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 had polynomial smoothed complexity. On a slight random perturbation of 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 $O(\max(d^5 \log^2 n,\ d^9 \log^4 d,\ d^3\sigma^{-4}))$. This shows that the tight Hirsch conjecture $n - d$ on the 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. 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.

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" [2]. In an attempt to explain this behavior, Spielman and Teng [2] introduced the concept of *smoothed analysis* of algorithms, in which one measured 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

$$\begin{aligned} &\text{maximize } \langle z, x \rangle \\ &\text{subject to } Ax \le b, \end{aligned} \tag{LP}$$

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 $Ax \le b$, 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 it chooses a 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 [3]. 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

Partially supported by NSF grant DMS 0401032 and Alfred P. Sloan Foundation.

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 with standard deviations $\sigma \max_i \|(\bar{a}_i, \bar{b}_i)\|$. Spielman and Teng proved

Theorem 1.1. [2] *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:

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

where the logarithmic factors are disregarded. The subsequent work of Deshpande and Spielman [1] improved on the exponents of d and σ; however, it doubled the exponent of n. We shall prove the following estimate:

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

$$\mathcal{P}(n, d, \sigma^{-1}) \le O(\max(d^5 \log^2 n,\ d^9 \log^4 d,\ d^3 \sigma^{-4})).$$

Perhaps the most surprising feature of Theorem 1.2 is that *the number of pivot steps is polylogarithmic in the number of constraints n*, while the previous bounds were polynomial in n.

This can change our intuition of what Linear Programming can achieve. Hirsch conjecture states that the diameter of the polytope $Ax \le b$ (the maximal number of steps in the shortest walk between any pair of vertices) is at most $n - d$. Hirsch conjecture is tight, so it is natural to think of it as a lower bound on the worst case complexity of *any* variant of the simplex method – any walk on vertices must be at least $n - d$ long. Theorem 1.2 (and Theorem 6.1 below) claim that a random perturbation destroys this obstacle by creating short paths between 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!

The reason why a random perturbation creates a short path between vertices is *not* that it destroys most of them. Even in the average case, when A is a matrix with independent i.i.d. Gaussian entries, the expected number of vertices of the random polytope asymptotically equals $2^d d^{-1/2} (d-1)^{-1} (\pi \log n)^{(d-1)/2}$ ([6], see [4]). This is exponential in d and sublinear but not polylogarithmic in n (compare to $\log^2 n$ in Theorem 1.2).

The smoothed complexity (expected running time) of the simplex method is $O(\mathcal{P}(n, d, \sigma^{-1})\ t_{\text{pivot}})$, 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 nesessary, which makes t_{pivot} much smaller. In this case Theorem 1.2 shows that *the shadow-vertex simplex method can solve very large scale problems* (with exponentially many constraints).

2. Outline of the approach

Our smoothed analysis of the simplex method is largely inspired by that of Spielman and Teng [2], but we have to resolve a few conceptual difficulties of [2]. Eventually this simplifies and improves the overall picture.

2.1. **Interpolation: reduction to unit linear programs.** First, we reduce arbitrary linear program (LP) to a *unit linear program*, a program in which $b = \mathbf{1}$. This is done by a simple interpolation. One more variable is introduced, and (LP) reduces 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, which is the convex hull of the origin and of the constraint vectors a_i. 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 $E = \text{span}(z_0, z)$ 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 this simplex method is bounded by the number of facets of P the plane E meets. This is the *size of the planar section of the random polytope* P, the number of edges of the polygon $P \cap E$. Under a hypothetical assumption that E is fixed or is statistically independent of P, estimating the size of $P \cap E$ becomes a solvable problem in asymptotic convex geometry. Indeed, Spielman and Teng [2] and later Deshpande and Spielman [1] showed that this size, called the shadow size in these papers, is polynomial in n, d and σ^{-1}.

The main complication in the analysis in [2] was that plane $E = \text{span}(z_0, z)$ 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 achieve this. The randomized Phase-I proposed in [2] 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 become 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 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 facet(z), the affine half-space below the affine span of 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 $1/2$. A standard concentration of measure argument shows that such a random point is at distance $\Omega(d^{-1/2})$ from the numb half-space with constant probability. (This distance is the *observable diameter* of the sphere, see [5] 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 of

$$M = \max_{i=1,\dots,n} \|a_i\|. \tag{2.1}$$

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 independently generated sets of added constraints, until the solution is correct. Because of constant probability of success at every step, this phase-I terminates in expectation after a 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 the main difficulty of the approach in [2]. 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 the order of $\min(\sigma, d^{-1/2}\log^{-1/2} n, d^{-3/2}\log^{-2.5} n)$, see (5.1). In the previous approach [2], such reduction was made with the standard deviation of order $\sigma^5 d^{-8.5} n^{-14}\log^{-2.5} n$.

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 canonical basis $(e_1, \dots, e_d)$ 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 sa 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. **Remaining difficulties.** There remain two problems though. One is a good estimate of the size of the polygon $P \cap E$ for a random polytope P and a fixed plane E. Known bounds ([2] Theorem 4.0.1 and [1]) are not quite sufficient for us, for they are at least linear in n, while we need a polylogarithmic dependence in our main Theorem 1.2. A loss of the factor of n occurs in estimating the angle of incidence ([2] 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_0 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). There is also an alternative method, which avoids estimating the angle of incidence.

The last and the least important problem is that the the dilation factor M, introduced to ensure that the random simplex is a facet of the new polytope, depends on the magnitudes of the constraint vectors a_i. This makes the added facet somewhat correlate with P. The same problem arose in [2] and was resolved there by quantizing M on an exponential scale, so that there were few choices for M, while the probability of success for each given choice of M is big enough. We prefer to retain the original definition of M, to keep phase-I most natural. We are still able to write out and analyze the joint density of the new constraints, even though it does not factor into a product of independent densities.

3. Preliminaries

3.1. **Notation.** The positive cone of a set K in a vector space will be denoted by $\mathrm{cone}(K)$, and its convex hull by $\mathrm{conv}(K)$ or $\triangle(K)$. A half-space in $\mathbb{R}^d$ is a set of the form $\{x : \langle z, x\rangle \le 0\}$ for some vector z. An affine half space takes the form $\{x : \langle z, x\rangle \le a\}$ for some number a. The definitions of hyperplane and affine hyperplane are similar, with equalities in place of inequalities. The normal to an affine hyperplane H which 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 $\langle h, x\rangle \le 1$.

Throughout the paper, we will assume that the vectors (a_i, b_i) that define the linear program (LP) are in a 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 equalities on x. It is easy to obtain x from I by inverting A_I on x. So we sometimes call the index I a solution of (LP).

For a polytope $P = \mathrm{conv}\{0, a_i\}_{i=1}^n$ and a vector z, we denote by $\mathrm{facet}(z) = \mathrm{facet}_P(z)$ the basis of the facet of P pierced by the ray $\{tz : t \ge 0\}$. More precisely, $\mathrm{facet}(z)$ is the family of all d-sets I such that $\triangle(a_i)_{i\in I}$ is a facet of the polytope P and $z \in \mathrm{cone}(a_i)_{i\in I}$. If z is in general position, $\mathrm{facet}(z)$ is an empty set or contains exactly one set I.

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 in 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 $I \in \mathrm{facet}(z)$ 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 method is described in detail in [2] Section 3.2. It works on unit linear programs of type (LP) with $b = \mathbf{1}$. A solution of such program is a member of $\mathrm{facet}(z)$ of the polytope $P = \mathrm{conv}\{0, a_i\}_{i=1}^n$. The program is unbounded iff $\mathrm{facet}(z) = \emptyset$. (See [2] Section 3.2).

Its input of the polar shadow vertex simplex method is the objective vector z, an *initial objective vector* z_0 and its initial objective facet $\mathrm{facet}(z_0)$ (for the polytope P as above), provided that $\mathrm{facet}(z_0)$ consists of only one set of indices. The simplex method rotates z_0 toward z and computes $\mathrm{facet}(q)$ for all vectors q between z_0 and z. At the end, it outputs the limit of $\mathrm{facet}(q)$ as q approaches z. This is the last running $\mathrm{facet}(q)$ before q reaches z_0.

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

If z and z_0 are linearly dependent, $z_0 = -cz$ for some $c > 0$, one can specify 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 $c_2 \ge 0$.

4. Interpolation on Linear Programs

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

$$\begin{aligned} &\text{maximize } \langle z, x\rangle \\ &\text{subject to } Ax \le \mathbf{1}. \end{aligned} \tag{Unit LP}$$

This reduction is general and independent from a particular method to solve linear programs. We will 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:

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

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 $\mathbb{R}^{d+1}$:

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

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 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 intuitive and 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.*

(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.* ∎

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 $t = 1$ is not satisfied by 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 $\mathrm{span}(z,u)$. 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 $\mathrm{span}(z,u)$. 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. Phase-I: Adding constraints

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

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

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 $\langle a, x\rangle \le 1$ 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 $\max_{Ax \le \mathbf{1}} \langle z, x\rangle$ 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

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

$0 \in P$ implies that $h \ge 0$. We define the affine hyperplane

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

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 $\mathrm{conv}(P \cup a) \subset H^-$, thus

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

where the first two inequalities follow from the inclusion $P \subset \mathrm{conv}(P \cup a) \subset H^-$, and the last equality follows from the definition of H^-. So, we have shown that $\|z\|_{\mathrm{conv}(P \cup a)} = \|z\|_P$, which says that the a and thus the affine half-space H^- is in the numb set of (Unit LP). Since $h \ge 0$, 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 $a_{n+1}, \dots, a_{n+d}$ 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 $\langle a_i, x\rangle \le 1$, $i = n+1, \dots, n+d$. ∎

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 $a_{n+1}, \dots, a_{n+d}$ 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: $M > 0$ and $U \in O(d)$.
Output: "Failure" or vectors $a_{n+1}, \dots, a_{n+d}$ and $z_0 \in \mathrm{cone}(a_{n+1}, \dots, a_{n+d})$.

(1) *Form a regular simplex:* let z_0' be a fixed unit vector in $\mathbb{R}^d$ and $\bar a_{n+1}', \dots, \bar a_{n+d}'$ 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 $z_0 = 2MUz_0'$, $\bar a_i = 2MU\bar a_i'$, $i = n+1, \dots, n+d$.
(3) *Smooth:* let a_i be independent Gaussian random variables with mean $\bar a_i$ and standard deviation $2M\sigma_1$, for $i = n+1, \dots, n+d$.
(4) *Check* if
  (a) $z_0 \in \mathrm{cone}(a_{n+1}, \dots, a_{n+d})$ and
  (b) Normal h to $\mathrm{aff}(a_{n+1}, \dots, a_{n+d})$ satisfies $\|h\| \le \frac{1}{M}$.
  If not, return "Failure".

The crucial property of Adding Constraints is

Theorem 5.4. *Let (Unit LP) be a unit linear program with a numb half-space H, and M be as in* (2.1)*. Then:*
1. Let $U \in O(n)$ 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 $\{n+1, \dots, n+d\}$.*
2. With probability at least $1/4$ in the choice of a random $U \in O(n)$, the algorithm Adding Constraints *does not return "Failure" and generates vectors $a_{n+1}, \dots, a_{n+d}$ that lie in the numb half-space H.*

Proof. See Appendix A. ∎

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.
Thus we can solve (Unit LP) by repeatedly solving (Unit LP$^+$) with independently added constraints until no "Failure" 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" returned and the solution I^+ contains none of the indices $n+1,\ldots,n+d$:

Phase-I: Apply ADDING CONSTRAINTS with M as in (2.1) and with the rotation U chosen randomly and independently in the orthogonal group $O(d)$ according to the Haaar 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^+.

6. NUMBER OF PIVOTS AND SECTIONS OF RANDOM POLYTOPES

Now we analyze the running time of SOLVER FOR (LP). In its first phase, it uses SOLVER FOR (UNIT LP). For any unit linear program, the expected number of iterations (i.e. calls to phase-I and phase-II) 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 once for (Int LP) in SOLVER FOR (LP) and repeatedly for (Unit LP$^+$) in SOLVER FOR (UNIT LP).

As explained in Section 2.2, the number of pivot steps made by 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.

As in [2], we can work under the assumptions

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

When we apply the polar shadow-vertex simplex method for (Int LP) in phase-II of SOLVER FOR (LP), the plane $E = \mathrm{span}((z,0),(0,1))$ is fixed, and the constraint vectors are $(0,1)$, $(0,-\infty)$, and $(a_i, 1-b_i)$ for $i = 1,\ldots,n$. The vectors $(a_i, 1-b_i)$ are independent Gaussian with centers of norm at most 2, and with standard deviation $\sqrt{2}\,\sigma$. The other two vertices 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, and we can assume that P is the convex hull of the points $(a_i, 1-b_i)$ only. Let $\Psi(a_1,\ldots,a_n)$ denote the joint density of independent Gaussian vectors in $\mathbb{R}^n$ with some centers of norm at most 2, and with standard deviation $\sqrt{2}\,\sigma$, where σ satisfies (6.1).

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 $E = \mathrm{span}(\frac{z_0}{\|z_0\|}, z) = \mathrm{span}(Uz_0', z)$ is fixed. The constraint vectors are $a_1,\ldots,a_{n+d}$, of which first n are independent Gaussian vectors with centers of norm at most 1 and with standard deviation σ, which satisfies (6.1). The last d of the constraint vectors are also Gaussian vectors chosen independently with centers $2M\tilde{a}_i$ and variance $2M\sigma_1$, where $\tilde{a}_i (= U\bar{a}_i')$ are fixed vectors of norm $\sqrt{1+\ell^2} \le 1.01$, and where M is as in (2.1) and σ_1 is as in (5.1). Thus the last d of the constraint vectors correlate with the first n vectors, but only through the random variable M. Let $\Phi(a_1,\ldots,a_{n+d})$ denote the density of such constraint vectors. Then we need an estimate of what was called the *shadow size bound* in [2].

Theorem 6.1 (Sections of random polytopes)**.** *Let $a_1, \dots, a_{n+d}$ be random vectors in $\mathbb{R}^d$, and E be a plane in $\mathbb{R}^d$. Then the following holds with $C = 10^{11}$.*

1. If $(a_1, \dots, a_n)$ have density $\Psi(a_1, \dots, a_n)$, then the random polytope $P = \operatorname{conv}(a_i)_{i=1}^n$ satisfies

$$\mathbb{E}\,|\operatorname{edges}(P \cap E)| \le Cd^3\sigma^{-4}.$$

2. If $(a_1, \dots, a_{n+d})$ have density $\Phi(a_1, \dots, a_{n+d})$, then the random polytope $P = \operatorname{conv}(a_i)_{i=1}^{n+d}$ satisfies

$$\mathbb{E}\,|\operatorname{edges}(P \cap E)| \le Cd^3 \min(\sigma, \sigma_1)^{-4}.$$

The shadow bound of Spielman-Teng ([2] Theorem 4.0.1) was $O(nd^3\sigma^{-6})$ for part 1, which was not quite sufficient for us because of the polynomial, rather than polylogarithmic, dependence on n.

Proof. See Appendix B. ∎

The desired estimate in Theorem 1.2 on the total expected number of pivot steps can be put in the form $O(d^3 \min(\sigma, \sigma_1)^{-4})$. Hence Theorem 6.1 yields the desired expected number of the pivot steps in phase-II of Solver for (LP), and also the expected number of pivot steps, conditioned on a choice of U, in one call of phase-II of Solver for (Unit LP).

It remains to bound the expected *total* number of pivot steps in Solver for (Unit LP), over all iterations it makes. This is a 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

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

To bound the expectation of Z, we denote by $\mathbb{E}_0$ the expectation with respect to random (smoothed) vectors $(a_1, \dots, a_n)$, 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 $(a_{n+1}, \dots, a_{n+d})$.

Let us first condition on the choice of $(a_1, \dots, a_n)$. 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

$$\mathbb{E}Z = \mathbb{E}_0 \sum_{k=1}^{\infty} (\mathbb{E}_k Z_k) \prod_{j=1}^{k-1} \mathbb{E}_j F_j. \tag{6.2}$$

As observed above, $\mathbb{E}_j F_j = \mathbb{P}(F_j = 1) \le 3/4$, which bounds the product in (6.2) by $(3/4)^k$. Moreover, $\mathbb{E}_0 \mathbb{E}_k Z_k \le \max_U \mathbb{E}_\Phi Z_1$, where E_Φ is the expectation with respect to the random vectors $(a_1, \dots, a_{n+d})$ conditioned on a choice of U in k-th iteration. As we mentioned, this random vectors have joint density Φ. Hence Theorem 6.1 bounds $\max_U \mathbb{E}_\Phi Z_1$. Summarizing, we have shown that $\mathbb{E}Z \le O(d^3 \min(\sigma, \sigma_1)^{-4})$. This proves Theorem 1.2 and completes the smoothed analysis of the simplex method.

References

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

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

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

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

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

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

[7] S. Szarek, *Spaces with Large Distance to* ℓ_∞^n *and Random Matrices*, American Journal of Mathematics 112 (1990), 899–942.

A. Appendix A. Proof of Theorem 5.4

A.1. **Part 1.** We need to prove that (4a) and (4b) in Adding Constraints imply that in the polytope $P^+ = \mathrm{conv}(0, a_1, \dots, a_{n+d})$, one has $\mathrm{facet}(z) = \{n+1, \dots, n+d\}$. By (4a), it will be enough to show that all points $a_1, \dots, a_n$ lie below the affine span $\mathrm{aff}(a_{n+1}, \dots, a_{n+d}) =: H$. 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 $\langle h, x\rangle \le 1$. Thus x is indeed below H. This completes the proof.

A.2. **Part 2.** By homogeneity, we can assume throughout the proof 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

$$\mathbb{P}\{z_0 \in \mathrm{cone}(a_{n+1}, \dots, a_{n+d})\} \ge 0.99; \tag{A.1}$$

$$\mathbb{P}\{\text{normal } h \text{ to } \mathrm{aff}(a_{n+1}, \dots, a_{n+d}) \text{ satisfies } \|h\| \le \frac{1}{M}\} \ge 0.99; \tag{A.2}$$

$$\mathbb{P}\{a_{n+1}, \dots, a_{n+d} \text{ are in } H\} \ge 1/3. \tag{A.3}$$

Events in (A.1) and (A.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 $i = n+1, \dots, n+d$. 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 $\ell_2 \to \ell_2$ norm of a random $d \times d$ matrix with independent Gaussian random entries of mean 0 and variance σ_1:

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

see e.g. [7] In particular,

$$\mathbb{P}\{\|G\| \le 3\sigma_1\sqrt{d}\} \ge 0.99. \tag{A.4}$$

We will view the vectors $a_{n+1}, \dots, a_{n+d}$ as images of some fixed orthonormal vector basis $e_{n+1}, \dots, e_{n+d}$ 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

$$\bar a_i = Te_i, \quad a_i = (T+G)e_i, \quad j = n+1, \dots, n+d.$$

We first show that

$$\|T^{-1}\| \le 1/\ell. \tag{A.5}$$

Indeed, $\triangle(e_{n+1}, \dots, e_{n+d})$ 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, $\triangle(\bar a_{n+1}, \dots, \bar a_{n+d})$ 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 $V \in O(n)$, 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 $T_1 x = d^{1/2} x + \ell(1-1/d)^{-1/2} x_2$. Thus $\|T^{-1}\| = \|T_1^{-1}\| = \ell^{-1}(1-1/d)^{1/2} \le 1/\ell$. Ths proves (A.5).

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

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

Recall that $a_i = (T+G)e_i$ and invert $T+G$; we can then compute the coefficients c_i as

$$c_i = \langle (T+G)^{-1} z_0, e_i \rangle. \tag{A.6}$$

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

$$\frac{1}{d} = \langle T^{-1} z_0, e_i \rangle. \tag{A.7}$$

Thus to bound c_i below, it suffices to show that the right sides of (A.6) and (A.7) are close. To this end, we use the identity $(T+G)^{-1} - T^{-1} = (1 + T^{-1}G)^{-1} T^{-1} G T^{-1}$ and the estimate $\|1+S\| \le (1 - \|S\|)^{-1}$ valid for operators of norm $\|S\| < 1$. Thus the inequality

$$\|(T+G)^{-1} - T^{-1}\| \le \frac{\|T^{-1}\|^2 \, \|G\|}{1 - \|T^{-1}\| \|G\|} \le \frac{1}{2d} \tag{A.8}$$

holds with probability at least 0.99, where the last inequality follows from (A.4), (A.5) and from our choice of ℓ and σ_1 made in (5.1). Since z_0 and e_i are unit vectors, (A.8) implies that the right sides of (A.6) and (A.7) are within $\frac{1}{2d}$ from each other. Thus $c_i \ge \frac{1}{2d} > 0$ for all i. This completes the proof of (A.1).

A.2.2. *Proof of* (A.2). The normal z_0' to $\mathrm{aff}(\bar{a}_{n+1}, \dots, \bar{a}_{n+d})$ and the normal h to $\mathrm{aff}(a_{n+1}, \dots, a_{n+d})$ can be computed as

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

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

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

By (A.8) and using $\|\mathbf{1}\| = d^{-1/2}$, with probability at least 0.99 one has $\|h - z_0'\| \le \frac{1}{2} d^{-3/2} \le 1$. Thus $\|h\| \le 2$, which completes the proof of (A.2).

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

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

We will write

$$\langle \nu, a_i \rangle = \langle \nu, z_0 \rangle + \langle \nu, \bar{a}_i - z_0 \rangle + \langle \nu, a_i - \bar{a}_i \rangle \tag{A.9}$$

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. [5] p.25) implies that

$$\mathbb{P}\Big\{ \langle \nu, z_0 \rangle \ge \frac{1}{60\sqrt{d}} \Big\} \ge \frac{1}{2} - 0.1. \tag{A.10}$$

This takes care of the first term in (A.9).

To bound the second term, we claim that

$$\mathbb{P}\Big\{ \max_{i=n+1,\dots,n+d} |\langle \nu, \bar{a}_i - z_0\rangle| \le \frac{1}{120\sqrt{d}} \Big\} \ge 0.99. \tag{A.11}$$

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 $L = \operatorname{span}(z_0)^\perp$ 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 $\operatorname{aff}(\bar{a}_i)$. 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 orhtonormal 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

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

$V^*\nu'$ 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 (A.11) it suffices to show that for x uniformly distributed on S^{d-2} and for any fixed vectors $\ell_1, \dots, \ell_d$ in $\mathbb{R}^{d-1}$ of norm ℓ, one has

$$\mathbb{P}\Big\{ \max_{i=1,\dots,d} |\langle x, \ell_i\rangle| \le \frac{1}{120\sqrt{d}} \Big\} \ge 0.99. \tag{A.12}$$

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

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

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 [5] (1.1)). We apply this for $h_i = \frac{1}{\ell}\ell_i$ and with $s = \frac{1}{120\ell}$, which makes $p \le \frac{1}{100d}$. This implies (A.12) and, ultimately, (A.11).

To estimate the third term in (A.9), 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 $\sigma_1 \le \frac{1}{120\sqrt{d}} =: s$. Then

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

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

$$\mathbb{P}\Big\{ \min_{i=n+1,\dots,n+d} \langle \nu, a_i - \bar{a}_i\rangle \ge -\frac{1}{120\sqrt{d}} \Big\} = 1 - \sum_{i=n+1}^{n+d} \mathbb{P}\{g_i > s\} \ge 0.99. \tag{A.13}$$

Combining (A.10), (A.11) and (A.13), we can now estimate (A.9):

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

This completes the proof of Theorem 5.4. ∎

B. Appendix B. Proof of Theorem 6.1

We will outline two approaches to Theorem 6.1. To be specific, we shall focus on Part 1. The other part is similar, except that one has to be careful when dealing with the density Ψ, which is not the product of the independent densities due to the factor M.

B.1. **First argument.** Our first approach is to improve upon the part of the argument of [2] where it looses a factor of n. Recall that we need a polylogarithmic dependence on n.

As in [2], we parametrize the one-dimensional torus $\mathbb{T} = E \cap S^{d-1}$ by $q = q(\theta) = z\sin(\theta) + t\cos(\theta)$ where z, t are orthonormal vectors in E. We quantize θ uniformly in $[0, 2\pi)$ with step $2\pi/m$, which yields a quantized torus $\mathbb{T}_m$ that consists of m equispaced points in $\mathbb{T}$.

The argument in the beginning of the proof of Theorem 4.0.1 in [2] reduces an upper estimate on $\mathbb{E}\,|\operatorname{edges}(P \cap E)|$ to the statement that a fixed vector $q \in \mathbb{T}_m$ is not likely to be too close to the boundary of its facet (of the polytope $P_0 = \operatorname{conv}(0, P)$). The closeness here is measured with respect to the *angular distance* $\operatorname{ang}(x, y)$, which is the angle formed by the vectors x and y. Then one needs to replace the angular distance with the usual, Euclidean, distance. This is easy whenever the angle at which q meets its facet, called *the angle of incidence*, is not too small (Lemma 4.0.2 in [2]). So [2] goes on to bound below the angle of incidence (Section 4.2 in [2]). This is where the loss of a factor of n occurs.

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). The following elementary observation on the plane is crucial.

Lemma B.1 (Three viewpoints)**.** *Let $K = \operatorname{conv}(b_1, \dots, b_N)$ 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 of side* 10 *centered at the origin. Denote $K_i = \operatorname{conv}(0, -0_i + K)$. Then for every edge (b_k, b_m) of K there exists a vector q such that $\operatorname{facet}_{K_i}(q) = \{k, m\}$ and*

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

In other words, 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 $\operatorname{conv}(0_i, K)$.

Here and in the sequel, we identify $\operatorname{facet}(q)$ with the index set it contains (since the polytopes in question are in general position, it contains at most one index set). For $I = \operatorname{facet}_P(q)$, we denote by $\operatorname{Facet}_P(q)$ or $\operatorname{Facet}_P(I)$ the corresponding *geometric facet* of P, the convex hull of the vertices of P with indices in I.

We consider the event

$$\mathcal{E} = \{\text{all } \|a_i\| \le 10,\ i = 1, \dots, n\}$$

which can be easily estimated using the concentration of Gaussian vectors: $\mathbb{P}(E^c) \le \binom{n}{d}^{-1}$. The random variable $\operatorname{edges}(P \cap E)$ is bounded above by $\binom{n}{d}$, which is the maximal number of facets of P. It follows that

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

We will apply the first part of Lemma B.1 for the random polygon $P \cap E$, whenever it satisfies $\mathcal{E}$. All of its points are then bounded by 10 in norm. Let 0_1, 0_2, 0_3 be the vertices of an equilateral triangle in the plane E, with side 100 and centered at the origin. Denote $P_i = \operatorname{conv}(0, -0_i + P)$. Lemma B.1 states in particular that 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

between the edges of $P \cap E$ and the set $\{\text{facet}_{P_i \cap E}(q) : \ q \in \mathbb{T}, \ i = 1,2,3\}$. We can further replace this set by $\{\text{facet}_{P_i}(q) : \ q \in \mathbb{T}, \ i = 1,2,3\}$, since each $\text{facet}_{P_i}(q)$ uniquely determines the edge $\text{facet}_{P_i \cap E}(q)$; and vice versa, each edge can belong to a unique facet. Therefore

$$\text{Exp} \le \mathbb{E}\,|\{\text{facet}_{P_i}(q) : \ q \in \mathbb{T}, \ i = 1,2,3\}| + 1 \tag{B.1}$$

and, by a discretization in limit as in [2] Lemma 4.0.6,

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

Moreover, by the same discretization argument, we may ignore in (B.1) all facets whose intersection with E have angular length no bigger than, say, $2\pi/m$. (The angular length of an interval $\mathcal{I}$ is the angular distance between its endpoints.) After this, we replace P_i by $P_i \cap E$ as we mentioned above, and intersect with the event $\mathcal{E}$ again, as before. This gives

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

We are going to apply the second part of Lemma B.1 for a realization of P for which the event $\mathcal{E}$ holds. Consider any facet from the set in (B.3). So let $I = \text{facet}_{P_i \cap E}(q)$ for some $i \in \{1,2,3\}$ and some $q \in \mathbb{T}_m$. Now by Lemma B.1 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,

$$\text{dist}(0, \text{aff}(\mathcal{I})) \ge 1. \tag{B.4}$$

Since all viewpoints 0_i have equal norm, this also maximizes the angular length of $\mathcal{I}$. Because only facets of angular length $> 2\pi/m$ were included into the set in (B.3), we conclude that the angular lengh of $\mathcal{I}$ must also be bigger than $2\pi/m$. It follows that $\mathcal{I}$ contains a point q'' in $\mathbb{T}_m$.

Summarizing, we have realized every facet $I = \text{facet}_{P_i \cap E}(q)$ from (B.3) as $I = \text{facet}_{P_{i_0} \cap E}(q'')$ for some i_0 and some $q'' \in \mathbb{T}_m$. Moreover, this facet (i.e. $\mathcal{I} = \text{facet}_{P_{i_0} \cap E}(I)$) has a good angle of incidence (B.4). With this bound on the distance, the angular distance and the usual distance on $\mathcal{I}$ are equivalent. This is another simple observation on the plane.

Lemma B.2 (Angular and Euclidean distances)**.** *Let $\mathcal{I}$ be an interval on the plane, such that* (B.4) *holds. Then for every two points $x, y \in \mathcal{I}$ of norm at most* 200 *one has*

$$c\,\text{dist}(x,y) \le \text{ang}(x,y) \le \text{dist}(x,y)$$

where $c = 10^{-5}$ (which can be easily improved).

Recall that because of the event $\mathcal{E}$, all points of P have norm at most 10, thus all points of P_i have norm at most $\|0_i\| + 10 \le 200$. Hence the same bound holds for all points in our interval $\mathcal{I}$. Then Lemma (B.2) shows that for the angular and the usual distances are equivalent on $\mathcal{I}$ up to a factor of c. Call such a facet nondegenerate. We have shown that

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

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 to the boundary of $\mathrm{Facet}_{P_i \cap E}(q)$, thus also to the boundary of $\mathrm{Facet}_{P_i}(q)$, is at most $\frac{1}{c} \cdot 2\pi/m =: C/m$. Therefore

$$\begin{aligned}\mathrm{Exp} &\le \lim_{m\to\infty} \mathbb{E}\,|\{\mathrm{facet}_{P_i}(q) \text{ such that } \operatorname{dist}(q, \partial\,\mathrm{Facet}_{P_i}(q)) \le C/m,\ q \in \mathbb{T}_m,\ i = 1,2,3\}| + 2 \\ &\le 3 \max_{i=1,2,3} \lim_{m\to\infty} \mathbb{E}\,|\{\mathrm{facet}_{P_i}(q) \text{ such that } \operatorname{dist}(q, \partial\,\mathrm{Facet}_{P_i}(q)) \le C/m,\ q \in \mathbb{T}_m\}| + 2.\end{aligned}$$

For a fixed k, the polytope P_k is the polytope P translated by a fixed vector 0_k of norm at most 100. This reduces the problem to estimating

$$\lim_{m\to\infty} \mathbb{E}\,|\{\mathrm{facet}_P(q) \text{ such that } \operatorname{dist}(q, \partial\,\mathrm{Facet}_P(q)) \le C/m,\ q \in \mathbb{T}_m\}| \tag{B.5}$$

for a random polytope P as in the statement of Theorem 6.1 (and where we allow the centers of the distribution Φ to be of norm at most 100 rather than 1).

This step allowed us to replace the angular distance by the usual distance in the beginning of Spielman-Teng's proof of Theorem 4.0.1 in [2]. Now one can essentially continue with the argument of [2]. Then (B.5) gets bounded by the quantity $O(d^2\sigma^{-4})$ in Distance Lemma 4.1.2 [2] multiplied by d (the number of $(d-2)$-dimensional facets of $\mathrm{Facet}_P(q)$ that make up its boundary). This gives

$$\mathrm{Exp} = O(d^3\sigma^{-4}),$$

which completes the proof. ∎

B.2. **Alternative argument.** There is an alternative argument for Theorem 6.1. It also gives a polylogarithmic dependence on n, but presently yields bigger exponent of d. Its main advantage that it is more elegant and much more flexible. It completely avoids estimating the angle of incidence. It also does not use Combination Lemma 2.3.5 of [2]. This liberates the argument from a necessity of choosing the good event P_I^j in a delicate way (Definition 4.0.8 in [2]). It does not start with the discretization in the limit, so it only requires estimates on the likelihood of being ε-close to the boundary of a random simplex for a *fixed* ε, rather than for all small ε (as Lemma 4.1.2 [2] does.)

We count the edges via indicator functions as

$$|\operatorname{edges}(P \cap E)| = \sum_I \mathbf{1}_{\{X_I > 0\}}$$

where the sum is over all d-element subsets of $\{1, \dots, n\}$ and X_I is the length of the intersection of the facet determined by I with E:

$$X_I = \begin{cases} |\triangle(a_i)_{i\in I} \cap E| & \text{if } \triangle(a_i)_{i\in I} \text{ is a facet of } P, \\ 0 & \text{otherwise.} \end{cases}$$

Then by the linearity of the expectation,

$$\mathrm{Exp} = \mathbb{E}|\operatorname{edges}(P \cap E)| = \sum_I \mathbb{P}\{X_I > 0\}.$$

Now we want to replace 0 here by some positive quantity ε. This is possible if there is a lower bound of the form

$$\mathbb{P}\Big\{|\triangle(a_i)_{i\in I} \cap E| > \varepsilon \;\Big|\; \triangle(a_i)_{i\in I} \text{ is a facet of } P \text{ and it intersects with } E\Big\} \ge p, \tag{B.6}$$

which in other words is

$$\mathbb{P}\{X_I > \varepsilon \mid X_I > 0\} \ge p.$$

If we have such a bound, then

$$\text{Exp} \le \frac{1}{p} \sum_I \mathbb{P}\{X_I > \varepsilon\}$$

which by Markov's inequality is bounded by

$$\le \frac{1}{p} \sum_I \frac{\mathbb{E}X_I}{\varepsilon} = \frac{1}{p\varepsilon} \mathbb{E}\Big(\sum_I X_I \Big).$$

The sum $\sum_I X_I$ is the perimeter of the random polygon $P \cap E$. Since P is nicely bounded in expectation (by the standard concentration inequalities), this perimeter is nicely bounded, too. Thus

$$\text{Exp} = O(1/p\varepsilon).$$

It only remains to prove (B.6). This is a one-dimensional version of the zero-dimensional results on the distance to the boundary of a random simplex (Lemma 4.1.1 in [2]).

One can reduce the problem to the zero-dimensional case by proving that (a) the facet $F = \triangle(a_i)_{i \in I}$ is non-degenerate with high probability (contains a nontrivial Euclidean ball); (b) the endpoints of the line segment $F \cap E$ are not too close to the boundaries of the $(d-2)$-dimensional facets of F which it pierces. Statements (a) and (b) together clearly imply a lower bound on the length of $F \cap E$.

Both (a) and (b) are essentially proved in [2]. Indeed, (a) follows from the Height of the Simplex Lemma 4.1.3 of [2], while (b) follows from Distance Bound Lemma 4.1.2 in [2]. Note that our statements involve the usual, rather than the angular, distance. ∎

In proving (b), one needs to take a union bound over d facets of F, which incurs an extra multiplicative factor of d. The dependence on n remains polylogarithmic. It would be nice to see if this argument can be modified to prevent the loss of the factor of d.

Department of Mathematics, University of California, Davis, CA 95616, U.S.A.
E-mail address: vershynin@math.ucdavis.edu