, namely for $t \in [0, t_0]$ with $t_0 := 1/\|b - \mathbf{1}\|_\infty$, the right hand side of (Int LP) is nonnegative:

$$tb + (1 - t)\mathbf{1} \geq 0.$$

Hence (LP$_t$) is feasible for all $t \in [0, t_0]$ (0 is a feasible point).

For a fixed $\lambda < 0$, we can write (Int LP) as the optimization problem

$$\max_{t \in \mathrm{Dom}(g)} g(t), \quad \text{where } g(t) = \max_{x \in P_t} \frac{1}{-\lambda} \langle z, x \rangle - t.$$

The domain $\mathrm{Dom}(g)$ consists of all $t \in [0, 1]$ for which P_t is nonempty, i.e. (LP$_t$) is feasible. In particular, $[0, t_0] \subseteq \mathrm{Dom}(g)$.

By (A.3), we have

$$|g(t) + t| \leq \lambda_0/(-\lambda) \quad \text{for all } t \in \mathrm{Dom}(g).$$

¿From this and from the fact that $\mathrm{Dom}(g)$ contains a neighborhood $[0, t_0]$ of 0 it follows that

$$\underset{t \in \mathrm{Dom}(g)}{\mathrm{argmax}}\, g(t) \leq -2\lambda_0/\lambda,$$

which converges to 0 as $\lambda \to -\infty$.

We have shown that a solution (x_λ, t_λ) of (Int LP) with parameter λ satisfies $t_l \to 0$ as $\lambda \to -\infty$. Similarly to our proof of (iii), we deduce that $t_\lambda = 0$ for all sufficiently small λ. This proves (ii).

Proof of (iv). This follows directly from (iii) since (Int LP) for $t = 1$ is (LP).

The proof of Proposition 4.1 is complete.

B. APPENDIX B. PROOF OF THEOREM 5.4

B.1. **Part 1.** We need to prove that (4a) and (4b) in ADDING CONSTRAINTS imply that in the polytope $P^+ = \mathrm{conv}(0, a_1, \ldots, a_{n+d})$, one has facet$(z) = \{n+1, \ldots, n+d\}$. By (4a), it will be enough to show that all points $a_1, \ldots, a_n$ lie below the affine span $\mathrm{aff}(a_{n+1}, \ldots, a_{n+d}) =: H$. Since all these points have norm at most M, it will suffice to show that all vectors x of norm at most M are below H. By (4b), the normal h to H has norm at most $1/M$, thus $\langle h, x \rangle \leq 1$. Thus x is indeed below H. This completes the proof.

B.2. **Part 2.** Without loss of generality, we can assume that $M_0 = M$. Also, by the homogeneity, we can assume that $M = 1/2$. Thus there is no dilation in step 2 of ADDING

CONSTRAINTS. Let H be a numb half-space. It suffices to show that

$$\mathbb{P}\{z_0 \in \mathrm{cone}(a_{n+1}, \ldots, a_{n+d})\} \geq 0.99; \tag{B.1}$$

$$\mathbb{P}\{\text{normal } h \text{ to } \mathrm{aff}(a_{n+1}, \ldots, a_{n+d}) \text{ satisfies } \|h\| \leq \frac{1}{M}\} \geq 0.99; \tag{B.2}$$

$$\mathbb{P}\{a_{n+1}, \ldots, a_{n+d} \text{ are in } H\} \geq 1/3. \tag{B.3}$$

Events in (B.1) and (B.2) are invariant under the rotation U. So, in proving these two estimates we can assume that U is the identity, which means that $z_0 = z_0'$ and $\bar{a}_i = \bar{a}_i'$ for $i = n+1, \ldots, n+d$. We can also assume that d is bigger than some suitable absolute constant (100 will be enough).

We will use throughout the proof the known estimate on the spectral norm of random matrices. Recall that the spectral norm of a $d \times d$ matrix B is defined as

$$\|B\| = \max_{x \in \mathbb{R}^d} \frac{\|Bx\|}{\|x\|}.$$

For a random $d \times d$ matrix B with independent standard normal random entries, one has

$$\mathbb{P}\{\|B\| > 2t\sqrt{d}\} \leq 2^d(d-1)t^{d-2}e^{-d(t^2-1)/2} \quad \text{for } t \geq 1 \tag{B.4}$$

see e.g. [14]. Therefore, for a random $d \times d$ matrix G with independent entries of mean 0 and variance σ_1, we have

$$\mathbb{P}\{\|G\| > 2\sigma_1 t\sqrt{d}\} \leq 2^d(d-1)t^{d-2}e^{-d(t^2-1)/2} \quad \text{for } t \geq 1.$$

(this follows from (B.4) for $B = G/\sigma_1$).

In particular,

$$\mathbb{P}\{\|G\| \leq 3\sigma_1\sqrt{d}\} \geq 0.99. \tag{B.5}$$

We will view the vectors $a_{n+1}, \ldots, a_{n+d}$ as images of some fixed orthonormal vector basis $e_{n+1}, \ldots, e_{n+d}$ of $\mathbb{R}^d$. Denote

$$\mathbf{1} = \sum_{i=n+1}^{n+d} e_i.$$

We define the linear operator T in $\mathbb{R}^d$ so that

$$\bar{a}_i = Te_i, \quad a_i = (T+G)e_i, \quad i = n+1, \ldots, n+d.$$

We first show that

$$\|T^{-1}\| \leq 1/\ell. \tag{B.6}$$

Indeed, $\triangle(e_{n+1}, \ldots, e_{n+d})$ is a simplex with center $d^{-1}\mathbf{1}$ of norm $d^{-1/2}$ and radius $\|d^{-1}\mathbf{1} - e_i\| = \sqrt{1 - 1/d}$. Similarly, $\triangle(\bar{a}_{n+1}, \ldots, \bar{a}_{n+d})$ is a simplex with center z_0 of norm 1 and radius $\|z_0 - \bar{a}_i\| = \ell$. Therefore we can write $T = VT_1$ with a suitable $V \in O(d)$, and where T acts as follows: if $x = x_1 + x_2$ with $x_1 \in \mathrm{span}(\mathbf{1})$ and $x_2 \in \mathrm{span}(\mathbf{1})^\perp$, then

$T_1 x = d^{1/2} x_1 + \ell(1 - 1/d)^{-1/2} x_2$. Thus $\|T^{-1}\| = \|T_1^{-1}\| = \ell^{-1}(1 - 1/d)^{1/2} \leq 1/\ell$. This proves (B.6).

B.2.1. *Proof of* (B.1). An equivalent way to state (B.1) is that

$$z_0 = \sum_{i=n+1}^{n+d} c_i a_i \quad \text{where all } c_i \geq 0. \tag{B.7}$$

Recall that $a_i = (T + G)e_i$. Multiplying both sides of (B.7) by the operator $(T + G)^{-1}$, we obtain

$$(T + G)^{-1} z_0 = \sum_{i=n+1}^{n+d} c_i e_i.$$

Taking the inner product of both sides of this equation with the vectors e_i, we can compute the coefficients c_i as

$$c_i = \langle (T + G)^{-1} z_0, e_i \rangle. \tag{B.8}$$

On the other hand, z_0 is the center of the simplex $\triangle(\bar{a}_{n+1}, \ldots, \bar{a}_{n+d})$, so $z_0 = \sum_{i=n+1}^{n+d} (1/d) \bar{a}_i$. Since $\bar{a}_i = T e_i$, a similar argument shows that

$$\frac{1}{d} = \langle T^{-1} z_0, e_i \rangle. \tag{B.9}$$

Thus to bound c_i below, it suffices to show that the right sides of (B.8) and (B.9) are close. To this end, we use the identity $T^{-1} - (T+G)^{-1} = (1+T^{-1}G)^{-1}T^{-1}GT^{-1}$ and the estimate $\|1 + S\| \leq (1 - \|S\|)^{-1}$ valid for operators of norm $\|S\| < 1$. Thus the inequality

$$\|(T + G)^{-1} - T^{-1}\| \leq \frac{\|T^{-1}\|^2 \|G\|}{1 - \|T^{-1}\|\|G\|} \leq \frac{1}{2d} \tag{B.10}$$

holds with probability at least 0.99, where the last inequality follows from (B.5), (B.6) and from our choice of ℓ and σ_1 made in (5.1). Since z_0 and e_i are unit vectors, (B.10) implies that the right sides of (B.8) and (B.9) are within $\frac{1}{2d}$ from each other. Thus $c_i \geq \frac{1}{2d} > 0$ for all i. This completes the proof of (B.1).

B.2.2. *Proof of* (B.2). We claim that the normal z_0' to $\text{aff}(\bar{a}_{n+1}, \ldots, \bar{a}_{n+d})$ and the normal h to $\text{aff}(a_{n+1}, \ldots, a_{n+d})$ can be computed as

$$z_0' = (T^*)^{-1} \mathbf{1}, \quad h = ((T + G)^*)^{-1} \mathbf{1}. \tag{B.11}$$

Indeed, for every $i \in \{n + 1, \ldots, n + d\}$ we have

$$\langle (T^*)^{-1} \mathbf{1}, \bar{a}_i \rangle = \langle \mathbf{1}, T^{-1} \bar{a}_i \rangle = \langle \mathbf{1}, e_i \rangle = 1.$$

Hence by the definition of the normal, the vector $(T^*)^{-1} \mathbf{1}$ is the normal to $\text{aff}(\bar{a}_{n+1}, \ldots, \bar{a}_{n+d})$. The second identity in (B.11) is proved in a similar way.

Since z_0' is a unit vector, to bound the norm of h it suffices to estimate

$$\|h - z_0'\| \leq \|((T + G)^*)^{-1} - (T^*)^{-1}\|\|\mathbf{1}\| = \|(T + G)^{-1} - T^{-1}\|\|\mathbf{1}\|.$$

By (B.10) and using $\|\mathbf{1}\| = d^{-1/2}$, with probability at least 0.99 one has $\|h - z_0'\| \leq \frac{1}{2} d^{-3/2} \leq 1$. Thus $\|h\| \leq 2$, which completes the proof of (B.2).

B.2.3. *Proof of* (B.3). Let ν be a unit vector such that the half-space is $H = \{x : \langle \nu, x \rangle \geq 0\}$. Then (B.3) is equivalent to saying that

$$\mathbb{P}\{\langle \nu, a_i \rangle \geq 0, \quad i = n + 1, \ldots, n + d\} \geq 1/3.$$

We will write

$$\langle \nu, a_i \rangle = \langle \nu, z_0 \rangle + \langle \nu, \bar{a}_i - z_0 \rangle + \langle \nu, a_i - \bar{a}_i \rangle \tag{B.12}$$

and estimate each of the three terms separately.

Since z_0 is a random vector uniformly distributed on the sphere S^{d-1}, a known calculation of the measure of a spherical cap (see e.g. [11] p.25) implies that

$$\mathbb{P}\left\{ \langle \nu, z_0 \rangle \geq \frac{1}{60\sqrt{d}} \right\} \geq \frac{1}{2} - 0.1. \tag{B.13}$$

This takes care of the first term in (B.12).

To bound the second term, we claim that

$$\mathbb{P}\left\{ \max_{i=n+1,\ldots,n+d} |\langle \nu, \bar{a}_i - z_0 \rangle| \leq \frac{1}{120\sqrt{d}} \right\} \geq 0.99. \tag{B.14}$$

To prove this, we shall use the rotation invariance of the random rotation U. Without changing its distribution, we can compose U with a further rotation in the hyperplane orthogonal to Uz_0'. More precisely, U is distributed identically with VW. Here $W \in O(d)$ is a random rotation; denote $z_0 := Wz_0'$. Then V is a random rotation in $L = \mathrm{span}(z_0)^\perp$ and for which L is an invariant subspace, that is $Vz_0 = z_0$.

Then we can write $\bar{a}_i - z_0 = V\ell_i$, where $\ell_i := W(\bar{a}_i' - z_0') = W\bar{a}_i' - z_0$. The vectors ℓ_i are in L because $\langle \ell_i, z_0 \rangle = \langle W(\bar{a}_i' - z_0'), Wz_0' \rangle = \langle \bar{a}_i' - z_0', z_0' \rangle = 0$ since z_0' is a unit vector and, moreover, the normal of $\mathrm{aff}(\bar{a}_i)$. Since L is an invariant subspace of V, it follows that $V\ell_i \in L$. Furthermore, $\|\ell_i\| = \|\bar{a}_i' - z_0'\| = \ell$.

Let P_L denote the orthogonal projection onto L. Then $P_L\nu$ is a vector of norm at most one, so denoting $\nu' = P_L\nu/\|P_L\nu\|$ we have

$$|\langle \nu, \bar{a}_i - z_0 \rangle| = |\langle \nu, V\ell_i \rangle| = |\langle P_L\nu, V\ell_i \rangle| = |\langle V^*P_L\nu, \ell_i \rangle| \leq |\langle V^*\nu', \ell_i \rangle|$$

$V^*\nu'$ is a random vector uniformly distributed on the sphere of L, and ℓ_i are fixed vectors in L of norm ℓ.

Then to prove (B.14) it suffices to show that for x uniformly distributed on S^{d-2} and for any fixed vectors $\ell_1, \ldots, \ell_d$ in $\mathbb{R}^{d-1}$ of norm ℓ, one has

$$\mathbb{P}\left\{ \max_{i=1,\ldots,d} |\langle x, \ell_i \rangle| \leq \frac{1}{120\sqrt{d}} \right\} \geq 0.99. \tag{B.15}$$

This is well known as the estimate of the mean width of the simplex. Indeed, for any choice of unit vectors $h_1, \ldots, h_d$ in $\mathbb{R}^{d-1}$ and any $s > 0$,

$$\mathbb{P}\Big\{ \max_{i=1,\ldots,d} |\langle x, h_i \rangle| > \frac{s}{\sqrt{d}} \Big\} \leq \sum_{i=1}^{d} \mathbb{P}\Big\{ |\langle x, h_i \rangle| > \frac{s}{\sqrt{d}} \Big\}$$

and each probability in the right hand side is bounded by $p := \exp(-(d-3)^2 s^2 / 4d)$ by the concentration of measure on the sphere (see [11] (1.1)). We apply this for $h_i = \frac{1}{\ell} \ell_i$ and with $s = \frac{1}{120\ell}$, which makes $p \leq \frac{1}{100d}$. This implies (B.15) and, ultimately, (B.14).

To estimate the third term in (B.12), we can condition on any choice of U, so that $\bar{a}_i$ become fixed. Then $g_i = -\langle \nu, a_i - \bar{a}_i \rangle$ are independent Gaussian random variables with mean 0 and variance $\sigma_1 \leq \frac{1}{120\sqrt{d}} =: s$. Then

$$\mathbb{P}\{g_1 > s\} \leq \frac{1}{\sqrt{2\pi}} \exp(-s^2 / 2\sigma_1^2) \leq \frac{1}{100d}$$

by a standard estimate on the Gaussian tail and by our choice of σ_1 and s. Hence

$$\mathbb{P}\Big\{ \min_{i=n+1,\ldots,n+d} \langle \nu, a_i - \bar{a}_i \rangle \geq -\frac{1}{120\sqrt{d}} \Big\} = 1 - \sum_{i=n+1}^{n+d} \mathbb{P}\{g_i > s\} \geq 0.99. \qquad \text{(B.16)}$$

Combining (B.13), (B.14) and (B.16), we can now estimate (B.12):

$$\mathbb{P}\{\langle \nu, a_i \rangle \geq 0, \quad i = n+1, \ldots, n+d\} \geq \frac{1}{2} - 0.1 - 0.01 - 0.01 > \frac{1}{3}.$$

This completes the proof of Theorem 5.4. ∎

C. Appendix C. Change of variables

The following basic change of variables is useful when dealing with the hyperplane E spanned by linearly independent $a_1, \ldots, a_d$ in $\mathbb{R}^d$. It is explained in more detail in [6].

We specify E by choosing $r \in \mathbb{R}_+$ and $\omega \in S^{d-1}$ in such a way that $\langle \omega, a_i \rangle = r$ for all i. Thus ω is the unit vector in the direction orthogonal to E, and r is the distance from the origin to E.

We choose a reference unit vector h in $\mathbb{R}^d$. The hyperplane H orthogonal to h will be identified with $\mathbb{R}^{d-1}$. For every $\omega \neq -h$, we denote by R_ω the linear transformation that rotates h to ω in the two-dimensional subspace through h and ω and that is the identity in the orthogonal subspace. Then one can map a point $b \in \mathbb{R}^{d-1}$ to $a \in E$ by $a = R_\omega b + r\omega$.

Let $b_1, \ldots, b_n$ be vectors in $\mathbb{R}^{d-1}$, and $a_1, \ldots, a_n$ be the corresponding vectors in $\mathbb{R}^d$ under this change of variables, i.e.

$$a_i = R_\omega b_i + r\omega, \qquad i = 1, \ldots, d. \qquad \text{(C.1)}$$

The Jacobian of this change of variables $(\omega, r, b_1, \ldots, b_d) \mapsto (a_1, \ldots, a_d)$ equals

$$(d-1)! \operatorname{Vol}(\triangle(b_1, \ldots, b_d)).$$

This is a well known formula in the integral geometry due to Blaschke (see [13]).

Lemma C.1 (Change of variables). *Let $a_1, \ldots, a_d$ be independent Gaussian vectors in $\mathbb{R}^d$ with centers of norm at most 1 and standard deviation σ. Let $E = \operatorname{aff}(a_1, \ldots, a_d)$. Then the vectors $b_1, \ldots, b_d$ obtained by change of variables (C.1) satisfy the following.*

    1. $\|b_i\| \le \|a_i\|$ for all i.

    2. Let us condition on a realization of r and ω (i.e. on E). Then the density of $(b_1, \ldots, b_d)$ is proportional to

$$\operatorname{Vol}(\triangle(b_1, \ldots, b_d)) \prod_{i=1}^{d} \nu_i(b_i), \tag{C.2}$$

where ν_i are the densities of Gaussian vectors in $\mathbb{R}^{d-1}$ with centers of norm at most 1 and standard deviation σ.

Proof. 1. First note that $a_i - r\omega$ is orthogonal to ω. Indeed, $\langle a_i - r\omega, \omega \rangle = \langle a_i, \omega \rangle - r = 0$. Therefore

$$\|a_i\|^2 = \|a_i - r\omega\|^2 + \|r\omega\|^2 = \|R_\omega b_i\|^2 + r^2 = \|b_i\|^2 + r^2.$$

This proves part 1.

2. The density of $(a_1, \ldots, a_d)$ is $\prod_{i=1}^{d} \mu_i(a_i)$, where μ_i are Gaussian densities. Denote the center of the distribution of a_i by $\bar{a}_i$. Let P_E denote the orthogonal projection in $\mathbb{R}^d$ onto E.

Since the realization of E is fixed, the induced distribution $\mu_i(a_i \,|\, a_i \in E)$ is a distribution of a Gaussian vector in E with center $P_E \bar{a}_i$ and standard deviation σ. (See e.g. Proposition 2.4.3 in [6].)

Since the change of variables (C.1) is an isometry $\mathbb{R}^{d-1} \to E$, part 2 is proved except the upper bound 1 the norms of the centers of b_i. These centers, denoted by $\bar{b}_i$, are the vectors in $\mathbb{R}^{d-1}$ that correspond to $P_e \bar{a}_i$ under change of variables (C.1):

$$P_E \bar{a}_i = R_\omega \bar{b}_i + r\omega.$$

Since $E = \{x \in \mathbb{R}^d : \langle x, \omega \rangle = r\}$, we have $P_E 0 = r\omega$. Therefore

$$\|\bar{b}_i\| = \|R_\omega \bar{b}_i\| = \|P_E \bar{a}_i - r\omega\| = \|P_E \bar{a}_i - P_E 0\| \le \|\bar{a}_i - 0\| = \|\bar{a}_i\|.$$

This proves part 2 of the lemma.   ∎

REFERENCES

[1] I. Adler, *The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method.* Technical Report, University of California at Berkeley, May 1983

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

[3] I. Adler, N. Megiddo, *A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension*, Journal of the ACM 32 (1985), 871–895

[4] A. Ben-Tal, A. Nemirovski, *On polyhedral approximations of the second-order cone*, Math. Oper. Res. 26 (2001), 193–205.

[5] A. Deshpande, D. A. Spielman, *Improved smoothed analysis of the shadow vertex simplex method*, 46th IEEE FOCS, 349–356, 2005

[6] D. A. Spielman, S.-H. Teng, *Smoothed analysis: why the simplex algorithm usually takes polynomial time*, Journal of the ACM 51 (2004), 385–463

[7] S. Gaas, T. Saaty, *The computational algorithm for the parametric objective function*, Naval Research Logistics Quarterly 2 (1955), 39–45

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

[9] I. Hueter, *Limit theorems for the convex hull of random points in higher dimensions*, Transactions of the AMS 351 (1999), 4337–4363

[10] G. Kalai, D. J. Kleitman, *A quasi-polynomial bound for the diameter of graphs of polyhedra*, Bulletin Amer. Math. Soc. 26 (1992), 315–316

[11] M. Ledoux, *The concentration of measure phenomenon*, AMS Math. Surveys and Monographs 89, 2001

[12] H. Raynaud, *Sur l'enveloppe convexe des nuages de points aléatoires dans $\mathbb{R}^n$*, Journal of Applied Probability 7 (1970), 35–48

[13] L. Santalo, *Integral geometry and geometric probability.* Encyclopedia of Mathematics and its Applications. Addison-Wesley, 1976

[14] S. Szarek, *Spaces with large distance to ℓ_∞^n and random matrices*, American Journal of Mathematics 112 (1990), 899–942.

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

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, DAVIS, CA 95616, U.S.A.

E-mail address: vershynin@math.ucdavis.edu