A FRIENDLY SMOOTHED ANALYSIS OF THE SIMPLEX METHOD*

DANIEL DADUSH † AND SOPHIE HUIBERTS ‡

Abstract. Explaining the excellent practical performance of the simplex method for linear programming has been a major topic of research for over 50 years. One of the most successful frameworks for understanding the simplex method was given by Spielman and Teng (JACM '04), who developed the notion of smoothed analysis. Starting from an arbitrary linear program (LP) with d variables and n constraints, Spielman and Teng analyzed the expected runtime over random perturbations of the LP, known as the smoothed LP, where variance σ^2 Gaussian noise is added to the LP data. In particular, they gave a two-stage shadow vertex simplex algorithm which uses an expected $\widetilde{O}(d^{55}n^{86}\sigma^{-30} + d^{70}n^{86})$ number of simplex pivots to solve the smoothed LP. Their analysis and runtime was substantially improved by Deshpande and Spielman (FOCS '05) and later Vershynin (SICOMP '09). The fastest current algorithm, due to Vershynin, solves the smoothed LP using an expected $O\big(\log^2 n \cdot \log\log n \cdot (d^3\sigma^{-4} + d^5\log^2 n + d^9\log^4 d)\big)$ number of pivots, improving the dependence on n from polynomial to poly-logarithmic.

While the original proof of Spielman and Teng has now been substantially simplified, the resulting analyses are still quite long and complex and the parameter dependencies far from optimal. In this work, we make substantial progress on this front, providing an improved and simpler analysis of shadow simplex methods, where our algorithm requires an expected

$$O(d^2\sqrt{\log n}\,\sigma^{-2} + d^3\log^{3/2} n)$$

number of simplex pivots. We obtain our results via an improved *shadow bound*, key to earlier analyses as well, combined with improvements on algorithmic techniques of Vershynin. As an added bonus, our analysis is completely *modular* and applies to a range of perturbations, which, aside from Gaussians, also includes Laplace perturbations.

Key words. Linear Programming, Shadow Vertex Simplex Method, Smoothed Analysis.

AMS subject classifications. 52B99, 68Q87, 68W40

1. Introduction.

The simplex method for linear programming (LP) is one of the most important algorithms of the 20th century. Invented by Dantzig in 1947 [26, 27], it remains to this day one of the fastest methods for solving LPs in practice. The simplex method is not one algorithm however, but a class of LP algorithms, each differing in the choice of *pivot rule*. At a high level, the simplex method moves from vertex to vertex along edges of the feasible polyhedron, where the pivot rule decides which edges to cross, until an optimal vertex or unbounded ray is found. Important examples include Dantzig's most negative reduced cost [27], the Gass and Saaty parametric objective [41] and Goldfarb's steepest edge [43] method. We note that for solving LPs in the context of branch & bound and cutting plane methods for integer programming, where the successive LPs are "close together", the dual steepest edge method [38] is the *dominant* algorithm in practice [12, 11], due its observed ability to quickly re-optimize.

The continued success of the simplex method in practice is remarkable for two reasons. Firstly, there is no known polynomial time simplex method for LP. Indeed, there are exponential examples for almost every major pivot rule starting with constructions based on *deformed products* [60, 54, 7, 48, 70, 44, 5], such as the Klee-Minty cube [60], which defeat most classical pivot rules, and more recently based on Markov decision processes (MDP) [40, 39], which notably defeat randomized and history dependent pivot rules. Furthermore, for an LP with d variables and n constraints,

*Submitted to the editors June 2018

†Centrum Wiskunde & Informatica, The Netherlands. Supported by NWO Veni grant 639.071.510. Email: dadush@cwi.nl.

‡Centrum Wiskunde & Informatica, The Netherlands. Email: sophie.huiberts@gmail.com.

the fastest provable (randomized) simplex method requires $2^{O(\sqrt{d \log(2+(n-d)/d)})}$ pivots [55, 65, 50], while the observed practical behavior is linear $O(d+n)$ [79]. Secondly, it remains the most popular way to solve LPs despite the tremendous progress for polynomial time methods [59], most notably, interior point methods [57, 75, 69, 62]. How can we explain the simplex method's excellent practical performance?

This question has fascinated researchers for decades. An immediate question is how does one model instances in "practice", or at least instances where simplex should perform well? The research on this subject has broadly speaking followed three different lines: the analysis of average case LP models, where natural distributions of LPs are studied, the smoothed analysis of arbitrary LPs, where small random perturbations are added to the LP data, and work on structured LPs, such as totally unimodular systems and MDPs (Markov Decision Processes). We review the major results for the first two lines in the next section, as they are the most relevant to the present work, and defer additional discussion to the related work section. To formalize the model, we consider LPs in d variables and n constraints of the following form:

$$\begin{aligned} \max\ & \mathbf{c}^\mathsf{T}\mathbf{x} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b}. \end{aligned} \tag{1.1}$$

We denote the feasible polyhedron $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ by P. We now introduce relevant details for the simplex methods of interest to this work.

Parametric Simplex Algorithms. While a variety of pivot rules have been studied, the most successfully analyzed in theory are the so-called parametric simplex methods, due to the useful geometric characterization of the paths they follow. The first such method, and the main one used in the context of smoothed analysis, is the parametric objective method of Gass and Saaty [41], dubbed the *shadow (vertex) simplex* method by Borgwardt [16]. Starting at a *known* vertex $\mathbf{v}$ of P maximizing an objective $\mathbf{c}'$, the parametric objective method computes the path corresponding to the sequence of maximizers for the objectives obtained by interpolating $\mathbf{c}' \to \mathbf{c}$ [1]. The name shadow vertex method is derived from the fact that the visited vertices are in correspondence with those on the projection of P onto $W := \text{span}(\mathbf{c}, \mathbf{c}')$, the 2D convex polygon known as the shadow of P on W (see figure 2.2 for an illustration). In particular, the number of vertices traversed by the method is bounded by the number of vertices of the projection, known as the *size of the shadow*.

An obvious problem, as with most simplex methods, is how to initialize the method at a feasible vertex if one exists. This is generally referred to as the Phase I problem, where Phase II then corresponds to finding an optimal solution. A common Phase I adds artificial variable(s) to make feasibility trivial and applies the simplex method to drive them to zero.

A more general method, popular in the context of average case analysis, is the self-dual parametric simplex method of Dantzig [28]. In this method, one *simultaneously* interpolates the objectives $\mathbf{c}' \to \mathbf{c}$ and right hand sides $\mathbf{b}' \to \mathbf{b}$ which has the effect of combining Phase I and II. Here $\mathbf{c}'$ and $\mathbf{b}'$ are chosen to induce a *known* initial maximizer. While the polyhedron is no longer fixed, the breakpoints in the path of maximizers (now a piecewise linear curve) can be computed via certain primal and dual pivots. This procedure was in fact generalized by Lemke [63] to solve linear complementarity problems. We note that the self dual method can roughly speaking be simulated in a higher dimensional space by adding an interpolation variable λ,

[1] This path is well-defined under mild non-degeneracy assumptions

i.e. $\mathbf{Ax} \leq \lambda\mathbf{b} + (1-\lambda)\mathbf{b}', 0 \leq \lambda \leq 1$, which has been the principal approach in smoothed analysis.

1.1. Prior Work. Here we present the main works in both average case and smoothed analysis which inform our main results, presented in the next section. A common theme in these works, which all study parametric simplex methods, is to first obtain a bound on the expected parametric path length, with respect to some distribution on interpolations and LPs, and then find a way to use the bounds algorithmically. This second step can be non-obvious, as it is often the case that one cannot directly find a starting vertex on the path in question. We now present the main random LP models that have been studied, presenting path length bounds and algorithms. Lastly, as our results are in the smoothed analysis setting, we explain the high level strategies used to prove smoothed (shadow) path bounds.

Average case Models. The first model, introduced in the seminal work of Borgwardt [16, 17, 18, 19], examined LPs of the form $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{Ax} \leq \mathbf{1}$, possibly with $\mathbf{x} \geq \mathbf{0}$ constraints (note that this model is always feasible at $\mathbf{0}$), where the rows of $\mathbf{A} \in \mathbb{R}^{n\times d}$ are drawn i.i.d. from a rotationally symmetric distribution (RSD) and $\mathbf{c} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ is fixed and non-zero. Borgwardt proved tight bounds on the expected shadow size of the feasible polyhedron when projected onto any fixed plane. For general RSD, he proved a sharp $\Theta(d^2 n^{1/(d-1)})$ [18, 19] bound, tight for rows drawn uniformly from the sphere, and for Gaussians a sharp $\Theta(d^{1.5}\sqrt{\log n})$ bound [18], though this last bound is only known to hold asymptotically as $n \to \infty$ (i.e., very large compared to d). On the algorithmic side, Borgwardt [17] gave a *dimension by dimension* (DD) algorithm which optimizes over such polytopes by traversing $d-2$ different shadow vertex paths. The DD algorithm proceeds by iteratively solving the restrictions $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{Ax} \leq 1, x_i = 0, i \in \{k+1, \ldots, d\}$, for $k \geq 2$, which are all of RSD type.

For the next class, Smale [80] analyzed the standard self dual method for LPs where $\mathbf{A}$ and $(\mathbf{c}, \mathbf{b})$ are chosen from independent RSM distributions, where Megiddo [68] gave the best known bound of $f(\min\{d, n\})$ iterations, for some exponentially large function f. Adler [2] and Haimovich [49] examined a much weaker model where the data is fixed, but where the signs of all the inequalities, including non-negativity constraints, are flipped uniformly at random. Using the combinatorics of hyperplane arrangements, they achieved a remarkable bound of $O(\min\{d, n\})$ for the expected length of parametric paths. These results were made algorithmic shortly thereafter [84, 4, 3], where it was shown that a lexicographic version of the parametric self dual simplex method[2] requires $\Theta(\min\{d, n\}^2)$ iterations, where tightness was established in [4]. While these results are impressive, a notable criticism of the symmetry model is that it results in infeasible LPs almost surely once n is a bit larger than d.

Smoothed LP Models. The *smoothed* analysis framework, introduced in the breakthrough work of Spielman and Teng [82], helps explain the performance of algorithms whose worst-case examples are in essence *pathological*, i.e., which arise from very brittle structures in instance data. To get rid of these structures, the idea is to add a small amount of noise to the data, quantified by a parameter σ, where the general goal is then to prove an expected runtime bound over any *smoothed instance* that scales inverse polynomially with σ. Beyond the simplex method, smoothed analysis has been successfully applied to many other algorithms such as interior point methods [81], Gaussian elimination [76], Lloyd's k-means algorithm [6], the 2-OPT heuristic for the

[2]These works use seemingly different algorithms, though they were shown to be equivalent to a lexicographic self-dual simplex method by Meggiddo [67].

TSP [36], and much more.

The smoothed LP model, introduced by [82], starts with any base LP

$$\max \mathbf{c}^\mathsf{T}\mathbf{x},\ \bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{b}}, \qquad \text{(Base LP)}$$

$\bar{\mathbf{A}} \in \mathbb{R}^{n\times d}$, $\bar{\mathbf{b}} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, where the rows of $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$ are normalized to have ℓ_2 norm at most 1. From the base LP, we generate the smoothed LP by adding Gaussian perturbations to both the constraint matrix $\bar{\mathbf{A}}$ and the right hand side $\bar{\mathbf{b}}$. Precisely, the data of the smoothed LP is

$$\mathbf{A} = \bar{\mathbf{A}} + \hat{\mathbf{A}},\ \mathbf{b} = \bar{\mathbf{b}} + \hat{\mathbf{b}},\ \mathbf{c}, \qquad \text{(Smoothed LP Data)}$$

where the perturbations $\hat{\mathbf{A}}$,$\hat{\mathbf{b}}$ have i.i.d. mean 0, variance σ^2 Gaussian entries. Note that the objective is not perturbed in this model, though we require that $\mathbf{c} \neq \mathbf{0}$. An LP algorithm is said to have polynomial smoothed complexity if for any base LP data $\bar{\mathbf{A}}, \bar{\mathbf{b}}, \mathbf{c}$ as above, we have

$$\mathbb{E}_{\hat{\mathbf{A}},\hat{\mathbf{b}}}[T(\mathbf{A}, \mathbf{b}, \mathbf{c})] = \text{poly}(n, d, 1/\sigma), \qquad \text{(Smoothed Complexity)}$$

where $T(\mathbf{A}, \mathbf{b}, \mathbf{c})$ is the runtime of the algorithm on a given smoothed instance. Crucially, this complexity measure allows for an inverse polynomial dependence on σ, the perturbation size. Focusing on the simplex method, T will measure the number of simplex pivots used by the algorithm as a proxy for the runtime.

Spielman and Teng [82] proved the first polynomial smoothed complexity bound for the simplex method. In particular, they gave a two phase shadow vertex method which uses an expected $\widetilde{O}(d^{55}n^{86}\sigma^{-30} + d^{70}n^{86})$ number of pivots. This bound was substantially improved by Deshpande and Spielman [30] and Vershynin [86], where Vershynin gave the fastest such method requiring an expected

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

number of pivots.

In all these works, the complexity of the algorithms is reduced in a black box manner to a shadow bound for *smoothed unit LPs*. In particular, a smoothed unit LP has a base system $\bar{\mathbf{A}}\mathbf{x} \leq \mathbf{1}$, where $\bar{\mathbf{A}}$ has row norms at most 1, and smoothing is performed only to $\bar{\mathbf{A}}$. Here the goal is to obtain a bound on the expected shadow size with respect to any fixed plane. Note that if $\bar{\mathbf{A}}$ is the zero matrix, then this is exactly Borgwardt's Gaussian model, where he achieved the asymptotically tight bound of $\Theta(d^{1.5}\sqrt{\log n})$ as $n \to \infty$ [18]. For smoothed unit LPs, Spielman and Teng [82] gave the first bound of $O(d^3 n\sigma^{-6} + d^6 n \log^3 n)$. Deshpande and Spielman [30] derived a bound of $O(dn^2 \log n\sigma^{-2} + d^2n^2 \log^2 n)$, substantially improving the dependence on σ while doubling the dependence on n. Lastly, Vershynin [86] achieved a bound of $O(d^3\sigma^{-4}+d^5 \log^2 n)$, dramatically improving the dependence on n to poly-logarithmic, though still with a worse dependence on σ than [30].

Before discussing the high level ideas for how these bounds are proved, we overview how they are used algorithmically. In this context, [82] and [86] provide two different reductions to the unit LP analysis, each via an interpolation method. Spielman and Teng first solve the smoothed LP with respect to an artificial "somewhat uniform" right hand side $\mathbf{b}'$, constructed to force a randomly chosen basis of $\mathbf{A}$ to yield a vertex of the artificial system. From here they use the shadow vertex method to compute a maximizer for right hand side $\mathbf{b}'$, and continue via interpolation to derive

an optimal solution for $\mathbf{b}$. Here the analysis is quite challenging, since in both steps the LPs are not quite smoothed unit LPs and the used shadow planes correlate with the perturbations. To circumvent these issues, Vershynin uses a *random vertex* (RV) algorithm, which starts with $\mathbf{b}' = \mathbf{1}$ (i.e. a unit LP) and adds a random additional set of d inequalities to the system to induce an "uncorrelated known vertex". From this random vertex, he proceeds similarly to Spielman and Teng, but now at every step the LP is of smoothed unit type and the used shadow planes are (almost) independent of the perturbations. In Vershynin's approach, the main hurdle was to give a simple shadow vertex algorithm to solve unit LPs, which correspond to the Phase 1 problem. An extremely simple method for this was in fact already given in the early work of Borgwardt [18], namely, the dimension by dimension (DD) algorithm. The application of the DD algorithm in the smoothed analysis context was however only discovered much later by Schnalzger [78]. As it is both simple and not widely known, we will describe the DD algorithm and its analysis in Section 4.

We note that beyond the above model, smoothed analysis techniques have been used to analyze the simplex method in other interesting settings. In [22], the successive shortest path algorithm for min-cost flow, which is a shadow vertex algorithm, was shown to be efficient when only the objective (i.e. edge costs) is perturbed. In [58], Kelner and Spielman used smoothed analysis techniques to give a "simplex like" algorithm which solves arbitrary LPs in polynomial time. Here they developed a technique to analyze the expected shadow size when only the right hand side of an LP is perturbed.

Shadow Bounds for Smoothed Unit LPs. Let $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d, i \in [n]$, denote the rows of the constraint matrix of the smoothed unit LP $\mathbf{Ax} \leq \mathbf{1}$. The goal is to bound the expected number of vertices in the projection of the feasible polyhedron P onto a fixed 2D plane W. As noticed by Borgwardt, by a duality argument, this number of vertices is upper bounded by the number of edges in the *polar polygon* (see figure 2.2 for an illustration). Letting $Q := \text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n)$, the convex hull of the rows, the polar polygon can be expressed as $D := Q \cap W$.

We overview the different approaches used in [82, 30, 86] to bound the number of edges of D. Let $\mathbf{u}_\theta$, $\theta \in [0, 2\pi]$, denote an angular parametrization of the unit circle in W, and let $\mathbf{r}_\theta = \mathbf{u}_\theta \cdot \mathbb{R}_{\geq 0}$ denote the corresponding ray. Spielman and Teng [82] bounded the probability that any two nearby rays $\mathbf{r}_\theta$ and $\mathbf{r}_{\theta+\varepsilon}$ intersect different edges of D by a linear function of ε. Summing this probability over any fine enough discretization of the circle upper bounds the expected number of edges of D [3]. Their probability bound proceeds in two steps, first they estimate the probability that the Euclidean distance between the intersection of $\mathbf{r}_\theta$ with its corresponding edge and the boundary of that edge is small (the *distance lemma*), and second they estimate the probability that angular distance is small compared to Euclidean distance (the *angle of incidence bound*). Vershynin [86] avoided the use of the angle of incidence bound by measuring the intersection probabilities with respect to the "best" of three different viewpoints, i.e. where the rays emanate from a well-chosen set of three equally spaced viewpoints as opposed to just the origin. This gave a much more efficient reduction to the distance lemma, and in particular allowed Vershynin to reduce the dependence on n from linear to poly-logarithmic. Deshpande and Spielman [30] bounded different probabilities to get their shadow bound. Namely, they bounded the probability that nearby objectives $\mathbf{u}_\theta$ and $\mathbf{u}_{\theta+\varepsilon}$ are maximized at different vertices of D. The corresponding discretized sum over the circle directly bounds the number of vertices of D,

[3] One must be a bit more careful when D does not contain the origin, but the details are similar.

which is the same as the number of edges.

Complexity in two dimensions. In two dimensions, the shadow size reduces to the complexity of the convex hull. The smoothed complexity of the convex hull was first considered by Damerow and Sohler [25], who obtained an upper bound of $O((1+\sigma^{-d})\log^{3d/2} n)$ expected non-dominated points in d variables, bounding the smoothed complexity of a two-dimensional convex hull by $O((1+\sigma^{-2})\log^{3/2} n)$ vertices. Schnalzger [78] proved a complexity bound of $O(\sigma^{-2}+\log n)$ based on the discretized sum approach of Spielman and Teng [82]. The best general bounds were proved by Devillers, Glisse, Goaoc, and Thomasse [32], who considered both i.i.d. Gaussian noise and i.i.d. uniform noise drawn from a scaling of the sphere $\delta\mathbb{S}^{d-1}$. In the Gaussian case, they prove an upper bound of $O((1+\sigma^{-1})\sqrt{\log n})$ and lower bounds of $\Omega(n)$ for $0 \le \sigma \le \frac{1}{n^2}$, $\Omega(\frac{1}{\sqrt{\sigma}}\sqrt[4]{\log(n\sqrt{\sigma})})$ for $\frac{1}{n^2} \le \sigma \le \frac{1}{\sqrt{\log n}}$ and $\Omega(\sqrt{\ln n})$ for $\frac{1}{\sqrt{\log n}} \le \sigma$.

1.2. Results. While the original proof of Spielman and Teng has now been substantially simplified, the resulting analyses are still complex and the parameter improvements have not been uniform. In this work, we give a "best of all worlds" analysis, which is both much simpler and improves all prior parameter dependencies. Our main contribution is a substantially improved shadow bound, presented below.

We note that some of the bounds below (including ours) only hold for $d \ge 3$. Recalling the models, the results in the Table 1.1 bound the expected number of vertices in the projection of a random polytope $\mathbf{Ax} \le \mathbf{1}$, $\mathbf{A} \in \mathbb{R}^{n\times d}$, onto any fixed 2-dimensional plane. The models differ in the class of distributions examined for $\mathbf{A}$. In the RSD model, the rows of $\mathbf{A}$ are distributed i.i.d. according to an *arbitrary* rotationally symmetric distribution. In the Gaussian model, the rows of $\mathbf{A}$ are i.i.d. mean zero standard Gaussian vectors. Note that this is a special case of the RSD model. In the smoothed model, the rows of $\mathbf{A}$ are d-dimensional Gaussian random vectors with standard deviation σ centered at vectors of norm at most 1, i.e. the expected matrix $\mathbb{E}[\mathbf{A}]$ has rows of ℓ_2 norm at most 1. The $n \to \infty$ in the table indicates that that bound only holds for n large enough (compared to d). The Gaussian, $n \to \infty$ model is a special case of the smoothed analysis model, and hence the $\Omega(d^{3/2}\sqrt{\log n})$ bound also holds in the smoothed model for n big enough.

As can be seen, our new shadow bound yields a substantial improvement over earlier smoothed bounds in all regimes of σ and is also competitive in the Gaussian model. For small σ, our bound improves the dependence on d from d^3 to d^2, achieves the same σ^{-2} dependence as [30], and improves the dependence on n to $\sqrt{\log n}$. For $\sigma \ge 1$, our bound becomes $O(d^{2.5}\log^{3/2} n)$, which in comparison to Borgwardt's optimal (asymptotic) Gaussian bound is only off by a $d\log n$ factor. Furthermore, our proof is substantially simpler than Borgwardt's and holds for all n and d. No interesting lower bounds for the small σ regime are known for $d \ge 3$, though the results of [32, 31, 18] suggest that the correct lower bound might be much lower than current upper bounds. We leave these questions as open problems.

An interesting point of our analysis is that it is *completely modular*, and that it gives bounds for perturbations other than Gaussians. In fact, in our approach it is easier to obtain bounds for Laplace perturbations (see Section 3) than for the Gaussian distribution. The range of analyzable perturbations still remains limited however, our analysis doesn't extend to bounded perturbations such as uniform $[-1/\sigma, 1/\sigma]$ for example. As is well known, LPs in practice tend to be sparse and hence don't follow

Works	Expected Number of Vertices	Model
[19]	$\Theta(d^2 n^{1/(d-1)})$	RSD
[18]	$\Theta(d^{3/2}\sqrt{\log n})$	Gaussian, $n \to \infty$
[82]	$O(d^3 n\sigma^{-6} + d^6 n \log^3 n)$	Smooth
[30]	$O(dn^2 \log n\ \sigma^{-2} + d^2 n^2 \log^2 n)$	Smooth
[86]	$O(d^3\sigma^{-4} + d^5 \log^2 n)$	Smooth
This paper	$O(d^2\sqrt{\log n}\ \sigma^{-2} + d^{2.5}\log^{3/2} n(1+\sigma^{-1}))$	Smooth

TABLE 1.1
Shadow Bounds. Logarithmic factors are simplified. The Gaussian, $n \to \infty$ lower bound applies in the smoothed model as well.

Works	Expected Number of Pivots	Model	Algorithm
[18, 78]	$d \cdot$ max shadow size	Multiple	DD + Int. LP
[18, 51, 19]	$O(d^{2.5} n^{1/(d-1)})$	RSD, $n \to \infty$	DD
[86]	$O\big(\log^3 n \cdot (d^3\sigma^{-4} + d^5 \log^2 n + d^9 \log^4 d)\big)$	Smooth	RV + Int.LP
This paper	$O(d^2\sqrt{\log n}\ \sigma^{-2} + d^3 \log^{3/2} n)$	Smooth	Symmetric RV + Int. LP

TABLE 1.2
Runtime bounds. Logarithmic factors are simplified.

a Gaussian distribution (which yields a totally dense constraint matrix). It is thus of considerable interest to understand the smoothed behavior of the simplex method under wider classes of perturbations, such as perturbations with restricted support.

From the algorithmic perspective, we describe the two phase interpolation approach of Vershynin [86], which we instantiate using two different Phase 1 algorithms to solve unit LPs. As a warmup, we first describe Schnalzger's application of the dimension by dimension (DD) algorithm [78], as it yields the simplest known Phase 1 algorithm and is not widely known. Following this, we introduce a new variant of Vershynin's random vertex (Symmetric RV) algorithm which induces an artificial (degenerate) random vertex by adding $2d-2$ inequalities placed symmetrically around a randomly chosen objective. The symmetry condition ensures that this random vertex optimizes the chosen objective with probability 1. Vershynin's original approach added d random inequalities, which only guaranteed that the induced vertex optimized the chosen objective with $1 - 1/\text{poly}(d)$ probability. Via a more careful analysis of the RV algorithm combined with the additional guarantees ensured by our variant, we derive a substantially improved complexity estimate. Specifically, our Symmetric RV algorithm runs in time $O(d^2\sqrt{\log n}\ \sigma^{-2} + d^3 \log^{3/2} n)$, which is faster than both the original RV algorithm and Borgwardt's dimension by dimension algorithm in all parameter regimes. We defer further discussion of this to section 4 of the paper.

1.3. Techniques: Improved Shadow Bound. We now give a detailed sketch of the proof of our improved shadow bound. Proofs of all claims can be found in section 3. The outline of the presentation is as follows. To begin, we explain our general

edge counting strategy, where we depart from the previously discussed analyses. In particular, we adapt the approach of Kelner and Spielman (KS) [58], who analyzed a smoothing model where only the right-hand side is perturbed, to the present setting. Following this, we present a parametrized shadow bound, which applies to any class of perturbations for which the relevant parameters are bounded. The main motivation of the abstraction in the parametrized model is us to clearly identify the relevant properties of the perturbations we need to obtain shadow bounds. Lastly, we give the high-level idea of how we estimate the relevant quantities in the KS approach within the parametrized model.

Edge Counting Strategy. Our goal is to compute a bound on the expected number of edges in the polygon $Q \cap W$, where W is the two-dimensional shadow plane, $Q := \text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ are the smoothed constraints of a unit LP. Recall that this is an upper bound on the shadow size.

In [58], Kelner and Spielman developed a very elegant and useful alternative strategy to bound the expected number of edges, which can be applied to many distributions over 2D convex polygons. Whereas they analyzed the geometry of the primal shadow polygon, the projection of P onto W, we will instead work with the geometry of the polar polygon $Q \cap W$. The analysis begins with the following elementary identity:

$$\mathbb{E}[\text{perimeter}(Q \cap W)] = \mathbb{E}[\sum_{\mathbf{e} \in \text{edges}(Q \cap W)} \text{length}(\mathbf{e})] \; . \tag{1.2}$$

Starting from the above identity, the approach first derives a good upper bound on the perimeter and a lower bound on the right-hand side in terms of the number of edges and the minimum edge length. The bound on the number of edges is then derived as the ratio of the perimeter bound and the minimum edge length.

We focus first on the perimeter upper bound. Since $Q \cap W$ is convex, any containing circle has larger perimeter. Furthermore, we clearly have $Q \cap W \subseteq \pi_W(Q)$, where π_W is the orthogonal projection onto W. Combining these two observations, we derive the first useful inequalities:

$$\mathbb{E}[\text{perimeter}(Q \cap W)] \leq \mathbb{E}[2\pi \max_{\mathbf{x} \in Q \cap W} \|\mathbf{x}\|] \leq \mathbb{E}[2\pi \max_{i \in [n]} \|\pi_W(\mathbf{a}_i)\|] \; . \tag{1.3}$$

To extract the expected number of edges from the right hand side of (1.2), we first note that every edge of $Q \cap W$ is derived from a facet of Q intersected with W (see Figure 2.2 for an illustration). Assuming non-degeneracy, the possible facets of Q are $F_I := \text{conv}(\mathbf{a}_i : i \in I)$, where $I \subseteq [n]$ is any subset of size d. Let E_I denote the event that F_I induces an edge of $Q \cap W$, more precisely, that F_I is a facet of Q and that $F_I \cap W \neq \emptyset$. From here, we get that

$$\begin{aligned} \mathbb{E}[\sum_{\mathbf{e} \in \text{edges}(Q \cap W)} \text{length}(\mathbf{e})] &= \sum_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \Pr[E_I] \\ &\geq \min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \cdot \sum_{|I|=d} \Pr[E_I] \\ &= \min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \cdot \mathbb{E}[|\text{edges}(Q \cap W)|] \; . \end{aligned} \tag{1.4}$$

Combining (1.2), (1.3), (1.4), we derive the following fundamental bound:

$$\mathbb{E}[|\text{edges}(Q \cap W)|] \leq \frac{\mathbb{E}[2\pi \max_{i\in[n]} \|\pi_W(\mathbf{a}_i)\|]}{\min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I]} \ . \tag{1.5}$$

In the actual proof, we further restrict our attention to potential edges having probability $\Pr[E_I] \geq 2\binom{n}{d}^{-1}$ of appearing, which helps control how extreme the conditioning on E_I can be. Note that the edges appearing with probability smaller than $2\binom{n}{d}^{-1}$ contribute at most 2 to the expectated number of edges, and hence can be ignored. Thus our task now directly reduces to showing that the maximum perturbation is not too large on average, an easy condition, while ensuring that the edges that are not too unlikely to appear are reasonably long on average, the more difficult condition.

We note that applying the KS approach already improves the situation with respect to the maximum perturbation size compared to earlier analyses, as [82, 30, 86] all require a bound to hold with high probability as opposed to on expectation. For this purpose, they enforced the condition $1/\sigma \geq \sqrt{d \log n}$ (for Gaussian perturbations), which we do not require here.

Bound for Parametrized Distributions. We now present the parameters of the pertubation distributions we use to obtain our bounds on the enumerator and denominator of 1.5. We also discuss how these parameters behave for the Gaussian and Laplace distribution.

Let us now assume that $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ are independently distributed. As before we assume that the centers $\bar{\mathbf{a}}_i := \mathbb{E}[\mathbf{a}_i]$, $i \in [n]$, have norm at most 1. We denote the perturbations by $\hat{\mathbf{a}}_i := \mathbf{a}_i - \bar{\mathbf{a}}_i$, $i \in [n]$. We will assume for simplicity of the presentation that all the perturbations $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_n$ are i.i.d. according to a distribution with probability density μ (in general, they could each have a distinct distribution).

At a high-level, the main properties we require of the distribution are that it be smooth and that it have sufficiently strong tail bounds. We formalize these requirements via the following 4 parameters, where we let $\mathbf{X} \sim \mu$ below:

1. μ is an L-log-Lipschitz probability density function, that is, $|\log \mu(\mathbf{x}) - \log \mu(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.
2. The variance of $\mathbf{X}$, when restricted to any line $l \subset \mathbb{R}^d$, is at least τ^2.
3. The cutoff radius $R_{n,d} > 0$ is such that $\Pr[\|\mathbf{X}\| \geq R_{n,d}] \leq \frac{1}{d\binom{n}{d}}$.
4. The n-th deviation r_n is such that, for all $\boldsymbol{\theta} \in \mathbb{R}^d$, $\|\boldsymbol{\theta}\| = 1$, and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ i.i.d., we have $\mathbb{E}[\max_{i\in[n]} |\langle \mathbf{X}_i, \boldsymbol{\theta}\rangle|] \leq r_n$.

We refer the reader to subsection 3.1.1 for more formal definitions of these parameters. We note that these parameters naturally arise from our proof strategy and directly expose the relevant quantities for our shadow bound.

The first two parameters are smoothness related while the last two relate to tail bounds. Using these four parameters, we will derive appropriate bounds for the enumerator and denominator in (1.5). Assuming the above parameter bounds for $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_n$, our main "plug and play" bound on the expected shadow size is as follows (see Theorem 3.10):

$$\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] = O\left(\frac{d^{1.5}L}{\tau}(1 + R_{n,d})(1 + r_n)\right) \ . \tag{1.6}$$

We can use this parametrized bound to prove the shadow bound for Gaussian and Laplace distributed noise. For the variance σ^2 Gaussian distribution in $\mathbb{R}^d$, it is direct

to verify that $\tau = \sigma$ for any line (since every line restriction results in a 1D variance σ^2 Gaussian), and from standard tail bounds that $R_{n,d} = O(\sigma\sqrt{d \log n})$ and $r_n = O(\sigma\sqrt{\log n})$. The only parameter that cannot be bounded directly is the log-Lipschitz parameter L, since $\|\mathbf{x}/\sigma\|^2/2$, the log of the Gaussian density, is quadratic. For Laplace distributed perturbations however, this last difficulty is completely avoided. Here a comparably sized Laplace perturbation (i.e. same expected norm) has density proportional to $e^{-(\sqrt{d}/\sigma)\|\mathbf{x}\|}$ which is by definition log-Lipshitz with $L = \sqrt{d}/\sigma$. The other parameters are somewhat worse, it can be shown that $R_{n,d} = O(\sigma\sqrt{d}\log n)$, $r_n = O(\sigma \log n)$ and $\tau \geq \sigma/\sqrt{d}$, where in particular τ is a $\sqrt{d}$-factor smaller than the Gaussian. Thus, for Laplace perturbations our parametrized bound applies directly and yields a bound of $O(d^{2.5}\sigma^{-2})$ is the small σ regime.

To apply our analysis to the Gaussian setting, we start with the fact, noted in all prior analyses, that the Gaussian is locally smooth within any fixed radius. In particular, within radius $R_{n,d}$ of the mean, the Gaussian density is $O(\sqrt{d \log n}/\sigma)$-log-Lipschitz. As events that happen with probability $\ll \binom{n}{d}^{-1}$ have little effect on the expected shadow bound (recall that the shadow is always bounded by $\binom{n}{d}$), one can hope to condition on each perturbation living inside the $R_{n,d}$ radius ball. This is in fact the approach taken in the prior analyses [82, 30, 86]. This conditioning however does not ensure full log-Lipshitzness and causes problems for points near the boundary. Furthermore, the conditioning may also decrease line variances for lines near the boundary.

To understand why this is problematic, we note that the main role of the smoothness parameters L and τ is to ensure enough "wiggle-room" to guarantee that edges induced by any fixed basis are long on expectation. Using the above conditioning, it is clear that edges induced by facets whose perturbations occur close to the $R_{n,d}$ boundary must be dealt with carefully. To avoid such difficulties altogether, we leverage the local log-Lipshitzness of the Gaussian in a "smoother" way. Instead of conditioning, we simply replace the Gaussian density with a globally $O(\sqrt{d \log n}/\sigma)$-log-Lipshitz density which has statistical distance $\ll \binom{n}{d}^{-1}$ to the Gaussian (thus preserving the shadow bound) and also yields nearly identical bounds for the other parameters. This distribution will consist of an appropriate gluing of a Gaussian and Laplace density, which we call the Laplace-Gaussian distribution (see section 3.3 for details). Thus, by slightly modifying the distribution, we are able to use our parametrized model to obtain shadow bounds for Gaussian perturbations in a black box manner.

Bounding the Perimeter and Edge Length. We now briefly describe how the perimeter and minimum edge length in (1.5) are bounded in our parametrized perturbation model to obtain (1.6). As this is the most technical part of the analysis, we refer the reader to the proofs in section 3 and give only a very rough discussion here. As above, we will assume that the perturbations satisfy the bounds given by $L, \tau, R_{n,d}, r_n$.

For the perimeter bound, we immediately derive the bound

$$\mathbb{E}[\max_{i\in[n]} \|\pi_W(\mathbf{a}_i)\|] \leq 1 + \mathbb{E}[\max_{i\in[n]} \|\pi_W(\hat{\mathbf{a}}_i)\|] \leq 1 + 2r_n$$

by the triangle inequality. From here, we must bound the minimum expected edge length, which requires the majority of the work. For this task, we provide a clean analysis, which shares high-level similarities with the Spielman and Teng distance lemma, though our task is simpler. Firstly, we only need to show that an edge is large on average, whereas the distance lemma has the more difficult task of proving that

an edge is unlikely to be small. Second, our conditioning is much milder. Namely, the distance lemma conditions a facet F_I on intersecting a specified ray $\mathbf{r}_\theta$, whereas we only condition F_I on intersecting W. This conditioning gives the edge much more "wiggle room", and is the main leverage we use to get the factor d improvement.

Let us fix $F := F_{[d]} = \text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ as the potential facet of interest, under the assumption that $E := E_{[d]}$, i.e., that F induces an edge of $Q \cap W$, has probability at least $2\binom{n}{d}^{-1}$. Our analysis of the edge length conditioned on E proceeds as follows:

1. Show that if F induces an edge, then under this conditioning F has *small diameter* with good probability, namely its vertices are all at distance at most $O(1 + R_{n,d})$ from each other (Lemma 3.17). This uses the tailbound defining $R_{n,d}$ and the fact that E occurs with non-trivial probability.
2. Condition on F being a facet of Q by fixing its containing affine hyperplane H (Lemma 3.20). This is standard and is achieved using a change of variables analyzed by Blaschke (see section 2.4 for details).
3. Let $l := H \cap W$ denote the line which intersects F to form an edge of $Q \cap W$. Show that on average the longest chord of F parallel to l is long. We achieve the bound $\Omega(\tau/\sqrt{d})$ (Lemma 3.29) using that the vertices of F restricted to lines parallel to l have variance at least τ^2.
4. Show that on average F is pierced by l through a chord that is not too much shorter than the longest one. Here we derive the final bound on the expected edge length of

$$\mathbb{E}[\text{length}(F \cap W) \mid E] = \Omega((\tau/\sqrt{d}) \cdot 1/(dL(1 + R_{n,d}))) \text{ (Lemma 3.27)}$$

using the fact that the distribution of the vertices is L-log-Lipschitz and that F has diameter $O(1 + R_{n,d})$.

This concludes the high-level discussion of the proof.

1.4. Related work.

Structured Polytopes. An important line of work has been to study LPs with good geometric or combinatorial properties. Much work has been done to analyze primal and dual network simplex algorithms for fundamental combinatorial problems on flow polyhedra such as bipartite matching [53], shortest path [33, 47], maximum flow [45, 42] and minimum cost flow [72, 46, 73]. Generalizing on the purely combinatorial setting, LPs where the constraint matrix $\mathbf{A} \in \mathbb{Z}^{n \times d}$ is totally unimodular (TU), i.e. the determinant of any square submatrix of $\mathbf{A}$ is in $\{0, \pm 1\}$, were analyzed by Dyer and Frieze [34], who gave a random walk based simplex algorithm which requires $\text{poly}(d, n)$ pivots. Recently, an improved random walk approach was given by Eisenbrand and Vempala [35], which works in the more general setting where the subdeterminants are bounded in absolute value by Δ, who gave an $O(\text{poly}(d, \Delta))$ bound on the number of Phase II pivots (note that there is no dependence on n). Furthermore, randomized variants of the shadow vertex algorithm were analyzed in this setting by [23, 24], where in particular [24] gave an expected $O(d^5\Delta^2 \log(d\Delta))$ bound on the number of Phase I and II pivots. Another interesting class of structured polytopes comes from the LPs associated with Markov Decision Processes (MDP), where simplex rules such as Dantzig's most negative reduced cost correspond to variants of policy iteration. Ye [87] gave polynomial bounds for Dantzig's rule and Howard's policy iteration for MDPs with a fixed discount rate, and Ye and Post [74] showed that Dantzig's rule converges in strongly polynomial time for deterministic MDPs with variable discount rates.

Diameter Bounds. Another important line of research has been to establish diameter bounds for polyhedra, namely to give upper bounds on the shortest path length between any two vertices of a polyhedron as a function of the dimension d and the number of inequalities n. For any simplex method pivoting on the vertices of a fixed polytope, the diameter is clearly a lower bound on the worst-case number of pivots. The famous Hirsch conjecture from 1957, posited that for polytopes (bounded polyhedra) the correct bound should be $n - d$. This precise bound was recently disproven by Santos [77], who gave a 43 dimensional counter-example, improved to 20 in [66], where the diameter is about $1.05(n - d)$ (these counter-examples can also be extended to infinite families). However, the possibility of a polynomial (or even linear) bound is still left open, and is known as the polynomial Hirsch conjecture. From this standpoint, the best general results are the $O(2^d n)$ bound by Barnette [10] and Larman [61], and the quasi-polynomial bound of Kalai and Kleitman [56], recently refined by Todd [85] and Sukegawa [83] to $(n-d)^{\log_2 O(d/\log d)}$. As above, such bounds have been studied for structured classes of polytopes. In particular, the diameter of polytopes with bounded subdeterminants was studied by various authors [34, 14, 24], where the best known bound of $O(d^3\Delta^2 \log(d\Delta))$ was given in [24]. The diameters of other classes such as 0/1 polytopes [71], transportation polytopes [8, 21, 29, 20] and flag polytopes [1] have also been studied.

1.5. Organization. Section 2 contains basic definitions and background material. The proofs of our shadow bounds are given in section 3, 3.2 and 3.3 respectively. The details regarding the two phase shadow vertex algorithm we use, which rely in an almost black box way on the shadow bound, are presented in section 4.

2. Preliminaries.

2.1. Notation and basic definitions.

- Vectors are printed in bold to contrast with scalars: $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The space $\mathbb{R}^d$ comes with a standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_d$. We abbreviate $\mathbf{1} := (1, 1, \ldots, 1)$ and $\mathbf{0} := (0, 0, \ldots, 0)$. Vector inequalities are defined coordinate-wise: $\mathbf{v} \leq \mathbf{w}$ if and only if $v_i \leq w_i$ for all $i \leq d$.
- We abbreviate $[n] := \{1, \ldots, n\}$ and $\binom{[n]}{d} = \{I \subset [n] \mid |I| = d\}$. For $a, b \in \mathbb{R}$ we have intervals $[a, b] = \{r \in \mathbb{R} : a \leq r \leq b\}$ and $(a, b) = \{r \in \mathbb{R} : a < r < b\}$.
- For $x > 0$, we define $\log x$ to be the logarithm base e of x.
- For a set $C \subseteq \mathbb{R}^n$, we denote its topological boundary by ∂C.
- For a set A, we use the notation $\mathbb{1}[\mathbf{x} \in A]$ to denote the indicator function of A, i.e., $\mathbb{1}[\mathbf{x} \in A] = 1$ if $\mathbf{x} \in A$ and 0 otherwise.
- For $A, B \subset \mathbb{R}^d$ we write the Minkowski sum $A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\}$. For a vector $\mathbf{v} \in \mathbb{R}^d$ we write $A + \mathbf{v} = A + \{\mathbf{v}\}$. For a set of scalars $S \subset \mathbb{R}$ we write $\mathbf{v} \cdot S = \{s\mathbf{v} : s \in S\}$.
- The inner product of $\mathbf{x}$ and $\mathbf{y}$ is written with as $\mathbf{x}^\intercal \mathbf{y} = \sum_{i=1}^d x_i y_i$. We use the ℓ_2-norm $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\intercal \mathbf{x}}$ and the ℓ_1-norm $\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$. Every norm without subscript is the ℓ_2-norm. The unit sphere in $\mathbb{R}^d$ is $\mathbb{S}^{d-1} = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right\}$ and the unit ball is $\mathcal{B}_2^d = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\right\}$.
- A set $V + \mathbf{p}$ is an affine subspace if $V \subset \mathbb{R}^d$ is a linear subspace. If $S \subset \mathbb{R}^d$ then the affine hull $\mathrm{aff}(S)$ is the smallest affine subspace containing S. We say $\dim(S) = k$ if $\dim(\mathrm{aff}(S)) = k$.
- For any linear or affine subspace $V \subset \mathbb{R}^d$ the orthogonal projection onto V is denoted by π_V.

- For a linear subspace $V \subseteq \mathbb{R}^d$, we denote its orthogonal complement by $V^\perp = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{v}^\mathsf{T}\mathbf{x} = 0, \forall\, \mathbf{v} \in V\right\}$. For $\mathbf{v} \in \mathbb{R}^d$ we abbreviate $\mathbf{v}^\perp := \text{span}(\mathbf{v})^\perp$.
- We write $\text{vol}_k(S)$ for the k-dimensional volume of S. The 1-dimensional volume of a line segment l will also be written as $\text{length}(l)$.
- We say vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ in $\mathbb{R}^d$ are *affinely independent* if there is no $(k-2)$-dimensional affine subspace containing all of $\mathbf{a}_1, \ldots, \mathbf{a}_k$. Algebraically, $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are affinely independent if the system $\sum_{i \le k} \lambda_i \mathbf{a}_i = \mathbf{0}, \sum_{i \le k} \lambda_i = 0$ has no non-trivial solution.
- For $\mathbf{A} \in \mathbb{R}^{n \times d}$ a matrix and $B \subset [n]$ we write $\mathbf{A}_B \in \mathbb{R}^{|B| \times d}$ for the submatrix of $\mathbf{A}$ consisting of the rows indexed in B, and for $\mathbf{b} \in \mathbb{R}^n$ we write $\mathbf{b}_B$ for the restriction of $\mathbf{b}$ to the coordinates indexed in B.

2.2. Convexity. A polyhedron P is of the form $P = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \le \mathbf{b}\right\}$ for $\mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{b} \in \mathbb{R}^n$. A face $F \subseteq P$ is a convex subset such that if $\mathbf{x}, \mathbf{y} \in P$ and for $\lambda \in (0,1)$ $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in F$, then $\mathbf{x}, \mathbf{y} \in F$. In particular, a set F is a face of the polyhedron P if and only if there exists $I \subset [n]$ such that F coincides with P intersected with $\mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i, \forall i \in I$. A zero-dimensional face is called a vertex, one-dimensional face is called an edge, and a $\dim(P)-1$-dimensional face is called a facet. We use the notation $\text{vertices}(P)$ to denote the set of vertices of P and $\text{edges}(P)$ for the set of edges of P.

A set $S \subset \mathbb{R}^d$ is convex if for all $\mathbf{x}, \mathbf{y} \in S, \lambda \in [0,1]$ we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S$. We write $\text{conv}(S)$ to denote the convex hull of S, which is the intersection of all convex sets $T \supset S$. In a d-dimensional vector space, the convex hull equals

$$\text{conv}(S) = \left\{ \sum_{i=1}^{d+1} \lambda_i \mathbf{s}_i : \lambda_1, \ldots, \lambda_{d+1} \ge 0, \sum_{i=1}^{d+1} \lambda_i = 1, \mathbf{s}_1, \ldots, \mathbf{s}_{d+1} \in S \right\}.$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ the line segment between $\mathbf{x}$ and $\mathbf{y}$ is denoted $[\mathbf{x}, \mathbf{y}] = \text{conv}(\{\mathbf{x}, \mathbf{y}\})$.

We will need the following classical comparison inequality for surface areas of convex sets (see for example [15, Chapter 7]).

LEMMA 2.1 (Monotonicity of Surface Area). *If $K_1 \subseteq K_2 \subset \mathbb{R}^d$ are compact full-dimensional convex sets, then* $\text{vol}_{d-1}(\partial K_1) \le \text{vol}_{d-1}(\partial K_2)$.

2.3. Random Variables.

For a random variable $X \in \mathbb{R}$, we denote its expectation (mean) by $\mathbb{E}[X]$ and its variance by $\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$. For a random vector $\mathbf{X} \in \mathbb{R}^n$, we define its expectation (mean) $\mathbb{E}[\mathbf{X}] := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])$ and its variance (expected squared distance from the mean) $\text{Var}(\mathbf{X}) := \mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X}]\|^2]$.

For jointly distributed $X \in \Omega_1, Y \in \Omega_2$, we will often minimize the expectation of X over instantiations $y \in A \subset \Omega_2$. For this, we use the notation

$$\min_{Y \in A} \mathbb{E}[X \mid Y] := \min_{y \in A} \mathbb{E}[X \mid Y = y].$$

If μ is a probability density function, we write $x \sim \mu$ to denote that x is a random variable distributed with probability density μ.

For an event $E \subseteq \Omega$ in a measure space, we write $E^c := \Omega \setminus E$ to denote its complement.

2.3.1. Gaussian distribution.

DEFINITION 2.2. *The* Gaussian distribution *or* normal distribution $N_d(\bar{\mathbf{a}}, \sigma)$ *in d variables with mean $\bar{\mathbf{a}}$ and standard deviation σ has density* $(2\pi)^{-d/2} e^{-\|\mathbf{x} - \bar{\mathbf{a}}\|^2/(2\sigma^2)}$. *We abbreviate $N_d(\sigma) = N_d(\mathbf{0}, \sigma)$.*

Important facts about the Gaussian distribution include:

- Given a k-dimensional affine subspace $W \subseteq \mathbb{R}^d$, if $\mathbf{X}$ is $N_d(\bar{\mathbf{a}}, \sigma)$-distributed then both the orthogonal projection $\pi_W(\mathbf{X})$ and the restriction of $\mathbf{X}$ to W are $N_k(\pi_W(\bar{\mathbf{a}}), \sigma)$-distributed in W.
- For $\mathbf{X} \sim N_d(\bar{\mathbf{a}}, \sigma)$ we have $\mathbb{E}[\mathbf{X}] = \bar{\mathbf{a}}$ and $\mathbb{E}[((\mathbf{X} - \bar{\mathbf{a}})^\mathsf{T}\boldsymbol{\theta})^2] = \sigma^2$ for all $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$.
- The expected squared distance to the mean is $\mathbb{E}[\|\mathbf{X} - \bar{\mathbf{a}}\|^2] = d\sigma^2$.
- The moment generating function of $X \sim N_1(0, \sigma)$ is $\mathbb{E}[e^{\lambda X}] = e^{\lambda^2\sigma^2/2}$, for all $\lambda \in \mathbb{R}$, and that of X^2 is $\mathbb{E}[e^{\lambda X^2}] = 1/\sqrt{1 - 2\lambda\sigma}$ for $\lambda < 1/(2\sigma)$.

We will need the following tail bound for Gaussian random variables. We include a proof for completeness.

LEMMA 2.3 (Gaussian tail bounds). *For $\mathbf{X} \in \mathbb{R}^d$ distributed as $N_d(\mathbf{0}, \sigma)$, $t \geq 1$,*

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-(d/2)(t-1)^2} \ . \tag{2.1}$$

For $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ and $t \geq 0$,

$$\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t\sigma] \leq 2e^{-t^2/2} \ . \tag{2.2}$$

Proof. By homogeneity, we may without loss of generality assume that $\sigma = 1$.

Proof of (2.1).

$$\begin{aligned}
\Pr[\|\mathbf{X}\| \geq \sqrt{d}t] &= \min_{\lambda \in (0,1/2)} \Pr[e^{\lambda\|\mathbf{X}\|^2} \geq e^{\lambda t^2 d}] \\
&\leq \min_{\lambda \in (0,1/2)} \mathbb{E}[e^{\lambda\|\mathbf{X}\|^2}]e^{-\lambda t^2 d} \qquad \text{(Markov's inequality)} \\
&= \min_{\lambda \in (0,1/2)} \left(\prod_{i=1}^{d} \mathbb{E}[e^{\lambda X_i^2}]\right) e^{-\lambda t^2 d} \qquad \text{(Independence of coefficients)} \\
&= \min_{\lambda \in (0,1/2)} \left(\frac{1}{1-2\lambda}\right)^{d/2} e^{-\lambda t^2 d} \\
&\leq e^{-(d/2)(t^2 - 2\log t - 1)} \qquad \text{(setting } \lambda = \frac{1}{2}(1 - 1/t^2)\text{)} \\
&\leq e^{-(d/2)(t-1)^2} \qquad \text{(since } \log t \leq t - 1 \text{ for } t \geq 1\text{)}.
\end{aligned}$$

Proof of (2.2).

$$\begin{aligned}
\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] &= 2\Pr[\mathbf{X}^\mathsf{T}\boldsymbol{\theta} \geq t] \\
&\leq 2\min_{\lambda > 0} \mathbb{E}[e^{\lambda \mathbf{X}^\mathsf{T}\boldsymbol{\theta}}]e^{-\lambda t} \\
&= 2\min_{\lambda > 0} e^{\lambda^2/2 - \lambda t} \leq 2e^{-t^2/2} \text{ , setting } \lambda = t.
\end{aligned}$$

∎

2.3.2. Laplace distribution. Our shadow bounds will hold for a general class of distributions with bounds on certain parameters. We illustrate this for the d-dimensional Laplace distribution.

DEFINITION 2.4. *The* Laplace distribution $L_d(\bar{\mathbf{a}}, \sigma)$ *or* exponential distribution *in $\mathbb{R}^d$ with mean vector $\bar{\mathbf{a}}$ has probability density function*

$$\frac{\sqrt{d}^d}{(d-1)!\sigma^d \mathrm{vol}_{d-1}(\mathbb{S}^{d-1})} e^{-\|\mathbf{x} - \bar{\mathbf{a}}\|\sqrt{d}/\sigma}.$$

We abbreviate $L_d(\sigma) = L_d(\mathbf{0}, \sigma)$. We have normalized the distribution to have expected norm $\sqrt{d}\sigma$. Additionally, the variance along any direction is $\sigma^2(1 + \frac{1}{d})$.

The norm of a Laplace distributed random variable follows a *Gamma distribution*.

DEFINITION 2.5. *The Gamma distribution $\Gamma(\alpha, \beta), \alpha \in \mathbb{N}, \beta \in \mathbb{R}$, on the non-negative real numbers has probability density $\frac{\beta^\alpha}{(\alpha-1)!} t^{\alpha-1} e^{-\beta t}$. The moment generating function of the Gamma distribution is $\mathbb{E}_{X\sim\Gamma(\alpha,\beta)}[e^{\lambda X}] = (1 - \lambda/\beta)^{-\alpha}$ for $\lambda < \beta$.*

One can generate a d-dimensional Laplace distribution $L_d(\sigma)$ as the product of an independent scalar and vector. The vector $\boldsymbol{\theta}$ is sampled uniformly from the sphere $\mathbb{S}^{d-1}$. The scalar $s \sim \Gamma(d, \sqrt{d}/\sigma)$ is sampled from the Gamma distribution. The product $s\boldsymbol{\theta}$ has a $L_d(\sigma)$-distribution.

We will need the following tail bound for Laplace distributed random variables. We include a proof for completeness.

LEMMA 2.6 (Laplace tail bounds). *For $\mathbf{X} \in \mathbb{R}^d$, $d \geq 2$, distributed as $(\mathbf{0}, \sigma)$-Laplace and $t \geq 1$,*

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-d(t-\log t-1)} \ . \tag{2.3}$$

In particular, for $t \geq 2$,

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-dt/7} \ . \tag{2.4}$$

For $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $t \geq 0$,

$$\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t\sigma] \leq \begin{cases} 2e^{-t^2/16} : 0 \leq t \leq 2\sqrt{d} \\ e^{-\sqrt{d}t/7} : t \geq 2\sqrt{d} \end{cases} \ . \tag{2.5}$$

Proof. By homogeneity, we may without loss of generality assume that $\sigma = 1$.

Proof of (2.3).

$$\begin{aligned} \Pr[\|\mathbf{X}\| \geq \sqrt{d}t] &= \min_{\lambda\in(0,\sqrt{d})} \Pr[e^{\lambda\|\mathbf{X}\|} \geq e^{\lambda\sqrt{d}t}] \\ &\leq \min_{\lambda\in(0,\sqrt{d})} \mathbb{E}[e^{\lambda\|\mathbf{X}\|}]e^{-\lambda\sqrt{d}t} \qquad \text{(Markov's inequality)} \\ &\leq \min_{\lambda\in(0,\sqrt{d})} (1 - \lambda/\sqrt{d})^{-d} e^{-\lambda\sqrt{d}t} \\ &= e^{-d(t-\log t-1)} \text{ , setting } \lambda = \sqrt{d}(1 - 1/t). \end{aligned}$$

For the case $t \geq 2$, the desired inequality follows from the fact that $t - \log t - 1 \geq t/7$ for $t \geq 2$, noting that $(t - \log t - 1)/t$ is an increasing function on $t \geq 1$.

Proof of (2.5). For $t \geq 2\sqrt{d}$, we directly apply equation (2.4):

$$\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t\sigma] \leq \Pr[\|\mathbf{X}\| \geq t\sigma] \leq e^{-\sqrt{d}t/7}.$$

For $t \leq 2\sqrt{d}$, express $\mathbf{X} = s \cdot \boldsymbol{\omega}$ for $s \sim \Gamma(d, \sqrt{d}/\sigma)$, $\boldsymbol{\omega} \in \mathbb{S}^{d-1}$ uniformly sampled.

$$\begin{aligned} \Pr[|s\boldsymbol{\omega}^\mathsf{T}\boldsymbol{\theta}| \geq t\sigma] &\leq \Pr[|\boldsymbol{\omega}^\mathsf{T}\boldsymbol{\theta}| \geq t/(2\sqrt{d})] + \Pr[|s| \geq 2\sqrt{d}\sigma] \\ &\leq \Pr[|\boldsymbol{\omega}^\mathsf{T}\boldsymbol{\theta}| \geq t/(2\sqrt{d})] + e^{-d/4}. \end{aligned}$$

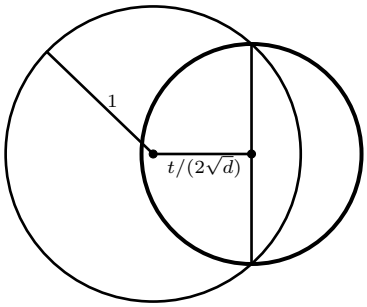

FIGURE 2.1. *The small sphere has at least as much surface area as combined surface area of the enclosed sphere cap and the opposite cap together by the monotonicity of surface area (Lemma 2.1).*

For the first term we follow [9, Lemma 2.2], where the second line is illustrated in Figure 2.1:

$$\begin{aligned}\Pr[|\boldsymbol{\omega}^\mathsf{T}\boldsymbol{\theta}| \geq t/(2\sqrt{d})] &= \frac{\mathrm{vol}_{d-1}(\left\{\boldsymbol{\omega} \in \mathbb{S}^{d-1} : |\boldsymbol{\omega}^\mathsf{T}\boldsymbol{\theta}| \geq t/(2\sqrt{d})\right\})}{\mathrm{vol}_{d-1}(\mathbb{S}^{d-1})} \\ &\leq \frac{\mathrm{vol}_{d-1}(\sqrt{1-\frac{t^2}{4d}}\mathbb{S}^{d-1})}{\mathrm{vol}_{d-1}(\mathbb{S}^{d-1})} \\ &= (1-\frac{t^2}{4d})^{(d-1)/2} \\ &\leq e^{-t^2(d-1)/(8d)} \leq e^{-t^2/16}.\end{aligned}$$

The desired conclusion follows since $e^{-t^2/16} + e^{-d/4} \leq 2e^{-t^2/16}$ for $0 \leq t \leq 2\sqrt{d}$. ∎

2.4. Change of variables. In section 3 we make use of a change of variables that was analyzed by Blaschke [13], and is standard in the study of convex hulls.

Recall that a change of variables affects a probability distribution. Let the vector $\mathbf{y} \in \mathbb{R}^d$ be a random variable with density μ. If $\mathbf{y} = \phi(\mathbf{x})$ and ϕ is invertible, then the induced density on $\mathbf{x}$ is

$$\mu(\phi(\mathbf{x}))\left|\det\left(\frac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}}\right)\right|,$$

where $\left|\det\left(\frac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}}\right)\right|$ is the Jacobian of ϕ. We describe a particular change of variables which has often been used for studying convex hulls, and, in particular, by Borgwardt [18] and Spielman and Teng [82] for deriving shadow bounds.

For affinely independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_d \in \mathbb{R}^d$ we have the coordinate transformation

$$(\mathbf{a}_1, \ldots, \mathbf{a}_d) \mapsto (\boldsymbol{\theta}, t, \mathbf{b}_1, \ldots, \mathbf{b}_d),$$

where $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ and $t \geq 0$ satisfy $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_i = t$ for every $i \in \{1, \ldots, d\}$ and the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ parametrize the positions of $\mathbf{a}_1, \ldots, \mathbf{a}_d$ within the hyperplane $\left\{\mathbf{x} \in \mathbb{R}^d \mid \boldsymbol{\theta}^\mathsf{T}\mathbf{x} = t\right\}$. We coordinatize the hyperplanes as follows:

Fix a reference vector $\mathbf{v} \in \mathbb{S}^{d-1}$, and pick an isometric embedding $h : \mathbb{R}^{d-1} \to \mathbf{v}^\perp$. For any unit vector $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, define the map $R'_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}^d$ as the unique map that

rotates $\mathbf{v}$ to $\boldsymbol{\theta}$ along $\operatorname{span}(\mathbf{v}, \boldsymbol{\theta})$ and fixes the orthogonal subspace $\operatorname{span}(\mathbf{v}, \boldsymbol{\theta})^\perp$. We define $R_{\boldsymbol{\theta}} = R'_{\boldsymbol{\theta}} \circ h$. The change of variables from $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, \mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ to $\mathbf{a}_1, \ldots, \mathbf{a}_d$ takes the form

$$(\mathbf{a}_1, \ldots, \mathbf{a}_d) = (R_{\boldsymbol{\theta}}\mathbf{b}_1 + t\boldsymbol{\theta}, \ldots, R_{\boldsymbol{\theta}}\mathbf{b}_d + t\boldsymbol{\theta}).$$

The change of variables as specified above is not uniquely defined when $\mathbf{a}_1, \ldots, \mathbf{a}_d$ are affinely dependent, when $t = 0$ or when $\boldsymbol{\theta} = -\mathbf{v}$.

THEOREM 2.7. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ be a unit vector, $t > 0$ and $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$. Consider the map*

$$(\boldsymbol{\theta}, t, \mathbf{b}_1, \ldots, \mathbf{b}_d) \mapsto (\mathbf{a}_1, \ldots, \mathbf{a}_d) = (R_{\boldsymbol{\theta}}\mathbf{b}_1 + t\boldsymbol{\theta}, \ldots, R_{\boldsymbol{\theta}}\mathbf{b}_d + t\boldsymbol{\theta}).$$

The Jacobian of this map equals

$$\left|\det\left(\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}}\right)\right| = (d-1)!\operatorname{vol}_{d-1}(\operatorname{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d)).$$

2.5. Shadow vertex algorithm. We briefly introduce the shadow vertex algorithm. For proofs of the statements below, see [52]. An alternative exposition about the shadow vertex algorithm can be found in [18].

Let $P = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ be a polyhedron, and let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ correspond to the rows of $\mathbf{A}$. We call a set $B \subseteq [n]$ a basis of $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ if $\mathbf{A}_B$ is invertible. This implies that $|B| = d$. We say B is a feasible basis if $\mathbf{x}_B = \mathbf{A}_B^{-1}\mathbf{b}_B$ satisfies $\mathbf{A}\mathbf{x}_B \leq \mathbf{b}$. The point $\mathbf{x}_B$ is a vertex of P. We say a feasible basis B is optimal for an objective $\mathbf{c} \in \mathbb{R}^d$ if $\mathbf{c}^\mathsf{T}\mathbf{A}_B^{-1} \geq \mathbf{0}$, which happens if and only if $\max_{\mathbf{x} \in P} \mathbf{c}^\mathsf{T}\mathbf{x} = \mathbf{c}^\mathsf{T}\mathbf{x}_B$.

Algorithm 2.1 Shadow vertex algorithm for non-degenerate polyhedron and shadow.

Require: $P = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$, $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$, feasible basis $B \subseteq [n]$ optimal for $\mathbf{d}$.
Ensure: Return optimal basis $B \subseteq [n]$ for $\mathbf{c}$ or *unbounded*.

```
λ_0 ← 0.
i ← 0.
loop
  i ← i + 1.
  λ_i := maximum λ ≤ 1 such that c_λ^T A_B^{-1} ≥ 0.
  if λ_i = 1 then
    return B.
  end if
  k := k ∈ B such that (c_{λ_i}^T A_B^{-1})_k = 0.
  x_B := A_B^{-1} b_B.
  s_i := supremum s > 0 such that A(x_B − s A_B^{-1} e_k) ≤ b.
  if s_i = ∞ then
    return unbounded.
  end if
  j := j ∈ [n] − B such that a_j^T (x_B − s_i A_B^{-1} e_k) = b_j.
  B ← B ∪ {j} \ {k}.
end loop
```

The shadow vertex algorithm is a pivot rule for the simplex method. Given a feasible basis $B \subseteq [n]$, an objective $\mathbf{d} \in \mathbb{R}^d$ for which B is optimal, and an objective

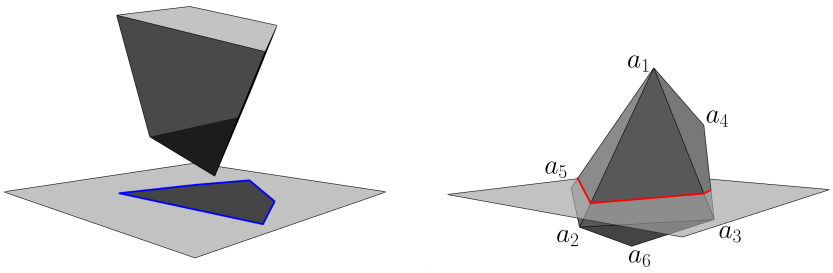

FIGURE 2.2. *On the left, a polytope and its shadow. On the right, the corresponding polar polytope intersected with the plane. There are as many edges marked blue as there are edges marked red.*

function $\mathbf{c} \in \mathbb{R}^d$ to optimize, where $\mathbf{c}$ and $\mathbf{d}$ are linearly independent, the shadow vertex algorithm (Algorithm 2.1) specifies which pivot steps to take to reach an optimal basis for $\mathbf{c}$. We note that there are many possible choices for starting objective $\mathbf{d}$.

We parametrize $\mathbf{c}_\lambda := (1-\lambda)\mathbf{d} + \lambda\mathbf{c}$ and start at $\lambda = 0$. The shadow vertex rule increases λ until there are $j \neq k \in [n]$ such that a new feasible basis $B \cup \{j\} \setminus \{k\}$ is optimal for $\mathbf{c}_\lambda$, and repeat with increased λ and new basis B until $\lambda = 1$.

The index $k \in B$ is such that the coordinate for k in $\mathbf{c}_\lambda^\intercal \mathbf{A}_B^{-1}$ first lowers to 0, and $j \notin B$ is such that $B \cup \{j\} \setminus \{k\}$ is a feasible basis: we follow the edge $\mathbf{A}_B^{-1}\mathbf{b}_B - \mathbf{A}_B^{-1}\mathbf{e}_k\mathbb{R}_+$ until we hit the first constraint $\mathbf{a}_j^\intercal \mathbf{x} \leq b_j$, and then replace k by j to get the new basis $B \cup \{j\} \setminus \{k\}$.

Changing the current basis from B to $B \cup \{j\} \setminus \{k\}$ is called a pivot step. As soon as $\lambda = 1$ we have $\mathbf{c}_\lambda = \mathbf{c}$, at which moment the current basis is optimal for our objective $\mathbf{c}$. If at some point no choice of j exists, then an unbounded ray has been found.

DEFINITION 2.8. *We say that the system* $\mathbf{Ax} \leq \mathbf{b}$ *is* non-degenerate *if* $n \geq d$, *any* $B \in \binom{[n]}{d}$ *is a basis, and every vertex of the corresponding polyhedron* P *is tight at exactly* d *linearly independent inequalities. When the description* $\mathbf{Ax} \leq \mathbf{b}$ *is clear, we say that* P *is non-degenerate to mean that its describing system is.*

DEFINITION 2.9. *We say that the shadow of* P *on a two-dimensional linear subspace* W *is* non-degenerate *if* $\dim(\pi_W(P)) = 2$ *and for every face* F *of* P *such that* $\pi_W(F)$ *is a face of* $\pi_W(P)$ *and* $\dim(\pi_W(F)) \leq 1$, *we have that* $\dim(\pi_W(F)) = \dim(F)$.

If both the polyhedron and the shadow are non-degenerate, each pivot step can be performed in $O(nd)$ time (see the pseudo-code for Algorithm 2.1). Under the distribution models we examine, degeneracy occurs with probability 0.

The shadow vertex rule is called as such because the visited vertices are in correspondence with vertices on the relative boundary of the orthogonal projection $\pi_W(P)$ of P onto $W = \text{span}(\mathbf{d}, \mathbf{c})$, where we denote $\pi_W(P)$ as the shadow of P on W. See Figure 2.2. We call the total number of vertices of the projection the *shadow size*, and it is the key geometric estimate in our analysis of the simplex method.

LEMMA 2.10. *For a polyhedron* $P = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{Ax} \leq \mathbf{b}\}$ *having a non-degenerate shadow on* W, *the vertices of* P *optimizing objectives in* $W \setminus \{\mathbf{0}\}$ *are in one-to-one correspondence with the vertices of* $\pi_W(P)$ *under the map* π_W.

We will consider non-degenerate polyhedra of the form $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{Ax} \leq \mathbf{1}\}$, in which case $\mathbf{0}$ is always contained in the polyhedron. The problem thus has a known

feasible solution. It is instructive to look at the geometry of shadow paths on such polyhedra from a *polar perspective*. For any non-degenerate polyhedron $P = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{1}\right\}$, we look at the polar polytope, defined as the convex hull $Q := \text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ of the constraint vectors. For any index-set $I \subseteq [n], |I| = d$, if the (unique) solution $\mathbf{x}_I$ to the equations

$$\mathbf{a}_i{}^{\mathsf{T}}\mathbf{x} = 1 \qquad \forall i \in I$$

is a vertex of the original polyhedron P, then the set $\text{conv}(\mathbf{a}_i : i \in I)$ forms a facet of the polytope $\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Conversely, if $\text{conv}(\mathbf{a}_i : i \in I)$ induces a facet of $Q := \text{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ (note the inclusion of $\mathbf{0}$), then $\mathbf{x}_I$ is a vertex of P. The addition of $\mathbf{0}$ to the polar of P allows us to detect unboundedness. Precisely, the facets of the extended polar $\text{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ containing $\mathbf{0}$ are in one to one correspondence with unbounded edges of P. P is bounded, i.e. a polytope, if and only if $\mathbf{0}$ is in the interior of Q. In this case $Q = Q'$, and hence every facet of Q is associated to a vertex of P.

In the polar perspective, a pivot step moves from one facet of Q' to a neighboring facet. The shadow vertex algorithm moves the objective $\mathbf{c}_\lambda$ along the line segment $[\mathbf{d}, \mathbf{c}]$ and keeps track of which facet of Q' is intersected by the ray $\mathbf{c}_\lambda\mathbb{R}^+$. If we move to a facet of Q' containing $\mathbf{0}$, we may conclude that the LP with objective $\mathbf{c}$ is in fact unbounded. Since we can only visit such facets at the end of a shadow path, we will be able to control the length of shadow paths using only the geometry of Q, which will help simplify our analyses. The main bound on the size of the shadow we will use is given in the following lemma.

LEMMA 2.11. *Let $P = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{1}\right\}$ be a non-degenerate polyhedron with a non-degenerate shadow on W. Then*

$$|\text{vertices}(\pi_W(P))| \leq |\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|.$$

The number of pivot steps taken in a shadow path is bounded from above by the number of edges in the intersection $\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap \text{span}(\mathbf{d}, \mathbf{c})$. Hence it suffices that we prove an upper bound on this geometric quantity. The following theorem summarizes the properties we will use of the shadow vertex algorithm.

THEOREM 2.12. *Let $P = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{b}\right\}$ denote a non-degenerate polyhedron, and let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ be the rows of $\mathbf{A}$. Let $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$ denote two objectives inducing a non-degenerate shadow on P, and let $W = \text{span}(\mathbf{d}, \mathbf{c})$. Given a feasible basis $I \in \binom{[n]}{d}$ for $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ which is optimal for $\mathbf{d}$, Algorithm 2.1 (shadow vertex) finds a feasible basis $J \in \binom{[n]}{d}$ optimal for $\mathbf{c}$ or declares unboundedness in a number of pivot steps bounded by $|\text{vertices}(\pi_W(P))|$, where π_W is the orthogonal projection onto W. In particular, when $\mathbf{b} = \mathbf{1}$, the number of pivots is at most*

$$|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)| \ .$$

3. Shadow Bounds. In this section, we derive our new and improved shadow bounds for Laplace and Gaussian distributed perturbations. We achieve these results by first proving a shadow bound for parametrized distributions as described in the next section, and then specializing to the case of Laplace and Gaussian perturbations. The bounds we obtain are described below.

THEOREM 3.1. *Let $W \subset \mathbb{R}^d$ be a fixed two-dimensional subspace, $n \geq d \geq 3$ and let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, be independent Gaussian random vectors with variance σ^2 and*

centers of norm at most 1. *Then the expected number of edges is bounded by*

$$\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] \leq \mathcal{D}_g(n, d, \sigma),$$

where the function $\mathcal{D}_g(d, n, \sigma)$ *is defined as*

$$\mathcal{D}_g(d, n, \sigma) := O(d^2 \sqrt{\log n}\ \sigma^{-2} + d^{2.5} \log n\ \sigma^{-1} + d^{2.5} \log^{1.5} n).$$

Our bound applies more generally for distributions satisfying certain parameters. We illustrate this with a shadow bound for perturbations distributed according to the Laplace distribution. This will serve as a good warm-up exercise for the slightly more involved analysis of the Gaussian distribution.

THEOREM 3.2. *Let* $W \subset \mathbb{R}^d$ *be a fixed two-dimensional subspace,* $n \geq d \geq 3$ *and let* $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, *be independent Laplace distributed random vectors with parameter* σ *and centers of norm at most* 1. *Then the expected number of edges is bounded by*

$$\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] = O(d^{2.5}\sigma^{-2} + d^3 \log n\ \sigma^{-1} + d^3 \log^2 n).$$

The proofs of Theorems 3.1 and 3.2 are given in respectively subsections 3.3 and 3.2.

3.1. Parametrized Shadow Bound. In this section, we prove a shadow bound theorem for any noise distribution that has non-trivial bounds on certain parameters. The parameters we will use are defined below.

3.1.1. Distribution parameters.

DEFINITION 3.3. *A distribution with density* μ *on* $\mathbb{R}^d$ *is* L-log-Lipschitz *if for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ *we have* $|\log(\mu(\mathbf{x})) - \log(\mu(\mathbf{y}))| \leq L\|\mathbf{x} - \mathbf{y}\|$. *Equivalently,* μ *is* L*-log-Lipschitz if* $\mu(\mathbf{x})/\mu(\mathbf{y}) \leq \exp(L\|\mathbf{x} - \mathbf{y}\|)$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

DEFINITION 3.4. *Given a probability distribution with density* μ *on* $\mathbb{R}^d$, *we define the line variance* τ^2 *as the infimum of the variances when restricted to any fixed line* $l \subset \mathbb{R}^d$*:*

$$\tau^2 = \inf_{\text{line } l \subset \mathbb{R}^d} \text{Var}(\mathbf{X} \sim \mu \mid \mathbf{X} \in l).$$

Both the log-Lipschitz constant and the minimal line variance relate to how "spread out" the probability mass is. The log-Lipschitzness of a random variable gives a lower bound on the line variance, which we prove in Lemma 3.8.

DEFINITION 3.5. *Given a distribution with probability density* μ *on* $\mathbb{R}^d$ *with expectation* $\mathbb{E}_{\mathbf{X}\sim\mu}[\mathbf{X}] = \mathbf{y}$ *we define the* n*-th deviation* r_n *to be the smallest number such that for any unit vector* $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$\int_{r_n}^{\infty} \Pr_{\mathbf{X}\sim\mu}[|(\mathbf{X} - \mathbf{y})^\mathsf{T}\boldsymbol{\theta}| \geq t]dt \leq r_n/n.$$

Note that as r_n *increases to* ∞, *the left-hand side goes to* 0 *and the right-hand side goes to* ∞. *We see that there must exist a number satisfying this inequality, so* r_n *is well-defined.*

The n-th deviation will allow us to give bounds on the expected maximum size $\mathbb{E}[\max_{i\leq n}|\mathbf{x}_i{}^\mathsf{T}\boldsymbol{\theta}|]$ of n separate perturbations in a given direction $\boldsymbol{\theta}$. We formalize this in Lemma 3.7.

DEFINITION 3.6. *Given a distribution with probability density μ on $\mathbb{R}^d$ with expectation $\mathbb{E}_{\mathbf{x}\sim\mu}[\mathbf{x}] = \mathbf{y}$, we define, for all $1 > p > 0$, the cutoff radius $R(p)$ as the smallest number satisfying*

$$\Pr_{\mathbf{x}\sim\mu}[\|\mathbf{x} - \mathbf{y}\| \geq R(p)] \leq p.$$

The cutoff radius of interest is $R_{n,d} := R(\frac{1}{d\binom{n}{d}})$. The cutoff radius tells us how concentrated the probability mass of the random variable is, while the log-Lipschitzness tells us how spread out the probability mass is. These quantities cannot both be arbitrarily good (small) at the same time. We formalize this notion in Lemma 3.9.

LEMMA 3.7. *If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are each distributed with mean $\mathbf{0}$ and n-th deviation at most r_n, then for any $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$,*

$$\mathbb{E}[\max_{i\in[n]} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i|] \leq 2r_n.$$

Proof. We rewrite the expectation as

$$\mathbb{E}[\max_{i\in[n]} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i|] = \int_0^\infty \Pr[\max_{i\in[n]} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i| \geq t] \, \mathrm{d}t.$$

We separately bound the integral up to r_n and from r_n to ∞. Since a probability is at most 1 we have

$$\int_0^{r_n} \Pr[\max_{i\in[n]} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i| \geq t] \, \mathrm{d}t \leq r_n,$$

and by definition of the n-th deviation and the union bound:

$$\begin{aligned}\int_{r_n}^\infty \Pr[\max_{i\in[n]} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i| \geq t] \, \mathrm{d}t &\leq \sum_{i\in[n]} \int_{r_n}^\infty \Pr[|\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i| \geq t] \\ &\leq r_n.\end{aligned}$$

Together these estimates yield the desired inequality,

$$\mathbb{E}[\max_{i\leq n} |\boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i|] \leq 2r_n.$$

∎

LEMMA 3.8. *If a distribution with probability density μ is L-log-Lipschitz, then its line variance satisfies $\tau \geq 1/(\sqrt{e}L)$.*

Proof. Let $\mathbf{v} + \mathbf{w}\mathbb{R}$ be a line and assume that $\mathbb{E}[\mathbf{x} \mid \mathbf{x} \in \mathbf{v} + \mathbf{w}\mathbb{R}] = \mathbf{v}$ and $\|\mathbf{w}\| = 1$. We show that with probability at least $1/e$, $\mathbf{x}$ has distance at least $1/L$ from $\mathbf{v}$. Conditioning on $\mathbf{x} \in \mathbf{v} + \mathbf{w}\mathbb{R}$, the induced probability mass is proportional to $\mu(\mathbf{x})$. We can bound the fraction of the induced probability mass that is far away from the expectation by the following calculation:

$$\begin{aligned}\int_{-\infty}^\infty \mu(\mathbf{v} + \gamma\mathbf{w}) \, \mathrm{d}\gamma &= \int_{-\infty}^0 \mu(\mathbf{v} + \gamma\mathbf{w}) \, \mathrm{d}\gamma + \int_0^\infty \mu(\mathbf{v} + \gamma\mathbf{w}) \, \mathrm{d}\gamma \\ &= \int_{-\infty}^{-1/L} \mu(\mathbf{v} + (\gamma + 1/L)\mathbf{w}) \, \mathrm{d}\gamma + \int_{1/L}^\infty \mu(\mathbf{v} + (\gamma - 1/L)\mathbf{w}) \, \mathrm{d}\gamma \\ &\leq e\int_{-\infty}^{-1/L} \mu(\mathbf{v} + \gamma\mathbf{w}) \, \mathrm{d}\gamma + e\int_{1/L}^\infty \mu(\mathbf{v} + \gamma\mathbf{w}) \, \mathrm{d}\gamma.\end{aligned}$$

The integral on the first line exists because it is the integral of a continuous non-negative function, and, if the integral were infinite, then the integral along every parallel line would be infinite by log-Lipschitzness, contradicting the fact that μ has integral 1 over $\mathbb{R}^d$.

Hence, $\Pr[\|\mathbf{x}-\mathbf{v}\| \geq 1/L \mid \mathbf{x} \in \mathbf{v}+\mathbf{w}\mathbb{R}] = \frac{\int_{-\infty}^{-1/L} \mu(\mathbf{v}+\gamma\mathbf{w})\ \mathrm{d}\gamma + \int_{1/L}^{\infty} \mu(\mathbf{v}+\gamma\mathbf{w})\ \mathrm{d}\gamma}{\int_{-\infty}^{\infty} \mu(\mathbf{v}+\gamma\mathbf{w})\ \mathrm{d}\gamma} \geq 1/e$, and we can lower bound the variance

$$\mathrm{Var}(\mathbf{x} \mid \mathbf{x} \in \mathbf{v} + \mathbf{w}\mathbb{R}) \geq \frac{1}{e}(1/L)^2.$$

Since the line $\mathbf{v} + \mathbf{w}\mathbb{R}$ was arbitrary, it follows that $\tau \geq 1/(\sqrt{e}L)$. ∎

LEMMA 3.9. *For a d-dimensional distribution with probability density μ, where $d \geq 3$, with parameters L, R as described above, we have the inequality $LR(1/2) \geq d/3$.*

Proof. Let $\bar{R} := R(1/2)$. If $L\bar{R} \geq d$, we are already done, so we may assume that $L\bar{R} < d$. Also, without loss of generality, we may assume that μ has mean $\mathbf{0}$. For $\alpha > 1$ to be chosen later we know

$$\begin{aligned}
1 &\geq \int_{\alpha\bar{R}\mathcal{B}_2^d} \mu(\mathbf{x})\ \mathrm{d}\mathbf{x} \\
&= \alpha^d \int_{\bar{R}\mathcal{B}_2^d} \mu(\alpha\mathbf{x})\ \mathrm{d}\mathbf{x} \\
&\geq \alpha^d e^{-(\alpha-1)L\bar{R}} \int_{\bar{R}\mathcal{B}_2^d} \mu(\mathbf{x})\ \mathrm{d}\mathbf{x} \\
&= \frac{\alpha^d}{2} e^{-(\alpha-1)L\bar{R}}.
\end{aligned}$$

Taking logarithms, we find

$$0 \geq d\log(\alpha) - (\alpha-1)L\bar{R} - \log(2).$$

We choose $\alpha = \frac{d}{L\bar{R}} > 1$ and look at the resulting inequality:

$$0 \geq d\log(\frac{d}{L\bar{R}}) - d + L\bar{R} - \log(2).$$

For $d \geq 3$, this inequality can only hold if $L\bar{R} \geq d/3$, as needed. ∎

3.1.2. Proving a shadow bound for parametrized distributions. The main result of this subsection is the following parametrized shadow bound.

THEOREM 3.10 (Parametrized Shadow Bound). *Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, where $n \geq d \geq 3$, be independently distributed according to L-log-Lipschitz distributions with centers of norm at most 1, line variances at least τ^2, cutoff radii at most $R_{n,d}$ and n-th deviations at most r_n. For any fixed two-dimensional linear subspace $W \subset \mathbb{R}^d$, the expected number of edges satisfies*

$$\mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] \leq O(\frac{d^{1.5}L}{\tau}(1 + R_{n,d})(1 + r_n)).$$

The proof is given at the end of the subsection. It will be derived from the sequence of lemmas given below. We refer the reader to subsection 1.3 for a high-level overview of the proof.

In the rest of the subsection, $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, where $n \geq d \geq 3$, will be as in Theorem 3.10. We use $Q := \operatorname{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ to denote the convex hull of the constraint vectors and W to denote the two-dimensional shadow plane.

The following non-degeneracy conditions on $\mathbf{a}_1, \ldots, \mathbf{a}_n$ will hold with probability 1, because $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are independently distributed with continuous distributions.

1. Every $d+1$ vectors from $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are affinely independent. Thus, every facet of Q is the convex hull of exactly d vectors from $\mathbf{a}_1, \ldots, \mathbf{a}_n$.
2. Any d distinct vectors $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_d}$, $i_1, \ldots, i_d \in [n]$, have a unique hyperplane through them. This hyperplane intersects W in a one-dimensional line, does not contain the origin $\mathbf{0}$, and its unit normal vector pointing away from the origin is not $-\mathbf{e}_1$. Note that the last two conditions imply that Blaschke's transformation on $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_d}$ is uniquely defined with $\mathbf{e}_1$ as reference vector.
3. For every edge $e \subset Q \cap W$ there is a unique facet F of Q such that $e = F \cap W$.

In what follows we will always assume the above conditions hold.

For our first lemma, in which we bound the number of edges in terms of two different expected lengths, we make a distinction between possible edges with high probability of appearing versus edges with low probability of appearing. The sets with probability at most $2\binom{n}{d}^{-1}$ to form an edge, together contribute at most 2 to the expected number of edges, as there are only $\binom{n}{d}$ bases.

For a basis with probability at least $2\binom{n}{d}^{-1}$ of forming an edge, we can safely condition on it forming an edge without forcing very unlikely events to happen. Because of this, we will later be able to condition on the vertices not being too far apart.

DEFINITION 3.11. *For $I \in \binom{[n]}{d}$, let E_I denote the event that $\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W$ forms an edge of $Q \cap W$.*

DEFINITION 3.12. *We define the set $B \subseteq \binom{[n]}{d}$ to be the set of those $I \subseteq [n]$ satisfying $|I| = d$ and $\Pr[E_I] \geq 2\binom{n}{d}^{-1}$.*

The next lemma is inspired by Theorem 3.2 of [58].

LEMMA 3.13. *The expected number of edges in $Q \cap W$ satisfies*

$$\mathbb{E}[|\text{edges}(Q \cap W)|] \leq 2 + \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{\min_{I \in B} \mathbb{E}[\text{length}(\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W) \mid E_I]}.$$

Proof. We give a lower bound on the perimeter of the intersection $Q \cap W$ in terms of the number of edges. By our non-degeneracy assumption, every edge can be uniquely represented as $\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W$, for $I \in \binom{[n]}{d}$. From this we derive the first equality, and we continue from that:

$$\begin{aligned}
\mathbb{E}[\text{perimeter}(Q \cap W)] &= \sum_{I \in \binom{[n]}{d}} \mathbb{E}[\text{length}(\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W) \mid E_I] \Pr[E_I] \\
&\geq \sum_{I \in B} \mathbb{E}[\text{length}(\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W) \mid E_I] \Pr[E_I] \\
&\geq \min_{I \in B} \mathbb{E}[\text{length}(\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W) \mid E_I] \sum_{J \in B} \Pr[E_J].
\end{aligned}$$

The first line holds because whenever E_I holds, $\operatorname{conv}(\mathbf{a}_i : i \in I) \cap W$ is an edge of $Q \cap W$, and every edge of $Q \cap W$ is formed by exactly one face F_J, by the non-degeneracy conditions we have assumed. By construction of B and linearity of expectation,

$\sum_{J\in B}\Pr[E_J] \geq \sum_{J\in\binom{[n]}{d}}\Pr[E_J] - 2 = \mathbb{E}[|\text{edges}(Q\cap W)|] - 2$. By dividing on both sides of the inequality, we can now conclude

$$\mathbb{E}[|\text{edges}(Q\cap W)|] \leq 2 + \frac{\mathbb{E}[\text{perimeter}(Q\cap W)]}{\min_{I\in B}\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i : i\in I)\cap W) \mid E_I]}.$$ ∎

Given the above, we may now restrict our task to proving an upper bound on the expected perimeter and a lower bound on the minimum expected edge length, which will be the focus on the remainder of the subsection.

The perimeter is bounded using a standard convexity argument. A convex shape has perimeter no more than that of any circle containing it. We exploit the fact that all centers have norm at most 1 and the expected perturbation sizes are not too big along any fixed axis.

LEMMA 3.14. *The expected perimeter of $Q\cap W$ is bounded by*

$$\mathbb{E}[\text{perimeter}(Q\cap W)] \leq 2\pi(1+4r_n),$$

where r_n is the n-deviation bound for $\mathbf{a}_1,\ldots,\mathbf{a}_n$.

Proof. By convexity, the perimeter is bounded from above by 2π times the norm of the maximum norm point. Let $\hat{\mathbf{a}}_i := \mathbf{a}_i - \mathbb{E}[\mathbf{a}_i]$ denote the perturbation of $\mathbf{a}_i$ from the center of its distribution, recalling that $\|\mathbb{E}[\mathbf{a}_i]\| \leq 1$ by assumption. We can now derive the bound

$$\begin{aligned}
\mathbb{E}[\text{perimeter}(Q\cap W)] &\leq 2\pi\mathbb{E}[\max_{\mathbf{x}\in Q\cap W}\|\mathbf{x}\|] \\
&= 2\pi\mathbb{E}[\max_{\mathbf{x}\in Q\cap W}\|\pi_W(\mathbf{x})\|] \\
&\leq 2\pi\mathbb{E}[\max_{\mathbf{x}\in Q}\|\pi_W(\mathbf{x})\|] \\
&= 2\pi\mathbb{E}[\max_{i\in[n]}\|\pi_W(\mathbf{a}_i)\|] \\
&\leq 2\pi\left(1+\mathbb{E}[\max_{i\leq n}\|\pi_W(\hat{\mathbf{a}}_i)\|]\right),
\end{aligned}$$

where the last inequality follows since $\mathbf{a}_1,\ldots,\mathbf{a}_n$ have centers of norm at most 1. Pick an orthogonal basis $\mathbf{v}_1,\mathbf{v}_2$ of W. By the triangle inequality the expected perturbation size satisfies

$$\mathbb{E}[\max_{i\leq n}\|\pi_W(\hat{\mathbf{a}}_i)\|] \leq \sum_{j\in\{1,2\}}\mathbb{E}[\max_{i\leq n}|\mathbf{v}_j{}^{\mathsf{T}}\hat{\mathbf{a}}_i|].$$

Each of the two expectations satisfies, by Lemma 3.7, $\mathbb{E}[\max_{i\leq n}|\mathbf{v}_j{}^{\mathsf{T}}\hat{\mathbf{a}}_i|] \leq 2r_n$, thereby concluding the proof. ∎

The rest of this subsection will be devoted to finding a suitable lower bound on the denominator $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i : i\in I)\cap W) \mid E_I]$ uniformly over all choices of $I\in B$. Without loss of generality we assume that $I=[d]$ and write $E := E_{[d]}$.

DEFINITION 3.15 (Containing hyperplane). *Define $H = \text{aff}(\mathbf{a}_1,\ldots,\mathbf{a}_d) = t\boldsymbol{\theta}+\boldsymbol{\theta}^{\perp}$, where $\boldsymbol{\theta}\in\mathbb{S}^{d-1}$, $t>0$ to be the hyperplane containing $\mathbf{a}_1,\ldots,\mathbf{a}_d$. Define $l = H\cap W$. From our non-degeneracy conditions we know that l is a line. Express $l = \mathbf{p}+\boldsymbol{\omega}\cdot\mathbb{R}$, where $\boldsymbol{\omega}\in\mathbb{S}^{d-1}$ and $\mathbf{p}\in\boldsymbol{\omega}^{\perp}$.*

To lower bound the length $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)\cap W)\mid E]$ we will need the pairwise distances between the different $\mathbf{a}_i$'s for $i\in\{1,\dots,d\}$ to be small along $\boldsymbol{\omega}^\perp$. This will allow us to get "wiggle room" around each vertex of $\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)$ that is proportional to the size of the facet.

Definition 3.16 (Bounded diameter event). *We define the event D to hold exactly when $\|\pi_{\boldsymbol{\omega}^\perp}(\mathbf{a}_i)-\pi_{\boldsymbol{\omega}^\perp}(\mathbf{a}_j)\|\le 2+2R_{n,d}$ for all $i,j\in[d]$.*

We will condition on the event D. This will not change the expected length by much, because the probability that D does not occur is small compared to the probability of E by our assumption that $\Pr[E]\ge\frac{2}{d\binom{n}{d}}$.

Lemma 3.17. *The expected edge length satisfies*

$$\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)\cap W)\mid E]\ge\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)\cap W)\mid D,E]/2.$$

Proof. Let the vector $\hat{\mathbf{a}}_i$ denote the perturbation $\mathbf{a}_i-\mathbb{E}[\mathbf{a}_i]$. Since distances can only decrease when projecting, the complementary event D^c satisfies

$$\begin{aligned}\Pr[D^c]&=\Pr[\max_{i,j\le d}\|\pi_{\boldsymbol{\omega}^\perp}(\mathbf{a}_i-\mathbf{a}_j)\|\ge 2+2R_{n,d}]\\&\le\Pr[\max_{i,j\le d}\|\mathbf{a}_i-\mathbf{a}_j\|\ge 2+2R_{n,d}],\end{aligned}$$

by the triangle inequality and the bound of 1 on the norms of the centers, the line above is at most

$$\begin{aligned}&\le\Pr[\max_{i\le d}\|\mathbf{a}_i\|\ge 1+R_{n,d}]\\&\le\Pr[\max_{i\le d}\|\hat{\mathbf{a}}_i\|\ge R_{n,d}]\\&\le\binom{n}{d}^{-1}.\end{aligned}$$

By our assumption that $[d]\in B$, we know that $\Pr[E]\ge 2\binom{n}{d}^{-1}$. In particular, it follows that $\Pr[E\cap D]\ge\Pr[E]-\Pr[D^c]\ge\Pr[E]/2$. Thus, we may conclude that

$$\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)\cap W)\mid E]\ge\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1,\dots,\mathbf{a}_d)\cap W)\mid D,E]/2.$$

∎

For the rest of this section, we use a change of variables on $\mathbf{a}_1,\dots,\mathbf{a}_d$. The nondegeneracy conditions we have assumed at the start of this section make the change of variables well-defined.

Definition 3.18 (Change of variables). *Recall the change of variables mapping $(\mathbf{a}_1,\dots,\mathbf{a}_d)\mapsto(\boldsymbol{\theta},t,\mathbf{b}_1,\dots,\mathbf{b}_d)$ for $\boldsymbol{\theta}\in\mathbb{S}^{d-1},t>0,\mathbf{b}_1,\dots,\mathbf{b}_d\in\mathbb{R}^{d-1}$ from Theorem 2.7. We abbreviate $\bar{\mu}_i(\boldsymbol{\theta},t,\mathbf{b}_i)=\mu_i(R_\theta(\mathbf{b}_i)+t\boldsymbol{\theta})$ and we write $\bar{\mu}_i(\mathbf{b}_i)$ when the values of $\boldsymbol{\theta},t$ are clear.*

By Theorem 2.7 of Blaschke [13] we know that for any fixed values of $\boldsymbol{\theta},t$ the vectors $\mathbf{b}_1,\dots,\mathbf{b}_d$ have joint probability density proportional to

$$\text{vol}_{d-1}(\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d))\prod_{i=1}^d\bar{\mu}_i(\mathbf{b}_i)\ . \tag{3.1}$$

We assumed that $\mathbf{a}_1,\dots,\mathbf{a}_d$ are affinely independent, so $\mathbf{b}_1,\dots,\mathbf{b}_d$ are affinely independent as well.

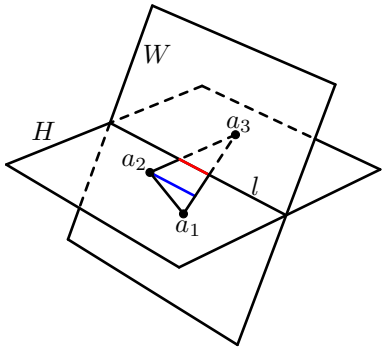

FIGURE 3.1. $\mathbf{a}_1, \ldots, \mathbf{a}_d$ *are conditioned for* $\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ *to intersect* W *and lie in* H. *The red line corresponds to induced edge. The blue line represents the longest chord parallel to* ℓ.

In the next lemma, we condition on the hyperplane $H = t\boldsymbol{\theta} + \boldsymbol{\theta}^\perp$ and from then on we restrict our attention to what happens inside H. Conditioned on $\mathbf{a}_1, \ldots, \mathbf{a}_d$ lying in H, the set $\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ is a facet of Q if and only if all of $\mathbf{a}_{d+1}, \ldots, \mathbf{a}_n$ lie on one side of H. This means that the shape of $\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ in H does not influence the event that it forms a facet, so in studying this convex hull we can then ignore $\mathbf{a}_{d+1}, \ldots, \mathbf{a}_n$.

We identify the hyperplane H with $\mathbb{R}^{d-1}$ and define $\bar{l} = \bar{\mathbf{p}} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R} \subset \mathbb{R}^{d-1}$ corresponding to $l = \mathbf{p} + \boldsymbol{\omega} \cdot \mathbb{R}$ by $\bar{\mathbf{p}} = R_{\boldsymbol{\theta}}^{-1}(\mathbf{p} - t\boldsymbol{\theta})$, $\bar{\boldsymbol{\omega}} = R_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\omega})$. We define $\bar{E}$ as the event that $\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l} \neq \emptyset$. Notice that E holds if and only if $\bar{E}$ and $\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ is a facet of Q. See Figure 3.1.

We will condition on the shape of the projected simplex.

DEFINITION 3.19 (Projected shape). *We define the projected shift variable by* $\mathbf{x} := \mathbf{x}_{\boldsymbol{\omega}}(\mathbf{b}_1) = \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_1)$ *and shape variable* $S := S_\omega(\mathbf{b}_1, \ldots, \mathbf{b}_d)$ *by*

$$S_{\boldsymbol{\omega}}(\mathbf{b}_1, \ldots, \mathbf{b}_d) = (\mathbf{0}, \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_2) - \mathbf{x}, \ldots, \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_d) - \mathbf{x}) \ .$$

We index $S = (\mathbf{s}_1, \ldots, \mathbf{s}_d)$, *so* $\mathbf{s}_i \in \bar{\boldsymbol{\omega}}^\perp$ *is the* i*-th vector in* S, *and furthermore define the diameter function* $\mathrm{diam}(S) = \max_{i,j \in [d]} \|\mathbf{s}_i - \mathbf{s}_j\|$. *We will condition on the shape being in the set of allowed shapes*

$$\mathcal{S} := \left\{(\mathbf{s}_1, \ldots, \mathbf{s}_d) \in (\bar{\boldsymbol{\omega}}^\perp)^d : \mathbf{s}_1 = 0, \mathrm{diam}(S) \leq 2 + 2R_{n,d}, \mathrm{rank}(\mathbf{s}_2, \ldots, \mathbf{s}_d) = d - 2\right\}.$$

Observe that $S \in \mathcal{S}$ if and only if the event D holds. To justify the rank condition on $\mathbf{s}_2, \ldots, \mathbf{s}_d$, note that by our non-degeneracy conditions, we have that $\mathbf{b}_1, \ldots, \mathbf{b}_d$ are affinely independent. In particular, they do not all lie in a $d-2$-dimensional affine subspace. This means that $\mathbf{s}_1, \ldots, \mathbf{s}_d$ do not all lie in a $d-3$-dimensional affine subspace, from which it follows that $\mathrm{rank}(\mathbf{s}_2, \ldots, \mathbf{s}_d) = d - 2$ (recalling that $\mathbf{s}_1 = \mathbf{0}$).

LEMMA 3.20. *Let* $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, \mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ *denote the change of variables of* $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ *as in Definition 3.18. Then, the expected length satisfies*

$$\mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d) \cap W) \,|\, D, E] \geq \inf_{\boldsymbol{\theta}, t, S \in \mathcal{S}} \mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \,|\, \boldsymbol{\theta}, t, S, \bar{E}].$$

Proof. To derive the desired inequality, we first understand the effect of conditioning on E. Let E_0 denote the event that $F := \mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ induces a facet of Q. Note that E is equivalent to $E_0 \cap \bar{E}$, where $\bar{E}$ is as above. We now perform the

change of variables from $\mathbf{a}_1, \dots, \mathbf{a}_d \in \mathbb{R}^d$ to $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \in \mathbb{R}_+$, $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ as in Definition 3.18. The set F is a facet of Q if and only if $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_{d+i} \le t$ for all $i \in [n-d]$ or $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_{d+i} \ge t$ for all $i \in [n-d]$. Given this, we see that

$$\begin{aligned}
&\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid D, E] \\
&= \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid D, E_0, \bar{E}] \\
&= \frac{\mathbb{E}[\ \mathbb{1}[E_0] \cdot \text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid D, \bar{E}]}{\Pr[E_0 \mid D, \bar{E}]} \qquad (3.2) \\
&= \frac{\mathbb{E}_{\boldsymbol{\theta},t}[\ \mathbb{E}[\mathbb{1}[E_0] \cdot \text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[E_0 \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}
\end{aligned}$$

Since $\mathbf{a}_1, \dots, \mathbf{a}_n$ are independent, conditioned on $\boldsymbol{\theta}, t$, the random vectors $\mathbf{b}_1, \dots, \mathbf{b}_d$ are independent of $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_{d+1}, \dots, \boldsymbol{\theta}^\mathsf{T}\mathbf{a}_n$. Since the events D and $\bar{E}$ only depend on $\mathbf{b}_1, \dots, \mathbf{b}_d$, continuing from (3.2), we get that

$$\begin{aligned}
&\frac{\mathbb{E}_{\boldsymbol{\theta},t}[\ \mathbb{E}[\mathbb{1}[E_0] \cdot \text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[E_0 \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]} \\
&= \frac{\mathbb{E}_{\boldsymbol{\theta},t}[\Pr[E_0 \mid \boldsymbol{\theta}, t] \cdot \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[E_0 \mid \boldsymbol{\theta}, t]]} \\
&\ge \inf_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t>0} \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}].
\end{aligned}$$

The last inequality uses that $\frac{\int f(x)g(x)\ \mathrm{d}x}{\int f(x)\ \mathrm{d}x} \ge \inf g(x)$ if f is non-negative and has finite integral.

Lastly, since the event D is equivalent to $S := S_\omega(\mathbf{b}_1, \dots, \mathbf{b}_d) \in \mathcal{S}$ as in Definition 3.19, we have that

$$\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \,|\, D, E] \ge \inf_{\boldsymbol{\theta}, t, S \in \mathcal{S}} \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \,|\, \boldsymbol{\theta}, t, S, \bar{E}].$$

□

DEFINITION 3.21 (Kernel combination). *For $S \in \mathcal{S}$, define the combination $\mathbf{z} := \mathbf{z}(S)$ to be the unique (up to sign) $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{R}^d$ satisfying*

$$\sum_{i=1}^d z_i \mathbf{s}_i = \mathbf{0},\ \sum_{i=1}^d z_i = 0,\ \|\mathbf{z}\|_1 = 1.$$

To justify the above definition, it suffices to show that the system of equations (i) $\sum_{i=1}^d z_i \mathbf{s}_i = \mathbf{0}, \sum_{i=1}^d z_i = 0$ has a one-dimensional solution space. Since $\mathbf{s}_1, \dots, \mathbf{s}_d$ live in a $d-2$ dimensional space, the solution space has dimension at least 1 by dimension counting. Next, note that $\mathbf{z}$ is a solution to (i) iff $z_1 = -\sum_{i=2}^d z_i$ and (ii) $\sum_{i=2}^d z_i \mathbf{s}_i = \mathbf{0}$ (since $\mathbf{s}_1 = \mathbf{0}$). Thus, the solution space of (i) and (ii) have the same dimension. Given our assumption that $\text{rank}(\mathbf{s}_2, \dots, \mathbf{s}_d) = d-2$, it follows that (ii) is one-dimensional, as needed.

Observe that for $S := S_{\boldsymbol{\omega}}(\mathbf{b}_1, \dots, \mathbf{b}_d)$, $\mathbf{z}$ satisfies $\pi_{\bar{\boldsymbol{\omega}}^\perp}(\sum_{i=1}^d z_i \mathbf{b}_i) = \mathbf{0}$.

The vector $\mathbf{z}$ provides us with a unit to measure lengths in "convex combination space". We make this formal with the next definition:

DEFINITION 3.22 (Chord combinations). *We define the set of convex combinations of the shape $S = (\mathbf{s}_1, \dots, \mathbf{s}_d) \in \mathcal{S}$ that equal $\mathbf{q} \in \bar{\boldsymbol{\omega}}^\perp$ by*

$$C_S(\mathbf{q}) := \left\{ (\lambda_1, \dots, \lambda_d) \ge \mathbf{0} \ :\ \sum_{i=1}^d \lambda_i = 1,\ \sum_{i=1}^d \lambda_i \mathbf{s}_i = \mathbf{q} \right\} \subset \mathbb{R}^d.$$

When S is clear we drop the subscript.

Observe that $C(\mathbf{q})$ is a line segment of the form $C(\mathbf{q}) = \boldsymbol{\lambda}_{\mathbf{q}} + \mathbf{z} \cdot [0, d_{\mathbf{q}}]$. We write $\|C(\mathbf{q})\|_1$ for the ℓ_1-diameter of $C(\mathbf{q})$. Since $C(\mathbf{q})$ is a line segment, $\|C(\mathbf{q})\|_1 = d_{\mathbf{q}}$. We prove two basic properties of $\|C(\mathbf{q})\|_1$ as a function of $\mathbf{q}$.

LEMMA 3.23 (Properties of chord combinations). *Let $\mathbf{y} := \mathbf{y}(S) = \sum_{i=1}^{d} |z_i| \mathbf{s}_i$, with $\mathbf{z} := \mathbf{z}(S)$ as in Definition 3.22. Then the following holds:*

- *$\|C(\mathbf{q})\|_1$ is a concave function for $\mathbf{q} \in \operatorname{conv}(S)$.*
- *$\max_{\mathbf{q} \in \operatorname{conv}(S)} \|C(\mathbf{q})\|_1 = \|C(\mathbf{y})\|_1 = 2$.*

Proof. For the first claim, take $\mathbf{x}, \mathbf{y} \in \operatorname{conv}(S)$. Let $\boldsymbol{\alpha} \in C(\mathbf{x})$ and $\boldsymbol{\beta} \in C(\mathbf{y})$. Then we see that, for all $\gamma \in [0,1]$,

$$\gamma\boldsymbol{\alpha} + (1-\gamma)\boldsymbol{\beta} \geq \mathbf{0}, \quad \sum_{i=1}^{d} \gamma\alpha_i + (1-\gamma)\beta_i = 1, \quad \sum_{i=1}^{d} (\gamma\alpha_i + (1-\gamma)\beta_i)\mathbf{s}_i = \gamma\mathbf{x} + (1-\gamma)\mathbf{y},$$

from which we derive that

$$\gamma C(\mathbf{x}) + (1-\gamma) C(\mathbf{y}) \subseteq C(\gamma\mathbf{x} + (1-\gamma)\mathbf{y}),$$

and hence $\|C(\gamma\mathbf{x}+(1-\gamma)\mathbf{y})\|_1 \geq \|\gamma C(\mathbf{x}) + (1-\gamma) C(\mathbf{y})\|_1 = \gamma\|C(\mathbf{x})\|_1 + (1-\gamma)\|C(\mathbf{y})\|_1$.

For the second claim, we look at the combination $\mathbf{y} := \sum_{i=1}^{n} |z_i| \mathbf{s}_i \in \operatorname{conv}(S)$. For all $\gamma \in [-1,1]$, we have $\sum_{i=1}^{d} (|z_i| + \gamma z_i)\mathbf{s}_i = \mathbf{y}$, $\sum_{i=1}^{d} |z_i| + \gamma z_i = \|\mathbf{z}\|_1 = 1$ and $|z_i| + \gamma z_i \geq 0$, $\forall i \in [d]$. Hence, $\|C(\mathbf{y})\|_1 \geq 2$. Now suppose there is some $\mathbf{y}'$ with $\|C(\mathbf{y}')\|_1 > 2$. That means there is some convex combination $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \geq \mathbf{0}$, $\|\boldsymbol{\lambda}\|_1 = 1$, with $\sum_{i=1}^{d} \lambda_i \mathbf{s}_i = \mathbf{y}'$ such that $\boldsymbol{\lambda} + \mathbf{z} > \mathbf{0}$ and $\boldsymbol{\lambda} - \mathbf{z} > \mathbf{0}$. Let $I \cup J$ be a partition of $[d]$ such that $z_i \geq 0$ for $i \in I$ and $z_j < 0$ for $j \in J$. We know that $\sum_{i=1}^{d} z_i = 0$, so $\sum_{i \in I} z_i = -\sum_{j \in J} z_j$. This makes $1 = \|\mathbf{z}\|_1 = \sum_{i \in I} z_i + \sum_{j \in J} -z_i = 2\sum_{i \in I} z_i$, so $\sum_{i \in I} z_i = 1/2$. The combination $\boldsymbol{\lambda}$ satisfies

$$\sum_{i \in I} \lambda_i > \sum_{i \in I} z_i = 1/2, \quad \sum_{j \in J} \lambda_j > \sum_{j \in J} -z_j = 1/2,$$

so $\|\boldsymbol{\lambda}\|_1 > 1$. By contradiction we conclude that $\max_{\mathbf{q} \in \operatorname{conv}(S)} \|C(\mathbf{q})\|_1 = 2$. ∎

The ℓ_1-diameter $\|C(\mathbf{q})\|_1$ specified by $\mathbf{q} \in \operatorname{conv}(S(\mathbf{b}_1, \ldots, \mathbf{b}_d))$ directly relates to the length of the chord $(\mathbf{q} + \mathbf{x} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R}) \cap \operatorname{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, which projects to $\mathbf{q} + \mathbf{x}$ under $\pi_{\bar{\boldsymbol{\omega}}^\perp}$. Specifically, $\|C(\mathbf{q})\|_1$ measures how long the chord is compared to the longest chord through the simplex. The exact relation is given below.

LEMMA 3.24. *Let $(h_1, \ldots, h_d) = (\bar{\boldsymbol{\omega}}^\mathsf{T}\mathbf{b}_1, \ldots, \bar{\boldsymbol{\omega}}^\mathsf{T}\mathbf{b}_d)$, $(\mathbf{s}_1, \ldots, \mathbf{s}_d) = S(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, $\mathbf{x} = \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_1)$. For any $\mathbf{q} \in \operatorname{conv}(S)$ the following equality holds:*

$$\operatorname{length}((\mathbf{x} + \mathbf{q} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R}) \cap \operatorname{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d)) = \|C(\mathbf{q})\|_1 \cdot |\sum_{i=1}^{d} z_i h_i|.$$

Proof. By construction there is a convex combination $\lambda_1, \ldots, \lambda_d \geq 0$, $\sum_{i=1}^{d} \lambda_i = 1$ satisfying $\sum_{i=1}^{d} \lambda_i \mathbf{s}_i = \mathbf{q}$ such that $C(\mathbf{q}) = [\boldsymbol{\lambda}, \boldsymbol{\lambda} + \|C(\mathbf{q})\|_1 \mathbf{z}]$ and hence

$$(\mathbf{x} + \mathbf{q} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R}) \cap \operatorname{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) = [\sum_{i=1}^{d} \lambda_i \mathbf{b}_i, \sum_{i=1}^{d} (\lambda_i + \|C(\mathbf{q})\|_1 z_i)\mathbf{b}_i].$$

From this we deduce

$$\begin{aligned}\text{length}((\mathbf{x}+\mathbf{q}+\bar{\boldsymbol{\omega}}\cdot\mathbb{R})\cap\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d)) &= \left\|\sum_{i=1}^{d}(\lambda_i+\|C(\mathbf{q})\|_1 z_i)\mathbf{b}_i-\sum_{i=1}^{d}\lambda_i\mathbf{b}_i\right\| \\ &= \left\|\sum_{i=1}^{d}\|C(\mathbf{q})\|_1 z_i\mathbf{b}_i\right\| \\ &= \|C(\mathbf{q})\|_1\cdot|\sum_{i=1}^{d}z_ih_i|.\end{aligned}$$

The third equality follows from the definition of $z_1,\dots,z_d$: as $\pi_{\bar{\boldsymbol{\omega}}^\perp}(\sum_{i=1}^d z_i\mathbf{b}_i)=\mathbf{0}$, we must have $\|\sum_{i=1}^d z_i\mathbf{b}_i\|=\|\sum_{i=1}^d z_ih_i\bar{\boldsymbol{\omega}}\|=|\sum_{i=1}^d z_ih_i|$. ∎

We can view the terms in the above product as follows: the length of the longest chord of $\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d)$ parallel to $\bar{l}$ is $2|\sum_{i=1}^d z_ih_i|$, and the ratio of the length of the chord $\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d)\cap\bar{l}$ to the length of the longest chord parallel to $\bar{l}$ equals $\|C(\mathbf{q})\|_1/2$. This follows from Lemma 3.23 since $\|C(\mathbf{q})\|_1$ achieves a maximum value of 2 at $\mathbf{q}=\mathbf{y}$. As discussed in the high-level description, we will bound the expected values of these two quantities separately.

The term $|\sum_{i=1}^d z_ih_i|$ can also be used to simplify the volume term in the probability density of $\mathbf{b}_1,\dots,\mathbf{b}_d$ after we condition on the shape S. We prove this in the next lemma.

LEMMA 3.25. *For fixed $\boldsymbol{\theta}\in\mathbb{S}^{d-1}, t>0, S\in\mathcal{S}$, define $\mathbf{x}\in\bar{\boldsymbol{\omega}}^\perp, h_1,\dots,h_d\in\mathbb{R}$ conditioned on $\boldsymbol{\theta},t,S$ to have joint probability density function proportional to*

$$|\sum_{i=1}^{d}z_ih_i|\cdot\prod_{i=1}^{d}\bar{\mu}_i(\mathbf{x}+\mathbf{s}_i+h_i\bar{\boldsymbol{\omega}}),$$

where $\mathbf{z}:=\mathbf{z}(S)$ is as in Definition 3.21. Then for $\mathbf{b}_1,\dots,\mathbf{b}_d\in\mathbb{R}^{d-1}$ distributed as in Lemma 3.20, conditioned on $\boldsymbol{\theta},t$ and the shape $S=(\mathbf{s}_1,\dots,\mathbf{s}_d)$, where $\mathbf{s}_1=\mathbf{0}$, we have equivalence of the distributions

$$(\mathbf{b}_1,\dots,\mathbf{b}_d)\mid\boldsymbol{\theta},t,S\equiv(\mathbf{x}+\mathbf{s}_1+h_1\bar{\boldsymbol{\omega}},\dots,\mathbf{x}+\mathbf{s}_d+h_d\bar{\boldsymbol{\omega}})\mid\boldsymbol{\theta},t,S.$$

Proof. By Definition 3.18, the variables $\mathbf{b}_1,\dots,\mathbf{b}_d$ conditioned on $\boldsymbol{\theta},t$ have density proportional to

$$\text{vol}_{d-1}(\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d))\prod_{i=1}^{d}\bar{\mu}_i(\mathbf{b}_i).$$

We make a change of variables from $\mathbf{b}_1,\dots,\mathbf{b}_d$ to $\mathbf{x},\mathbf{s}_2,\dots,\mathbf{s}_d\in\bar{\boldsymbol{\omega}}^\perp, h_1,\dots,h_d\in\mathbb{R}$, defined by

$$(\mathbf{b}_1,\dots,\mathbf{b}_d)=(\mathbf{x}+h_1\bar{\boldsymbol{\omega}},\mathbf{x}+\mathbf{s}_2+h_d\bar{\boldsymbol{\omega}},\dots,\mathbf{x}+\mathbf{s}_d+h_d\bar{\boldsymbol{\omega}}).$$

Recall that any invertible linear transformation has constant Jacobian. We observe that

$$\text{vol}_{d-1}(\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d))=\int_{\text{conv}(S)}\text{length}((\mathbf{x}+\mathbf{q}+\bar{\boldsymbol{\omega}}\cdot\mathbb{R})\cap\text{conv}(\mathbf{b}_1,\dots,\mathbf{b}_d))\,\mathrm{d}\mathbf{q}.$$

By Lemma 3.24 we find

$$\mathrm{vol}_{d-1}(\mathrm{conv}(\mathbf{b}_1,\ldots,\mathbf{b}_d)) = |\sum_{i=1}^{d} z_i h_i| \int_{\mathrm{conv}(S)} \|C(\mathbf{q})\|_1 \, \mathrm{d}\mathbf{q}.$$

The integral of $\|C(\mathbf{q})\|_1$ over $\mathrm{conv}(S)$ is independent of $\mathbf{x}, h_1, \ldots, h_d$. Thus, for fixed $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, S \in \mathcal{S}$, the random variables $\mathbf{x}, h_1, \ldots, h_d$ have joint probability density proportional to

$$|\sum_{i=1}^{d} z_i h_i| \cdot \prod_{i=1}^{d} \bar{\mu}_i(\mathbf{x} + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}).$$

∎

Recall that $\bar{l} = \bar{\mathbf{p}} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R}$. The event $\bar{E}$ that $\mathrm{conv}(\mathbf{b}_1,\ldots,\mathbf{b}_d) \cap \bar{l} \neq \emptyset$ occurs if and only if $\bar{\mathbf{p}} \in \mathbf{x} + \mathrm{conv}(S)$, hence if and only if $\bar{\mathbf{p}} - \mathbf{x} \in \mathrm{conv}(S)$.

LEMMA 3.26. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, S \in \mathcal{S}$ be fixed, and have random variables $\mathbf{b}_1,\ldots,\mathbf{b}_d \in \mathbb{R}^{d-1}$,$h_1,\ldots,h_d \in \mathbb{R}$, $\mathbf{x} \in \omega^{\perp}$ be distributed as in Lemma 3.25. Define $\mathbf{q} := \bar{\mathbf{p}} - \mathbf{x}$. Then, the expected edge length satisfies*

$$\mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1,\ldots,\mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, S, \bar{E}] \geq \mathbb{E}[\|C(\mathbf{q})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \cdot \inf_{\mathbf{x} \in \bar{\boldsymbol{\omega}}^{\perp}} \mathbb{E}[|\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}].$$

Proof. We start with the assertion of Lemma 3.24:

$$\mathrm{length}((\mathbf{x} + \mathbf{q} + \bar{\boldsymbol{\omega}} \cdot \mathbb{R}) \cap \mathrm{conv}(\mathbf{b}_1,\ldots,\mathbf{b}_d)) = \|C(\mathbf{q})\|_1 \cdot |\sum_{i=1}^{d} z_i h_i|.$$

We take expectation on both sides to derive the equality

$$\mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1,\ldots,\mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, S, \bar{E}] = \mathbb{E}[\|C(\mathbf{q})\|_1 \cdot |\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \bar{E}].$$

Since $\|C(\mathbf{q})\|_1$ and $|\sum_{i=1}^{d} z_i h_i|$ do not share any of their variables, we separate the two expectations:

$$\begin{aligned}
\mathbb{E}[\|C(\mathbf{q})\|_1 \cdot |\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \bar{E}] &= \mathbb{E}_{\mathbf{x}, h_1, \ldots, h_d}[\|C(\mathbf{q})\|_1 \cdot |\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \bar{E}] \\
&= \mathbb{E}_{\mathbf{x}}[\|C(\mathbf{q})\|_1 \mathbb{E}_{h_1,\ldots,h_d}[|\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}] \mid \boldsymbol{\theta}, t, S, \bar{E}] \\
&\geq \mathbb{E}_{\mathbf{x}}[\|C(\mathbf{q})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \inf_{\mathbf{x} \in \bar{\boldsymbol{\omega}}^{\perp}} \mathbb{E}_{h_1,\ldots,h_d}[|\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}].
\end{aligned}$$

∎

We will first bound the expected ℓ_1-diameter of $C(\mathbf{q})$, where $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$, which depends on where $\bar{\mathbf{p}} - \mathbf{x}$ intersects the projected simplex $\mathrm{conv}(S)$: where this quantity tends to get smaller as we approach to boundary of $\mathrm{conv}(S)$. We recall that $\bar{E}$ occurs if and only if $\mathbf{q} \in \mathrm{conv}(S)$.

LEMMA 3.27 (Chord combination bound). *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0$ and $S \in \mathcal{S}$ be fixed. Let $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$ be distributed as in Lemma 3.26. Then, the expected ℓ_1-diameter of $C(\mathbf{q})$ satisfies*

$$\mathbb{E}[\|C(\mathbf{q})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \geq \frac{e^{-2}}{dL(1 + R_{n,d})}$$

Proof. To get a lower bound on the expected value of $\|C(\mathbf{q})\|_1$, we will use the concavity of $\|C(\mathbf{q})\|_1$ over $\operatorname{conv}(S) = \operatorname{conv}(\mathbf{s}_1, \ldots, \mathbf{s}_d)$ and that $\max_{\mathbf{q} \in \operatorname{conv}(S)} \|C(\mathbf{q})\|_1 = 2$. These facts are proven in Lemma 3.23. We show that shifting the projected simplex does not change the probability density too much (using log-Lipschitzness), and use the properties of $\|C(\mathbf{q})\|_1$ mentioned above.

Let $\hat{\mu}$ denote the probability density of $\mathbf{q}$ conditioned on $\boldsymbol{\theta}, t, S, \bar{E}$. Note that $\hat{\mu}$ is supported on $\operatorname{conv}(S)$ and has density proportional to

$$\int \cdots \int \prod_{i=1}^{d} \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{q} + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}) \, \mathrm{d}h_1 \cdots \, \mathrm{d}h_d.$$

We claim that $\hat{\mu}$ is dL-log-Lipschitz. To see this, note that since $\bar{\mu}_1, \ldots, \bar{\mu}_d$ are L-log-Lipschitz, for $\mathbf{v}, \mathbf{v}' \in \operatorname{conv}(S)$ we have that

$$\begin{aligned}
&\int \cdots \int \prod_{i=1}^{d} \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{v} + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}) \, \mathrm{d}h_1 \cdots \, \mathrm{d}h_d \\
&\leq \int \cdots \int \prod_{i=1}^{d} e^{L\|\mathbf{v}' - \mathbf{v}\|} \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{v}' + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}) \, \mathrm{d}h_1 \cdots \, \mathrm{d}h_d \\
&= e^{dL\|\mathbf{v}' - \mathbf{v}\|} \int \cdots \int \prod_{i=1}^{d} \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{v}' + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}) \, \mathrm{d}h_1 \cdots \, \mathrm{d}h_d \text{ , as needed .}
\end{aligned}$$

Let $\alpha \in (0, 1)$ be a scale factor to be chosen later, and let $\mathbf{y} = \mathbf{y}(S) \in \operatorname{conv}(S)$ be as in Lemma 3.23. Now we can write

$$\begin{aligned}
\mathbb{E}[\|C(\mathbf{q})\| \mid \boldsymbol{\theta}, t, S, \bar{E}] &= \int_{\operatorname{conv}(S)} \|C(\mathbf{q})\|_1 \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q} \\
&\geq \int_{\alpha \operatorname{conv}(S) + (1-\alpha)\mathbf{y}} \|C(\mathbf{q})\|_1 \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q}, \qquad (3.3)
\end{aligned}$$

because the integrand is non-negative. By concavity of $\|C(\mathbf{q})\|_1$ we have the lower bound $\|C(\alpha\mathbf{q} + (1-\alpha)\mathbf{y})\| \geq 2(1-\alpha)$ for all $\mathbf{q} \in \operatorname{conv}(S)$. Therefore, (3.3) is lower bounded by

$$\begin{aligned}
&\geq \int_{\alpha \operatorname{conv}(S) + (1-\alpha)\mathbf{y}} 2(1-\alpha) \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q} \\
&= 2\alpha^d (1-\alpha) \int_{\operatorname{conv}(S)} \hat{\mu}(\alpha\mathbf{q} + (1-\alpha)\mathbf{y}) \, \mathrm{d}\mathbf{q} \\
&\geq 2\alpha^d (1-\alpha) e^{-\max_{\mathbf{q} \in \operatorname{conv}(S)} (1-\alpha)\|\mathbf{q} - \mathbf{y}\| \cdot dL} \int_{\operatorname{conv}(S)} \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q}, \\
&= 2\alpha^d (1-\alpha) e^{-\max_{i \in [d]} (1-\alpha)\|\mathbf{s}_i - \mathbf{y}\| \cdot dL}, \qquad (3.4)
\end{aligned}$$

where we used a change of variables in the first equality, the dL-log-Lipschitzness of $\hat{\mu}$ in the second inequality, and the convexity of the ℓ_2 norm in the last equality. Using the diameter bound of $2 + 2R_{n,d}$ for $\text{conv}(S)$, (3.4) is lower bounded by

$$\geq 2\alpha^d(1-\alpha)e^{-(1-\alpha)dL(2+2R_{n,d})}. \tag{3.5}$$

Setting $\alpha = 1 - \frac{1}{dL(2+2R_{n,d})} \geq 1 - 1/d$ (by Lemma 3.9) gives a lower bound for (3.5) of

$$\geq e^{-2}\frac{1}{dL(1+R_{n,d})} \ . \qquad \square$$

Recall that we have now fixed the position $\mathbf{x}$ and shape S of the projected simplex. The randomness we have left is in the positions $h_1, \ldots, h_d$ of $\mathbf{b}_1, \ldots, \mathbf{b}_d$ along lines parallel to the vector $\bar{\boldsymbol{\omega}}$. As $\boldsymbol{\theta}$ and t are also fixed, restricting $\mathbf{b}_i$ to lie on a line is the same as restricting $\mathbf{a}_i$ to lie on a line.

Thus, were it not for the correlation between $h_1, \ldots, h_d$, i.e., the factor $|\sum_{i=1}^d z_i h_i|$ in the joint probability density function, each h_i would be independent and have variance τ^2 by assumption, and thus one might expect $\mathbb{E}[|\sum_{i=1}^d z_i h_i|] = \Omega(\|\mathbf{z}\|\tau)$. The following lemmas establish this, and show that in fact, the correlation term only helps.

LEMMA 3.28. *Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \tau^2$. Then X satisfies*

$$\frac{\mathbb{E}\left[X^2\right]}{\mathbb{E}\left[|X|\right]} \geq (|\mu| + \tau)/2.$$

Proof. By definition one has $\mathbb{E}\left[X^2\right] = \mu^2 + \tau^2$. We will show that $\mathbb{E}\left[|X|\right] \leq |\mu| + \tau$ so that we can use the fact that $\mu^2 + \tau^2 \geq 2|\mu|\tau$ to derive that $\mu^2 + \tau^2 \geq (|\mu| + \tau)^2/2$. It then follows that $\mathbb{E}\left[X^2\right]/\mathbb{E}\left[|X|\right] \geq (|\mu| + \tau)/2$.

The expected absolute value $\mathbb{E}[|X|]$ satisfies

$$\mathbb{E}\left[|X|\right] \leq |\mu| + \mathbb{E}\left[|X - \mu|\right] \leq |\mu| + \mathbb{E}\left[(X-\mu)^2\right]^{1/2}$$

by Cauchy-Schwarz, hence $\mathbb{E}\left[|X|\right] \leq |\mu| + \tau$. $\square$

LEMMA 3.29 (Height of simplex bound). *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0$, $S \in \mathcal{S}, \mathbf{x} \in \bar{\omega}^\perp$ be fixed and let $\mathbf{z} := \mathbf{z}(S)$ be as in Definition 3.21. Then for $h_1, \ldots, h_d \in \mathbb{R}$ distributed as in Lemma 3.26, the expected inner product satisfies*

$$\inf_{\mathbf{x} \in \bar{\boldsymbol{\omega}}^\perp} \mathbb{E}[|\sum_{i=1}^d z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}] \geq \tau/(2\sqrt{d}).$$

Proof. For fixed $\theta, t, S, \mathbf{x}$, let $g_1, \ldots, g_d \in \mathbb{R}$ be independent random variables with respective probability densities $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$, where $\tilde{\mu}_i$, $i \in [d]$, is defined by

$$\tilde{\mu}_i(g_i) := \bar{\mu}(\mathbf{x} + \mathbf{s}_i + g_i\bar{\boldsymbol{\omega}}) = \mu(R_{\boldsymbol{\theta}}(\mathbf{x} + \mathbf{s}_i + g_i\bar{\boldsymbol{\omega}}) + t\boldsymbol{\theta}) \ .$$

Note that, by assumption, the variables $g_1, \ldots, g_d$ each have variance at least τ^2. We recall from Lemma 3.25 that the joint probability density of $h_1, \ldots, h_d$ is proportional to $|\sum_{i=1}^d z_i h_i| \prod_{i=1}^d \tilde{\mu}_i(h_i)$. Thus, we may rewrite the above expectation as

$$\begin{aligned}\mathbb{E}[|\sum_{i=1}^d z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}] &= \frac{\int \cdots \int_{\mathbb{R}} |\sum_{i=1}^d z_i h_i|^2 \prod_{i=1}^d \tilde{\mu}_i(h_i) \ \mathrm{d}h_1 \cdots \ \mathrm{d}h_d}{\int \cdots \int_{\mathbb{R}} |\sum_{i=1}^d z_i h_i| \prod_{i=1}^d \tilde{\mu}_i(h_i) \ \mathrm{d}h_1 \cdots \ \mathrm{d}h_d} \\ &= \frac{\mathbb{E}[|\sum_{i=1}^d z_i g_i|^2]}{\mathbb{E}[|\sum_{i=1}^d z_i g_i|]} \ ,\end{aligned}$$

where $g_1, \dots, g_d$ are distributed independently with densities $\tilde{\mu}_1, \dots, \tilde{\mu}_d$. By the additivity of variance for independent random variables, we see that

$$\mathrm{Var}(\sum_{i=1}^{d} z_i g_i) = \sum_{i=1}^{d} z_i^2 \mathrm{Var}(g_i) \geq \tau^2 \|\mathbf{z}\|^2 \geq \tau^2 \|\mathbf{z}\|_1^2 / d = \tau^2 / d.$$

We reach the desired conclusion by applying Lemma 3.28:

$$\frac{\mathbb{E}[|\sum_{i=1}^{d} z_i g_i|^2]}{\mathbb{E}[|\sum_{i=1}^{d} z_i g_i|]} \geq \frac{|\mathbb{E}[\sum_{i=1}^{d} z_i g_i]| + \sqrt{\mathrm{Var}(\sum_{i=1}^{d} z_i g_i)}}{2} \geq \tau/(2\sqrt{d}).$$

∎

Using the bounds from the preceding lemmas, the proof of our main theorem is now given below.

Proof of Theorem 3.10 (Parametrized Shadow Bound). By Lemma 3.13, we derive the shadow bound by combining an upper bound on $\mathbb{E}[\text{perimeter}(Q \cap W)]$ and a uniform lower bound on $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i : i \in I) \cap W) \mid E_I]$ for all $I \in B$. For the perimeter upper bound, by Lemma 3.14 we have that

$$\mathbb{E}[\text{perimeter}(Q \cap W)] \leq 2\pi(1 + 4r_n). \tag{3.6}$$

For the edge length bound, we assume w.l.o.g. as above that $I = [d]$. Combining prior lemmas, we have that

$$\begin{aligned}
&\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid E] \\
&\geq \frac{1}{2} \cdot \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid D, E] \quad (\text{ Lemma 3.17 }) \\
&\geq \frac{1}{2} \cdot \inf_{\substack{\boldsymbol{\theta} \in \mathbb{S}^{d-1} \\ t>0}} \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, S \in \mathcal{S}, \bar{E}] \quad (\text{ Lemma 3.20 }) \\
&\geq \frac{1}{2} \cdot \inf_{\substack{\boldsymbol{\theta} \in \mathbb{S}^{d-1} \\ t>0, S \in \mathcal{S}}} \left(\mathbb{E}[\|C(\bar{\mathbf{p}} - \mathbf{x})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \cdot \inf_{\mathbf{x} \in \bar{\boldsymbol{\omega}}^{\perp}} \mathbb{E}[|\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}] \right) \\
&\qquad\qquad (\text{ Lemma 3.26 }) \\
&\geq \frac{1}{2} \cdot \frac{e^{-2}}{dL(1 + R_{n,d})} \cdot \frac{\tau}{2\sqrt{d}} \quad (\text{ Lemmas 3.27 and 3.29 }).
\end{aligned} \tag{3.7}$$

The theorem now follows by taking the ratio of (3.6) and (3.7). ∎

3.2. Shadow bound for Laplace perturbations. Theorem 3.10 is most naturally used to prove shadow bounds or distributions where all parameters are bounded, which we illustrate here for Laplace-distributed perturbations. The Laplace distribution is defined in section 2.3.2. To achieve the shadow bound, we use the abstract shadow bound as a black box, and we bound the necessary parameters of the Laplace distribution below.

LEMMA 3.30. *For $n \geq d \geq 3$, the Laplace distribution $L_d(\bar{\mathbf{a}}, \sigma)$, satisfies the following properties:*

1. *The density is $\sqrt{d}/\sigma$-log-Lipschitz.*
2. *Its cutoff radius satisfies $R_{n,d} \leq 14\sigma\sqrt{d}\log n$.*
3. *The n-th deviation satisfies $r_n \leq 7\sigma \log n$.*
4. *The variance after restricting to any line satisfies $\tau \geq \sigma/\sqrt{de}$.*

Proof. By shift invariance of the parameters, we may assume w.l.o.g. that $\bar{\mathbf{a}} = \mathbf{0}$. Let $\mathbf{X}$ be distributed as $L_d(\mathbf{0}, \sigma)$ for use below.

1.. The density of the Laplace distribution is proportional to $e^{-\|\mathbf{x}\|\sqrt{d}/\sigma}$, for $\mathbf{x} \in \mathbb{R}^d$, and thus the logarithm of the density differs an additive constant from $-\|\mathbf{x}\|\sqrt{d}/\sigma$, which is clearly $\sqrt{d}/\sigma$-Lipschitz.

2.. The second property follows from Lemma 2.6:

$$\begin{aligned}\Pr[\|\mathbf{X}\| \geq 14\sigma\sqrt{d}\log n] &\leq e^{-2d\log n} = n^{-2d} \\ &\leq \frac{1}{d\binom{n}{d}}.\end{aligned}$$

3.. Again from Lemma 2.6. If $7\log n \geq 2\sqrt{d}$, we get that

$$\begin{aligned}\int_{7\sigma\log n}^{\infty} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t]\,\mathrm{d}t &\leq \int_{7\sigma\log n}^{\infty} e^{-\sqrt{d}t/(7\sigma)}\,\mathrm{d}t \\ &= \frac{7\sigma}{\sqrt{d}} n^{-\sqrt{d}\log n} \leq \frac{7\sigma\log n}{n}.\end{aligned}$$

If $7\log n \leq 2\sqrt{d}$, then

$$\begin{aligned}\int_{7\sigma\log n}^{\infty} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t]\,\mathrm{d}t &= \int_{7\sigma\log n}^{2\sigma\sqrt{d}} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t]\,\mathrm{d}t + \int_{2\sigma\sqrt{d}}^{\infty} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t]\,\mathrm{d}t \\ &\leq \int_{7\sigma\log n}^{2\sigma\sqrt{d}} 2e^{-t^2/(16\sigma^2)}\,\mathrm{d}t + \int_{2\sigma\sqrt{d}}^{\infty} e^{-\sqrt{d}t/(7\sigma)}\,\mathrm{d}t \\ &\leq 4\sigma\sqrt{d}e^{-(7\log n)^2/16} + \frac{7\sigma}{\sqrt{d}}e^{-2d/7} \\ &\leq 4\sigma\sqrt{d}/n^3 + 7\sigma/(\sqrt{d}n^{\sqrt{d}}) \leq \frac{7\sigma\log n}{n}.\end{aligned}$$

4.. This follows from the $\sqrt{d}/\sigma$-log-Lipschitzness and Lemma 3.8. ∎

Proof of Theorem 3.2 (Shadow bound for Laplace perturbations). We get the desired result by plugging in the bounds from Lemma 3.30 for $L, R_{n,d}, r_n$ and τ into the upper bound $O((d^{1.5}L/\tau)(1+R_{n,d})(1+r_n))$ from Theorem 3.10. ∎

3.3. Shadow bound for Gaussian perturbations. In this subsection, we prove our shadow bound for Gaussian perturbations.

The Gaussian distribution is not log-Lipschitz, so we can not directly apply Theorem 3.10. We will define a *smoothed out* version of the Gaussian distribution to remedy this problem, which we call the Laplace-Gaussian distribution. The Laplace-Gaussian distribution, defined below, matches the Gaussian distribution in every meaningful parameter, while also being log-Lipschitz. We will first bound the shadow size for Laplace-Gaussian perturbations, and then show that the expected number of edges of $Q \cap W$ for Gaussian perturbations is at most 1 larger.

DEFINITION 3.31. *We define a random variable $\mathbf{X} \in \mathbb{R}^d$ to be (σ, r)-*Laplace-Gaussian distributed with mean $\bar{\mathbf{a}}$*, or $\mathbf{X} \sim LG_d(\bar{\mathbf{a}}, \sigma, r)$, if its density is proportional to $f_{(\bar{\mathbf{a}},\sigma,r)} : \mathbb{R}^d \to \mathbb{R}_+$ given by*

$$f_{(\bar{\mathbf{a}},\sigma,r)}(\mathbf{x}) = \begin{cases} e^{-\|\mathbf{x}-\bar{\mathbf{a}}\|^2/(2\sigma^2)} & \text{if } \|\mathbf{x}-\bar{\mathbf{a}}\| \leq r\sigma \\ e^{-\|\mathbf{x}-\bar{\mathbf{a}}\|r/\sigma + r^2/2} & \text{if } \|\mathbf{x}-\bar{\mathbf{a}}\| \geq r\sigma. \end{cases}$$

Note that at $\|\mathbf{x}-\bar{\mathbf{a}}\| = r\sigma$, both cases give the density $e^{-r^2/2}$, and hence $f_{(\bar{\mathbf{a}},\sigma,r)}$ is well-defined and continuous on $\mathbb{R}^d$. For distributions with mean $\mathbf{0}$, we abbreviate $f_{(\sigma,r)} := f_{(\mathbf{0},\sigma,r)}$ and $LG_d(\sigma, r) := LG_d(\mathbf{0}, \sigma, r)$.

Just like for the shadow size bound for Laplace perturbations, we need strong enough tail bounds. We state these tail bounds here, and defer their proofs till the end of the section.

LEMMA 3.32 (Laplace-Gaussian tail bounds). *Let $\mathbf{X} \in \mathbb{R}^d$ be (σ, r)-Laplace-Gaussian distributed with mean $\mathbf{0}$, where $r := c\sqrt{d \log n}$, $c \geq 4$. Then for $t \geq r$,*

$$\Pr[\|\mathbf{X}\| \geq \sigma t] \leq e^{-(1/4)rt} \ . \tag{3.8}$$

For $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $t \geq 0$,

$$\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq \sigma t] \leq \begin{cases} e^{-(1/4)rt} & : t \geq r \\ 3e^{-t^2/4} & : 0 \leq t \leq r. \end{cases} \tag{3.9}$$

LEMMA 3.33. *For $n \geq d \geq 3$, the $(\sigma, 4\sqrt{d \log n})$-Laplace-Gaussian distribution in $\mathbb{R}^d$ with mean $\bar{\mathbf{a}}$ satisfies the following properties:*

1. *The density is $4\sigma^{-1}\sqrt{d \log n}$-log-Lipschitz.*
2. *Its cutoff radius satisfies $R_{n,d} \leq 4\sigma\sqrt{d \log n}$.*
3. *The n-th deviation is $r_n \leq 4\sigma\sqrt{\log n}$.*
4. *The variance after restricting to any line satisfies $\tau \geq \sigma/4$.*

Proof. As before, by shift invariance, we may assume w.l.o.g that $\bar{\mathbf{a}} = \mathbf{0}$. Let $\mathbf{X} \sim LG_d(\sigma, 4\sqrt{d \log n})$ and let $r := 4\sqrt{d \log n}$.

1.. The gradient of the function $\log(f_{(\sigma,r)}(\mathbf{x}))$ has norm bounded by $4\sigma^{-1}\sqrt{d \log n}$ wherever it is defined, which by continuity implies $f_{(\sigma,r)}$ is $4\sigma^{-1}\sqrt{d \log n}$-log-Lipschitz.

2.. Applying the tail bound from Lemma 3.32, we get that

$$\Pr[\|\mathbf{X}\| \geq 4\sigma\sqrt{d \log n}] \leq e^{-4d \log n} \leq \frac{1}{d\binom{n}{d}}.$$

3.. Again using Lemma 3.32,

$$\begin{aligned} \int_{4\sigma\sqrt{\log n}}^{\infty} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] \, \mathrm{d}t &= \int_{4\sigma\sqrt{\log n}}^{r\sigma} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] \, \mathrm{d}t + \int_{r\sigma}^{\infty} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] \, \mathrm{d}t \\ &\leq \int_{4\sigma\sqrt{\log n}}^{r\sigma} 3e^{-t^2/(4\sigma^2)} \, \mathrm{d}t + \int_{r\sigma}^{\infty} e^{-\sqrt{d \log n}t/\sigma} \, \mathrm{d}t \\ &\leq 4\sigma\sqrt{d \log n}(3n^{-4}) + \frac{\sigma}{\sqrt{d \log n}} n^{-4d} \\ &\leq 4\sigma\sqrt{\log n}/n. \end{aligned}$$

4.. For the line variance, by rotational symmetry, we may without loss of generality assume that $l := (\mathbf{y}, 0) + \mathbf{e}_d\mathbb{R}$, where $\mathbf{y} \in \mathbb{R}^{d-1}$, and so $(\mathbf{y}, 0)$ is the point on l closest to the origin. Since $f_{(\sigma,r)}(\mathbf{y}, \lambda) = f_{(\sigma,r)}(\mathbf{y}, -\lambda)$ for every $\lambda \in \mathbb{R}$, the expectation $\mathbb{E}[\mathbf{X} \mid \mathbf{X} \in l] = (\mathbf{y}, 0)$. Thus, $\mathrm{Var}(\mathbf{X} \mid \mathbf{X} \in l) = \mathbb{E}[X_d^2 \mid \mathbf{X} \in l]$.

Let $\bar{l} = (\mathbf{y}, 0) + [-\sigma, \sigma] \cdot \mathbf{e}_d$. Since $|X_d|$ is larger on $l \setminus \bar{l}$ than on $\bar{l}$, we clearly have $\mathbb{E}[X_d^2 \mid \mathbf{X} \in l] \geq \mathbb{E}[X_d^2 \mid \mathbf{X} \in \bar{l}]$, so it suffices to lower bound the latter quantity.

For each $\mathbf{y}$ with $\|\mathbf{y}\| \leq \sigma r$ we have for all $\lambda \in [-\sigma, \sigma]$ the inequality

$$1 \geq \frac{f_{(\sigma,r)}(\mathbf{y}, \lambda)}{f_{(\sigma,r)}(\mathbf{y}, 0)} \geq \frac{e^{-\|(\mathbf{y},\lambda)\|^2/(2\sigma^2)}}{e^{-\|(\mathbf{y},0)\|^2/(2\sigma^2)}} = e^{-\lambda^2/(2\sigma^2)} \geq e^{-1/2} \ . \tag{3.10}$$

Given the above, we have that

$$\begin{aligned}\mathbb{E}[X_d^2 \mid \mathbf{X} \in \bar{l}] &\geq (\sigma^2/4)\Pr[|X_d| \geq \sigma/2 \mid \mathbf{X} \in \bar{l}] \\ &= (\sigma^2/4)\frac{\int_{\sigma/2}^{\sigma} f_{(\sigma,r)}(\mathbf{y},t)\ \mathrm{d}t}{\int_0^{\sigma} f_{(\sigma,r)}(\mathbf{y},t)\ \mathrm{d}t} \\ &\geq (\sigma^2/4)\frac{\int_{\sigma/2}^{\sigma} f_{(\sigma,r)}(\mathbf{y},0)e^{-1/2}\ \mathrm{d}t}{\int_0^{\sigma} f_{(\sigma,r)}(\mathbf{y},0)\ \mathrm{d}t} \quad (\text{ by (3.10) }) \\ &= (\sigma^2/4)(e^{-1/2}/2) \geq \sigma^2/16 \quad , \text{as needed .}\end{aligned} \tag{3.11}$$

For $\mathbf{y}$ with $\|\mathbf{y}\| \geq \sigma r$, $\lambda \in [-\sigma, \sigma]$, we similarly have

$$\begin{aligned}\|(\mathbf{y},\lambda)\| &= \sqrt{\|\mathbf{y}\|^2 + \lambda^2} \\ &\leq \|\mathbf{y}\| + \frac{\lambda^2}{2\|\mathbf{y}\|} \leq \|\mathbf{y}\| + \frac{\lambda^2}{2r\sigma}.\end{aligned}$$

In particular, we get that

$$1 \geq \frac{f_{(\sigma,r)}(\mathbf{y},\lambda)}{f_{(\sigma,r)}(\mathbf{y},0)} = \frac{e^{-\|(\mathbf{y},\lambda)\|(r/\sigma)}}{e^{-\|(\mathbf{y},0)\|(r/\sigma)}} \geq e^{-\lambda^2/(2\sigma^2)} \geq e^{-1/2}\ . \tag{3.12}$$

The desired lower bound now follows by combining (3.11), (3.12). ∎

Given any unperturbed unit LP given by $\mathbf{c}, \bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_n$, we denote by $\mathbb{E}_{N_d(\sigma)}$ the expectation when its vertices are perturbed with noise distributed according to the Gaussian distribution of standard deviation σ and we write $\mathbb{E}_{LG_d(\sigma,r)}$ for the expectation when its vertices are perturbed by (σ, r)-Laplace-Gaussian noise. In the next lemma we prove that, for $r := 4\sqrt{d \log n}$, the expected number of edges for Gaussian distributed perturbations is not much bigger than the expected number for Laplace-Gaussian perturbations. We use the strong tail bounds we have on the two distributions along with the knowledge that restricted to a ball of radius $r\sigma$ the probability densities are equal. Recall that we use $\hat{\mathbf{a}}_i$ to denote the perturbation $\mathbf{a}_i - \mathbb{E}[\mathbf{a}_i]$.

LEMMA 3.34. *For $d \geq 3$, the number of edges in $\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W$ satisfies*

$$\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n))|] \leq 1 + \mathbb{E}_{LG_d(\sigma, 4\sqrt{d \log n})}[|\mathrm{edges}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n))|].$$

Proof. Let us abbreviate edges := $\mathrm{edges}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n))$ and let $r := 4\sqrt{d \log n}$. We make use of the fact that $N_d(\sigma)$ and $LG_d(\sigma, r)$ are equal when restricted to distance at most σr from their centers.

$$\begin{aligned}\mathbb{E}_{N(\sigma)}[|\mathrm{edges}|] = &\Pr_{N_d(\sigma)}[\exists i \in [n]\ \|\hat{\mathbf{a}}_i\| > \sigma r]\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \exists i \in [n]\ \|\hat{\mathbf{a}}_i\| > \sigma r] \\ &+ \Pr_{N_d(\sigma)}[\forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r]\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r].\end{aligned} \tag{3.13}$$

By Lemma 3.32, the first probability is at most $n^{-4d} \leq n^{-d}/4$, so we upper bound the first number of edges by $\binom{n}{d}$ making a total contribution of less than $1/4$. Now we use the fact that within radius $4\sigma\sqrt{d \log n}$ we have equality of densities between $N_d(\sigma)$ and $LG_d(\sigma, r)$. Continuing from (3.13),

$$\begin{aligned}&\leq 1/4 + \mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r] \\ &= 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r] \\ &\leq 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}|]/\Pr_{LG_d(\sigma,r)}[\forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r].\end{aligned} \tag{3.14}$$

The inequality above is true by non-negativity of the number of edges. Next we lower bound the denominator and continue (3.14),

$$\begin{aligned} &\leq 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\text{edges}|]/(1 - n^{-d}/4) \\ &\leq 1/4 + (1 + n^{-d}/2)\mathbb{E}_{LG_d(\sigma,r)}[|\text{edges}|]. \end{aligned} \tag{3.15}$$

The last inequality we deduce from the fact that $(1-\varepsilon)(1+2\varepsilon) = 1+\varepsilon-2\varepsilon^2$, which is bigger than 1 for $0 < \varepsilon < 1/2$. Again using the trivial upper bound of $\binom{n}{d}$ edges, we arrive at our desired conclusion that

$$\mathbb{E}_{N_d(\sigma)}[|\text{edges}|] \leq 1 + \mathbb{E}_{LG_d(\sigma,r)}[|\text{edges}|]. \qquad \square$$

We now have all the ingredients to prove our bound on the expected number of edges for Gaussian perturbations.

Proof of Theorem 3.1 (Shadow bound for Gaussian perturbations). By Lemma 3.34, we know that

$$\mathbb{E}_{N_d(\sigma)}[|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n))|] \leq 1 + \mathbb{E}_{LG_d(\sigma, 4\sqrt{d \log n})}[|\text{edges}(\text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n))|].$$

We now derive the shadow bound for Laplace-Gaussian perturbations by combining the parameter bounds in Lemma 3.33 with the parameterized shadow bound in Theorem 3.10. $\square$

We now prove the tail bounds for Laplace-Gaussian distributions. Recall that we set $r := c\sqrt{d \log n}$ with $c \geq 4$.

Proof of Lemma 3.32 (Tail bound for Laplace-Gaussian distribution). By homogeneity, we may w.l.o.g. assume that $\sigma = 1$. Define auxiliary random variables $\mathbf{Y} \in \mathbb{R}^d$ distributed as $(\mathbf{0}, 1/(c\sqrt{\log n}))$-Laplace and $\mathbf{Z} \in \mathbb{R}^d$ be distributed as $N_d(\mathbf{0}, 1)$.

Since $\mathbf{X}$ has density proportional to $f_{(1,r)}(\mathbf{x})$, which equals $e^{-\|\mathbf{x}\|^2/2}$ for $\|\mathbf{x}\| \leq r$ and $e^{-r\|\mathbf{x}\| + r^2/2}$ for $\|\mathbf{x}\| \geq r$, we immediately see that

$$\begin{aligned} \mathbf{Z} \mid \|\mathbf{Z}\| \leq r &\equiv \mathbf{X} \mid \|\mathbf{X}\| \leq r \\ \mathbf{Y} \mid \|\mathbf{Y}\| \geq r &\equiv \mathbf{X} \mid \|\mathbf{X}\| \geq r \end{aligned} \tag{3.16}$$

Proof of (3.8). By the above, for any $t \geq r$, we have that

$$\Pr[\|\mathbf{X}\| \geq t] = \Pr[\|\mathbf{Y}\| \geq t] \cdot \frac{\Pr[\|\mathbf{X}\| \geq r]}{\Pr[\|\mathbf{Y}\| \geq r]}\,. \tag{3.17}$$

For the first term, by the Laplace tail bound (2.3), we get that

$$\Pr[\|\mathbf{Y}\| \geq t] \leq e^{-rt - d\log(\frac{c\sqrt{\log n}\, t}{\sqrt{d}}) - d}\,. \tag{3.18}$$

For the second term,

$$\begin{aligned} \frac{\Pr[\|\mathbf{X}\| \geq r]}{\Pr[\|\mathbf{Y}\| \geq r]} &= e^{r^2/2} \frac{\int_{\mathbb{R}^n} e^{-r\|\mathbf{x}\|}\, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} f_{(\sigma,r)}(\mathbf{x})\, \mathrm{d}\mathbf{x}} \leq e^{r^2/2} \frac{\int_{\mathbb{R}^n} e^{-r\|\mathbf{x}\|}\, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} e^{-\|\mathbf{x}\|^2/2}\, \mathrm{d}\mathbf{x}} \\ &\leq e^{r^2/2} \frac{r^{-d} d! \mathrm{vol}_d(\mathcal{B}_2^d)}{\sqrt{2\pi}^d} \leq e^{(dc^2 \log n)/2} (\frac{\sqrt{e}}{c\sqrt{\log n}})^d \\ &\leq e^{(dc^2 \log n)/2}\,, \end{aligned} \tag{3.19}$$

where we have used the upper bound $\mathrm{vol}_d(\mathcal{B}_2^d) \leq (2\pi e/d)^{d/2}$, $r = c\sqrt{d \log n}$ and $c \geq \sqrt{e}$. Combining (3.18), (3.19) and that $t \geq r$, $c \geq 4$, we get

$$\begin{aligned} \Pr[\|\mathbf{X}\| \geq t] &\leq e^{-rt - d\log(\frac{c\sqrt{\log n}t}{\sqrt{d}}) - d} \cdot e^{(dc^2 \log n)/2} \\ &\leq e^{-rt/2 - d\log(\frac{c\sqrt{\log n}t}{\sqrt{d}}) - d} = e^{-d(\frac{rt}{2d} - \log(\frac{rt}{d}) - 1)} \\ &\leq e^{-d(\frac{rt}{4d})} = e^{-rt/4}, \end{aligned} \tag{3.20}$$

where the last inequality follows from $x/2 - \log(x) - 1 \geq x/4$, for $x \geq rt/d \geq c^2 \geq 16$.

Proof of (3.9). For $t \geq r$, using the bound (3.8), we get

$$\Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] \leq \Pr[\|\mathbf{X}\| \geq t] \leq e^{-c\sqrt{d \log n}t/4} \ . \tag{3.21}$$

For $t \leq r$, we see that

$$\begin{aligned} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t] &\leq \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t, \|\mathbf{X}\| \leq r] + \Pr[\|\mathbf{X}\| \geq r] \\ &\leq \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t, \|\mathbf{X}\| \leq r] + e^{-r^2/4} \ . \end{aligned} \tag{3.22}$$

By the identity (3.16), for the first term, using the Gaussian tail bound (2.2), we have that

$$\begin{aligned} \Pr[|\mathbf{X}^\mathsf{T}\boldsymbol{\theta}| \geq t, \|\mathbf{X}\| \leq r] &= \Pr[|\mathbf{Z}^\mathsf{T}\boldsymbol{\theta}| \geq t, \|\mathbf{Z}\| \leq r] \cdot \frac{\Pr[\|\mathbf{X}\| \leq r]}{\Pr[\|\mathbf{Z}\| \leq r]} \\ &= \Pr[|\mathbf{Z}^\mathsf{T}\boldsymbol{\theta}| \geq t, \|\mathbf{Z}\| \leq r] \cdot \frac{\int_{\mathbb{R}^n} e^{-\|\mathbf{x}\|^2/2} \, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} f_{(1,r)}(\mathbf{x}) \, \mathrm{d}\mathbf{x}} \\ &\leq \Pr[|\mathbf{Z}^\mathsf{T}\boldsymbol{\theta}| \geq t] \leq 2e^{-t^2/2} \ . \end{aligned} \tag{3.23}$$

The desired inequality (3.9) now follows directly by combining (3.21), (3.22), (3.23), noting that $2e^{-t^2/2} + e^{-r^2/4} \leq 3e^{-t^2/4}$ for $0 \leq t \leq r$. ∎

4. Simplex algorithms. In this section, we describe how to use the shadow bound to bound the complexity of a complete shadow vertex simplex algorithm. We restrict our attention here to Gaussian perturbations, as the details for Laplace perturbations are similar. We will follow the two-stage interpolation strategy given by Vershynin in [86]. The Random Vertex algorithm of [86] was shown in the same paper to work for any $\sigma \leq \min(\frac{c_1}{\sqrt{d \log n}}, \frac{c_1}{d^{3/2} \log d})$ for some $c_1 > 0$. The constraint on σ is always achievable by scaling down the matrix $\mathbf{A}$, though it will be reflected in the running time of the algorithm.

We will describe a modification of the RV algorithm that further relaxes the condition on the perturbation size to $\sigma \leq O(\frac{1}{\sqrt{d \log n}})$, for an expected $O(d^2\sqrt{\log n}\, \sigma^{-2} + d^3 \log^{1.5} n)$ pivot steps. Hence our algorithm is faster than both Vershynin's [86] RV algorithm and Borgwardt's dimension-by-dimension algorithm [18] for our shadow bound.

To recall, our goal is to solve the smoothed LP

$$\begin{aligned} \max \ & \mathbf{c}^\mathsf{T}\mathbf{x} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned} \qquad \text{(Smooth LP)}$$

where $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ and $n \geq d \geq 3$. Here each row $(\mathbf{a}_i, b_i)$, $i \in [n]$, of $(\mathbf{A}, \mathbf{b})$ is a variance σ^2 Gaussian random vector with mean $(\bar{\mathbf{a}}_i, \bar{b}_i) := \mathbb{E}[(\mathbf{a}_i, b_i)]$

of ℓ_2 norm at most 1. We will say that (Smooth LP) is unbounded (bounded) if the system $\mathbf{c}^\mathsf{T}\mathbf{x} > 0, \mathbf{A}\mathbf{x} \leq \mathbf{0}$ is feasible (infeasible). Note that (Smooth LP) can be unbounded and infeasible under this definition. If (Smooth LP) is bounded and feasible, then it has an optimal solution.

For the execution of the algorithms as stated, we assume the non-degeneracy conditions listed in Theorem 2.12. That is, we assume both the feasible polyhedron and shadows to be non-degenerate. These conditions hold with probability 1.

THEOREM 4.1. *(Smooth LP) can be solved by a two-phase shadow simplex method using an expected number of pivots of* $O(d^2\sqrt{\log n}\ \sigma^{-2} + d^3 \log^{1.5} n)$.

Proof. Combining Lemma 4.2 and Theorem 4.11, the expected number of simplex pivots is bounded by

$$10 + \mathcal{D}_g(d+1, n, \sigma/2) + 5\mathcal{D}_g(d, n+2d-2, \min\{\sigma, \bar{\sigma}\}/5) \ ,$$

where $\bar{\sigma}$ is as defined in (4.1). Noting that $1/\bar{\sigma} = O(\sqrt{d \log n})$, by the smoothed Gaussian shadow bound (Theorem 3.1), the above is bounded by

$$O(\mathcal{D}_g(d, n, \sigma) + \mathcal{D}_g(d, n, (\sqrt{d \log n})^{-1})) = O(d^2\sqrt{\log n}\sigma^{-2} + d^3 \log^{1.5} n) \ ,$$

as needed. ∎

Two-Phase Interpolation Method. Define the Phase I Unit LP:

$$\begin{aligned} \max\ & \mathbf{c}^\mathsf{T}\mathbf{x} && \text{(Unit LP)} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{1} \end{aligned}$$

and the Phase II interpolation LP with parametric objective for $\theta \in (-\pi/2, \pi/2)$:

$$\begin{aligned} \max\ & \cos(\theta)\mathbf{c}^\mathsf{T}\mathbf{x} + \sin(\theta)\lambda && \text{(Int. LP)} \\ & \mathbf{A}\mathbf{x} + (\mathbf{1} - \mathbf{b})\lambda \leq \mathbf{1} \\ & 0 \leq \lambda \leq 1. \end{aligned}$$

The above form of interpolation was first introduced in the context of smoothed analysis by Vershynin [86].

Let us assume for the moment that (Smooth LP) is bounded and feasible (i.e., has an optimal solution). Since boundedness is a property of $\mathbf{A}$ and not $\mathbf{b}$, note that this implies that (Unit LP) is also bounded (and clearly always feasible).

To understand the Phase II interpolation LP, the key observation is that for θ sufficiently close to $-\pi/2$, the maximizer will be the optimal solution to (Unit LP), i.e., will satisfy $\lambda = 0$, and for θ sufficiently close to $\pi/2$ the maximizer will be the optimal solution to (Smooth LP), i.e., will satisfy $\lambda = 1$. Thus given an optimal solution to the Phase I unit LP one can initialize a run of shadow vertex starting at θ just above $-\pi/2$, moving towards $\pi/2$ until the optimal solution to (Smooth LP) is found. The corresponding shadow plane is generated by $(\mathbf{c}, 0)$ and $(\mathbf{0}, 1)$ (associating λ with the last coordinate), and as usual the size of the shadow bounds the number of pivots.

If (Smooth LP) is unbounded (i.e., the system $\mathbf{c}^\mathsf{T}\mathbf{x} > 0, \mathbf{A}\mathbf{x} \leq \mathbf{0}$ is feasible), this will be detected during Phase I as (Unit LP) is also unbounded. If (Smooth LP) is infeasible but bounded, then the shadow vertex run will terminate at a vertex having $\lambda < 1$. Thus, all cases can be detected by the two-phase procedure (see [86, Proposition 4.1] for a formal proof).

We bound the number of pivot steps taken to solve (Int. LP) given a solution to (Unit LP), and after that we describe how to solve (Unit LP).

Consider polyhedron $P' = \left\{(\mathbf{x}, \lambda) \in \mathbb{R}^{d+1} : \mathbf{A}\mathbf{x} + (\mathbf{1} - \mathbf{b})\lambda \leq \mathbf{1}\right\}$, the slab $H = \left\{(\mathbf{x}, \lambda) \in \mathbb{R}^{d+1} : 0 \leq \lambda \leq 1\right\}$ and let $W = \text{span}(\mathbf{c}, \mathbf{e}_\lambda)$. In this notation, $P' \cap H$ is the feasible set of (Int. LP) and W is the shadow plane of (Int. LP). We bound the number of vertices in the shadow $\pi_W(P' \cap H)$ of (Int. LP) by relating it to $\pi_W(P')$.

The constraints defining P' are of smoothed unit type. Namely, the rows of $(\mathbf{A}, \mathbf{1} - \mathbf{b})$ are variance σ^2 Gaussians centered at means of norm at most 2. We derive this from the triangle inequality. Thus, we know $\pi_W(P')$ has at most $\mathcal{D}_g(d+1, n, \sigma/2)$ expected vertices. We divide σ by 2 because the centers have norm at most 2.

Since the shadow plane contains the normal vector $(\mathbf{0}, 1)$ to the inequalities $0 \leq \lambda \leq 1$, these constraints intersect the shadow plane W at right angles. It follows that $\pi_W(P' \cap H) = \pi_W(P') \cap H$. Adding 2 constraints to a 2D polyhedron can add at most 2 new edges, hence the constraints on λ can add at most 4 new vertices. By combining these observations, we directly derive the following lemma of Vershynin [86].

LEMMA 4.2. *If (Unit LP) is unbounded, then (Smooth LP) is unbounded. If (Unit LP) is bounded, then given an optimal solution to (Unit LP) one can solve (Smooth LP) using at most an expected* $\mathcal{D}_g(d+1, n, \sigma/2)+4$ *shadow vertex pivots over (Int. LP).*

Given the above, our main task is now to solve (Unit LP), i.e., either to find an optimal solution or to determine unboundedness. The simplest algorithm is Borgwardt's dimension-by-dimension (DD) algorithm, which was first used in the context of smoothed analysis by Schnalzger [78]. Due to its simplicity, we describe it briefly below as a warm-up.

DD algorithm. As outlined in the introduction, the DD algorithm solves Unit LP by iteratively solving the restrictions:

$$\begin{aligned} \max\ & \mathbf{c}_k^\mathsf{T}\mathbf{x} && \text{(Unit LP}_k\text{)} \\ & \mathbf{A}\mathbf{x} \leq \mathbf{1} \\ & x_i = 0,\ \forall i \in \{k+1, \ldots, d\}, \end{aligned}$$

where $k \in \{1, \ldots, d\}$ and $\mathbf{c}_k := (c_1, \ldots, c_k, 0, \ldots, 0)$. The main idea here is that the solution of (Unit LP$_k$), $k \in \{1, \ldots, d-1\}$, is generically on an edge of the shadow of (Unit LP$_{k+1}$) on the span of $\mathbf{c}_k$ and $\mathbf{e}_{k+1}$, which is sufficient to initialize the shadow simplex path in the next step. We note that Borgwardt's algorithm can be applied to any LP with a known feasible point as long as appropriate non-degeneracy conditions hold (which occur with probability 1 for smoothed LPs). To avoid degeneracy, we will assume that $\mathbf{c}_k \neq \mathbf{0}$ for all $k \in \{1, \ldots, d\}$, which can always be achieved by permuting the coordinates. Note that (Unit LP$_1$) can be trivially solved, as the feasible region is an interval whose endpoints are easy to compute.

THEOREM 4.3 ([17]). *Let W_k, $k \in \{2, \ldots, d\}$, denote the shadow of (Unit LP$_k$) on the span of $\mathbf{c}_{k-1}$ and $\mathbf{e}_k$. Then, if each (Unit LP$_k$) and shadow W_k is non-degenerate, for $k \in \{2, \ldots, d\}$, there exists a shadow simplex method which solves (Unit LP) using at most $\sum_{k=2}^{d} |\text{vertices}(W_k)|$ number of pivots.*

The shadow bounds of Theorem 3.1 only hold for $d \geq 3$, though sufficiently good bounds have been proven by [78, 32] to derive the following immediate corollary.

COROLLARY 4.4. *The smoothed (Unit LP) can be solved by the DD algorithm using an expected* $\sum_{k=2}^{d} \mathcal{D}_g(k, n, \sigma) = O(d^3\sqrt{\log n}\ \sigma^{-2} + d^{3.5}\sigma^{-1}\log n + d^{3.5}\log^{3/2} n)$ *number of shadow vertex pivots.*

Random vertex method. Vershynin's approach for initializing the shadow simplex method on (Unit LP) is to add a random smoothed system of d linear constraints to its description. These constraints are meant to induce a *known random* vertex $\mathbf{v}$ and corresponding maximizing objective $\mathbf{d}$ which are effectively uncorrelated with the original system. Starting at this vertex $\mathbf{v}$, we then follow the shadow path induced by rotating $\mathbf{d}$ towards $\mathbf{c}$. The main difficulty with this approach is to guarantee that the randomly generated system (i) adds a vertex which (ii) is optimized at $\mathbf{d}$ and (iii) does not cut off the optimal solution or all unbounded rays. Fortunately, each of these conditions is easily checkable, and hence if they fail (which will occur with constant probability), the process can be attempted again.

One restriction imposed by this approach is that the perturbation size needs to be rather small, namely

$$\sigma \leq \sigma_1 := \frac{c_1}{\max\left\{\sqrt{d \log n}, d^{1.5} \log d\right\}}$$

in [86] for some $c_1 > 0$. A more careful analysis of Vershynin's algorithm can relax the restriction to

$$\sigma \leq \sigma_2 := \frac{c_2}{\max\left\{\sqrt{d \log n}, \sqrt{d} \log d\right\}}$$

for some $c_2 > 0$. This restriction is necessary due to the fact that we wish to predict the effect of smoothing the system, in particular, the smoothing operation should not negate (i), (ii), or (iii). Recall that one can always artificially decrease σ by scaling down the matrix $\mathbf{A}$ as this does not change the structure of (Unit LP). The assumption on σ is thus without loss of generality. When stating running time bounds however, this restriction will be reflected by a larger additive term that does not depend on σ.

We adapt the Random Vertex algorithm to make (ii) guaranteed to hold, allowing us to relax the constraint on the perturbation size to

$$\sigma \leq \bar{\sigma} := \frac{1}{36\sqrt{d \log n}}. \tag{4.1}$$

Instead of adding d constraints, each with their own perturbation, we add $d-1$ pairs of constraints with mirrored perturbations. This forces the desired objective to be maximized at the random vertex whenever this vertex exists.

We begin with some preliminary remarks for Algorithm 4.1. First, the goal of defining $\mathbf{V}$ is to create a new artificial LP, (Unit LP') $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{A}\mathbf{x} \leq \mathbf{1}, \mathbf{V}\mathbf{x} \leq \mathbf{1}$, such that $\mathbf{x_0}$ is a vertex of the corresponding system which maximizes $\mathbf{d}$. On line 9 and 10, the algorithm checks if $\mathbf{x}_0$ is feasible and whether it is not the optimizer of $\mathbf{c}$ on (Unit LP'). Having passed these checks, (Unit LP') is solved via shadow vertex initialized at vertex $\mathbf{x}_0$ with objective $\mathbf{d}$. An unbounded solution to (Unit LP') is always an unbounded solution to (Unit LP). Lastly, it is checked on line 13 whether the bounded solution (if it exists) to (Unit LP') is a solution to (Unit LP). Correctness of the algorithm's output is thus straightforward. We do have to make sure that every step of the algorithm can be executed as described.

LEMMA 4.5. *In (Unit LP') as defined on lines 3-11 of Algorithm 4.1, with probability* 1*,* $\mathbf{x}_0$ *is well-defined, and, when entering the shadow simplex routine, the point* $\mathbf{x}_0$ *is a shadow vertex and the edge defined by* $\mathbf{B}_0$ *is a shadow edge on (Unit LP'). Moreover,* $\mathbf{x}_0$ *is the only degenerate vertex.*

Algorithm 4.1 Symmetric Random Vertex algorithm

Require: $\mathbf{c} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{A}$ is standard deviation $\sigma \leq \bar{\sigma}$ Gaussian with rows having centers of norm at most 1.

Ensure: Decide whether (Unit LP) $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{A}\mathbf{x} \leq \mathbf{1}$ is unbounded or return an optimal solution.

1: If some row of $\mathbf{A}$ has norm greater than 2, solve $\max \mathbf{c}^\mathsf{T}\mathbf{x}$,st.$\mathbf{A}\mathbf{x} \leq \mathbf{1}$ using any simplex method that takes at most $\binom{n}{d}$ pivot steps.
2: **loop**
3: Let $l = 1/6\sqrt{\log d}$.
4: Sample a rotation matrix $\mathbf{R} \in O(d)$ uniformly at random.
5: Sample $\mathbf{g}_1, \ldots, \mathbf{g}_{d-1} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ independently.
6: Set $\mathbf{v}_i^+ = \mathbf{R}(4\mathbf{e}_d + l\mathbf{e}_i + \mathbf{g}_i)$, $\mathbf{v}_i^- = \mathbf{R}(4\mathbf{e}_d - l\mathbf{e}_i - \mathbf{g}_i)$ for all $i \in [d-1]$.
7: Put $\mathbf{V} = (\mathbf{v}_1^+, \mathbf{v}_1^-, \mathbf{v}_2^+, \ldots, \mathbf{v}_{d-1}^+, \mathbf{v}_{d-1}^-)^\mathsf{T}$, $\mathbf{d} = \mathbf{R}\mathbf{e}_d$.
8: Find $\mathbf{x}_0$ such that $\mathbf{V}\mathbf{x}_0 = \mathbf{1}$.
9: If not $\mathbf{A}\mathbf{x}_0 < \mathbf{1}$, restart the loop.
10: Solve $\sum_{i=1}^{d-1} \lambda_i \mathbf{R}(l\mathbf{e}_i + \mathbf{g}_i) = \mathbf{c} + \lambda_d \mathbf{d}$. If $\lambda_d + \sum_{i=1}^{d-1} 4|\lambda_i| \leq 0$, restart the loop. (This corresponds to $\mathbf{x}_0$ being optimal for $\mathbf{c}$.)
11: Follow the shadow path from $\mathbf{d}$ to $\mathbf{c}$ on

$$\text{(Unit LP')} \qquad \begin{aligned} &\max \mathbf{c}^\mathsf{T}\mathbf{x} \\ &\mathbf{A}\mathbf{x} \leq \mathbf{1} \\ &\mathbf{V}\mathbf{x} \leq \mathbf{1}, \end{aligned}$$

starting from the vertex $\mathbf{x}_0$. For the first pivot, follow the edge which is tight at the constraints in $\mathbf{B}_0 = (\mathbf{v}_1^{\text{sign}(\lambda_1)}, \ldots, \mathbf{v}_{d-1}^{\text{sign}(\lambda_{d-1})})$. All other pivot steps are as in Algorithm 2.1.
12: If (Unit LP') is unbounded, return "unbounded".
13: If (Unit LP') is bounded and the optimal vertex $\mathbf{x}^*$ satisfies $\mathbf{V}\mathbf{x}^* < \mathbf{1}$, return $\mathbf{x}^*$ as the optimal solution to (Unit LP).
14: Otherwise, restart the loop.
15: **end loop**

Proof. Without loss of generality, we assume $\mathbf{R} = \mathbf{I}_{d\times d}$. With probability 1, the coefficients $\lambda_1, \ldots, \lambda_d$ exist and are uniquely defined.

We now show that $\mathbf{x}_0$ is well-defined. Let $\mathbf{x}_0^+$ be the solution to the following system of d equalities

$$\mathbf{v}_1^{+\mathsf{T}}\mathbf{x}_0^+ = 1, \mathbf{v}_2^{+\mathsf{T}}\mathbf{x}_0^+ = 1, \ldots, \mathbf{v}_{d-1}^{+\mathsf{T}}\mathbf{x}_0^+ = 1, \qquad 8\mathbf{e}_d^\mathsf{T}\mathbf{x}_0^+ = 2. \tag{4.2}$$

This system of equations almost surely has a single solution. We claim that $\mathbf{V}\mathbf{x}_0^+ = \mathbf{1}$. By writing $\mathbf{v}_i^- = 8\mathbf{e}_d - \mathbf{v}_i^+$, we find that $\mathbf{v}_i^{-\mathsf{T}}\mathbf{x}_0^+ = 1$ for all $i \in [d-1]$. Therefore, $\mathbf{x}_0 = \mathbf{x}_0^+$ is indeed well-defined.

By definition, upon entering the shadow simplex routine, $\mathbf{x}_0$ satisfies $\mathbf{A}\mathbf{x} < \mathbf{1}$, $\mathbf{V}\mathbf{x} \leq \mathbf{1}$ and is thus a vertex.

For all $t > 0$, define $\mathbf{x}_t$ to be the solution to $8\mathbf{e}_d^\mathsf{T}\mathbf{x}_t = 2 - t, \mathbf{B}_0\mathbf{x}_t = 1$. For any $\mathbf{v}_i^s \notin \mathbf{B}_0$, we have $\mathbf{v}_i^{s\mathsf{T}}\mathbf{x}_t = 1 - t < 1$. As the $\mathbf{x}_t$ lie on a line and $\mathbf{A}\mathbf{x}_0 < \mathbf{1}$, there exists some $\varepsilon > 0$ such that $\mathbf{x}_t$ is feasible for all $t \leq \varepsilon$. Hence the constraints in $\mathbf{B}_0$ define an edge of the feasible set.

The point $\mathbf{x}_0$ is tight at the inequalities $\mathbf{V}\mathbf{x} \leq \mathbf{1}$, and $\frac{1}{8d-8}\sum_{i=1}^{d-1}(\mathbf{v}_i^+ + \mathbf{v}_i^-) = \mathbf{d}$ is a corresponding dual solution, so we know that $\mathbf{x}_0$ is optimal for objective $\mathbf{d}$ and thus a shadow vertex.

Assume that $\mathbf{x}_0$ is not optimal for objective $\mathbf{c}$. One outgoing edge of $\mathbf{x}_0$ is tight at the inequalities $\mathbf{v}_i^{s\mathsf{T}}\mathbf{x} \leq 1$ for all $\mathbf{v}_i^s \in \mathbf{B}_0$ and that edge is on the shadow path exactly if the cone spanned by $\mathbf{B}_0$ intersects $\mathrm{cone}(\mathbf{c}, \mathbf{d})$ outside $\{\mathbf{0}\}$. This intersection is exactly the ray spanned by

$$\begin{aligned}\sum_{i=1}^{d-1}|\lambda_i|\mathbf{v}_i^{\mathrm{sign}(\lambda_i)} &= \sum_{i=1}^{d-1}\lambda_i(l\mathbf{e}_i + \mathbf{g}_i) + 4|\lambda_i|\mathbf{d} \\ &= \mathbf{c} + \lambda_d\mathbf{d} + \sum_{i=1}^{d-1}4|\lambda_i|\mathbf{d},\end{aligned}$$

and we know that $\lambda_d + \sum_{i=1}^{d-1}4|\lambda_i| > 0$ as otherwise we would have a certificate that $\mathbf{c} \in \mathrm{cone}(\mathbf{V}\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^d)$. We conclude that $\sum_{i=1}^{d-1}|\lambda_i|\mathbf{v}_i^{\mathrm{sign}(\lambda_i)}$ is a non-negative linear combination of $\mathbf{c}, \mathbf{d}$ and hence our description of the first shadow vertex pivot step is correct.

Lastly, we show that any vertex other than $\mathbf{x}_0$ is tight at exactly d independently distributed constraint vectors. Fix any basis B such that there exists an $i \in [d-1]$ with $\mathbf{v}_i^+, \mathbf{v}_i^- \in B$ and which does not define the vertex $\mathbf{x}_0$. Let $\mathbf{x}_B$ be such that $\mathbf{a}^\mathsf{T}\mathbf{x}_B = 1$ for all $\mathbf{a} \in B$. There exists some $j \in [d-1]$ such that both $\mathbf{v}_j^+, \mathbf{v}_j^- \notin B$, for otherwise we would have $\mathbf{x}_B = \mathbf{x}_0$. We show that, almost surely, $\mathbf{v}_j^{+\mathsf{T}}\mathbf{x}_B > 1$ or $\mathbf{v}_j^{-\mathsf{T}}\mathbf{x}_B > 1$, which implies that $\mathbf{x}_B$ is almost surely not feasible. We know that $\mathbf{v}_i^{+\mathsf{T}}\mathbf{x}_B = \mathbf{v}_i^{-\mathsf{T}}\mathbf{x}_B = 1$, and hence $4\mathbf{d}^\mathsf{T}\mathbf{x}_B = 1$. It follows that $\mathbf{v}_j^{+\mathsf{T}}\mathbf{x}_B = 2 - \mathbf{v}_j^{-\mathsf{T}}\mathbf{x}_B$, The only way to have both $\mathbf{v}_j^{+\mathsf{T}}\mathbf{x}_B \leq 1$ and $\mathbf{v}_j^{-\mathsf{T}}\mathbf{x}_B \leq 1$ would be if $\mathbf{v}_j^{+\mathsf{T}}\mathbf{x}_B = 1$. However, $\mathbf{x}_B$ and $\mathbf{v}_j^+$ are independently distributed and $\mathbf{v}_j^+$ has a continuous probability distribution, so $\mathbf{x}_B$ is a vertex with probability 0. ∎

To bound the expected running time of Algorithm 4.1, we bound the expected number of pivot steps per iteration of the loop, and the expected number of iterations of the loop.

First, we bound the expected shadow size in a single iteration. Because the constraint vectors $\mathbf{v}_i^+, \mathbf{v}_i^-$ are not independently distributed for any $i \in [d-1]$, we are unable to apply Theorem 3.1 in a completely black-box way. As we show below, in this new setting, the proof of Theorem 3.1 still goes through essentially without modification.

In the rest of this section, we abbreviate

$$\mathrm{conv}(\mathbf{A}, \mathbf{V}) := \mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{v}_1^+, \ldots, \mathbf{v}_{d-1}^+, \mathbf{v}_1^-, \ldots, \mathbf{v}_{d-1}^-).$$

LEMMA 4.6. *Let* $\mathbf{A}$ *have independent standard deviation* σ *Gaussian rows with centers of norm at most* 1 *and let* $\mathbf{V}$ *be sampled, independently from* $\mathbf{A}$*, as in lines 4-7 of Algorithm 4.1 with* $l \leq 1$*. The shadow size* $\mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{A}, \mathbf{V}) \cap \mathrm{span}(\mathbf{c}, \mathbf{d}))|]$ *is bounded by* $\mathcal{D}_g(d, n + 2d - 2, \min(\sigma, \bar{\sigma})/5) + 1$*.*

Proof. We fix the choice of $\mathbf{R}$. The distribution of constraint vectors is now independent of the two-dimensional plane.

$$\mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{A}, \mathbf{V}) \cap \mathrm{span}(\mathbf{c}, \mathbf{d}))|] \leq \max_{\mathbf{R}} \mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{A}, \mathbf{V}) \cap \mathrm{span}(\mathbf{c}, \mathbf{R}\mathbf{e}_d))| \mid \mathbf{R}].$$

The rows of $\mathbf{A}$ have centers of norm at most 1 and the rows of $\mathbf{V}$ have centers of norm at most $4 + l \leq 5$. After an appropriate rescaling, we can assume all $n + 2d - 2$ constraints have expectations of norm at most 1 and standard deviation $\sigma \leq \bar{\sigma}/5$.

To get the desired bound, we bound the number of edges other than the one induced by $\mathbf{x}_0$, $W \cap \{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}^\mathsf{T}\mathbf{x}_0 = 1\}$, which yields the $+1$ in the final bound. The proof is essentially identical to that of Theorem 3.1, i.e. we bound the ratio of the expected perimeter divided by the minimum expected edge of the polar polygon. We sketch the key points below. Firstly, notice that the perimeter bound in Lemma 3.14 does not require independence of the perturbations, so it still holds. For the minimum edge length, we restrict to the bases B as in Lemma 3.13 (which also does not require independence) after removing those which induce $\mathbf{x}_0$ as a vertex (it has already been counted). By Lemma 4.5, the remaining bases in B contain at most one of each pair $\{\mathbf{v}_i^-, \mathbf{v}_i^+\}$, $i \in [d-1]$, since bases containing two such vectors correspond to an edge different from the one induced by $\mathbf{x}_0$ with probability 0. In particular, every basis in B consists of only independent random vectors.

From here, the only remaining detail for the bound to go through is to to check that the conclusion of Lemma 3.20 still holds, i.e., that the position of vectors within their containing hyperplane does not affect the probability that these vectors form a facet of the convex hull. Without loss of generality, we consider the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_i$, $\mathbf{v}_1^+, \ldots, \mathbf{v}_j^+$ with $i + j = d$. Define $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0$ by $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_k = t$ for all $k \in [i]$, $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^+ = t$ for all $k \in [j]$. The set $\operatorname{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{v}_1^+, \ldots, \mathbf{v}_j^+)$ is a facet of the convex hull of the constraint vectors when either (1) $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_k < t$ for all $k > i$, $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^- < t$ for all $k \in [j]$ and $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^\pm < t$ for all $k > j$ or (2) when $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}_k > t$ for all $k > i$, $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^- > t$ for all $k \in [j]$ and $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^\pm > t$ for all $k > j$. The only one of these properties that is not independent of $\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{v}_1^+, \ldots, \mathbf{v}_j^+$ is whether $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^- < t$ or $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^- > t$ for $k \in [j]$, but we know that $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^- = 8\boldsymbol{\theta}^\mathsf{T}\mathbf{d} - \boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^+ = 8\boldsymbol{\theta}^\mathsf{T}\mathbf{d} - t$ for all $k \in [j]$, and so the value $\boldsymbol{\theta}^\mathsf{T}\mathbf{v}_k^-$ does not depend on the positions of $\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{v}_1^+, \ldots, \mathbf{v}_j^+$ within their containing hyperplane. We conclude that the expected number of edges is bounded by $\mathcal{D}_g(d, n + 2d - 2, \min(\sigma, \bar{\sigma})/5) + 1$. ∎

All that is left, is to show that the success probability of each loop is lower bounded by a constant.

DEFINITION 4.7. *For a matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$, we define its operator norm by*

$$\|\mathbf{M}\| = \max_{\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}} \frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|}$$

and its maximum and minimum singular values by

$$s_{\max}(\mathbf{M}) = \|\mathbf{M}\|, \qquad s_{\min}(\mathbf{M}) = \min_{\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}} \frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Using the Gaussian tailbound (2.1) together with a 1/2-net on the sphere (which has size at most 8^d, see e.g., [64], page 314), we immediately obtain the following tail bound for the operator norm of random Gaussian matrices.

LEMMA 4.8. *For a random $d \times d$ matrix G with independent standard normal entries, one has*

$$\Pr[\|\mathbf{G}\| > 2t\sqrt{d}] \leq 8^d e^{-d(t-1)^2/2}.$$

LEMMA 4.9. *Let $\mathbf{A} \in \mathbb{R}^{n \times d}$ have rows of norm at most 2 and $\sigma \leq \frac{l}{6\sqrt{d}}$. For $\mathbf{x}_0$ sampled as in lines 4-8 of Algorithm 4.1, with probability at least 0.98, the point $\mathbf{x}_0$ satisfies $\mathbf{A}\mathbf{x}_0 < \mathbf{1}$.*

Proof. Without loss of generality, we assume $\mathbf{R} = \mathbf{I}_{d\times d}$. We claim that, with sufficient probability, $\|\mathbf{x}_0 - \mathbf{e}_d/4\| < 1/4$. Together with the triangle inequality and the assumption that $\|\mathbf{a}_i\| \le 2$ for all $i \in [n]$, this suffices to show $\mathbf{A}\mathbf{x}_0 < \mathbf{1}$.

Elementary calculations show that $\mathbf{x}_0 - \mathbf{e}_d/4$ satisfies $\mathbf{e}_d{}^\mathsf{T}(\mathbf{x}_0 - \mathbf{e}_d/4) = 0$ and, for every $i \in [d-1]$, $(l\mathbf{e}_i + \mathbf{g}_i)^\mathsf{T}(\mathbf{x}_0 - \mathbf{e}_d/4) = -\mathbf{g}_i{}^\mathsf{T}\mathbf{e}_d/4$. Let $\mathbf{G}$ be the matrix with rows consisting of the first $d-1$ entries of each of $\mathbf{g}_1, \dots, \mathbf{g}_{d-1}$, and $\mathbf{g}$ be the vector consisting of the d'th entries of $\mathbf{g}_1, \dots, \mathbf{g}_{d-1}$. From the above equalites we derive

$$\begin{pmatrix} l\mathbf{I}_{d-1} + \mathbf{G} & \mathbf{g} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} (\mathbf{x}_0 - \mathbf{e}_d/4) = \frac{1}{4}\begin{pmatrix} -\mathbf{g} \\ 0 \end{pmatrix}$$
$$\begin{pmatrix} l\mathbf{I}_{d-1} + \mathbf{G} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} (\mathbf{x}_0 - \mathbf{e}_d/4) = \frac{1}{4}\begin{pmatrix} -\mathbf{g} \\ 0 \end{pmatrix}$$
$$\mathbf{x}_0 - \mathbf{e}_d/4 = \frac{1}{4}\begin{pmatrix} -\left(l\mathbf{I}_{d-1} + \mathbf{G}\right)^{-1}\mathbf{g} \\ 0 \end{pmatrix}.$$

Note that the matrix is almost surely invertible. We abbreviate $\mathbf{M} = l\mathbf{I}_{d-1} + \mathbf{G}$ and bound $\|\mathbf{x}_0 - \mathbf{e}_d/4\| \le \|\mathbf{M}^{-1}\|\|\mathbf{g}\|/4$. Using that $\sigma \le \frac{l}{6\sqrt{d}}$, we apply (2.1) to get $\|\mathbf{g}\| \le l/2$ with probability at least 0.99.

The operator norm of the inverse matrix satisfies $\|\mathbf{M}^{-1}\| = \frac{1}{s_{\min}(l\mathbf{I}+\mathbf{G})}$, and by the triangle inequality we derive

$$s_{\min}(l\mathbf{I} + \mathbf{G}) \ge s_{\min}(l\mathbf{I}) - s_{\max}(\mathbf{G}) = l - s_{\max}(\mathbf{G}).$$

By Lemma 4.8, we have $\|\mathbf{G}\| \le 3\sqrt{d}\sigma \le l/2$ with probability at least 0.99. Putting the pieces together, we conclude that

$$\frac{1}{4}\|\mathbf{M}^{-1}\|\|\mathbf{g}\| \le \frac{1}{4}\cdot\frac{1}{l - l/2}\cdot\frac{l}{2} \le 1/4.$$

We take the union bound over the two bad events and thus conclude that $\mathbf{a}_i{}^\mathsf{T}\mathbf{x}_0 \le \|\mathbf{a}_i\|\|\mathbf{x}_0\| < 1$ for all $i \in [n]$ with probability at least 0.98. ∎

Lastly, we need to prove that the conditionals on lines 10, 12 and 13 of Algorithm 4.1 succeeds with sufficient probability.

LEMMA 4.10 (Adapted from [86]). *Let $l \le 1/6\sqrt{\log d}$ and $\sigma \le 1/8\sqrt{d\log d}$. For fixed $\mathbf{A}$ and $\mathbf{V}$ sampled as in lines 4-7 of Algorithm 4.1, let $\mathbf{x}^*$ be the optimal solution to (Unit LP') if it exists. With probability at least 0.24, (Unit LP) being unbounded implies that (Unit LP') is unbounded and (Unit LP) being bounded implies $\mathbf{V}\mathbf{x}^* < \mathbf{1}$.*

Proof. Let $\mathbf{x}$ be the maximizer of (Unit LP) if it exists, or otherwise a generator for an unbounded ray in (Unit LP), and let $\boldsymbol{\omega} = \mathbf{x}/\|\mathbf{x}\|$. We aim to prove that $\mathbf{V}\boldsymbol{\omega} < \mathbf{0}$ with probability at least 0.24 over the randomness in $\mathbf{V}$, which is sufficient for the lemma to hold.

We fix $\mathbf{A}$, and hence $\boldsymbol{\omega}$ as well. We decompose

$$\mathbf{v}_i^+{}^\mathsf{T}\boldsymbol{\omega} = 4\mathbf{d}^\mathsf{T}\boldsymbol{\omega} + (l\mathbf{R}\mathbf{e}_i)^\mathsf{T}\boldsymbol{\omega} + (\mathbf{R}\mathbf{g}_i)^\mathsf{T}\boldsymbol{\omega}, \tag{4.3}$$

for all $i \in [d-1]$ and similarly for $\mathbf{v}_i^-$, and we will bound the different terms separately.

The inner product $\mathbf{d}^\mathsf{T}\boldsymbol{\omega}$ has probability density proportional to $\sqrt{1-t^2}^{d-3}$, as it is the one-dimensional marginal distribution over the sphere $\mathbb{S}^{d-1}$ (see e.g., [37],

equation 1.26). which can differ over the interval $[-\sqrt{\frac{2}{d-1}}, \sqrt{\frac{2}{d-1}}]$ by at most a factor $1/e$. We lower bound the probability that $\mathbf{d}^\mathsf{T}\boldsymbol{\omega}$ is far from being positive:

$$\begin{aligned}\Pr[\mathbf{d}^\mathsf{T}\boldsymbol{\omega} < -\frac{1}{4}\sqrt{\frac{2}{d-1}}] &= \frac{1}{2}\Pr[\mathbf{d}^\mathsf{T}\boldsymbol{\omega} < -\frac{1}{4}\sqrt{\frac{2}{d-1}} \mid \mathbf{d}^\mathsf{T}\boldsymbol{\omega} \leq 0] \\ &\geq \frac{1}{2}\Pr[\mathbf{d}^\mathsf{T}\boldsymbol{\omega} < -\frac{1}{4}\sqrt{\frac{2}{d-1}} \mid \mathbf{d}^\mathsf{T}\boldsymbol{\omega} \in [-\sqrt{\frac{2}{d-1}}, 0]] \\ &\geq \frac{1}{2} \cdot \frac{\frac{3}{4e}}{\frac{3}{4e} + \frac{1}{4}} \\ &\geq 0.26.\end{aligned}$$

Hence, for $\mathbf{d}$ a randomly chosen unit vector independent of ω, we have $4\mathbf{d}^\mathsf{T}\boldsymbol{\omega} < -\sqrt{\frac{2}{d-1}}$ with probability at least 0.26. Now we will give an upper bound on the second and third terms in (4.3) with sufficient probability.

By the same measure concentration argument as in the proof of (2.5) we know that $\Pr[|\mathbf{e}_i^\mathsf{T}\mathbf{R}^\mathsf{T}\boldsymbol{\omega}| > t/\sqrt{d-1}] \leq e^{-t^2/2}$. We apply the above statement with $t = 3\sqrt{\log d}$ and find that

$$|l\mathbf{e}_i^\mathsf{T}\mathbf{R}^\mathsf{T}\boldsymbol{\omega}| < tl/\sqrt{d-1} \leq 1/2\sqrt{d-1}$$

with probability at least $1 - \frac{0.01}{d}$.

For the last part, fix $\mathbf{R} = \mathbf{I}$ without loss of generality. The inner product ${\mathbf{g}_i}^\mathsf{T}\boldsymbol{\omega}$ is $N(0, \sigma^2)$ distributed, hence $\Pr[|{\mathbf{g}_i}^\mathsf{T}\boldsymbol{\omega}| < 4\sigma\sqrt{\log d}] \geq 1 - \frac{0.01}{d}$ by standard Gaussian tail bounds. Recall that $4\sigma\sqrt{\log d} \leq 1/2\sqrt{d-1}$.

Putting it all together, we take the union bound over the three terms in (4.3) and all $\mathbf{v}_i^+, \mathbf{v}_i^-$ with $i \in [d-1]$ and find that ${\mathbf{v}_i^+}^\mathsf{T}\boldsymbol{\omega} < 0$ and ${\mathbf{v}_i^-}^\mathsf{T}\boldsymbol{\omega} < 0$ for all $i \in [d-1]$ with probability at least $0.26 - (d-1)\frac{0.01}{d} - (d-1)\frac{0.01}{d} \geq 0.24$. ∎

THEOREM 4.11. *For $\sigma \leq \bar{\sigma}$, Algorithm 4.1 solves (Unit LP) in at most an expected $6 + 5\mathcal{D}_g(d, n+2d-2, \sigma/5)$ number of shadow vertex pivots.*

Proof. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ denote the rows of $\mathbf{A}$, where we recall that the centers $\bar{\mathbf{a}}_i := \mathbb{E}[\mathbf{a}_i]$, $i \in [n]$, have norm at most 1.

Pivots from line 1. Let L denote the event that the rows of $\mathbf{a_1}, \ldots, \mathbf{a_n}$ all have norm at most 2. Noting that each $\mathbf{a}_i$, $i \in [n]$, is a variance σ^2 Gaussian and $1/\sigma \geq 5\sqrt{d \log n}$, by Lemma 2.3 (Gaussian concentration), we have that

$$\begin{aligned}\Pr[L^c] &= \Pr[\exists i \in [n] : \|\mathbf{a}_i\| \geq 2] \leq n \Pr[\|\mathbf{a}_1 - \bar{\mathbf{a}}_1\| \geq 1] \\ &\leq n \Pr[\|\mathbf{a}_1 - \bar{\mathbf{a}}_1\| \geq 5\sqrt{d \log n}\sigma] \leq e^{-(d/2)(5\sqrt{\log n}-1)^2} \leq n^{-d}\,.\end{aligned}$$

Therefore, the simplex run on line 1 is executed with probability at most n^{-d} incurring at most $n^{-d}\binom{n}{d} \leq 1$ pivots on expectation.

Pivots from the main loop. Let $\mathbf{V}_1, \mathbf{V}_2, \ldots$ be independent samples of $\mathbf{V}$ as described in lines 3-7 of Algorithm 4.1. Define the random variable $N = N(A, \mathbf{V}_i : i \in \mathbb{N}) \geq 0$ as the number of iterations of the main loop if Algorithm 4.1 were run on input $\mathbf{A}, \mathbf{c}$ and the value of $\mathbf{V}$ in iteration i equals $\mathbf{V}_i$. Note that $N = 0$ exactly if L^c. Note that the value of $\mathbf{V}_i$ unique specifies the value of $\mathbf{d}_i$. Define the event F_i that the checks on lines 9 and 10 would pass on data $\mathbf{V}_i$. Lastly, let $P(\mathbf{A}, \mathbf{V}_i)$ denote the number of pivot steps that an iteration of the main loop would perform on the data $\mathbf{A}, \mathbf{V}_i$. In particular, $P(\mathbf{A}, \mathbf{V}_i) > 0$ exactly when L and F_i.

The total number of pivot steps is given by the expectation

$$\begin{aligned}\mathbb{E}[\sum_{k=1}^{N} P(\mathbf{A}, \mathbf{V}_k)] &= \mathbb{E}[\sum_{k=1}^{\infty} P(\mathbf{A}, \mathbf{V}_k)\mathbb{1}[N \geq k]] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[P(\mathbf{A}, \mathbf{V}_k)\mathbb{1}[N \geq k]].\end{aligned}$$

For any k, the event $N \geq k$ depends solely on $\mathbf{V}_1, \ldots, \mathbf{V}_{k-1}$, hence we get

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{E}[P(\mathbf{A}, \mathbf{V}_k)\mathbb{1}[N \geq k]] &= \sum_{k=1}^{\infty} \mathbb{E}_{\mathbf{A},\mathbf{V}_k}[P(\mathbf{A}, \mathbf{V}_k)\mathbb{E}_{\mathbf{V}_1,\ldots,\mathbf{V}_{k-1}}[\mathbb{1}[N \geq k \mid \mathbf{A}]] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[P(\mathbf{A}, \mathbf{V}_k)\Pr[N \geq k \mid \mathbf{A}]] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[P(\mathbf{A}, \mathbf{V}_k)\Pr[N > 1 \mid \mathbf{A}]^{k-1}],\end{aligned}$$

where the last line follows from the observation that the seperate trials are independent when $\mathbf{A}$ is fixed. When $\mathbf{A}$ is such that L^c holds, then $\Pr[N > 1 \mid \mathbf{A}] = 0$. Now we appeal to Lemma 4.9, Lemma 4.10. The first shows that the Algorithm 4.1 does not restart on line 9 with probability at least 0.98 and the second shows that the algorithm does not restart on lines 10 and 14 with probability at least 0.24. By the union bound, this implies that $\Pr[N > 1|\mathbf{A}] \leq 1 - 0.22$ for any $\mathbf{A}$ such that L holds. Hence we get

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{E}_{\mathbf{A},\mathbf{V}_k}[P(\mathbf{A}, \mathbf{V}_k)\Pr[N > 1 \mid \mathbf{A}]^{k-1}] &\leq \sum_{k=1}^{\infty} \mathbb{E}[P(\mathbf{A}, \mathbf{V}_k)(1 - 0.22)^{k-1}] \\ &= \frac{1}{0.22}\mathbb{E}[P(\mathbf{A}, \mathbf{V}_1)].\end{aligned}$$

The number of pivot steps $P(\mathbf{A}, \mathbf{V}_1)$ is nonzero exactly when L and F_1 hold, and is always bounded by the shadow size according to Theorem 2.12. We bound this quantity using Lemma 4.6 and get

$$\begin{aligned}\frac{1}{0.22}\mathbb{E}[P(\mathbf{A}, \mathbf{V}_1)] &\leq 5\mathbb{E}[\mathbb{1}_{F_1 \cap L}|\text{edges}(\text{conv}(\mathbf{A}, \mathbf{V}_1) \cap \text{span}(\mathbf{c}, \mathbf{d}_1))|] \\ &\leq 5\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{A}, \mathbf{V}_1) \cap \text{span}(\mathbf{c}, \mathbf{d}_1))|] \\ &\leq 5\mathcal{D}_g(d, n + 2d - 2, \min(\sigma, \bar{\sigma})/5) + 5.\end{aligned}$$

Final Bound. Combining the results from the above paragraphs, we get that the total expected number of simplex pivots in Algorithm 4.1 is bounded by:

$$\Pr[L^c]\binom{n}{d} + \mathbb{E}[\sum_{k=1}^{N} P(\mathbf{A}, \mathbf{V}_k)] \leq 6 + 5\mathcal{D}_g(d, n + 2d - 2, \sigma/5) \ ,$$

as needed. ∎

5. Conclusions and Open Problems. We have given a substantially simplified and improved shadow bound and used it to derive a faster simplex method. We are hopeful that our modular approach to the shadow bound will help spur the development of a more robust smoothed analysis of the simplex method, in particular, one that can deal with a much wider class of perturbations such as those coming from bounded distributions.

There remains a gap between upper and lower bounds on the smoothed shadow size. The lower bound of $\Omega(d^{3/2}\sqrt{\log n})$ by Borgwardt [18] does not depend on σ and is only proven for $n \to \infty$, and the lower bound of just over $\Omega(\min(n, \frac{1}{\sqrt{\sigma}}))$ by Devillers, Glisse, Goaoc, and Thomasse [32] is only proven to hold for $d = 2$. Nonetheless, both of these lower bounds are significantly lower than our upper bound of $O(d^2\sqrt{\log n}\ \sigma^{-2} + d^{2.5}(\sigma\sqrt{\log n} + \log n))$.

Lastly, there are no known diameter bounds for smoothed polytopes other than the general results mentioned above.

5.1. Acknowledgment. The authors would like to thank the Simons Institute program on "Bridging Continuous and Discrete Optimization" where some of this work was done.

REFERENCES

[1] K. A. Adiprasito and B. Benedetti, *The Hirsch conjecture holds for normal flag complexes*, Mathematics of Operations Research, 39 (2014), pp. 1340–1348, https://doi.org/10.1287/moor.2014.0661.

[2] I. Adler, *The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method*, tech. report, University of California, Dept. of IE and OR, 1983.

[3] I. Adler, R. M. Karp, and R. Shamir, *A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps*, J. Complexity, 3 (1987), pp. 372–387, https://doi.org/10.1016/0885-064X(87)90007-0, http://dx.doi.org/10.1016/0885-064X(87)90007-0.

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

[5] N. Amenta and G. M. Ziegler, *Deformed products and maximal shadows*, Contemporary Math., 223 (1998), pp. 57–90.

[6] D. Arthur, B. Manthey, and H. Röglin, *Smoothed analysis of the k-means method*, J. ACM, 58 (2011), pp. Art. 19, 31, https://doi.org/10.1145/2027216.2027217. Preliminary version in FOCS '09.

[7] D. Avis and V. Chvátal, *Notes on Blands pivoting rule*, in Polyhedral Combinatorics, Springer, 1978, pp. 24–34.

[8] M. L. Balinski, *The Hirsch conjecture for dual transportation polyhedra*, Math. Oper. Res., 9 (1984), pp. 629–633, https://doi.org/10.1287/moor.9.4.629, http://dx.doi.org/10.1287/moor.9.4.629.

[9] K. Ball, *An elementary introduction to modern convex geometry*, Flavors of geometry, 31 (1997), pp. 1–58.

[10] D. Barnette, *An upper bound for the diameter of a polytope*, Discrete Math., 10 (1974), pp. 9–13.

[11] R. E. Bixby, *A brief history of linear and mixed-integer programming computation*, Doc. Math., (2012), pp. 107–121.

[12] R. E. Bixby, M. Fenelon, Z. Gu, E. Rothberg, and R. Wunderling, *MIP: theory and practice—closing the gap*, in System modelling and optimization (Cambridge, 1999), Kluwer Acad. Publ., Boston, MA, 2000, pp. 19–49.

[13] W. Blaschke, *INTEGRALGEOMETRIE 2*, Bulletin mathématique de la Société Roumaine des Sciences, 37 (1935), pp. 3–11, http://www.jstor.org/stable/43769768.

[14] N. Bonifas, M. Di Summa, F. Eisenbrand, N. Hähnle, and M. Niemeier, *On sub-determinants and the diameter of polyhedra*, Discrete Comput. Geom., 52

(2014), pp. 102–115, https://doi.org/10.1007/s00454-014-9601-x, http://dx.doi.org/10.1007/s00454-014-9601-x. Preliminary version in SOCG '12.

[15] T. Bonnesen and W. Fenchel, *Theory of convex bodies*, BCS Associates, 1987. Translated from the German and edited by L. Boron, C. Christenson and B. Smith.

[16] K.-H. Borgwardt, *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung*, PhD thesis, Universität Kaiserslautern, 1977.

[17] K.-H. Borgwardt, *The average number of pivot steps required by the simplex-method is polynomial*, Z. Oper. Res. Ser. A-B, 26 (1982), pp. A157–A177.

[18] K.-H. Borgwardt, *The simplex method: A probabilistic analysis*, vol. 1 of Algorithms and Combinatorics: Study and Research Texts, Springer-Verlag, Berlin, 1987.

[19] K.-H. Borgwardt, *A sharp upper bound for the expected number of shadow vertices in LP-polyhedra under orthogonal projection on two-dimensional planes*, Math. Oper. Res., 24 (1999), pp. 544–603, https://doi.org/10.1287/moor.24.3.544, http://dx.doi.org/10.1287/moor.24.3.544. A corrected version of Figure 1 appears in Volume 24(4), pp. 925984.

[20] S. Borgwardt, J. A. De Loera, and E. Finhold, *The diameters of network-flow polytopes satisfy the Hirsch conjecture*, Mathematical Programming, (2017), https://doi.org/10.1007/s10107-017-1176-x, https://doi.org/10.1007/s10107-017-1176-x.

[21] G. Brightwell, J. van den Heuvel, and L. Stougie, *A linear bound on the diameter of the transportation polytope*, Combinatorica, 26 (2006), pp. 133–139, https://doi.org/10.1007/s00493-006-0010-5, http://dx.doi.org/10.1007/s00493-006-0010-5.

[22] T. Brunsch, K. Cornelissen, B. Manthey, H. Röglin, and C. Rösner, *Smoothed analysis of the successive shortest path algorithm*, SIAM J. Comput., 44 (2015), pp. 1798–1819, https://doi.org/10.1137/140989893. Preliminary version in SODA '13.

[23] T. Brunsch, A. Grosswendt, and H. Röglin, *Solving totally unimodular LPs with the shadow vertex algorithm*, in Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science, STACS '15, 2015.

[24] D. Dadush and N. Hähnle, *On the shadow simplex method for curved polyhedra*, Discrete Comput. Geom., 56 (2016), pp. 882–909, https://doi.org/10.1007/s00454-016-9793-3, http://dx.doi.org/10.1007/s00454-016-9793-3. Preliminary version in SOCG '15.

[25] V. Damerow and C. Sohler, *Extreme points under random noise*, in European Symposium on Algorithms, Springer, 2004, pp. 264–274.

[26] G. B. Dantzig, *Programming in a linear structure*, tech. report, U.S. Air Force Comptroller, USAF, Washington, D.C., 1948.

[27] G. B. Dantzig, *Maximization of a linear function of variables subject to linear inequalities*, in Activity Analysis of Production and Allocation, Cowles Commission Monograph No. 13, John Wiley & Sons, Inc., New York, N. Y.; Chapman & Hall, Ltd., London, 1951, pp. 339–347.

[28] G. B. Dantzig, *Linear programming and extensions*, Princeton university press, 1959.

[29] J. A. De Loera, E. D. Kim, S. Onn, and F. Santos, *Graphs of transportation polytopes*, J. Combin. Theory Ser. A, 116 (2009), pp. 1306–1325, https://doi.org/10.1016/j.jcta.2009.03.010, http://dx.doi.org/10.1016/j.jcta.2009.03.010.

[30] A. Deshpande and D. A. Spielman, *Improved smoothed analysis of the shadow vertex simplex method*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, FOCS '05, 2005, pp. 349–356, https://doi.org/10.1109/SFCS.2005.44, https://doi.org/10.1109/SFCS.2005.44.

[31] O. Devillers, M. Glisse, X. Goaoc, and R. Thomasse, *On the smoothed complexity of convex hulls*, in 31st International Symposium on Computational Geometry, Lipics, 2015.

[32] O. Devillers, M. Glisse, X. Goaoc, and R. Thomasse, *Smoothed complexity of convex hulls by witnesses and collectors*, Journal of Computational Geometry, 7 (2016), pp. 101–144.

[33] R. Dial, F. Glover, D. Karney, and D. Klingman, *A computational analysis of alternative algorithms and labeling techniques for finding shortest path trees*, Networks, 9 (1979), pp. 215–248.

[34] M. Dyer and A. Frieze, *Random walks, totally unimodular matrices, and a randomised dual simplex algorithm*, Mathematical Programming, 64 (1994), pp. 1–16.

[35] F. Eisenbrand and S. Vempala, *Geometric random edge*, Mathematical Programming, (2017), pp. 1–15.

[36] M. Englert, H. Röglin, and B. Vöcking, *Worst case and probabilistic analysis of the 2-opt algorithm for the TSP*, Algorithmica, 68 (2014), pp. 190–264, https://doi.org/10.1007/s00453-013-9801-4. Preliminary version in SODA '07.

[37] K. Fang, S. Kotz, and K. Ng, *Symmetric Multivariate and Related Distributions*, Monographs on Statistics and Applied Probability, Springer US, 1990.

[38] J. J. Forrest and D. Goldfarb, *Steepest-edge simplex algorithms for linear programming*,

Math. Programming, 57 (1992), pp. 341–374, https://doi.org/10.1007/BF01581089, http://dx.doi.org/10.1007/BF01581089.

[39] O. Friedmann, *A subexponential lower bound for Zadeh's pivoting rule for solving linear programs and games*, in Proceedings of 15th International Conference on Integer Programming and Combinatorial Optimization, O. Günlük and G. J. Woeginger, eds., IPCO '11, Berlin, Heidelberg, 2011, Springer Berlin Heidelberg, pp. 192–206, https://doi.org/10.1007/978-3-642-20807-2_16.

[40] O. Friedmann, T. D. Hansen, and U. Zwick, *Subexponential lower bounds for randomized pivoting rules for the simplex algorithm*, in Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing, STOC '11, New York, NY, USA, 2011, ACM, pp. 283–292, https://doi.org/10.1145/1993636.1993675, http://doi.acm.org/10.1145/1993636.1993675.

[41] S. Gass and T. Saaty, *The computational algorithm for the parametric objective function*, Naval Res. Logist. Quart., 2 (1955), pp. 39–45, https://doi.org/10.1002/nav.3800020106, http://dx.doi.org/10.1002/nav.3800020106.

[42] A. V. Goldberg, M. D. Grigoriadis, and R. E. Tarjan, *Use of dynamic trees in a network simplex algorithm for the maximum flow problem*, Math. Programming, 50 (1991), pp. 277–290, https://doi.org/10.1007/BF01594940.

[43] D. Goldfarb, *Using the steepest-edge simplex algorithm to solve sparse linear programs*, Sparse matrix computations, (1976), pp. 227–240.

[44] D. Goldfarb, *Worst case complexity of the shadow vertex simplex algorithm*, tech. report, Columbia University, New York, 1983.

[45] D. Goldfarb and J. Hao, *A primal simplex algorithm that solves the maximum flow problem in at most nm pivots and $O(n^2m)$ time*, Mathematical Programming, 47 (1990), pp. 353–365.

[46] D. Goldfarb and J. Hao, *Polynomial-time primal simplex algorithms for the minimum cost network flow problem*, Algorithmica, 8 (1992), pp. 145–160, https://doi.org/10.1007/BF01758840.

[47] D. Goldfarb, J. Hao, and S.-R. Kai, *Efficient shortest path simplex algorithms*, Oper. Res., 38 (1990), pp. 624–628, https://doi.org/10.1287/opre.38.4.624.

[48] D. Goldfarb and W. Y. Sit, *Worst case behavior of the steepest edge simplex method*, Discrete Appl. Math., 1 (1979), pp. 277–285, https://doi.org/10.1016/0166-218X(79)90004-0.

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

[50] T. D. Hansen and U. Zwick, *An improved version of the random-facet pivoting rule for the simplex algorithm*, in Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing, STOC '15, New York, NY, USA, 2015, ACM, pp. 209–218, https://doi.org/10.1145/2746539.2746557, http://doi.acm.org/10.1145/2746539.2746557.

[51] G. Höfner, *Lineare Optimierung mit dem Schatteneckenalgorithmus: Untersuchungen zum mittleren Rechenaufwand und Entartungsverhalten*, PhD thesis, Universität Augsburg, 1995.

[52] S. Huiberts, *How Large is the Shadow? Smoothed Analysis of the Simplex Method*, master's thesis, Utrecht University, the Netherlands, 2018, https://dspace.library.uu.nl/handle/1874/359199.

[53] M. S. Hung, *A polynomial simplex method for the assignment problem*, Oper. Res., 31 (1983), pp. 595–600, https://doi.org/10.1287/opre.31.3.595.

[54] R. G. Jeroslow, *The simplex algorithm with the pivot rule of maximizing criterion improvement*, Discrete Math., 4 (1973), pp. 367–377, https://doi.org/10.1016/0012-365X(73)90171-4.

[55] G. Kalai, *A subexponential randomized simplex algorithm (extended abstract)*, in Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing, STOC '92, New York, NY, USA, 1992, ACM, pp. 475–482, https://doi.org/10.1145/129712.129759, http://doi.acm.org/10.1145/129712.129759.

[56] G. Kalai and D. J. Kleitman, *A quasi-polynomial bound for the diameter of graphs of polyhedra*, Bull. Amer. Math. Soc. (N.S.), 26 (1992), pp. 315–316, https://doi.org/10.1090/S0273-0979-1992-00285-9, http://dx.doi.org/10.1090/S0273-0979-1992-00285-9.

[57] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395, https://doi.org/10.1007/BF02579150, http://dx.doi.org/10.1007/BF02579150.

[58] J. A. Kelner and D. A. Spielman, *A randomized polynomial-time simplex algorithm for linear programming*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing, STOC '06, ACM, New York, 2006, pp. 51–60, https://doi.org/10.1145/1132516.

1132524, http://dx.doi.org/10.1145/1132516.1132524.

[59] L. G. Khachiyan, *A polynomial algorithm in linear programming*, Dokl. Akad. Nauk SSSR, 244 (1979), pp. 1093–1096.

[60] V. Klee and G. J. Minty, *How good is the simplex algorithm*, tech. report, University of Washington, 1970.

[61] D. G. Larman, *Paths of polytopes*, Proc. London Math. Soc. (3), 20 (1970), pp. 161–178.

[62] Y. T. Lee and A. Sidford, *Path finding methods for linear programming: Solving linear programs in $\sqrt{\text{rank}}$ iterations and faster algorithms for maximum flow*, in Proceedings of the IEEE 55th Annual Symposium on Foundations of Computer Science, FOCS '14, Oct 2014, pp. 424–433, https://doi.org/10.1109/FOCS.2014.52.

[63] C. E. Lemke, *Bimatrix equilibrium points and mathematical programming*, Management science, 11 (1965), pp. 681–689.

[64] J. Matoušek, *Lectures on discrete geometry*, vol. 212, Springer New York, 2002.

[65] J. Matoušek, M. Sharir, and E. Welzl, *A subexponential bound for linear programming*, Algorithmica, 16 (1996), pp. 498–516, https://doi.org/10.1007/s004539900062.

[66] B. Matschke, F. Santos, and C. Weibel, *The width of five-dimensional prismatoids*, Proc. Lond. Math. Soc. (3), 110 (2015), pp. 647–672, https://doi.org/10.1112/plms/pdu064.

[67] N. Megiddo, *A note on the generality of the self-dual algorithm with various starting points*, Methods of operations research, 49 (1985), pp. 271–275.

[68] N. Megiddo, *Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm*, Math. Programming, 35 (1986), pp. 140–172, https://doi.org/10.1007/BF01580645.

[69] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601, https://doi.org/10.1137/0802028, http://dx.doi.org/10.1137/0802028.

[70] K. G. Murty, *Computational complexity of parametric linear programming*, Math. Programming, 19 (1980), pp. 213–219, https://doi.org/10.1007/BF01581642.

[71] D. Naddef, *The Hirsch conjecture is true for $(0,1)$-polytopes*, Math. Programming, 45 (1989), pp. 109–110, https://doi.org/10.1007/BF01589099, http://dx.doi.org/10.1007/BF01589099.

[72] J. B. Orlin, *Genuinely polynomial simplex and non-simplex algorithms for the minimum cost flow problem*, tech. report, Massachusetts Institute of Technology, 1984.

[73] J. B. Orlin, S. A. Plotkin, and E. Tardos, *Polynomial dual network simplex algorithms*, Math. Programming, 60 (1993), pp. 255–276, https://doi.org/10.1007/BF01580615.

[74] I. Post and Y. Ye, *The simplex method is strongly polynomial for deterministic markov decision processes*, Mathematics of Operations Research, 40 (2015), pp. 859–868. Preliminary version in SODA '14.

[75] J. Renegar, *A polynomial-time algorithm, based on Newton's method, for linear programming*, Math. Programming, 40 (1988), pp. 59–93, https://doi.org/10.1007/BF01580724, http://dx.doi.org/10.1007/BF01580724.

[76] A. Sankar, D. A. Spielman, and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 446–476, https://doi.org/10.1137/S0895479803436202.

[77] F. Santos, *A counterexample to the Hirsch conjecture*, Ann. of Math. (2), 176 (2012), pp. 383–412, https://doi.org/10.4007/annals.2012.176.1.7, http://dx.doi.org/10.4007/annals.2012.176.1.7.

[78] E. Schnalzger, *Lineare Optimierung mit dem Schatteneckenalgorithmus im Kontext probabilistischer Analysen*, PhD thesis, Universität Augsburg, 2014. Original in German. English translation by K.H. Borgwardt available at https://www.math.uni-augsburg.de/prof/opt/mitarbeiter/Ehemalige/borgwardt/Downloads/Abschlussarbeiten/Doc_Habil.pdf.

[79] R. Shamir, *The efficiency of the simplex method: a survey*, Management Sci., 33 (1987), pp. 301–334, https://doi.org/10.1287/mnsc.33.3.301.

[80] S. Smale, *On the average number of steps of the simplex method of linear programming*, Mathematical programming, 27 (1983), pp. 241–262.

[81] D. A. Spielman and S.-H. Teng, *Smoothed analysis of termination of linear programming algorithms*, Math. Program., 97 (2003), pp. 375–404, https://doi.org/10.1007/s10107-003-0448-9. ISMP, 2003 (Copenhagen).

[82] D. A. Spielman and S.-H. Teng, *Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time*, J. ACM, 51 (2004), pp. 385–463 (electronic), https://doi.org/10.1145/990308.990310, http://dx.doi.org/10.1145/990308.990310.

[83] N. Sukegawa, *An asymptotically improved upper bound on the diameter of polyhedra*, Discrete & Computational Geometry, (2018), pp. 1–10, https://doi.org/10.1007/s00454-018-0016-y,

https://doi.org/10.1007/s00454-018-0016-y.

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

[85] M. J. Todd, *An improved Kalai–Kleitman bound for the diameter of a polyhedron*, SIAM Journal on Discrete Mathematics, 28 (2014), pp. 1944–1947.

[86] R. Vershynin, *Beyond Hirsch conjecture: walks on random polytopes and smoothed complexity of the simplex method*, SIAM J. Comput., 39 (2009), pp. 646–678, https://doi.org/10.1137/070683386, http://dx.doi.org/10.1137/070683386. Preliminary version in FOCS '06.

[87] Y. Ye, *The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate*, Mathematics of Operations Research, 36 (2011), pp. 593–603.